CHAPTER 8

ATTENTION

General introduction, basic models and data

LEARNING OBJECTIVES

By the end of this chapter, you should be able to:

- Discuss why we need attention at all.
- Summarise the structural constraint models of attention by Broadbent (1958), Treisman (1964), and Deutsch and Deutsch (1963).
- Summarise the ideas of resource allocation set out by Kahneman (1973).
- Appreciate how both structural and processing constraints are incorporated in Pashler's (1998) framework for thinking.
- Discuss load theory (Lavie, 2000).

CHAPTER CONTENTS

A cognitive psychologist in the DJ booth

Anybody who has tried to DJ (and some of us are still trying) will appreciate how difficult it can be to listen to two things at once. Imagine the scene – you've just played a few great tunes to build up the crowd and now are getting ready to drop that track that will seal your reputation as a true master mixer (see Figure 8.1). The problem is that you've got to make sure that the two records are playing at the same speed and that those rhythms fit nicely. So, using a single left-ear headphone, you listen to the record already playing and cue up the vinyl you want to play before unleashing it onto the unsuspecting crowd. So now you're listening to both the playing and the to-be-played record through the same 'input channel'. But are we really able to listen to two things at once? Would it be more reasonable to assume we switch between sources of information? What

Figure 8.1 Attentional control and the modern world
Being a successful DJ requires phenomenal attentional capabilities (as well as being able to put your hand in the air).
Source: Katy Beswetherick/Rex Features.

would happen if one song were fed through the left headphone and the other through the right headphone? Would this make mixing easier or harder? And how about doing all of these things while ignoring those flashing lights and dry ice? Is someone calling your name from behind the DJ booth? Quick, your dance floor is waiting . . .

REFLECTIVE QUESTIONS

1. Think about an airline pilot and the various kinds of information that they have to deal with (the view out of the cockpit window, visual computer displays, auditory warning signals, communication with ground control, co-pilot and passengers). How can you apply the concepts from this chapter such that you might improve the way in which the pilot deals with these inputs?
2. Think about your immediate context right now as you read this book. Try to make a list of all the potential inputs in your environment (remember to think about all the senses). Are you being successful in attending? Is there an input that you would like to ignore that you can't? Why do you think this is? On the basis of what you've learnt in the chapter, how can you make your environment an easier one to work in?

Introduction and preliminary considerations

Psychologists use the concept of attention to explain how it is that we are able to select aspects of our environment that we are interested in while ignoring others. If we are being eyed up from across a busy wine bar, we might focus on picking up non-verbal cues from our potential suitor, while disregarding our friend's (currently irrelevant) demonstration of their new golf swing. How we might effectively juggle such concurrent and competing activities has become a question of central importance and serves as the focus of this chapter. We will look at a number of theories of attention and attentional control and discuss a number of experiments that have been put forward to help explain how we achieve attentional selection, that is, the ability to focus on the important stuff and ignore everything else. Generally speaking, attentional selection is apparently so effortless that you are probably taking it for granted right now.

In more formal terms, we will discuss contrasting views about whereabouts in the human information processing system exactly such selection takes places. Does selection occur very early on in processing or does it occur much later? Accordingly, we will discuss the division between so-called 'early selection' and 'late selection' theories. We will also discuss how selection has been thought of either in terms of *structural constraints* or in terms of *processing constraints*. Such mysterious possibilities will be thoroughly explored as we proceed.

But just for the moment, think about where you are right now. Perhaps you're reading this on the bus. How are you blocking out the noise of the other passengers? Can you be sure your attention will switch if your mobile suddenly starts ringing? How do you know when to get off if you're engrossed in this book – wait, you haven't gone past your stop, have you?

Out with the new and in with the old

Quite a lot of advertising capitalises on the naïvely held belief that this year's model is much better than last year's; modern living encourages the view that it is only right to discard the old version in favour of the new. Despite this, the disappointment can be palpable when it suddenly dawns that there really is very little to tell the new and old apart. As in life, this happens quite a lot in cognitive psychology. In many cases, it turns out that the only difference between this year's model and last year's is the packaging – old wine/new bottles. Indeed one of the mildly depressing characteristics of the history of cognitive psychology is how often old ideas are discarded, or otherwise forgotten, only to be rediscovered and repackaged at a later date.

So what are we to make of all this? Well, it does suggest that progress can only be made if the old ideas are properly evaluated. We need to be able to discriminate the useful from the useless so that we can set firm foundations for theorising. It is for this reason that the current discussion of attention starts with a consideration of Broadbent's filter theory of attention which dates back to 1958. Most of the ideas present in that model feature, in some form or other, in most models of attention ever since. So clearly these 'old' ideas have turned out to be incredibly useful. In general terms, the theory was based on a particular view about why attention is necessary. The basic idea is that attention is necessary because the central mechanisms cannot cope with the amount of sensory stimulation present at any given time. As Allport (1989) noted, by this view, the fundamental 'purpose of attentional mechanisms is to protect the brain's limited capacity system (or systems) from informational overload' (p. 633).

At the risk of revealing too much personal information, at the time of writing one of the authors is at home, there is a chirpy sparrow in a tree in the garden which is interfering with the CD currently playing, he has a slight headache from talking on the phone until 4am this morning, he is resting his foot on the corner of the dining table (not as comfortable as it sounds) and he is also digesting a sausage sandwich and a cup of tea. There are some thoughts about brewing up again, maybe taking the day off and reading the Saturday paper on the beach, and then there are decisions to be made about what to take to the barbeque he's been invited to tomorrow. Now the sparrow is back and chirping louder than usual, the phone has started to ring and the kettle has boiled and the CD needs changing and . . . Even in this state of relative inactivity, there is a lot going on in terms of sensory stimulation and clearly there are important decisions to be made about what ought to be attended to.

Early filtering accounts of selection

Broadbent's (1958) filter theory of attention is notable as being one of the most famous arrows-and-boxes accounts within the human information processing framework for thinking about cognition. It is an abstract, functional account (as described in Chapters 1 and 2) and describes putative component stages of processing that are assumed to intervene between a given stimulus and the corresponding response. There is, essentially, no discussion of physiology or neurophysiology. Even though there are clear physical constraints that set the boundaries on human information processing – for instance, we only have two eyes and two ears, that brain cells can only fire a certain number of times a second, etc. – the limitations of the human processing system were discussed in the model in abstract terms. More particularly, the idea is that the brain ought to be conceived as being akin to a communication system along the lines that we discussed in Chapter 2.

As with any communication system, there is the notion of a limited capacity channel that connects the sender with the receiver. In terms of human information processing, the sender is, essentially, the outside world (the bird song, whistling kettle, ringing telephone, etc., are being sent by the outside world) and the receiver is the participant as defined by the central interpretive mechanisms (i.e., the conscious perceiver). Limited capacity, in this context, refers to an upper bound placed on the amount of information that can be transmitted down the channel at any given time: as in a meat packing factory, there are only so many frozen chickens that can go down the conveyor belt at once. Intervening between the outside and the inside worlds are various other processes concerning sensory encoding/perceptual analysis and so on. What the filter theory does is flesh these intervening processes out in some detail. Central, though, is the idea that, at an abstract level, the human central nervous system could be conceived as comprising a limited capacity information transmission channel.

At this juncture, it is worth recalling Turvey's (1973) concurrent and contingent model (see Chapter 4). In that model, whereas the peripheral mechanisms can handle much information quickly and in parallel (concurrently), the central mechanisms operate more slowly and can only sensibly deal with one item at once (the central mechanisms are contingent on the peripheral mechanisms). Sensory information is initially registered in parallel, but further processing operates in a serial (one-at-a-time) mode. Moreover, because of the constraint of the **limited capacity channel**, only some of the information presented to the senses can be dealt with at once. Some of this information needs to be selected as being currently relevant – the rest is, essentially, ignored. It is this notion of selection that underpins the notion of attention in the theory.

To get a better handle on all this, it is simplest to consider the actual information processing account set out by Broadbent (1958). As can be seen from Figure 8.2, input to the senses is registered initially and

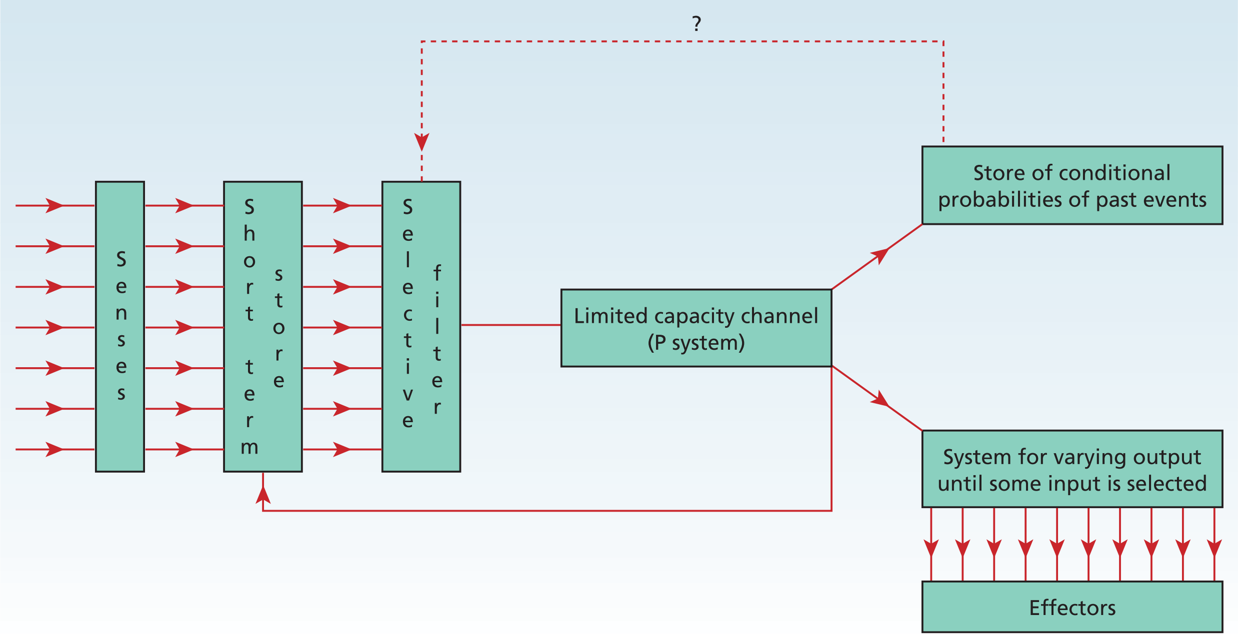

Figure 8.2 The now classic arrows-and-boxes account of the human information processing system as put forward by Broadbent (1958)

Source: Broadbent, D. E. (1958). *Perception and communication* (fig. 7, p. 299). Oxford, England: Pergamon Press.

is then passed forward to a short-term store or sensory memory system (the **S system**) in parallel (at the same time). One way to think of the S system is in terms of the body's total sensory apparatus – that is, the body's sensory receptors. As our previous example showed, there is an immense amount of sensory stimulation all of the time (the sight of the sausage sandwich, the whistling kettle, ringing telephone, etc.), and it is assumed that, because of this, the S system handles the various different kinds of information in parallel.

Rather unfortunately, the S system has also been discussed in terms of comprising parallel **input channels** which map onto the different senses such as vision, hearing and touch (Moray, 1969). A further complication, though, is that sometimes the auditory system is discussed as being composed of left and right ear channels (Broadbent, 1958, p. 58). So the mapping of input channels onto the different senses is not exactly correct, since we can also think of individual eyes and ears as input channels. So we must be careful of the distinction between input channels (so defined) and the central limited capacity processing channel. The basis idea is that the S system comprises many input channels that deliver information continuously (and in parallel) as stimulation from the outside world impinges on the body. These many input channels act as the front end to the human information processing system. The problem now is to decide which aspects of this stimulation are worthy of attention. Only a subset is selected for further processing and so enters the next stage which is characterised by the operations of the limited capacity processing channel. The basic issue is to decide which of the many input channels should be the focus of attention.

Selection by filtering

The basic idea here is that a selective filter affects the read-out or output of items from the short-term store so as to pass them on for further processing (for similar ideas see Chapter 3 and the discussion of information transfer from iconic memory). In this model, the operation of the filter acts as the first form of selection discussed by Broadbent (1958) and, unsurprisingly, this form of selection was known as *filtering*. A defining aspect of this **selection by filtering** is that it operates on the information stored in sensory memory, and as we have already discussed, such information specifies only the physical characteristics of stimuli. Sensory memory is said to represent the physical characteristics of stimuli, therefore selection from this store can only operate relative to physical features. In this way, selection by filtering is pre-categorical in nature; the identity of the actual stimulus remains unknown at this stage. (Again there are many parallels to be drawn here between this sort of sensory memory and the early discussions of iconic memory – see Chapter 3.) Select only those aspects of stimulation that possess a common physical attribute and filter out all the rest.

In addition, it is assumed that selection by filtering operates only upon items in the short-term store that share the currently pertinent common feature that the filter is set to. So read-out from your short-term store while standing in a field might deliver all the brown items (if the filter is set for brown items) *or* all the items on the right-hand side of space (if the filter is set for all the items to the right), but not all the brown items *and* all the items on the right-hand side of space. To use appropriate terminology, read-out from sensory memory operates at a **pre-categorical level**. Selection operates at a level prior to the stage at which the stimulus is identified, or, alternatively, assigned to some form of mental category. Having been confronted with a big brown shape while walking through the Rocky Mountains you now need to decide whether this actually warrants further action. Clearly the assignment of the big brown shape to the mental category of 'bear' would have significantly different consequences than assigning it to 'large rustling shrub', but at the earliest stages of processing such assignments cannot be made – only physical characteristics of the stimuli are registered.

Another important point about the short-term store is that, despite the fact that it has a very large capacity, items will decay from the store within a matter of seconds if they are not read out by the filter. For example, if two items arrive simultaneously in the S system, one will be passed on immediately, but the other will have to wait until the limited capacity channel frees up. It is also assumed that an item's probability of being selected increases directly as a function of its physical intensity. Just think about how hard it is to eavesdrop on a whispered conversation, in contrast to responding to someone shouting 'Fire!' in a doctor's waiting room. In terms of sound, a shout is more intense than a whisper, and so it is more likely that a shouted message will be passed on for further processing than a whispered message. In addition, it is also assumed the filter can switch between different selection criteria but that such switches take time. So you may feel that you can listen to one conversation while attending to another, but according to the theory, you can only do this if you switch the filter between them. How might this happen?

Switching the filter

Well, assume you are standing in a group of friends and high-pitched Janet is standing to your right and low-pitched John to your left and both are talking at the same time. There are two obvious physical things that differ between Janet and John; namely, the speakers are standing in different positions and their voices have different acoustic properties. So in attending to Janet, the filter could latch onto either a location cue (listen to the right) or an acoustic cue (listen to high-pitched voice). Switching between the conversations means **switching the filter** from either right to left or from high pitch to low pitch. The point is that you will not be able to process the two conversations at once because the filter can only follow one physical cue at once.

So although you may feel that you can process Janet and John at the same time, according to filter theory, the only way this is possible is to switch rapidly between the two conversations. However, according to the theory this switching takes time and so some information from the conversations will likely be lost during the time it takes to switch the filter. In this regard filtering is seen to be an all-or-nothing operation. Items are either selected for further processing or they are left in the S system to decay. In general, selection by filtering is assumed to be useful in cases when the participant is presented with a number of concurrent and complex events. In choosing to attend to (i.e., to select) one class of events it will be most likely that other events are less well processed and indeed may never be.

Rehearsal

All of these ideas are summarised in Figure 8.3. Items that are selected for read-out by the filter are passed on to the limited capacity channel, shown as the P (or perceptual) system in the figure. This **P system** is assumed to operate serially such that selected items are processed one at a time. A further important component of the model now may come into play and this is known as **rehearsal**. We have already seen that unless an item is selected for read-out from the short-term store it will be lost within a matter of seconds. If the system is in danger of information overload then any item selected for read-out can be passed back from the P system and rehearsed, such that it is recirculated into the short-term store. In this way the limitations

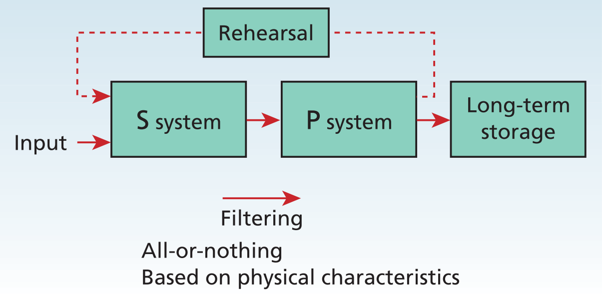

Figure 8.3 A skeletal account of Broadbent's (1958) model of processing

of the central channel can be overcome, to a certain extent. Of course, rehearsing items in this way carries an overhead that may, in turn, deflect other items from being selected from the short-term store. Critically, though, only items that have been selected and have exited from the P system stand any chance of entering the long-term memory system. In this way it is only items that are selected from the S system and are processed by the P system that can enter long-term storage.

In the original discussion of the model (Broadbent, 1958), not very much was given away about the long-term memory system. As can be seen in Figure 8.2 this system in part contains some form of record of the conditional probabilities of past events – in other words it is conceived as being an **associative memory**. In simple terms the idea is that this system keeps track of associations between items. So long as A and B co-occur then a record is kept of the strength of this co-occurrence in long-term memory. For example, the strength of association between salt and vinegar will be greater than that between salt and pepper if you live next to a chip shop rather than a fancy restaurant. The second component of long-term memory is contained on the route to the effectors and, by this, Broadbent meant the bodily mechanisms necessary for making a response. However, it is difficult to be more specific about this particular component because, again, very little else was given away about its properties (see Broadbent, 1971, p. 16).

The P system

In the original discussion of the model, the operation of the P (perceptual) system was also under-specified. Initially, some doubts were expressed about whether what came out of the limited capacity channel was the same as what went into it (Broadbent, 1958, p. 227). Was the same sort of information circulated around the system? Or was it the case that the P system actually transformed inputted information into a different sort of code? A more definite proposal was, eventually, put forward by Broadbent (1971, p. 178). Now he suggested that the P system affects perceptual–semantic categorisation, in that items in the P system are assigned to their corresponding mental categories. In other words, the outputs from this system coded **post-categorical information**: the previous brown shape is assigned to the rather alarming mental category of 'human predators' and not the friendly flora category!

Based on this discussion, it can be seen that the items that are not selected, and therefore do not enter the P system, are 'excluded from semantic categorisation or identification' (Allport, 1989, p. 633). On the understanding that the items exiting from the P system are represented in a code that specifies the post-categorical nature of the item, this was later interpreted as indicating that the P system used a different form of memory system to that of the original sensory buffer. Sensory information is initially registered in the S system in some raw form that codes the physical nature of the stimulus (e.g., large brown shape), while outputs from the P system code the categorical nature of the stimulus (e.g., bear) – semantic information is now made explicit. The representation of the stimulus entering the P system from the S system simply codes physical characteristics of the stimulus (i.e., its size, shape, position, colour, etc. (e.g., large brown shape) but it does not code its identity. At this stage the commitment to 'bear' has not been made. On leaving the P system the stimulus has been identified, hence 'bear' is now coded. The concern was with whether the so-called sensory codes and semantic codes demanded different forms of storage. Such concerns about different sorts of stimulus coding and different sorts of memory systems are covered in much more detail later (see Chapter 10). Here we merely introduce you to the sort of issue that the model throws up and we will be much more careful in our appraisal later.

Information processing constraints in the model

In Broadbent's (1958) original filter theory we can identify various sorts of processing constraints. The limited capacity channel, or P system in Figure 8.3, provides what is known as a **structural bottleneck** – the body of the bottle holds a lot of wine but only a small amount escapes through the neck of the bottle when the bottle is inverted. Whereas sensory stimulation is captured in parallel by the S system, the processing on the limited capacity channel is sequential in that only one item can be processed centrally at a time. You'll be aware of both the CD playing and the sparrow in the garden, but only actively able to listen to one at once. Similarly the limited capacity channel has a strict upper bound on how quickly any item can be cleared from the S system. It is quite possible for a logjam to occur in the S system if information is arriving too quickly for it all to enter the P system. This form of bottleneck is known as a **structural constraint** because it is part of the architecture of the model. Think in terms of a departmental store containing a lift (see Figure 8.4). The lift operates as a structural bottleneck because only a small number of people enter the lift at a given time. The speed with which the lift operates, though, would be akin to a *processing constraint* and the filter theory also contains various sorts of processing constraints.

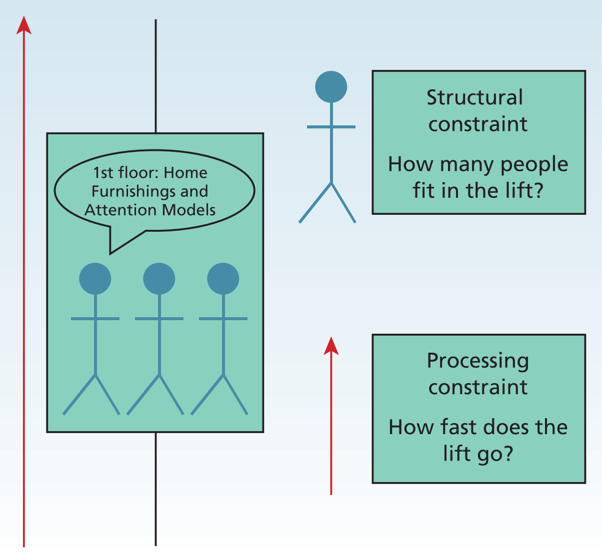

Figure 8.4 Structural vs. processing constraints
An illustrative example of the difference between structural constraints and processing constraints.

For instance, consider the filter: the filter selects items that are defined by a common physical property and items that do not possess this property are essentially ignored. The main processing constraint in dealing with complex events therefore is the time it takes the filter to switch from one pertinent feature to the next. Again information may be lost from the S system because this decays before the filter is switched to the appropriate stimulus characteristic.

So given this theory, what is the evidence? What are the relevant data? Despite the cursory treatment of the model here, the whole of Broadbent's 1958 book provided the empirical justification for the model, but in a kind of brave and foolish way we will only consider two sorts of experiment that set the foundations for the theory.

Pinpoint question 8.1

How might we avoid item decay if information is left in the sensory system in the Broadbent (1958) model?

Split-span experiments

The first kind of experiment that provides support for the filter theory are known as split-span experiments, initially reported by Broadbent (1954) and demonstrated in schematic form in Figure 8.5. On each trial participants were presented with a recording of a spoken list of digits over headphones, either under presentation conditions known as (i) a *conventional list* or under presentation conditions known as (ii) a *binaural list*. In a conventional list, the same series of digits were spoken in unison down the left and right headphone channels. So say the list was 734215, then both left and right input channels would carry 'seven', 'three', 'four', 'two', 'one' and 'five' in synchrony. All the participant

For example . . .

Structural constraints don't just occur in models of attention – you only need to think about a tennis ball serving machine for a real-world example. The top basket holds many balls waiting to be launched, but only one can be fired at once. The actual serving mechanism and its one-ball-at-once operations constitutes a structural bottleneck. Which in this case is pretty useful otherwise you'd never be able to practice your half volley. This is very much like the neck of a bottle of wine, and the limited capacity channel that Broadbent talked about in his model. You could also think of a processing constraint using the same example, the processing constraint being how fast the machine can reload and fire balls at you.

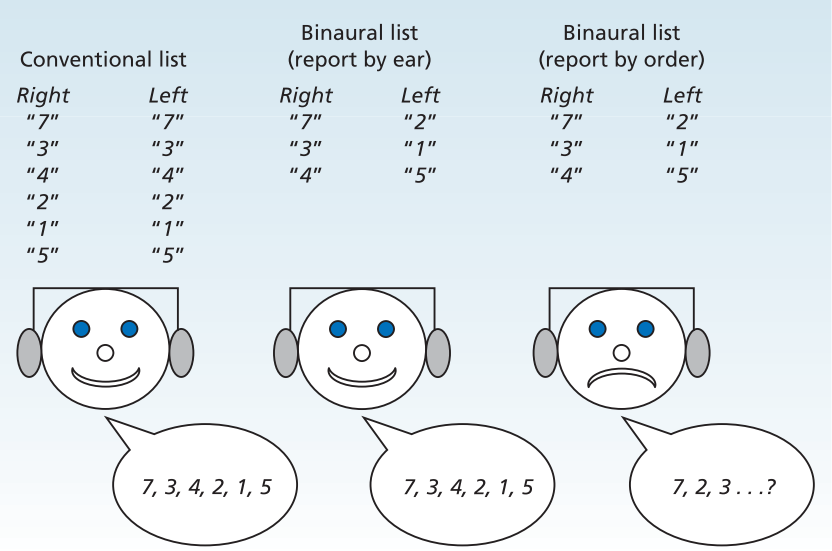

Figure 8.5 Split-span experiments
Illustrative examples of different conditions in the split-span experiments reported by Broadbent (1954).

had to do was to listen to the list and then write down the actual sequence of digits heard. With conventional list presentation over 90 per cent of the lists were reported correctly.

In the case of binaural presentation simultaneous pairs of digits were presented to the left and right channels but different items were presented to the two ears. So the sequence of items now for the left and right ears would be the pairs 'seven'/'two', 'three'/'one' and 'four'/'five' spoken in unison. For one group of participants they were told simply to report as many items as they could. The most striking observation was that they overwhelmingly chose to attempt to report all the items from one ear before they reported any of the items from the other. So, in our example, a typical response for the order of report would be 734215. Participants were able to report approximately 60 per cent of the lists correctly if they adopted this recall strategy. In contrast, performance dropped dramatically when participants were instructed to report the digits in an alternating report condition. Now they had to alternate between the ears of delivery. From our example, the participant was expected to report the digits in the order 723145. Under these conditions, performance was compromised.

Critically, performance also depended on the rate of presentation of the pairs. In the conditions we have been discussing, the rate of presentation of the pairs was one pair every 1/2 s. With this rate of presentation, in the alternating report condition, participants were only able to report approximately 20 per cent of the lists correctly. However, if the rate of presentation of the lists was slowed down to one pair every 2 s then report approximated the 50 per cent correct level.

To begin to explain these data we have to accept the notion that the ear of delivery is conceived as an input channel and that the filter selects items only from one of these channels. We choose to attend to either the left or right ear in a similar way to which we choose to attend to Janet or John. Furthermore it was also asserted that switching between these channels takes time. So when participants were presented with binaural lists and were free to use whichever strategy they wished in terms of reporting, they chose to report the content delivered to one ear and then report the content delivered to the other ear. Such a strategy led to the least amount of filter switching. Performance was relatively good because during list presentation the filter is set to one channel (the left ear) and forwards the items from this channel. The items from the other channel (the right ear) are registered and remain in the sensory buffer (the S system) until the filter is switched so that these items can be read out for further processing. Moreover, when the rate of presentation of the list was short enough, none of the items from the later selected channel decayed and were lost.

In contrast, in the alternating report condition, when participants were required to report alternating digits from both ears, performance was poor and this was taken to reflect limitations in having to switch the filter. Now during list presentation the participant attempted to switch the filter between the left and right ears rapidly so that alternate items from the different ears could be passed to the limited capacity channel. At high presentation rates (1/2 s) the filter was simply unable to fluctuate fast enough and items were therefore lost. However, if the presentation rate was slow enough (2 s) then it is possible to switch between the input channels and recover the items in the correct way.

Similar data were also collected in a later variant of the split-span procedure (Broadbent, 1956) in which the presentation of the items was divided across sight and sound (the visual and auditory modalities) in a cross-modal condition. This time, digit sequences were divided up so that pairs were presented simultaneously with one member being presented visually and the other being presented aurally. A within-modality condition was also run in which the digits were presented either in the visual or auditory modality. The central finding was that memory in the cross-modal task was equivalent to that in the within-modality task, but only if the participants adopted the correct strategy. To do well in the cross-modal task, participants had to attempt to report all the items from one modality before switching to the other. Attempting to alternate between the modalities resulted in inferior performance. In this regard, the eyes and, separately, the ears could be taken to correspond to different input channels, with switching between the two input channels taking time.

Pinpoint question 8.2

What are the two ways in which digit recall can be improved for the binaural condition in split-span experiments?

Shadowing experiments

The second type of experiment that provided data that motivated Broadbent's (1958) filter theory of attention can be traced to the shadowing studies of Cherry (1953). These experiments were similar to split-span studies in terms of bombarding participants with lots of acoustic information, but, as we will see, the tasks were quite different. Here we can again adopt the notion of the two ears (i.e., the left ear and the right ear) being defined as separate input channels, and distinguish between an **attended channel** and an **unattended channel** with this distinction becoming apparent in a moment. In a basic shadowing experiment, separate messages are played concurrently down the left and right speakers of headphones. Such an experiment is also known as a **dichotic listening** experiment because messages are being played down the two channels at the same time. The aim of the experiment is to force the participant to attend to one channel but the objective is to try to assess what information is being recovered from the unattended channel. How much of John can you perceive when paying attention to Janet? The ingenious aspect of these experiments is the manner in which participants were forced to repeat out loud the attended message as it was being played. They were asked to **shadow** (immediately repeat aloud) the attended message. As Moray (1969) noted, the messages were presented at a rate of about 150 words per minute and after some practice most participants were able to report the attended message correctly.

The beginning and end portions of the unattended message always comprised a male voice speaking English, but the middle portion of the message changed in different ways across different conditions such as changing the sex or nationality of the speaker. Using this shadowing technique, Cherry quizzed his participants afterwards on what they noticed about the message on the *unattended* channel. Cherry found that participants did notice a gross change in pitch (low male voice to high female voice, and whether a pure tone was played) but they did not notice the introduction of a foreign language or reversed speech. In short, participants could say whether the unattended message contained speech and whether it was spoken in a male or female voice, but they were unable to report anything about the content of the message.

A general conclusion therefore was that all that can be recovered from the unattended message is some specification of its physical characteristics and it seemed as if there was no semantic analysis (analysis of the meaning or content) of the unattended message at all. Such a conclusion is generally in line with the predictions of the original filter theory: unless items are selected for further processing they remain coded in terms of their pre-categorical characteristics. The content of the unattended message cannot be recovered because, according to filter theory, the filter prohibits the unattended message from entering the limited capacity channel where semantic analysis takes place. It is only the attended message that undergoes a full analysis.

Provocative data – challenges to the early filter account

Having concluded that participants in his study were unable to report a change to the unattended channel involving reversed speech, Cherry (1953) also noted that some participants did detect that something odd had occurred when the unattended message changed in this way. A possible conclusion therefore is that these participants may have picked up a distinction between a meaningful and meaningless message even though, according to the original filter theory, the unattended message never gains access to any kind of semantic analysis. However, according to Broadbent's (1958) filter theory, such a state of affairs might occur if participants either intentionally, or unintentionally, switched the filter to the unattended channel when the reversed speech was presented.

A more recent and careful appraisal of this particular experimental manipulation has been provided by Wood and Cowan (1995a). They were able to confirm that the detection of the reversed speech did depend on shifts of attention to the unattended channel, but the central issue was: why did such shifts occur? Were such switches random or did they reflect a more interesting possibility? Did the unattended message actually undergo some form of semantic analysis? We might assume that, in the normal course of events, the unattended message does undergo some form of semantic analysis and when this is disrupted – such as when it changes to reversed speech – attention is summoned over to the unattended channel by switching the filter. Clearly such a possibility fits rather uncomfortably with the original all-or-nothing conception of the filter in Broadbent's theory (see Figure 8.3).

However, other evidence began to emerge which suggested that this rather uncomfortable possibility might have some truth behind it. For instance, Moray (1959) contrived a rather devious shadowing experiment in which he embedded instructions to change ears or stop shadowing in the messages. Such commands were easily detected if presented on the attended channel but went undetected on the unattended channel. However, if the command was prefixed by the participant's own name, on the unattended channel, then it was heard on about 30 per cent of occasions. Yet, as Wood and Cowan (1995b) noted, participants hardly ever followed the commands. The only evidence that the commands were detected came from subjective recollections immediately after the shadowing trial.

Wood and Cowan (1995b) were able to replicate Moray's findings in much more tightly controlled conditions. They concluded that participants who were able to detect their own name on the unattended channel also showed evidence of shifting their attention to the unattended channel following registration of their own name. In this regard, presentation of the name did summon attention to the unattended channel. Importantly, this reveals that there was some semantic analysis of the unattended message even though this was not entirely consistent with the conception of attentional constraints described in the original filter theory.

Research focus 8.1

I think my ears are burning: why do I hear my name across a crowded room?

While our own conversation is clearly going to be the most riveting in any party setting, it seems as though we are also hopelessly drawn to the sound of our own name across a crowded room. Researchers have been particularly interested in the attentional demands of cocktail parties for about half a century now, which presumably has nothing do to with the alcohol involved. Conway, Cowan and Bunting (2001) wondered about individual variation in the ability to divide auditory attention across two channels and whether this would be related to something known as working memory (i.e., a short-term memory system – see Chapter 10 for a more thorough description). Working memory is thought to be involved in maintaining attentional focus upon task-relevant information (Baddeley & Hitch, 1974). This is why Conway et al. (2001) thought that it might be related to performance in studies where attention has the potential to be divided across multiple input channels. Therefore individuals with poor working memory might be able to hear their name across a crowded room better because they find it difficult to maintain attention on their own conversation in a distracting environment.

Forty participants were used, half of whom were identified as having low working memory span, while the other half had high working memory span. All participants were then presented over headphones with two different messages, one to each ear as in the original Cherry (1953) study. Participants were required to shadow one message in the right ear while ignoring a simultaneously presented message in the left ear. Unbeknown to the participants, their name had been inserted into the unattended message, in addition to a control name that was not their own. After shadowing, participants were asked whether they found anything unusual about the unattended message and if so, what it was.

▶

Sixty-five per cent of individuals characterised with poor working memory heard their own name on the unattended channel in contrast to only 20 per cent of those with good working memory capacity. The distracting effect of hearing one's own name in an unattended channel also persisted for one or two words after the event, as revealed by a larger number of shadowing errors committed by individuals who had been drawn to their own name.

At a basic level, Conway et al. (2001) support the idea that sometimes information on unattended channels can be semantically analysed just as Moray (1959) showed. However, this study also emphasises the importance of individual differences in performing these kinds of dichotic listening tasks, suggesting that any structural limited attentional capacity must be considered on a person-to-person basis. A final conclusion could be that your performance at a cocktail party might not have anything to do with drinking Long Island Iced Teas after all.

Source: Conway, A. R. A., Cowan, N., & Bunting, M. F. (2001). The cocktail party phenomenon revisited: The importance of working memory capacity. *Psychonomic Bulletin & Review, 8*, 331–335.

limited capacity channel A notion that is based on the view that information must be transmitted in human information processing between different processing modules – for instance, between the peripheral and the central mechanisms. The understanding is that processing modules are connected by processing channels, any one of which can only cope with a certain amount of information at any given time.

S system The early sensory storage and coding mechanisms in Broadbent's 1958 model of human information processing.

input channels The idea that the sensory apparatus can be divided into independent streams of processing. Descriptions of an input channel can include whole senses (e.g., touch) or specific sensory transducers (e.g., the left ear).

selection by filtering A specific kind of selection taken from Broadbent's (1958) filter theory in which information is chosen from sensory memory on the basis of physical characteristics of the input.

pre-categorical level A stage in processing at which the categorical identity of the stimulus is not known. For example, it's red and it's small but what is it?

switching the filter In Broadbent's (1958) account, whereby the focus of attention is changed from one input channel to another.

P system Strictly, the perceptual system in Broadbent's (1958) filter theory of attention. In this scheme the P system assigns meanings to signals propagated down the limited-capacity channel.

rehearsal Allowing information to be recirculated in a processing system. In Broadbent's (1958) account, information from the P system is returned to the S system.

associative memory Information storage that keeps track of the associative relations between two events. Typically the strength of an associative link is indexed by the co-occurrence between the two associates.

post-categorical information A specification of the categorical nature of a stimulus.

structural bottleneck A claim that not all information arriving at a particular stage of processing can be dealt with concurrently; only information associated with one task can pass through to the next stage at once.

structural constraint A limitation within the architecture of the system that limits the amount of information that can pass through at any one time. Also known as a structural bottleneck.

attended channel An input channel that participants concentrate on.

unattended channel An input channel that participants are asked to ignore.

dichotic listening Situations that involve different acoustic material being presented to left and right ears at the same time.

shadow To report back verbally the contents of an attended channel in real time.

Pinpoint question 8.3

How does the work of Wood and Cowan (1995b) challenge the filter theory of Broadbent (1958)?

The attenuated filter model of attention

In addition to the provocative data reported by Cherry (1953), other evidence also began to emerge that indicated that unattended messages may be analysed at the semantic level. Indeed, a now classic piece of evidence was reported by Treisman (1964) and this resulted in a modification to the original filter theory. Again participants were tested in shadowing tasks and again a rather devious manipulation was employed. Each message was 50 words long and each individual message was either comprehensible text or a stream of spoken words that came across as some kind of garbled English. As in standard shadowing experiments, different messages were simultaneously delivered to the left and right headphone speakers. On each trial though, the messages could switch channels. Schematically this is conveyed thus:

Attended channel	. . . I SAW THE GIRL/song was WISHING . . .
Unattended channel	. . . me that bird/JUMPING in the street . . .

Text on a given line corresponds to what was actually presented on the designated channel. The '/' indicates that point at which the messages were switched and the capitalised text indicates what the participant actually vocalised during the shadowing trial. What this indicates is that occasionally participants followed the content of the message across the channels for one or two words after the messages switched but they then immediately reverted back to the attended channel.

This particular pattern of performance was peculiar to cases where the messages comprised comprehensible text and significantly fewer shadowing switches occurred when the messages contained only approximations to English. This finding reinforced the conclusion that some form of semantic analysis of the unattended message must have been taking place, otherwise how could it be that the participants mistakenly followed the content of the messages across the channels when the switch occurred? If the messages were not semantically coherent then such channel switches tended not to occur. We shall cover Treisman's detailed account of her data later, but for now all we need note is the following. This sort of finding led Treisman (1964) to suggest that, in contrast to the original all-or-nothing filter in the original filter theory, unattended messages were attenuated (dampened down) and were not completely blocked out from further analysis. This suggestion forms the basis of the **attenuated filter account** of processing and in its general form this was accepted as a sensible modification to the original filter theory (Broadbent, 1971, 1982).

One way to think about the attenuated filter theory is to conceive of the unattended message as having a lower signal-to-noise ratio (see Chapter 4) than that of the attended message. The filter operates to dampen down the unattended message but it does not block it out completely. The reason that participants sometimes followed the content across onto the unattended channel is that the unattended message was, in some sense, being analysed at a semantic level. Semantically relevant items on the unattended channel were therefore easily recognised – they were in a sense primed by the context provided by the leading attended message.

Further revisions to the original filter theory

So far we have examined the original filter theory (Broadbent, 1958) and the attenuated filter theory (Treisman, 1964). Both of these were attempts to explain how it is we can selectively attend to some of the vast amount of sensory stimulation and discard the rest. However, we have only focused on one form of selection – selection by filtering – but two other forms of selection were introduced and discussed later by Broadbent (1971). These were **selection by categorisation** and **selection by pigeonholing** and it seems that discussion of these other forms of selection were an attempt to extend the theory so as to encompass the more challenging data.

Selection by categorisation

The first (new) form of selection, known as *categorising*, featured little in the extended theory (Broadbent, 1971). Nevertheless, from what is given away, it clearly is central to ideas about object recognition and identification in general. We shall return to these points when we discuss these topics in more detail later (see Chapter 13). Categorising was said to involve both input and output selection and refers to assigning stimuli to mental categories. The outputs from the limited capacity channel were termed *category states* because the purpose of the limited capacity channel was to take an item and assign it to an appropriate mental category. In this regard, the outputs from the limited capacity channel code post-categorical information.

At a further level of detail it was noted that many different stimuli may be assigned to the same category state; for example, pug, labrador, and poodle are all members of the category DOG. In addition, it was also recognised that each stimuli may evoke a variety of similar responses (e.g., saying 'Look at the dog', 'Fetch', etc.). This is why categorisation is said to involve both stimulus (input) selection and response (output) selection. The participant must (on the input side) discriminate dogs from all other stimuli and produce the most appropriate of a variety of different possible responses (on the output side).

In discussing input selection, the idea is that the perceptual system must be able to discriminate relevant from irrelevant stimuli. With filtering, such selection is relatively straightforward because it refers to cases where relevant stimuli share a common feature so any stimuli missing this feature can be ignored as irrelevant – when shopping for an engagement ring, for example, it might be the case that anything other than gold would be ignored by your potential spouse. Categorising, however, involves cases where there is no single discriminating feature; for instance, there is no such perceptual feature common to all dogs. Certainly all dogs are supposed to bark but their actual barks are quite different from one another. On these grounds, categorising applies when category assignment depends on analysis of more than one feature (barks, four legs, tail, etc.). However, this does not necessarily mean that the stimulus must undergo a complete analysis as long as discriminating features can be recovered.

A concrete example is provided by the work of Rabbitt (1964, 1967). Here participants were asked to base a response on whether a C or O was present among other non-target letters. Participants needed time to practise the task, and early on in practice the time to find the target letter lengthened as the display set size increased (i.e., as the number of items in the display increased). However, later on in practice the irrelevant letters had less of an effect on target detection times. Initially participants searched for C or O among non-targets taken from the set AEFHIKL. Participants improved over time because they learnt the C/O category. They 'learnt' that this category contains only curves and no straight-line features. Having acquired this category, non-target items could be eliminated from further processing on the basis of detecting a straight line.

Supporting evidence for this view was found by switching the non-target items to either a new set containing other angular letters (e.g., TZXY) or one containing curved letters (e.g., BDGQ). Performance dropped considerably when curved non-targets were introduced. In contrast, performance did not suffer so much when the other angular non-target set was introduced. What this indicates is that selection by categorisation can be more much more complex than simple filtering. Moreover, changing the setting of the filter can be very quickly accomplished as we have seen in the split-span experiments of Broadbent (1954). In contrast, changing the setting of the category rule can take an immense amount of time.

Selection by pigeonholing

The other (new) form of selection discussed by Broadbent (1971) is known as *pigeonholing* and this can only come about once mental categories, in the form described above, have been acquired. The general idea here is that the participant can bias certain category assignments over others. Instead of selecting items on the basis of a unique and common feature, as is the case with filtering, pigeonholing operates by changing the probabilities of the outputs (the category states) from the limited capacity channel. It is very much in the vein of a criterion shift as discussed in the context of signal detection theory (see Chapter 4). Think in terms of moving a threshold downwards such that only a minimal amount of information consistent with a favoured category will trigger the corresponding response.

For example, let's assume that you are waiting for a friend from a train at a busy station and she has told you that she will be carrying a red bag and a black umbrella. In the first instance you bias the category of red things so that you attempt to recognise your friend from her bag. If this is not working, now bias the black things category and attempt to recognise her on the basis of her umbrella. Switching the selection rule here can be carried out fairly quickly because it implies biasing a different but already established mental category. The general idea is that you will more readily recognise your friend because of these kinds of expectations than if you have no expectations at all.

The differences between stimulus set and response set

Despite the discussion of categorisation, much more interest was placed on the differences between selection by filtering and pigeonholing, and indeed these notions have endured and have been incorporated into a much more recent account of attentional processing (see Bundesen, 1990). Specifically Broadbent discussed filtering in terms of something he called **stimulus set**, and pigeonholing in terms of something else he called **response set**. Stimulus set refers to the common physical characteristic unique to the relevant items – e.g., all the red items, all the low-pitch items, all the squares, etc. In contrast, response set refers to cases where the assignment of stimuli to responses is necessarily defined with respect to some form of mental category: for example, classifying *x* as an arithmetic symbol rather than as being composed of two crossed diagonal lines.

In simple terms, whereas this stimulus set can be defined relative to the physical characteristic of the stimulus, this response set cannot. For instance, a stimulus set in a shadowing experiment might instruct the participant to report back everything spoken in a male voice and ignore anything spoken in a female voice, whereas a response set would instruct the participant to report back only the names of animals that are spoken in any voice. Similarly, in a display of red and black digits and letters, stimulus set might define the targets as all the black items or all the red items whereas response set might define the targets as all the digits or all the letters (Broadbent, 1970).

Filter theory predicts that selection by filtering should be easier than pigeonholing because of the manner and timing in which different aspects of the stimulus become available during processing. The attentional filter is said to operate early on in processing and latches onto a distinctive physical characteristic as registered in the sensory store (it can latch onto 'red things'). In contrast, selection by pigeonholing can only take effect once an item has entered and exited the limited capacity channel and has been assigned to a mental category (such as 'my friend'). Hence selection by stimulus set should be easier than selection by response set since, in order to decide whether an item is a target in a response set, the item must be assigned to an appropriate mental category. Broadly speaking, the data have confirmed this prediction (Broadbent, 1970). Indeed, the findings from the iconic memory literature, showing that selection by colour is far easier than selection by alpha-numeric class (see Chapter 3), provide converging support (Bundesen et al., 1984).

According to Broadbent (1971), by embellishing the original filter theory with the more detailed discussion of selection by filtering, categorisation and pigeonholing, this allowed him to provide a much more elaborate framework for thinking about the nature of attention and attentional constraints. For instance, by incorporating the notion of pigeonholing into the theory, a mechanism is provided for explaining how aspects of the unattended message might break through in shadowing experiments. For example, the biasing context used by Treisman (1964) on the attended channel lowered the thresholds for detecting relevant words. Hence recognition of contextually appropriate material was facilitated and participants unintentionally switched to the unattended message.

Pinpoint question 8.4

Why is it harder to find the letters C and O among BDGQ rather than TZXY?

attenuated filter account Treisman's (1964) revision of Broadbent's (1958) filter theory in which relevant portions of an unattended signal may be semantically processed. In this account filtering is partial and not all-or-nothing as in Broadbent's earlier theory.

selection by categorisation The ability to select items for further processing on the basis of the category membership.

selection by pigeonholing The ability to select items because of lowering of a category threshold – if it looks like a duck and walks like a duck, it must be a duck.

stimulus set The common physical characteristics shared by items that are designated targets.

response set Stimuli defined as targets because they share the same mental category. No one physical characteristic defines the targets.

Late filtering accounts of selection

Although it became increasingly apparent that the all-or-nothing account of filtering in the original theory was untenable, Broadbent retained the notion of an early filter that picked out material for further processing purely on the basis of their physical characteristics. Select the red items, select the loudest items, select the item moving very quickly towards your head! A concession had been made, however, in that the filter merely attenuated irrelevant (unattended) stimuli and did not block them out completely. In contrast to this notion of early selection, a radical alternative was put forward in terms of a late operating filter, in which selection takes place at a much later stage of processing and involves much more complex categories than physical stimulus properties. This view is generally associated with Deutsch and Deutsch (1963), and is typically described relative to the schematic presented in Figure 8.6.

Deutsch and Deutsch discussed data, such as Moray's (1959), that were taken to reveal semantic analysis of

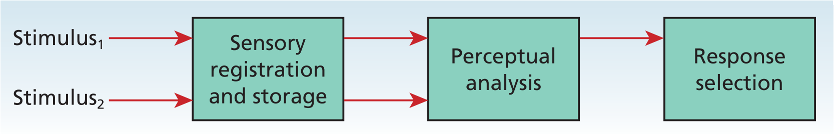

Figure 8.6 An arrows-and-boxes account of late attentional selection
Selection does not operate until all inputs have undergone perceptual analysis and category assignment. Selection for action does not operate until everything has been identified.

Source: Kahneman, D. (1973). *Attention and effort* (fig. 1.1b, p. 6). Englewood Cliffs, New Jersey: Prentice Hall. Reproduced with permission.

unattended material. By considering this sort of data, they dismissed the notion of an early filter situated at a stage where only the physical characteristics of stimuli were available. In contrast, they took the evidence to show that 'all sensory messages which impinge on the organism are perceptually analyzed at the highest level' (p. 85).

More recently such a view has been developed by Duncan (1980). According to him, prior to any selection taking place, each stimulus is fully identified and this means that a representation of the stimulus is derived that codes its physical characteristics, its name if this is applicable, and (critically) its category assignment (p. 284). A further elaboration on the Deutsch and Deutsch account was also provided by Kahneman (1973). According to him the late filtering model implies that 'the meanings of all concurrent stimuli are extracted in parallel and without interference' (pp. 6–7). Any form of selection therefore only takes place at a post-categorical stage of processing prior to response execution. It is as if everything is identified but that the critical constraints relate to responding.

If you have four pans of water all boiling at one time, then the one that is about to boil over will be the stimulus that is currently most important and the one that will require a response. You only have one pair of hands so only one of the pans can be dealt with at once. In this account the stimulus that will be responded to is the one that is momentarily designated the most important. It is the assignment of importance to the various stimuli that is what is implied by 'semantic analysis' in the Deutsch and Deutsch (1963) account. Moreover, as Kahneman (1973) notes, the weighting of importance can reflect momentary intentions (you reach for the mango chutney, not the raita) or enduring dispositions (you prefer beer to white wine).

There is a huge literature on the merits and shortcomings of early and late filter theories; however, one study is particularly relevant and will be discussed briefly here. Treisman and Geffen (1967) designed a dichotic listening task in which participants were given two tasks to complete. The primary task was simply to report back (i.e., shadow) the designated attended message. The secondary task was physically to tap when a designated target word occurred on either the attended or unattended channel. The argument went like this. If selection was located at an early perceptual stage then targets on the unattended channel should be missed and therefore participants should be less accurate in tapping to unattended than attended targets. However, if the problem lay at a response selection stage – as argued by Deutsch and Deutsch (1963) – then taps to both sorts of targets should be equally as bad, regardless of whether the target occurs on the attended or the unattended channel.

In greatly summarising the results, what Treisman and Geffen (1967) found was that participants tapped to 87 per cent of the attended targets but to only 8 per cent of the unattended targets. Hence the data fit more comfortably with the early (attenuating) filter view than the late response selection view (see Deutsch & Deutsch, 1967; Lindsay, 1967; Treisman, 1967; for an interesting interchange on these findings). However, to be fair, the experiment addressed a very particular aspect of one late selection theory and the findings therefore do not undermine late selection theories *in general*.

Evidence in support of late selection

So is there any convincing evidence in favour of late selection theories? Well, in fact there is, and some of the more intriguing evidence has been reported by Duncan (1980). Here performance was examined in some disarmingly simple, but subtle, visual search experiments (see Figure 8.7). Participants were presented with a series of four items on each trial (let's say the letters X, T, E and the number 3), and performance was compared across two main conditions. In the SIM (simultaneous) condition all four items were presented at once, whereas in the SUCC (successive) condition one pair of items was presented initially and then another pair followed this pair.

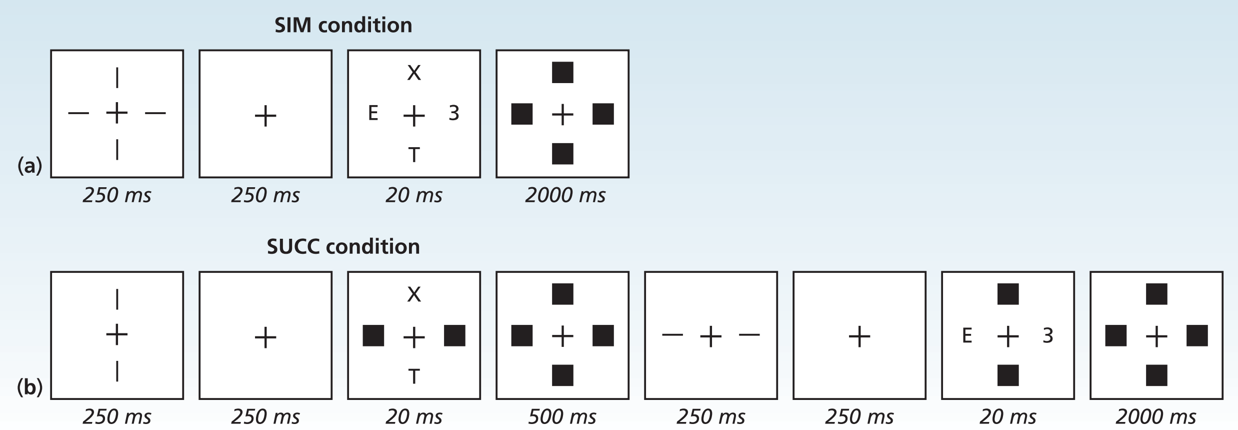

Figure 8.7 Schematic representation of the sorts of displays and conditions used by Duncan (1980)
(a) In the SIM(ultaneous) condition, participants were presented with stimuli on a vertical and horizontal limb at the same time. (b) In the SUCC(essive) condition, participants were presented with stimuli from one limb at a time. The task was to spot target digits presented among non-target letters. The start of a trial is at the leftmost side of the figure and the time line showing how events unfolded within a trial travels from left to right. In the SUCC condition the rows of events unfolded consecutively.

Any given item could appear at one of four positions that were located at the ends of an imaginary '+' centred on a display screen. The '+' was said to be composed of a horizontal and a vertical limb. Two items were presented top and bottom relative to centre, while two items were presented left and right relative to centre. In the SIM condition the four items were presented very briefly (e.g., for 20 ms) and were then immediately masked by individual rectangular light patches (this provides another example of backward masking; see Chapter 3). In the SUCC condition the items on the first limb were presented briefly and then masked; approximately a second later the items on the second limb were presented and then masked.

The participants' task was to look out for target digits presented among non-target letters (a 3, say, among the X, T and E in Figure 8.7). On a given trial the displays could contain either a single target (on one limb), one target on each limb (i.e., two targets in total), or no targets on either limb. Performance was also assessed in a so-called *combined task* and, separately, in a *separated task*. In the combined task, participants simply had to press a key if they detected the presence of a target anywhere. So even though, on some trials two targets were present, they could simply respond as soon as they detected either target. Participants merely had to respond if they saw a target – they did not have to report which target they saw. In contrast, in the separated task, participants had to make target present/absent judgements for each limb separately. Measures of accuracy were of central import for (i) comparisons between the SUCC and SIM conditions and (ii) comparisons across cases where either one target or two targets were presented.

In considering the combined task data (in which the participant could respond to any target), there was little difference in performance across the SIM and SUCC conditions when only a single target was present. Participants were marginally more accurate in detecting the single target when each limb was presented in turn (in the SUCC condition) than when the whole display was presented (in the SIM condition). So there was a small cost in having to divide attention across the whole display relative to the cases where they could focus attention on each pair of items in sequence. Remember, in the SIM condition, participants were trying to concentrate on both limbs of the display at once and hence divide their attention across the whole display. In contrast, in the SUCC condition, they were trying to focus their attention on each limb at a time as each limb was presented in sequence.

Of more interest, though, are the data in the separated task, in which participants had to make judgements whether a target was present on each limb. Now there was a large effect of having to divide attention, with performance being worse in the SIM than the SUCC case. Not surprisingly perhaps, participants performed worse when having to process four simultaneous items than they were when faced with a sequence of two pairs of items. The most intriguing effects, however, occurred in the separated task when two targets were present – one on each limb. The critical finding

➔ What have we learnt?

So far we have traced the history of research in attentional selection in terms of (i) the early all-or-nothing filter account of Broadbent (1958), (ii) the attenuated filter account of Treisman (1964), and, finally, (iii) the late selection account of Deutsch and Deutsch (1963). Broadbent extended his framework for thinking by discussing various ideas about how selection might operate. He discussed (i) selection by filtering, (ii) selection by categorising, and (iii) selection by pigeonholing. Each form of selection was discussed in terms of a general framework for thinking. Nevertheless, the data reported by Duncan (1980) are challenging and seemingly cannot be easily accommodated by the theory. In reply Broadbent (1982) insisted that until latency measures were taken as well as accuracy measures, then the costs in performance due to the presence of multiple targets should be interpreted with caution. Since then the arguments about where best to try to locate selection in the human information processing system have oscillated back and forth. More recently, various compromise positions have been put forward, as we will consider later in the chapter. For now, though, it is appropriate to discuss a quite different take on the issues. From this perspective, performance limitations should not be taken to reflect structural constraints present in the processing system, but the consequences of *resource constraints*.

was that participants were particularly impaired when two targets were concurrently present (in the SIM condition) and where they were expected to register and respond to both targets. In a later similar experiment, Duncan found that whereas the presence of more than one target disrupted performance, the number of concurrent non-targets was relatively unimportant.

So what does all this tell us about early vs. late theories of attentional selection? Well, the first point is that the limitations in processing that these experiments reveal seems to reflect problems the system has in processing concurrent targets – as Duncan stated, the data suggest that non-target elements 'do not compete for limited capacity processes' (p. 284). So how can this be? Well, as we have already noted, according to Duncan all items are fully analysed at an initial unlimited stage of processing. The critical constraint occurs at a subsequent stage in which items are selected for further processing. The identity of an item is recovered at the first stage and this means that each item will be identified as either a target (digit) or a non-target (letter). Non-targets can therefore be sensibly ignored but the targets must now compete for entry into a limited capacity system. Output from this limited capacity system eventuates in a perceptual record of the item in question, although problems arise if more than one target is present concurrently. Target items compete for entry into the limited capacity system and hence performance suffers accordingly. So the problems do not arise in identifying what is and what is not a target, they arise after the identity of the items has been recovered, and only when multiple targets compete for attention. ➔ See 'What have we learnt?', above.

Pinpoint question 8.5

What evidence is there in the Duncan (1980) experiment to suggest the existence of an information filter that occurs late on in processing, as Deutsch and Deutsch (1963) suggest?

No 'structural bottleneck' accounts of attention

All of the theories we have so far discussed in this chapter have assumed some kind of structural bottleneck in the processing architecture that constricts the amount of material that can be dealt with at once. If you remember our lift example earlier on in this chapter, this is analogous to how big the lift is and therefore how many passengers can fit in the lift at any one time (are we dealing with a dumb waiter here or an industrial shaft?). However, performance also critically depends on how fast the lift moves from floor to floor. We can therefore distinguish between models of attention that focus on structural constraints (akin to the number of passengers that the lift can take at any one time) and those that concentrate upon processing constraints (akin to the maximum speed at which the lift can travel). It is to these latter models that we now turn.

As Figure 8.3 reveals, a filter provides a so-called *structural constraint* and the difference between so-called early and late filter theories concerns where the central bottleneck is assumed to be located in the sequence of internal processes. Despite that for most of the latter half of the twentieth century, attentional researchers were preoccupied in trying to establish which of these two general theories is correct, many also commented that such a simple binary opposition failed to do justice to the complex nature of human information processing (see Allport, 1989, for review). For instance, if we accept the modularity view of mind (Fodor, 1983; see Chapter 2), in which the mind is composed of discrete compartments dealing with specific cognitive functions, then it is most appropriate to think in terms of multiple constraints distributed across the different modules. Indeed many possibilities suggest themselves, and we will consider some of the more plausible alternatives as the discussion proceeds.

For instance, some theorists have discussed accounts in which the notion of a filter is simply avoided altogether. One such example was discussed briefly by Hochberg (1970). The basic idea rests on assuming some form of analysis-by-synthesis account of processing (see Chapter 6). Put simply, any material that confirms an expectation of what the stimulus is, will be held in memory and will be recalled if necessary. Anything that does not confirm expectations is not encoded and will be lost. Unfortunately, this account was not very thoroughly spelt out, and given that it has provoked very little interest in the literature (but see Kahneman, 1973), we can move on to perhaps more useful ways of thinking.

One such way of thinking has been provided by Norman (1969), and a schematic representation of his ideas is provided in Figure 8.8. In this scheme every stimulus undergoes sensory encoding, and if the stimulus is familiar, then stored knowledge about the stimulus will be activated. For example, if we mention *American Psycho* then you might activate the fact that it was a book by Bret Easton Ellis, that the main character's name is Patrick Bateman, and when the original paperback was released it contained a Francis Bacon painting on the front cover which was replaced by a picture of Christian Bale when it got turned into a movie, etc. None of this will happen if you are unfamiliar with *American Psycho*.

However, to give proper justice to Norman's model, it is perhaps more appropriate to reconsider the notion of word detectors as discussed in the context of Broadbent and Gregory (1971; see Chapter 5). For simplicity's sake we have only discussed a word detector in abstract terms: it is a mechanism responsible for collecting evidence in favour of a particular familiar word – the word detector for NURSE will fire if 'NURSE', 'nUrSe', 'nurse', '*nurse*', etc. are encountered. In Norman's account, a more sophisticated notion of such a detector was implied. This mechanism is more than something like a mental light bulb that is switched on when a particular word is presented. On the contrary, in this scheme a given detector provides the entry point to all stored knowledge about the word. The implication is that when a detector becomes activated, then access to stored knowledge about the semantics of the stimulus is granted: for instance, access to the information that a nurse is a paramedic, typically encountered in a hospital setting, etc.

By this view the detector provides access to all knowledge about the stimulus that the participant possesses that is represented in long-term memory. In this way, and following sensory analysis, the long-term

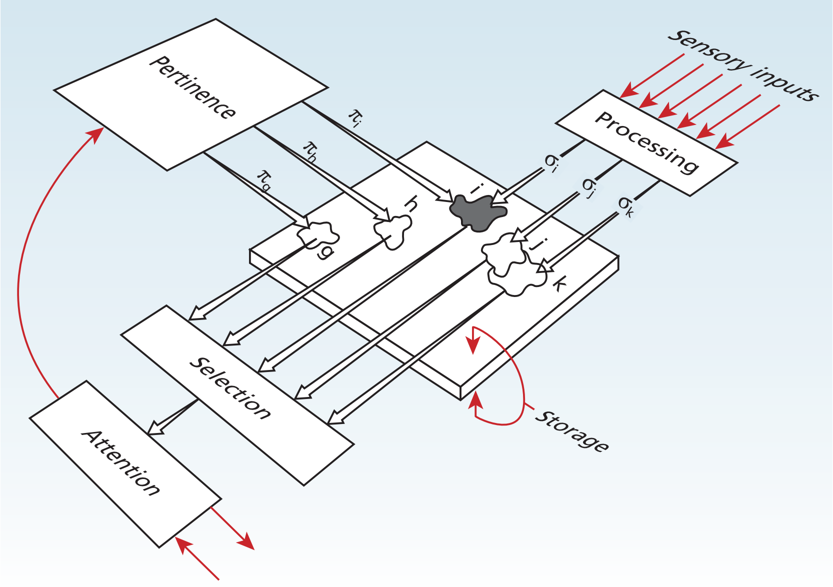

Figure 8.8 A simple arrows-and-boxes account of Norman's (1969) theory of attention
How attentional selection may operate in the absence of any structural constraints. 'Storage' refers to long-term memory (see text for further details).

Source: Norman, D. A. (1969). Towards a theory of memory and attention. *Psychological Review*, *75*, 522–536 (fig. 1, p. 526). Reproduced with permission from APA.

memory representation of a given stimulus is activated to a certain degree depending on what was actually presented and what was delivered by the sensory analysis. So whereas NURSE is the ideal stimulus for the NURSE detector, the detector could become partially activated by 'N*RSE', 'NARSE', 'NURS', and so on. To return to the railway station, the detector for 'your friend' would become partially activated by seeing someone wearing 'her' coat.

Moreover, the activation of such long-term representations was also assumed to reflect top-down influences. So for instance, activation could spread on the basis of contextual information: the LAMB detector becomes partially active by reading 'Mary had a little . . .' (or indeed 'The Silence of the . . .' if we are to continue our macabre literature theme). Back on the railway platform, you have a sensible expectation that your friend will be arriving off the next train, hence 'your friend's' detector will be primed or partially activated by this kind of top-down influence. In terms of the model, Norman discussed such activation in terms of what he called **pertinence**. This refers to the level of activation of any given stored stimulus representation (the 'my friend' detector will be partially activated – it will have a high pertinence value). Moreover, the pertinence of any given stimulus can change over time if the current state of affairs demands this. Swerve to avoid the pedestrian that steps off the kerb and then brake to avoid the on-coming traffic. The critical aspect, though, is that the current pertinence value reflects both the products of sensory analysis together with any top-down influences. In this way, the distinction between attention and memory is blurred because the processing constraints that exist in the model refer to operations concerning memory access and retrieval. Indeed, Norman (1969) happily drew comparisons between his account and analysis by synthesis (see Chapter 6).

In the current context, the critical point is that there is no structural bottleneck in this account. At any given moment items are selected for further processing on the basis of their being the most currently pertinent. In this way attention can be focused on certain items but it can also be shifted around, depending on how the pertinence values change. In discussing this account, Hirst (1986) cautioned against confusing it with the late selection account put forward by Deutsch and Deutsch (1963). Remember that according to that account all items are fully analysed; in contrast, according to Norman, how thoroughly an item is processed is determined by its current level of pertinence. All items do make contact with their long-term memory representations, but the level of analysis of any particular item is determined by its current pertinence value.

pertinence According to Norman (1969), the level of activation of any given stored stimulus representation calculated as a combination of sensory analysis and top-down influence.

The notion of attentional resources

Consideration of the models of Deutsch and Deutsch and of Norman forces us to think of alternatives of how best to construe attentional constraints outside the confines of the early filter theory. Perhaps the most radically different sorts of ideas can be found in the context of discussions of **attentional resources**. Here, we are moving away from issues about structural constraints – where to place a filter – towards a consideration of processing constraints. For instance, quite different views of performance limitations are now forthcoming and these are not best conveyed in terms of attempting to squeeze too much information down a processing bottleneck. Whereas bottleneck accounts of processing constraints assume that the central limitation is structural in nature – only so much information from the input can be forced through the bottleneck at any one time – resource theories take a quite different view. It is to these that we now turn.

A single pool of resources?

Perhaps the simplest analogy here is with the idea of a basic electric circuit connected up to a 12-volt battery (see Figure 8.9). The circuit has a finite electrical resource fixed at the upper bound of the current dissipated by the battery. If the circuit has one light bulb attached it will shine brightly because all of the resources in the circuit are devoted to powering this one bulb. If other light bulbs are all connected up to the same circuit, and they demand less than the upper bound of the available current, then they will all shine brightly.

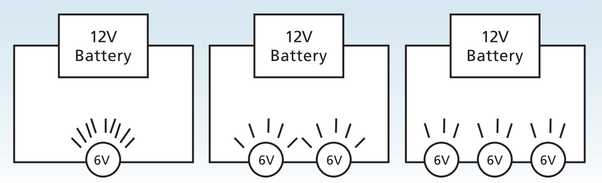

Figure 8.9 Mental resources by analogy with a simple electric circuit
A simple electric circuit illustrating how a fixed resource (the current provided by a battery) can be depleted as more components are added to the system – as more lights are added to the circuit each shines a little less brightly.

However, if more and more light bulbs are added then they all shine, but dimly. The light bulbs are demanding more from the battery than it can push out. The current provided by the battery is a fixed resource and adding components to the circuit can result in overload. At the limit none of the components in the circuit will operate properly.

A brief introduction to dual-task methodology

Here part of the analogy can be fleshed out in terms of cognitive psychology and the common-sense belief that it can be quite demanding to try to do more than one thing at once. More formally this sort of issue has been examined in terms of dual-task experiments. A dual-task experiment takes the following form:

- Stage 1 – have participants carry out one task (task A) and measure how well they do at this task in isolation.
- Stage 2 – do the same for a second task (task B) in isolation.
- Stage 3 – now have participants complete both tasks A and B at the same time.

The typical expectation is that participants will perform less well at either or both tasks when they are combined than when they are completed in isolation. Such a result is known as a **dual-task decrement**. It is not a caricature of single resource theory to state that it assumes that the whole of cognition places demands on a single pool of resources, in that the completion of any two concurrent tasks may cause problems – yes, walking and chewing gum at the same time may interfere with one another! However, stated in such a way, it is perhaps not surprising that some have found this to be a genuinely implausible description of the limits placed on human cognitive performance. Given this, we will consider the alternatives in more detail shortly.

Single resource accounts and the dual-task decrement

By a single resource account, the dual-task decrement arises because task A and task B compete for the same 'pool of resources' – there is only one 'battery' in the system. Of course the absence of a dual-task decrement can also be explained by single resource theory on the assumption that the combination of tasks does not exceed the upper bound on the available resources so both tasks can be completed without interference – there's plenty of juice in the battery to go round. So evidence for a single resource model can be derived from dual-task experiments. To flesh these ideas out, it is useful to concentrate on the most famous single resource model as described by Kahneman (1973).

Kahneman's model is shown in schematic form in Figure 8.10. The central premise in the account is that

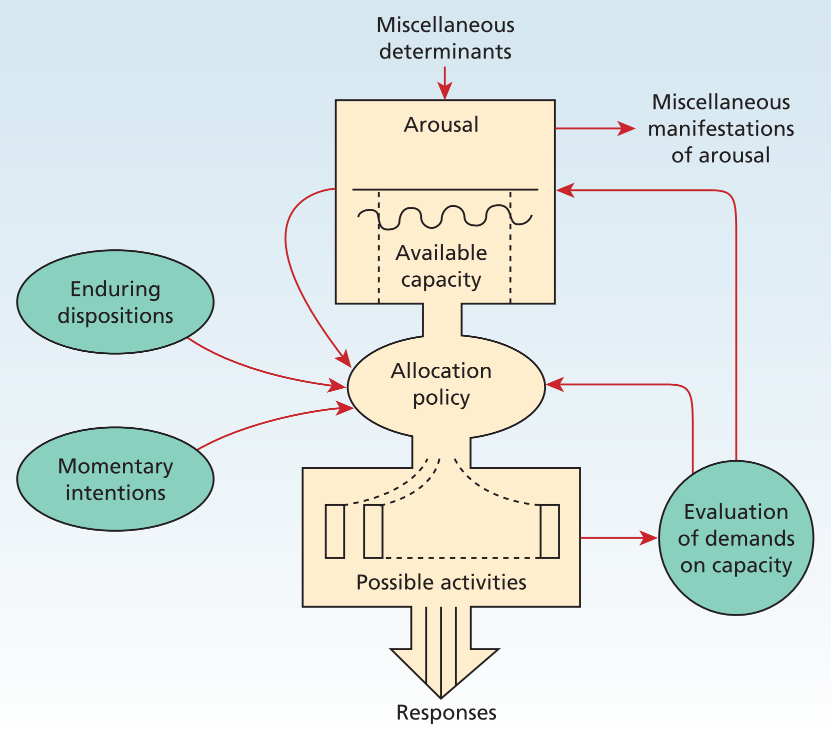

Figure 8.10 The arrows-and-boxes account of the human information processing system as set out by Kahneman (1973)

Source: Kahneman, D. (1973). *Attention and effort* (fig. 1.2, p. 10). Englewood Cliffs, New Jersey: Prentice Hall. Reproduced with permission.

Research focus 8.2

Patting my head and rubbing my belly: can I really do two things at once?

Patting the head while concurrently rubbing the stomach are two tasks whose joint mastery seems to be reserved only for those people with far too much time on their hands. Franz, Zelaznik, Swinnen and Walter (2001) referred to this special kind of dual-task interference as 'spatial coupling' which is observed when we try to get our hands to do two different things at once: typically, what we find is our head patting becomes a little more like rubbing and our stomach rubbing becomes a little more like patting. Franz et al. (2001) wondered whether dual-task interference could be reduced if these two tasks were somehow reconfigured into a single task.

Eight participants were asked to draw either the top or bottom half of a circle with their left or right hand. In uni-manual conditions, participants only used one hand to draw one half of the circle. In bi-manual conditions, participants drew with both hands, either two top halves, two bottom halves, or one top and one bottom half. Critically, these bi-manual conditions were categorised into cases where the two parts of the drawing completed a circle (top–bottom) and cases where the position of the two parts were reversed (bottom–top condition; like a letter u sitting on a letter n – see Figure 8.11).

Franz et al. (2001) reported that most participants found the bi-manual bottom–top condition to be the most difficult, with the top–bottom condition being easier because participants were able to bear in mind a circle. These qualitative comments were supported by quantitative analyses, in which a measurement of the spatial disruption of the bottom–top condition was larger than the top–bottom condition.

This study shows that there are both motor and cognitive constraints that influence dual-task performance. In the above example, the motor difficulty in performing bi-manual conditions may be alleviated by recasting the components of the task into a single image. The dual task essentially becomes a single task. The authors speculated about how this finding might extend to other examples of multi-tasking. For example, they stated that holding in mind two supposedly unrelated ideas is difficult but made easier if you are able to bridge the two ideas with a single concept. Therefore, perhaps patting your head and rubbing your belly might be difficult, but prubbing your helly could be less so.

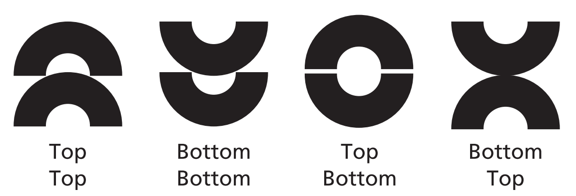

Figure 8.11 Schematics of the target figures that participants were expected to produce in the study by Franz et al. (2001)

Source: Franz, E. A., Zelaznik, H. N., Swinnen, S., & Walter, C. (2001). Spatial conceptual influences on the coordination of bimanual actions: When a dual task becomes a single task. *Journal of Motor Behavior*, *33*, 103–112.

there exists an upper bound on the capacity to complete so-called mental work – there is only so much attention that can be devoted to any given task at a time. Using the work analogy, 'paying attention' in this account is akin to 'exerting effort' or 'investing capacity' (Kahneman, 1973, p. 8). The schematic in Figure 8.10 resembles the familiar arrows-and-boxes diagram, but as you can see, there appears to be no way in to the model. In this regard, it is very much an abstract description of the internal workings of the mind. Central to the model is the notion of arousal. At the limit, you are more likely to devote attention to the televised football game if your team is playing in the Cup Final than if you are drowsing in front of a match between two teams you have never heard of. In general terms, the allocation of attention will depend partially on your current state of arousal.

Kahneman (1973) reviewed much of the evidence in support of such a suggestion and noted that it is well established that both too little and too much arousal can have detrimental impacts on performance. Under-arousal is generally explained in terms of lack of motivation at a task, whereas over-arousal will result in the participant being unable to discriminate between relevant and irrelevant aspects of the task. Over-arousal can be compared to that initial rush you might get when they open the doors to the January sales: 'A teapot descaler! A tweed jacket! A nose hair remover! Wait, I only came in for an egg timer.' However, Kahneman went on to make a more general point, namely that, as a task becomes more difficult, it may be possible to increase the supply of resources because of the operation of physiological arousal mechanisms. In this way arousal may be able to compensate for the increased task difficulty. However, when task difficulty becomes immense, any compensation due to arousal may not be sufficient to meet the task demands.

Independently of arousal, two other components are also assumed to influence the deployment of attentional resources – an *allocation policy* and a further *control mechanism* that evaluates the current demands on capacity. Failures to evaluate the current demands properly are typically associated with under-arousal and we will have more to say about the allocation policy in discussing dual-task decrements later.

Appraisal of single resource theories

Although the presence of a dual-task decrement fits comfortably with the idea of a fixed pool of common processing resources, Navon and colleagues in a series of papers (Navon, 1984, 1985, 1990; Gopher & Navon, 1980; Navon & Gopher, 1979, 1980) have thoroughly picked away at this idea and have arrived at a quite different view of processing constraints. First, though, let us be clear about what it is that single resource models actually predict. Single resource models assume that concurrent tasks compete for the same fixed pool of resources and therefore there should exist a trade-off in performing the tasks when they are combined (Navon, 1990). If both tasks draw on the same pool of resources there should be a dual-task decrement when the tasks are combined. Essentially, if task A demands more resources to complete than task B, then whereas performance on task A may improve or remain steady, performance on task B should decline.

Okay, you are sitting writing a course essay (task B – let's face it, it's always task B) and are also listening to the radio (task A) at the same time. Now by all accounts writing the essay is the more demanding exercise. However, let's assume your favourite programme is now aired and they begin to play an interview with your favourite celebrity/footballer/fashion icon (or indeed 'your favourite celebrity, footballer, fashion icon'). To attend fully to the interview has the consequence of demanding more resources, hence your ability to write the essay dries up.

In mapping out such possibilities, Norman and Bobrow (1975) introduced the notion of a **performance operating characteristic** (**POC**; examples are shown in Figure 8.12). Although the functions shown in the figure are idealised, it is possible to trace out a POC in a number of ways. For example, participants could be required to perform task A or task B to a particular level of competence. This could be achieved by instructing participants to make responses within a certain response deadline or make sure that they make few mistakes on the task. Alternatively, they could be instructed to prioritise task A over task B: concentrate primarily on task A but ensure not to ignore task B completely. Another possibility is to see what effect an increase in the difficulty of task A has on the ability to perform task B concurrently. In such cases, the assumption is that, as more resources are devoted to task A, performance on task A will improve, but performance on task B should reveal a concomitant decrease. More generally, it is assumed that the more of a drag task A places on the resource pool, the less resources can be devoted to task B.

Norman and Bobrow (1976) encapsulated these ideas in something they defined as the principle of **complementarity**. As before, it is assumed that there will be a dual-task decrement when two single tasks are combined – performance will be worse in this case than when either task is carried out in isolation. However, more generally there will be a trade-off between the two tasks. In other words, the more you try to listen to the radio, the worse the essay writing will get, and alternatively the more you try to write the essay, the less you will take in from the radio. In more formal terms, according to the principle of complementarity there is a more challenging prediction of single resource models. This is that there will be a reciprocal interaction between the resource demands of task A and of task B. The nature of this reciprocal relationship can best be tackled by examining performance across a range of combinations of different variations in the relative difficulties of task A and task B, and this means tracing out a corresponding POC. Compare writing down 'the the the . . .' with attempting to write an essay. Compare listening to ambient music with listening to a conversation on the radio. Now think of all the possible combinations, derive some measures of performance and then compute a POC.

Evidence in support of the single resource account of performance has indeed been forthcoming in the literature. For instance, Gopher and Navon (1980) examined performance in something called a manual tracking experiment, in which participants had control

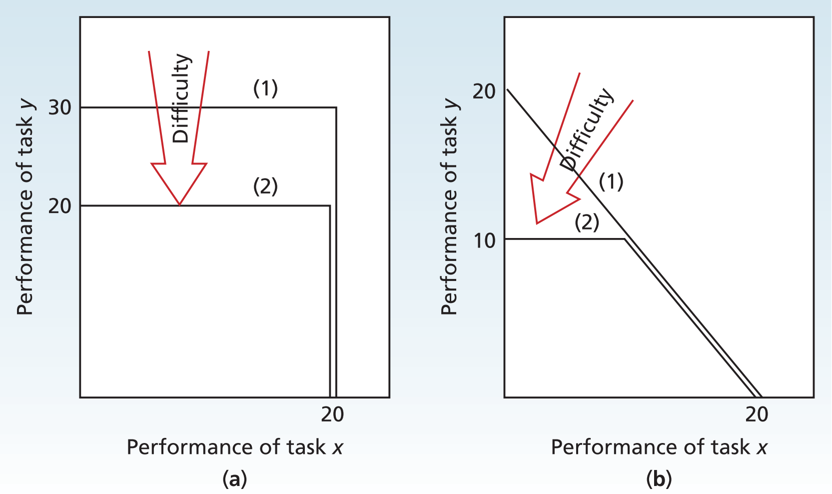

Figure 8.12 Examples of hypothetical performance operating characteristics (POCs)
The way to read these graphs is that large *y* values imply good performance on task *y* (i.e., task *y* is relatively easy). Likewise, large *x* values imply good performance on task *x* (i.e., task *x* is relatively easy). (a) shows the consequences of combining a more difficult task *y* with task *x*. As the difficulty of the task *y* increases, performance on *y* falls from a level of 30 (curve 1) to a level of 20 (curve 2) but there is no concomitant change in performance on task *x*. In other words, there are no common resources that task *y* and *x* draw upon. In (b) the situation is more complex. Curve 1 illustrates the principle of complementarity – there is a uniform trade-off in performance across the two tasks. A more difficult task *y* can only be dealt with in combination with a less difficult task *x*. Curve 2 shows the case where increases in task *y* difficulty can be tolerated up to a point (i.e., a value of 10 on the *y* axis) but after that there has to be compensation: task *x* must get easier. An implication is that there are sufficient resources to perform a relatively difficult task *y* up to a point, but further increases in difficulty do have implications for carrying out task *x*. In this case there is a mixture of resources that are shared by both tasks and resources that are unique to each task. These examples are taken from Navon and Gopher (1980), who consider even more complex cases of resource allocation.

Source: Navon, D., & Gopher, D. (1980). Task difficulty, resources, and dual-task performance. In R. S. Nickerson (Ed.), *Attention and performance VIII* (pp. 297–315, fig. 15.3 A and B, p. 309). Hillsdale, New Jersey: Erlbaum. Reproduced with permission.

over the movement of a visually presented *X* on a computer screen via the left/right and up/down movements of a games controller. The computer controlled the up/down and left/right movement of a target square boundary on the computer's screen and the aim was to try to ensure that the *X* remained within the square as the square moved about the screen. Here task A was defined as controlling the horizontal movements of the *X* and task B was defined as controlling the vertical movements of the *X*.

In one experiment Gopher and Navon systematically investigated what happened when the difficulty of control over the horizontal, and separately, the vertical movements of *X* were manipulated. Increasing task difficulty meant that slight movements of the controller greatly accelerated the moving *X*. What they found was that there was a trade-off between performing these tasks as control difficulty was systematically varied. Moreover this trade-off was clearly in line with the idea that both tasks placed demands on the same pool of resources. Given that both tasks involved trying to control the movement of a single object with the same hand, it is perhaps not so surprising that a single resource model provided an adequate fit to the data.

Despite the evidence consistent with single resource assumptions, Gopher and Navon reported other similar tracking experiments in which the results were not so supportive. In discussing their data, Gopher and Navon introduced a contrast between (i) single resource models, and (ii) models that assume task-specific or multiple resources. We shall turn to multiple resource models shortly, but more pressing is consideration of data that urge caution in trying to interpret POCs in general. The conclusion will be that we need to be very careful in making claims about attentional resources on the basis of POCs alone.

Pinpoint question 8.6

If participants have to complete two tasks which both make demands from the same pool of resources, what *two* predictions can we make about dual-task performance?

Resources and resource allocation in more detail

Navon (1990) ran a dual-task experiment in which task A involved the classification of a visually presented digit – participants had to respond as to whether the digit was odd or even. In task B participants had to respond to a concurrently visually presented Hebrew letter. They had to decide whether the letter came from the first or second half of the alphabet. Both tasks were carried out under reaction time (RT) conditions. The typical instructions to participants in such tasks are that they should try to respond as quickly as they can without making too many errors. Performance on the two tasks was also manipulated across two conditions.

In the *minimal requirements conditions*, participants were given continuous within-trial feedback on how well they were expected to perform on both tasks. Performance requirements were systematically varied within a range that the participants could easily achieve. Moreover, for each level assessed, slightly more weighting was to be attributed to one of the tasks over the other by setting the RT deadline at a slightly lower level for one of the tasks. What this means is that participants were pushed to respond within a certain time limit for one of the tasks. This was accomplished by shortening the RT deadline – either press the key within this deadline or the response will be treated as an error. In this way the hope was to achieve a performance trade-off for the two tasks. For example, as the RT deadline was lowered for task A, RTs on task B should lengthen.

In contrast, in the *optimum–maximum conditions*, participants were told to try to maximise their response speed on one task even though the difficulty of the other task was altered. Again, according to a single resource model, any variation in the difficulty of one task ought to produce a concomitant variation in performance in the other task – a more difficult task requires more resources. (Remember the *principle of complementarity*?) So even under the optimum–maximum requirements there ought to have been a trade-off in task performance.

A schematic representation of the data from the experiment is presented in Figure 8.13. Consider performance in the minimum requirements condition first. Here there is a clear performance trade-off between the two tasks such that improvement on one is associated with a decrement on the other. RTs on task A shorten as RTs on task B lengthen. Such a pattern is generally in line with single resource models. However, quite a different pattern emerged in the data for the optimum–maximum conditions: here there simply is no evidence of a similar performance trade-off. Now when participants attempted to maximise the speed of response on one task, performance was essentially unaffected by variation in the speed of response on the other task. So when maximising speed on task A, varying the speed of response on task B failed to produce any change in task A responses.

According to Navon (1990), this suggests that the trade-off seen in the minimum requirements condition apparently does not necessarily reflect any limit on being able to perform both tasks concurrently as described by single resource models. Rather, it seems that the pattern of performance reflects participants' compliance with what they consider to be the objectives of the experiment. The more general and important conclusion is that participants appear to have more control over dual-task performance than is implied by the single constraint defined by a finite pool of attentional resources. A performance trade-off may reflect nothing more than participants being able to perform in a manner that is asked of them. It therefore does not necessarily imply anything of significance about the nature and availability of attentional resources.

Indeed, Navon (1990) went further and argued that such demonstrations as these can undermine the faith that may be put in single resource models: merely demonstrating a performance trade-off in a dual-task experiment does not, in and of itself, indicate the constraints that apply when only a single pool of attentional resources is available. He went further, though, and in a detailed review began to question the general usefulness of thinking solely in terms of attentional resources. In some cases it might be better to think not in terms of scarcity of some form of limited resource but in terms of other forms of task interference: attentional resources provide just one way for thinking about how two concurrent tasks may interfere with one another. Evidence in support of this idea comes from a dual-task experiment reported by Navon and Miller (1987) (see Figure 8.14 for more details).

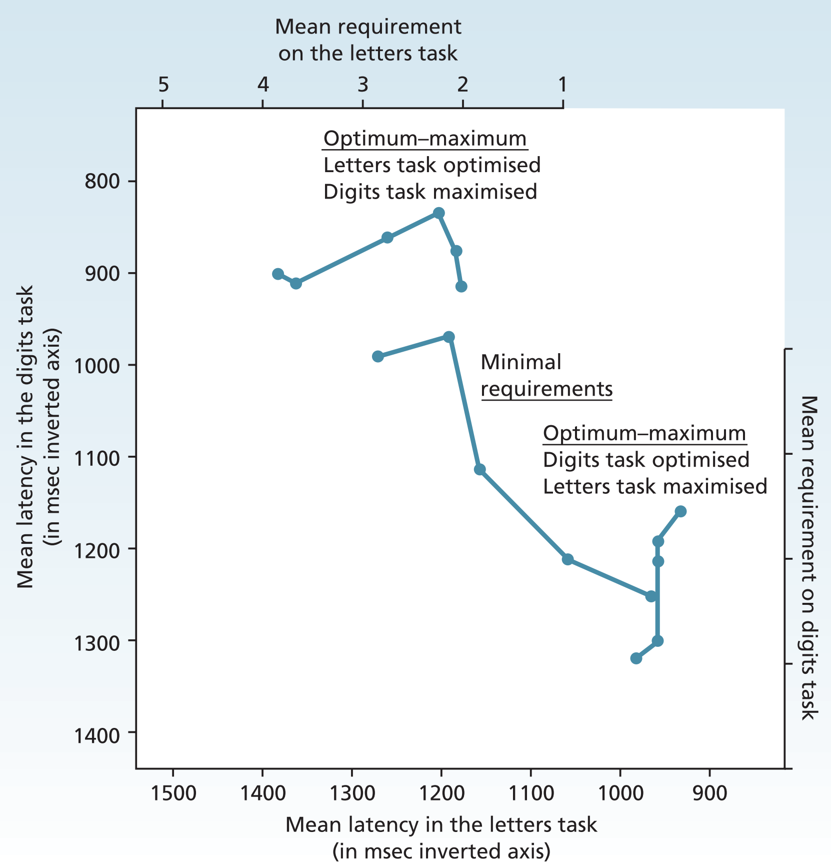

Figure 8.13 Data from Navon (1990)
In the minimum requirement task, that is, in the absence of any requirements to weight either task differentially, there is a clear performance trade-off. To meet a response deadline on one of the tasks, RTs on the other task suffered. In the optimum–maximum conditions, participants were instructed to maximise the speed of response on one task. Even though task difficulty was varied on the other task, there was no evidence of any forms of performance trade-offs (see text for further details).

Source: Navon, D. (1990). Do people allocate limited processing resources among concurrent activities? In L. Green, & J. H. Kagel (Eds.), *Advances in behavioral economics. Volume 2* (pp. 209–225, fig. 3, p. 219). Norwood, New Jersey: Ablex. Reproduced with permission of the Greenwood Publishing Group. Inc.

Attentional resources or something else?

Let us take a simplified case and consider just those trials in which four words are presented concurrently on the screen of a computer. One pair was vertically aligned and one pair was horizontally aligned at the end points of an invisible '+'. Task A was to judge whether a boy's name was presented in the vertical limb. Task B was to judge whether a city name was presented on the horizontal limb. Performance was assessed in each task separately (task A *or* task B), and in a dual-task situation in which judgements about both limbs had to be made (task A *and* task B). Five types of word stimuli were defined:

1. An *on-channel target* – a city name (e.g., 'London') on the horizontal city limb or a boy's name (e.g., 'George') on the vertical boy's limb.
2. An *on-channel associate* – a city-related word (e.g., 'England') on the city limb or a boy's name related word (e.g., 'Jane') on the boy's limb.
3. An *off-channel target* – a city name on the boy's limb or a boy's name on the city limb.
4. An *off-channel associate* – a city-related word on the boy's limb or a boy's name related word on the city limb.
5. A *neutral word* unrelated to both city names and boys' names.

For ease of exposition the words defined in 1–4 are known here as *response-related words*. Each limb could contain either no response-related words or only one response-related word.

Various predictions were set out. First, consider those concerning performance in the single-task case. On the assumption that participants could selectively attend to the designated pair of items (on a given attended limb), then it was assumed that there should be no effect of the type of item included on the other limb. This prediction was borne out by the data. So if you are only doing the boy's name task, in isolation, then you are unaffected by anything that occurs on the

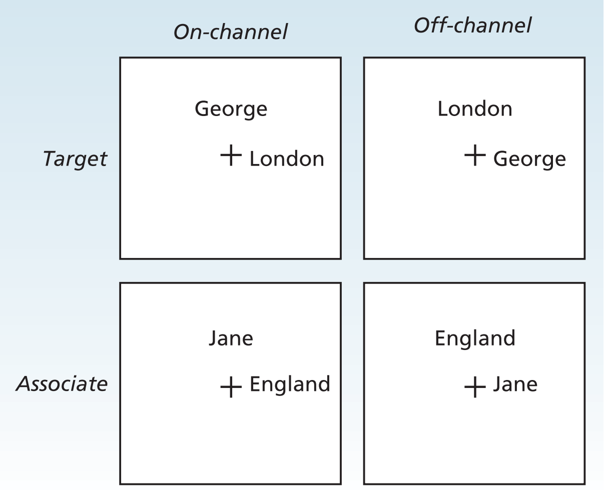

Figure 8.14 Examples of some of the stimuli in Navon and Miller's (1987) dual-task experiment
One task was to judge whether a boy's name was presented in the vertical limb and another task was to judge whether a city name was presented on the horizontal limb. As a result, on-channel targets were city names on the horizontal limb or boys' names on the vertical limb. On-channel associates were city-related names on the horizontal limb or boy-related names on the vertical limb. Off-channel targets were city names on the vertical limb or boys' names on the horizontal limb. Off-channel associates were city-related names on the vertical limb or boy-related names on the horizontal limb. Neutral words were also used, but are not shown here. In these cases displays of display size 2 are shown; discussion in the text focuses on cases where displays of display size 4 were used.

other horizontal limb. More involved, though, are the predictions concerning cases when a response-related word was included in the attended limb. Part of the rationale was based on the assumption that task difficulty relative to a neutral word would be increased in the presence of an on-channel associate. For example, the mere presence of 'Jane' on the vertical boy's limb should slow down the boy's name judgement. Again the data supported this prediction: participants were impaired in responding when an on-channel associate was present on the attended limb. Finally, items relevant to the other task should behave exactly as other neutral words even when included on the attended limb. Again the data were generally in line with this prediction.

More critical, though, is what happened in the dual-task situation. First it was predicted that there would be a dual-task decrement, and indeed overall it was found that average RTs were twice as slow in the dual- than the single-task cases. Moreover, responses on both tasks were affected by the presence of any response-relevant item. Indeed the most interesting effects concerned cases where items relevant to one task occurred on the wrong limb. Performance was particularly slow if, say, 'George' appeared on the city limb or 'London' appeared on the boy's limb. Performance was also affected when an associate relevant to the other task was presented on the wrong channel. In contrast to single-task performance, the presence of, say, 'Jane' on the city limb dramatically slowed performance.

Outcome conflict, not competition for a common resource

The conclusions that Navon and Miller drew from these results do not fit particularly comfortably with the notion of attentional resources. Something else must be going on in the experiments than mere competition for some form of common resource. What the data suggest is something quite different. In having to analyse the semantic content of both channels and make responses accordingly, performance reflected a particular type of task interference known as **outcome conflict**. As Navon and Miller (1987) stated, outcome conflict arises in dual-task situations when 'the outcome of the processing required for one task conflicts with the processing required for the other task' (p. 435).

In the context of their experiment, and from the participant's point of view, knowledge about boys' names and knowledge about city names is most pertinent to the tasks. The aims are (i) to try to avoid responding to any item that is associated with either of the target categories, so don't be fooled into responding to channel associates, and (ii) certainly to not respond to a target present on the wrong limb. Point (i) is made difficult by the fact that both targets ('George') and their associates ('Jane') will activate the corresponding mental category (boys' names). Point (ii) is made difficult by the fact that either target can appear on any limb but a target is only valid when it occurs on its particular limb. Any such disruptive effects can be gauged relative to neutral trials because neutral words will not activate either of the target categories.

It is as if the participant has to untangle and keep track of appropriate from inappropriate meanings invoked by the displayed words. The problem arises in trying to monitor semantic activation produced by the different response-related words in their conflicting categories. The data cannot be explained just in terms of scarcity of some fixed resource because the neutral words will also activate their relevant categories, yet these do not produce the interference found with the response-relevant items. For example, 'horse', on either the city or the boy's name limb would activate the ANIMAL category, but this category is completely irrelevant to either task. By this view the same amount of semantic activation is produced on all trials when the display contains the same number of words – every word produces semantic activation and as a consequence the same amount of this kind of word-processing resource is being used up on every trial. However, the response-relevant words produce an additional outcome conflict that is not produced when a neutral word is presented. Something other than competition for a common pool of resources is clearly going on.

Navon and Miller went on to discuss several sorts of outcome conflicts and in general they defined them as arising when a particular task produces 'outputs, throughputs, or side effects' that are harmful to the execution of another concurrent task (p. 435). The basic idea is that despite the fact that claims about attentional resources abound, there are other ways of construing processing limitations. Indeed, given the complexity of human cognition, it is perhaps only sensible to admit that there may be many different sorts of processing constraints that are quite different in nature to competition for a single scarce mental resource.

Pinpoint question 8.7

Why did the presentation of neutral words fail to impact upon resource allocation in the Navon and Miller (1987) study?

For example . . .

Despite the somewhat constrictive nature of the tasks that cognitive psychologists ask participants to perform in examining dual-task performance, real-life examples are all around us. Take the highly enjoyable case of taking notes in a lecture as an example of a real-life dual task. As Piolat, Olive and Kellogg (2005) pointed out, there are lots of problems faced by our intrepid note-taker. For example, there is a discrepancy between the average speed of writing and the average speed of talking which means you're always lagging behind the lecturer. More than that, you're having to monitor the speech continuously while writing so that you don't miss the thread of the argument. And then, of course, you've got to decide what's worth writing down and what isn't. It's a tricky business.

In detailing exactly how tricky note-taking is, Piolat et al. (2005) summarised a number of studies in which note-taking was compared with a number of other procedures such as playing chess and text copying, in a dual-task paradigm. As we have mentioned before, a dual-task decrement would be observed if participants were less able to complete both tasks at the same time, relative to only one task at a time. Under these conditions, it was shown that the dual-task decrement for note-taking was larger than the simple copying of text and equivalent to playing chess. And we all know how taxing checkmating can be. In the cases of both note-taking and chess, Piolat et al. (2005, p. 299) stated that such activities demand the 'retrieval of large amounts of knowledge, conceptual planning, and the development of solutions to the problems posed in each situation'.

While it may seem like an incredible feat that we can note-take at all, there are always ways to improve this rather handy skill. For example, Piolat et al. (2005) stated that students learn more when they do not try to record everything in a linear fashion. Rather, identifying the connections between concepts and producing mental maps facilitates learning. At a very basic level, note-taking acts essentially as a memory dump, allowing you to save and retrieve information at a later date. If you're passionate about the subject matter, then it is going to be a lot easier for you to learn – which is something you should bear in mind as you're reading this book!

Source: Piolat, A., Olive, T., & Kellogg, R. T. (2005). Cognitive effort during note taking. *Applied Cognitive Psychology*, *19*, 291–312.

attentional resources The equivalent of mental electricity. A finite amount of mental energy that must be distributed among competing stimuli and tasks.

dual-task decrement A reduction in performance when performing two tasks together, relative to performing the same two tasks separately.

performance operating characteristic (POC) A function relating performance on two tasks when the two tasks are combined. The POC functions can be traced out as the difficulty of the tasks is varied or as participants vary their commitment to either or both tasks.

complementarity The reciprocal interaction between the resource demands of two tasks. If the demands of one task increase, performance on a competing task suffers.

outcome conflict Conflicts that arise in cases where competing stimuli are present in a display and it is difficult to resolve which to respond to.

Multiple resources?

Although it is possible to come away from reading Navon's papers with the conclusion that the concept of an attentional resource is both poorly specified and of little use, he is generally sympathetic to the idea that different tasks may place demands on different pools of resources (see Navon & Gopher, 1979). By this view, different processing mechanisms have their own processing characteristics and capacity limitations. An early example of such a view was put forward by Treisman (1969), by which sensory analysis is seen to proceed in terms of the operation of *independent input analysers*. (The analogy is not exact, but in simple terms it is possible to think of the analysers in much the same way as Fodor (1983) discussed *input modules* – see Chapter 2. Indeed, the notion is also reminiscent of the *perceptual nets* taken from Turvey's concurrent and contingent model of processing – see Chapter 4).

In this scheme, different analysers (different input channels) operate in parallel with one another but processing within an analyser is serial in nature. If we take these different analysers to represent the different senses, then an experiment reported by Treisman and Davies (1973) is highly relevant. They ran an experiment in which participants were asked to monitor information streams presented to the auditory and the visual modalities. Two visual streams were defined relative to two separate visual screen locations and two auditory streams were defined relative to the left and right speakers, respectively, of a pair of headphones. On each trial, participants were required to monitor for the presence of a given target. In a so-called *surface condition*, the target was the syllable 'end' embedded in a word (e.g., 'Br<u>END</u>a'). So in the surface condition, target detection depended on an analysis of the physical or surface characteristics of the target (its written or spoken form). In contrast, in a so-called *semantic condition*, animal names acted as targets (e.g., 'baboon'). Target detection in this condition now depended on accessing word meanings. Overall, the patterns of performance were the same across these two conditions as gauged by measures of target detection accuracy.

Generally there was a dual-task decrement: participants performed worse when monitoring two streams than just a single stream. More importantly, though, participants were particularly poor when monitoring two streams within a given modality than when they were monitoring one stream in each modality. They were more likely to make errors when concentrating on both visual streams, or both auditory streams, than when they divided attention across one visual and one auditory stream.

Clearly such a pattern of results fits comfortably with the idea that two tasks within the same modality may compete for the same resources but concurrent tasks in different modalities do not. However, Kahneman (1973) urged some caution here. Even though participants performed better in the cross-modality than the within-modality condition, there still was a dual-task decrement relative to when participants were able to concentrate on a single task. So while there are resource implications for carrying out two tasks within the same modality, there is also a dual-task decrement when two tasks are carried out across different modalities relative to the completion of a single task.

Research focus 8.3

'Sorry, I can't speak now, I'm in the hospital': mobile phone use and driving as dual task

While, for some of us, walking and chewing gum may seem like a daunting dual task, others are under the illusion that their multi-tasking abilities know no bounds. Unfortunately, such multi-tasking can take place in environments that pose serious risks, not only for that individual but also for those around them. One fairly contentious example is attempting to hold a conversation via a non-hands-free mobile phone while driving (something that is now illegal – at least in the UK). Support for this legislation comes from Strayer, Drews and Johnston (2003) who investigated how driving performance changes as a result of attempting to hold a conversation over a mobile phone. Here is a case where we have a very real example of a dual task that demands the division of attentional resources between vision (driving) and audition (talking).

Clearly as a result of the ethical issues involved in trying to run an experiment of this nature in a real-world setting, Strayer et al. (2003) utilised a high-fidelity driving simulator in order to test participants' driving performance under single-task (i.e., just driving) and dual-task (i.e., driving and talking) conditions. Importantly, the researchers also used a hands-free mobile phone, thereby ensuring that any dual-task interference was due to the act of communication rather than the motor behaviour associated with the use of a mobile phone.

Strayer et al. (2003) found that the dual-task decrement associated with mobile phone use while driving was most prominent during high-density traffic. Participants were involved in more accidents, slower in terms of braking onset and braking offset and they also took longer to reach a required minimum speed in the dual-task condition relative to the single-task condition. Nattering participants also took less notice of billboards while driving, correctly recognising fewer billboards as 'old' after completing the dual task relative to the single task.

Under safe and controlled conditions, these researchers have demonstrated that this kind of dual-task interference has important – potentially life-threatening – consequences. Moreover, the study does not provide support for the argument that using a hands-free mobile phone eliminates dual-task interference and is somehow a safe option. Maybe some of us should just stick to walking and chewing gum after all.

Source: Strayer, D. L., Drews, F. A., & Johnston, W. A. (2003). Cell phone-induced failures of visual attention during simulated driving. *Journal of Experimental Psychology: Applied*, *9*, 23–32.

When doing two things at once is as easy as doing either alone

We have already seen (in Research focus 8.2) how doing two things at once might become easier if they are seen as one task. Perhaps more impressive, though, are the dual-task studies that reveal how proficient participants can become after intensive practice. To take just three examples: (i) Hirst, Spelke, Reaves, Chaharack and Neisser (1980) taught participants to read prose while simultaneously writing single words and short sentences from dictation; (ii) Allport, Antonis and Reynolds (1972) practised participants at piano playing from a score while shadowing an auditory message; and (iii) Shaffer (1975) combined copy typing and shadowing. In all cases and after extensive practice (e.g., after six weeks in the Hirst et al. study) participants were shown to have improved dramatically and became competent in the dual-task situation. The impression is that participants became as efficient at doing both tasks at once as they were at doing either task alone.

In reviewing these particular studies, though, Broadbent (1982) was particularly sceptical that the data were as convincing as they might appear on a superficial reading. He pointed out that even though performance did improve dramatically over the course of the training sessions, there still remained some evidence of dual-task interference even after extended practice. The issue boils down to whether the data provide cast-iron evidence for time-sharing, that is, the ability to complete two tasks at the same time by being able to divide attention between two concurrently demanding tasks. In line with the notion of filter switching,

For example . . .

If you've ever trained to become a majorette (if not, what are you waiting for?), you'll notice that your baton twirling becomes more and more fluid until you hardly notice that you're doing it at all. Therefore, it's not as though the finger movements demand the same amount of cognitive effort as you practise; rather, it seems that the skill becomes easier and actually demands less and less resources the better you get. So if anything, the suggestion is that one characteristic of skilled performance is that it depends on little or no attentional resources.

Broadbent was convinced that the data instead reflect strategies in task switching and scheduling that previously had not been thought possible. According to him, the data do not provide compelling evidence for time-sharing and for the claims about being able to divide attention in dual-task situations.

Regardless of what the 'right' conclusion is to draw from these studies (see Hirst, 1986, for a riposte) the dramatic improvements in performance are difficult to square with the notion of a single and fixed pool of attentional resources. This whole idea becomes untenable if the suggestion is that practice increases the pool of resources. As Hirst (1986) argued, the studies that do show effects of practice, are probably better interpreted in terms of skill acquisition than anything about the allocation of attentional resources.

In the absence of extensive practice (and indeed in some situations, despite extensive practice), it just is the case that some pairs of tasks will interfere more with one another than other pairs of tasks (cf. Treisman & Davies, 1973). In addressing the possible reasons for this, Wickens (1984) set out a three-dimensional space of possibilities within which multiple resources may be defined. The dimensions reflect (i) early vs. late processes, (ii) separate modalities and finally (iii) 'processing codes' (p. 302) (see Figure 8.15).

In (i) the division is between perceptual encoding/categorisation and response selection and execution. The implication here is that any two tasks that place demands on the same stage of processing are likely to interfere with one another. For example, attempting to turn down the volume on the TV via the remote control and pick up the ringing phone places demands on the same response components. In (ii) the implication is that tasks performed in different modalities are less likely to interfere than are tasks performed within the same modality (cf. Treisman & Davis, 1973). This may go some way to explain how people can continue to shop in a supermarket while continuing a conversation over a mobile phone. Finally the implication in (iii) is that tasks that give rise to different internal codes are less likely to interfere than tasks that generate the same sorts of codes. You might like to test this idea out by trying to complete a Sudoku puzzle at the same time as some long multiplication problems, and then doing Sudoku and a crossword. All other things being equal,

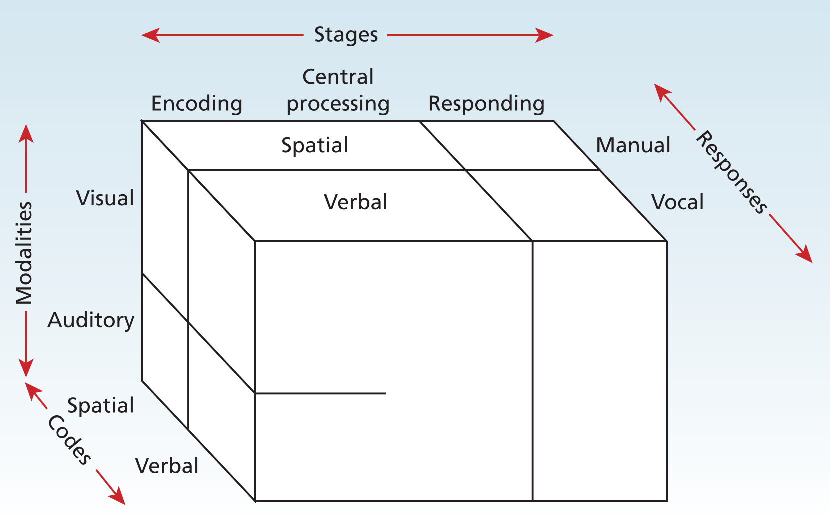

Figure 8.15 Wickens' (1984) framework for thinking about cases under which competing tasks may interfere with one another The 3D space is given by dimensions defined relative to (i) the stage in the human information processing system where conflicts may arise, (ii) the modalities in which the tasks are presented and (iii) the mental codes that are called upon by the tasks.

Source: Wickens, C. D. (1984). *Engineering psychology and human performance* (fig. 8.7, p. 302). Columbus, Ohio: Charles E. Merrill.

the Sudoku plus crossword condition should be easier because you're dealing with numbers and letters, relative to dealing with only numbers when trying to complete the first pair of puzzles.

We have discussed some of the evidence that strengthens the plausibility of this framework and Wickens (1984) discussed more. It is without doubt, though, that Wickens has provided an entirely sensible framework for thinking about when and how concurrent tasks may interfere. However, it does not settle the deeper issue about whether there is a central processing constraint on attempting to do more than one thing at once. This is a very important question that has given rise to a rather extensive literature and we shall discuss this possibility in more detail in the next chapter.

Pinpoint question 8.8

According to Wickens (1984), what are the three ways in which the interference between two tasks can be reduced?

Pulling it all together

In attempting to provide an integrative overview of the various kinds of theories of attention that have been considered in the literature, Pashler (1998) put forward the scheme shown in Figure 8.16. Here, the space of possible theories is defined by a 2 × 2 matrix whose dimensions define, respectively, (i) whether the theory allows for the possibility that unattended stimuli are identified and (ii) whether the theory allows for the possibility that multiple attended stimuli can be processed simultaneously. We have already considered: (i) early selection accounts in which multiple attended stimuli are not processed simultaneously and unattended stimuli are not identified, such as Broadbent's (1958) filter theory; and (ii) late selection accounts in which multiple attended stimuli are processed simultaneously and unattended stimuli are identified, such as Deutsch and Deutsch (1963). However, two cells of the matrix have yet to be discussed.

The '???' in the figure is taken to signify accounts in which all items are slavishly identified but in a sequential (one-at-a-time) manner. This would be like standing by a conveyor belt as freshly picked tomatoes are funnelled past one at a time: your job is to try to grade each for size and quality. Pashler termed this sort of account an **uncontrolled serial model** in which perceptual analysis is said to proceed serially but exhaustively – you can't leave the conveyor belt until all the tomatoes have gone past. In this sort of functional account there is no notion of selection, but processing is constrained to operate in a serial fashion. Where does such a strange notion come from?

The idea of **serial exhaustive processing** is most readily associated with the work of Sternberg (1967). In the now classic **Sternberg memory scanning task** participants are provided at the beginning of a trial with a **memory set** of characters, for instance the digits 3 and 7 presented on a visual display. The memory

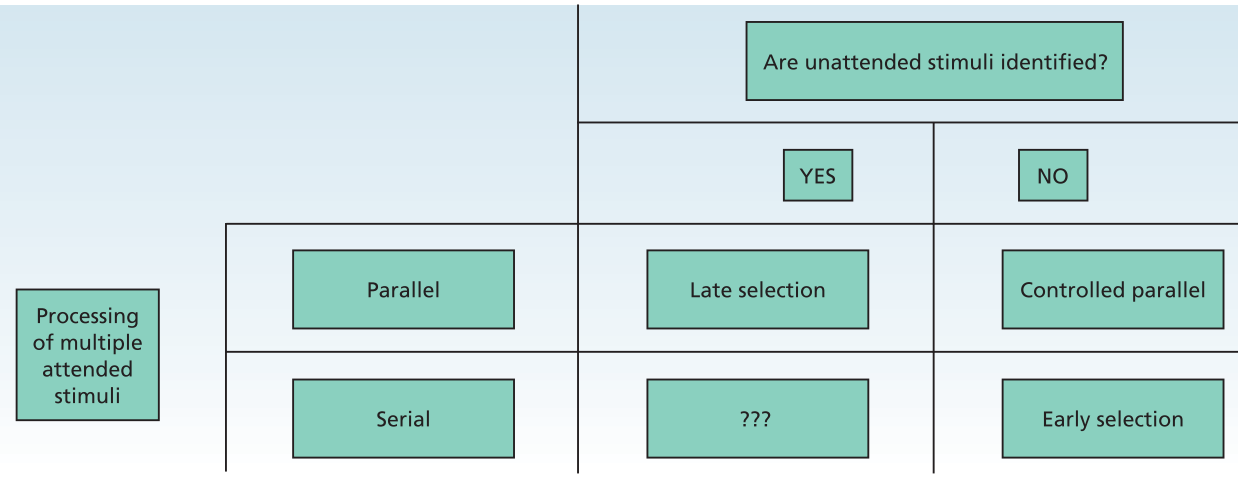

Figure 8.16 The 2 × 2 space of possible theories of attention (taken from Pashler, 1998)

Source: Pashler, H. E. (1998). *The psychology of attention* (fig. 1.6, p. 22). Cambridge, Massachusetts: The MIT Press. Copyright © 1998 Massachusetts Institute of Technology. Reproduced with permission of the MIT Press.

set is removed and replaced with a single **probe item** (a digit) and the participant has to press one response key if the probe is a member of the memory set (the participant must make a positive response) and another key if the digit is not (the participant must make a negative response). The experiments are run under RT conditions so response speed is the main measure of interest but accuracy is also recorded. Performance is measured for differing memory set sizes and the resulting averaged RT data are then plotted as a function of memory set size (see Figure 8.17 for an idealised example). As the figure shows, a central result is that RT increases linearly as memory set size increases but more particularly the rate of increase is the same for positive and negative responses. The standard theoretical analysis of the task is the following.

When the participants are provided with the memory set, they encode and store these in some form of temporary memory system. These items must be stored until the probe item is presented. The probe is then

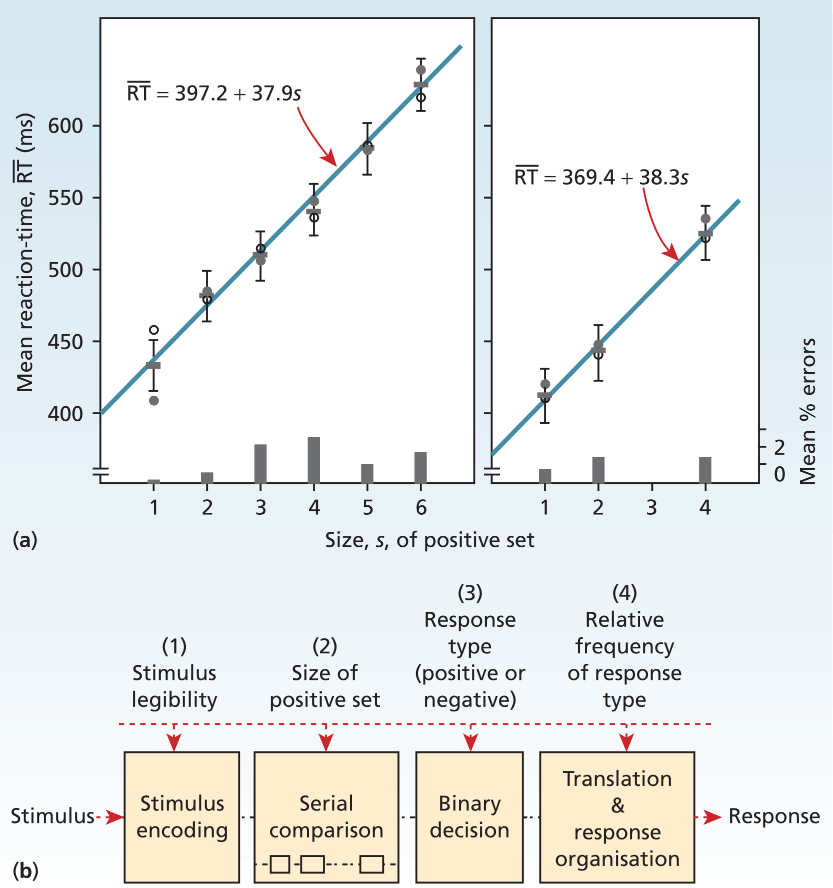

Figure 8.17 Sternberg's classic mental scanning experimental paradigm
(a) Shows typical data taken from a Sternberg memory scanning experiment. On the left, performance with memory sets that varied from 1 to 6 items – a different memory set was used prior to each trial. On the right, performance with a memory set (consisting of 1, 2, 3 or 4 items) that was kept constant over many trials. A critical finding is that response time increases linearly with memory set size – every added item to the memory set size incurs a constant increment in response time. Furthermore, identical functions are shown for present and absent responses. That is, on average a present response takes the same amount of time to make as an absent response (i.e., the data points overlap). Taken together, these two findings support the notion of serial exhaustive processing – don't make a response until you have considered every item in the memory set. (b) is a schematic arrows-and-boxes representation of the component stages of processing that Sternberg put forward for the memory scanning task. Of some additional interest is the fact that, over a series of experiments, he was able to infer whereabouts particular experimental variables were having their influence in this stages-of-processing account. For instance, stimulus legibility affects the earliest encoding stage of processing.

Source: Sternberg, S. (1975). Memory scanning: New findings and current controversies. *The Quarterly Journal of Experimental Psychology*, *27*, 1–32 (fig, 2., p. 5, fig. 4, p. 6). Reproduced with permission from Taylor & Francis Ltd.

encoded and the participant must compare this input representation against the stored representations of the memory set digits. If a match is found, a positive response can be made; if no match is found, then a negative response can be made.

It was the nature of the memory search and comparison process that gave rise to the notion of *serial exhaustive processing*. This is most easily understood in relation to Figure 8.17. As with any straight line graph the function can be decomposed into a slope (the steepness of the line) and an intercept (the point at which the line crosses the vertical axis). Sternberg took the intercept to reflect the sum of (i) the amount of time needed to encode the probe item, and (ii) the amount of time needed to convert the eventual decision into a response. The slope of the function was taken to reflect the rate of the memory comparison process, simply how quickly individual items could be processed. Two properties of these lines are important for our discussion here.

First, the linear nature of the function was taken to indicate the operation of a serial mechanism: each character added to the memory set carried an overhead of an additional unit of processing time. Second, the fact that the slope of the negative response function was the same as for that for the positive response function was taken to show that exactly the same sort of comparison process was going on in both positive and negative trials.

What the data therefore seemed to rule out is something known as **serial self-terminating processing**. In this context a serial self-terminating process would mean that the slope on the positive trials would be half that on the negative trials. Why? Well, on target present trials sometimes the target will be the first item in the memory set, sometimes it will be the last and sometimes it will be somewhere between the first and the last. On average this means that the target will be found after half the memory set items have been checked. Remember, the process is self-terminating – so as you find the target, stop checking and respond ‘present’. In contrast, on absent trials you can only press the ‘absent’ key once you have checked off all the items. In extending this analysis over the increasing memory set sizes tested, the conclusion is that the difference in slope between the positive and negative memory search functions should be 1:2. It should on average take twice as long to decide ‘absent’ as it does to decide ‘present’ for a given memory set size.

The problem with the memory scanning data, however, is that the memory search does not appear to reflect serial self-terminating processing because the slopes of the positive and negative functions are the same. As we have just argued, serial self-terminating processing would be most consistent with the negative slope being twice as steep as the positive slope, and it’s nowhere near. On these grounds, Sternberg argued that the data were most consistent with the notion of a serial exhaustive process in which all items in the memory set were checked on every trial regardless of whether the probe was part of the memory set or not. So even in the case when ‘3’ is the probe, and the memory includes both 3 and 7, the implication is that the stored 3 is checked and then the stored 7 is checked prior to a response being made. Although the adequacy of the reasoning has been challenged (see commentary by Sutherland, 1967), the notion of serial exhaustive processing has endured: hence the ‘???’ in Figure 8.16.

Pinpoint question 8.9

What support is there for serial exhaustive processing?

Controlled parallel processing

The final alternative in Figure 8.16 that remains unexamined is that termed controlled parallel. **Controlled parallel processing** allows some flexibility which is ruled out by the simple caricatures of early and late selection theories. By this view, parallel processing of several stimuli is possible but a possible downside is that unattended stimuli may go unprocessed. The eventual model that Pashler (1998) provided is set out in schematic form in Figure 8.18. The model incorporates both a filtering mechanism and resource limitations (as shown by the beaker of ‘capacity’ in the bottom-right corner of the figure). The filtering mechanism selects items for further processing and the selected stimuli may then compete for the limited capacity that underpins semantic analysis. If the demands placed on this pool of resources are within the capacity limitations of the system, then parallel processing of several stimuli is possible. However, if the system is overloaded then processing efficiency is compromised. That is, if too many stimuli are competing for resources, then processing will falter.

As with Broadbent’s (1958) early filter model, the controlled parallel processing model follows from a very careful review of the attentional literature – some of which we have examined here. According to Pashler (1998), overwhelmingly the evidence leads to such an account in which both structural (filtering

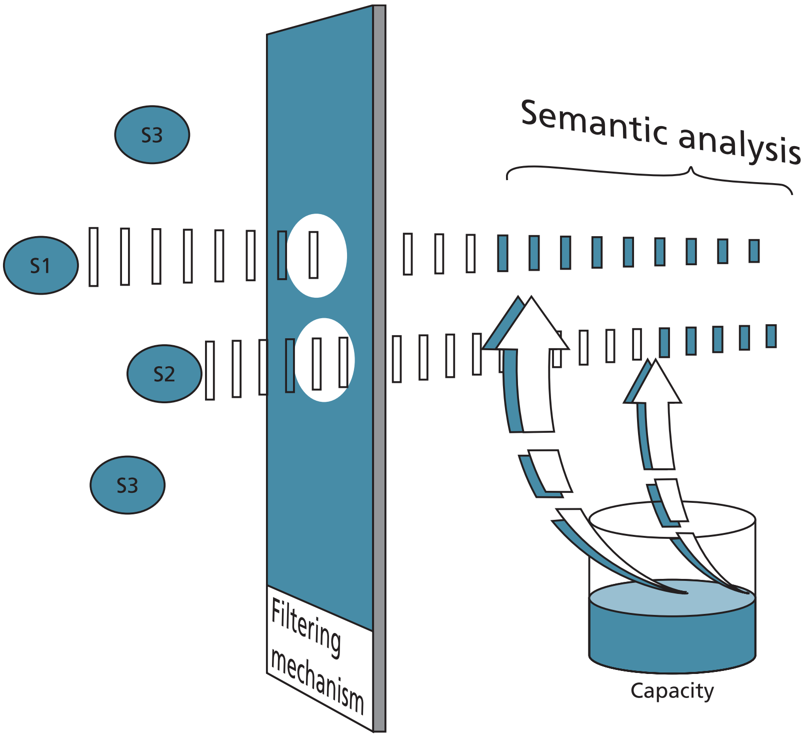

Figure 8.18 A schematic representation of the human information processing system as described by Pashler (1998) See text for further details.

Source: Pashler, H. E. (1998). *The psychology of attention* (fig. 5.1, p. 227). Cambridge, Massachusetts: The MIT Press. Copyright © 1998 Massachusetts Institute of Technology. Reproduced with permission. The MIT Press.

mechanisms) and processing (capacity/resource) constraints are needed.

uncontrolled serial model A framework of processing in which all items must be processed but in a sequential fashion.

serial exhaustive processing The sequential examination of a list of items, with testing continuing until all items have been considered.

Sternberg memory scanning task A task in which participants are initially given a set of possible target items to hold in mind; they are then presented with a single item and are asked to decide if it is a target.

memory set List of to-be-remembered items.

probe item In the Sternberg memory scanning task, an item to be compared with a memory set.

serial self-terminating processing The sequential examination of a list of items, with examination stopping after the item of interest has been found.

controlled parallel processing The general view, espoused by Pashler (1998), that parallel processing of stimuli is possible but this is not without restraint and is under executive control.

Perceptual load theory

Before moving on from this discussion of general theories of attention, it is useful to consider a more recent account that bears some resemblance to Pashler's, although neither is discussed in relation to the other. This is the perceptual load theory put forward by Lavie (2000). Lavie distinguished between two types of control mechanisms, namely passive and active mechanisms. While it may be a step too far, there does appear to be rather close similarities between Lavie's (2000) 'passive' mechanism and Pashler's filter, and her 'active' mechanism and Pashler's semantic analysis/central mechanisms. Both also agree on the distinction between early (perceptual) vs. later (executive) stages of processing. The main point of contention, though, is whether some form of capacity limitation governs the early perceptual stage of processing: according to Lavie, it does. She has claimed to have shown how it is possible to overload an early perceptual analysis stage of processing. Accordingly, selection by filtering can be compromised if the task places heavy demands on the initial perceptual analysis stage of processing.

Load theory and effects of varying perceptual load

Figure 8.19 provides examples of the sorts of visual displays used by Lavie (1995) to test her perceptual load theory. The task was to press one key if a target *z* was present and a different key if a target *x* was present. Response-relevant items were presented randomly at one of six pre-defined positions in a central row of the computer display. In the S1 condition there was only ever a single target present. In the S6 condition the target was embedded in a row of six characters. Also on every trial a non-target character was randomly positioned in the periphery of the display. As can be seen from Figure 8.19, the non-target could either be compatible with the central target (i.e., the central target was an *x* and the peripheral non-target was also an *X*), the non-target could be incompatible with the target (i.e., a *z*), or it could be neutral with respect to either response (i.e., a *P*).

In general terms two effects were of initial interest: first, whether the compatibility of the non-target would influence responses to the target; and second, whether there would also be an effect of display size (i.e., the number of items in the display) on responses. More interesting, though, was what the combined effect of these two factors would be – would the compatibility effect interact with the effect of display size?

The results of the experiment were relatively clear-cut. Responses were shorter on trials containing a small number of items than large. Although this factor has been referred to display size, Lavie termed the factor **perceptual load**. The understanding was that, by increasing the number of non-target items in the display, this would increase the load placed on perceptual encoding mechanisms. So a large display size was taken to reflect a high perceptual load being imposed, whereas a small display was taken to reflect a low perceptual load being imposed. The most interesting result, however, was that the effect of non-target compatibility varied across these two levels of perceptual load.

Let us stand back for a second so as to appreciate what the basic compatibility effect should look like. Here previous experiments had been able to establish firmly that responses in these sorts of task were heavily influenced by the nature of irrelevant non-target items concurrently present. The standard effect is that RTs are shortest on trials containing compatible items (both target and non-target are *x*), longer on neutral trials (target as *x* and non-target as *P*) and longest on incompatible trials (target as *x* and non-target as *z*).

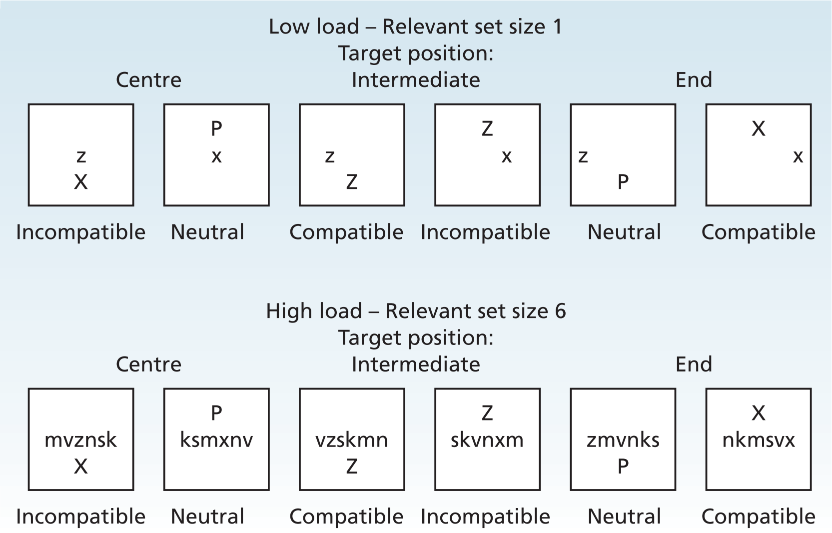

Figure 8.19 Examples of displays used by Lavie (1995)
The target letters are *z* (press one key) and *x* (press another key). Participants were told to focus attention on the central letters and decide which target was present there. The central letters were accompanied by the presence of a peripheral, non-target letter. As can be seen, the peripheral non-target could either be congruent with the central target, incongruent with the central target or response-irrelevant (i.e., neutral). Across the low and high perceptual load conditions the number of central letters varied. Perceptual load increased as the number of central letters increased.

Source: Lavie, N. (1995). Perceptual load as a necessary condition for selective attention. *Journal of Experimental Psychology: Human Perception and Performance*, *21*, 451–468 (fig. 1, p. 455). Reproduced with permission from APA.

Indeed, the lengthening of responses on incompatible trials is taken to reflect something known as **response competition**. Without going into too many details, the basic idea is that both the target and non-target are identified and as a consequence a conflict occurs at the level of response selection and execution. The target *x* primes the *x* response and the non-target *z* primes the *z* response. So which response is appropriate?

Now in Lavie's experiment the critical finding was that the standard compatibility effect only arose for the low perceptual load condition, when there was a small number of non-targets in the display. There was no corresponding effect for the high perceptual load condition, when there were several non-targets in the displays. Lavie explained this particular pattern of effects by reference to her perceptual load hypothesis. If the load on perceptual analysis is slight, then filtering is unnecessary and both targets and non-targets are processed. However, when perceptual analysis is made more demanding, then full (perceptual) capacity will be devoted to detecting targets. Given that all resources are dedicated towards target detection, none of the non-targets will be processed further. The perceptual analysis of the target is so demanding that the perception of the non-targets is said to be excluded.

Think of it this way. You are waiting to hear the football results on a pub's television set and the pub is crowded and very noisy. So the main perceptual task (hearing the telly) is very demanding. Under these conditions you are likely not to notice that your mate is asking whether you want another drink (he is the distracting non-target). Compare this with the situation in which the pub is empty, the telly is clearly audible and your mate asks whether you want another drink.

Load theory and effects of varying memory load

More intriguing still was a further extension to the theory tested in more recent experiments. On the assumption that sometimes both relevant and irrelevant items exit from the perceptual analysis stage, then the active mechanisms now come into play in an attempt to discriminate the relevant from the irrelevant. On these grounds, if the active mechanisms are overloaded then any detrimental effect of irrelevant items might then be exacerbated. So whereas overloading the passive (perceptual) mechanism results in a lessening of interference from non-targets, completely the opposite is predicted if the active (central) mechanisms are overloaded. If the central mechanisms are overloaded, then, even in cases where the perceptual load is slight, the prediction is that interference from non-targets should be substantial.

This second strand to the theory has been tested on a number of occasions but one example will suffice here. De Fockert, Rees, Frith and Lavie (2001) ran the following dual-task experiment using stimuli similar to those depicted in Figure 8.20. In the attentional task the stimulus was a photograph of a famous face superimposed by a typed name. Participants were required to make a response to the face, classifying it as belonging to a politician or a pop star. A congruent stimulus contained the name and face of the same person (e.g.,

Figure 8.20 Variations on a theme by de Fockert et al. (2001)

Sources: (top row): Getty Images; (bottom row): Hulton-Deutsch Collection/Corbis.

Mick Jagger's face with Mick Jagger's name); an incongruent stimulus contained the face of a politician and the name of a pop star (e.g., Bill Clinton's face with David Bowie's name) or the face of a pop star and the name of a politician). Finally a neutral stimulus contained the face of a politician or a pop star superimposed by an anonymous name (e.g., Jim Smith). On the basis of previous experiments it was predicted that RTs to congruent stimuli should be shorter than RTs to incongruent stimuli. However, the main interest was with how this effect might be influenced by the nature of a secondary memory task.

Following the initial presentation of a central fixation cross, a memory set of five digits was presented. Participants were instructed to retain this ordered sequence until the end of the trial. Okay, so retain this set of numbers and keep reading: 0 3 1 2 4. Next there followed a varied number of display frames containing the face stimuli defined before. When the face frames had completed, a probe digit was presented and participants had to respond with the digit that *followed* it in the original sequence. So what was the number after the probe digit 2? Imagine having to do that as well as the face judgement task and you will get a fair idea of what it was like to be in the experiment (see Figure 8.21 for a schematic representation of the events at each trial).

In order to manipulate the active central mechanisms posited by Lavie (2000), memory load was varied by either having participants retain a fixed digit sequence over a complete block of trials (0 1 2 3 4 in the *low memory load condition*) or by having them retain a different sequence on every trial (the *high memory load condition*). If the perceptual load theory was correct, then the effect of an incongruent stimulus should be greater in the high memory load condition.

Overall, the concurrent memory load affected responses on the attentional task with RTs being shorter in the low memory condition than the high memory condition. Critically, though, the congruency effect was considerably larger in the high memory load condition than the low memory load condition. The implication was that when memory was overloaded, participants experienced substantial difficulties in suppressing the irrelevant typed name information.

One central assumption here is that the memory task tapped into the same active central mechanisms responsible for suppressing irrelevant semantic information activated by the presentation of the face. By this view, the degree to which a given non-target will interfere with some other task is critically dependent on the availability of capacity in the temporary memory system. This hypothesis was supported because, as de Fockert et al. (2001) showed, the ability to ignore irrelevant information was *diminished* when memory load was increased. The conflict between responding to the face and the name was increased when the concurrent memory load was high.

The contrast between the earlier study by Lavie (1995) and the later work by de Fockert et al. (2001) is particularly interesting. In summary, overloading the

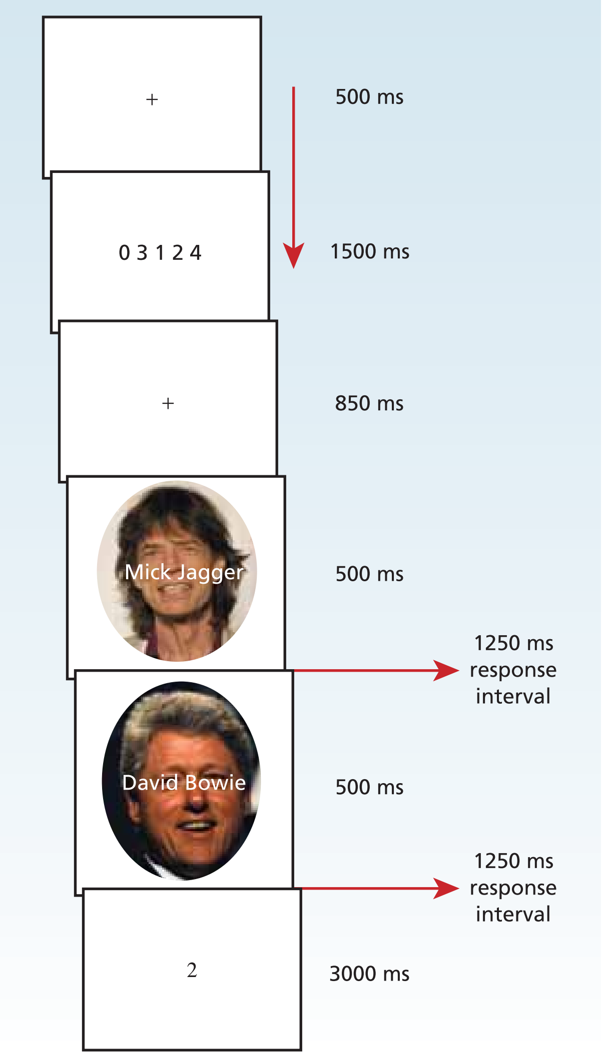

Figure 8.21 Schematic representation of the sequence of events in the experiment reported by de Fockert et al. (2001)
See text for further details.

Sources: de Fockert, J. W., Rees, G., Frith, C. D., & Lavie, N. (2001). The role of working memory in visual selective attention. *Science*, *291*, 1803–1806 (fig. 1, p. 1804). Reproduced with permission from AAAS.
Images: (top): Getty Images; (bottom): Corbis.

passive (perceptual) mechanism results in a lessening of interference from non-targets (as shown by Lavie, 1995), but overloading the active (central) mechanisms leads to an increase in interference from irrelevant information.

Pinpoint question 8.10

How do high demands placed on passive perceptual analyses contrast with high demands placed on active central analyses?

perceptual load The amount of perceptual analysis demanded by a given task.

response competition Conflict that occurs when more than one response is invoked by a stimulus.

Concluding comments

You may well now be feeling that the number of different theories of attention that are out there is the same as the number of people who consider themselves to be researchers in attention. To allay these fears it is important to bear in mind that certain themes recur in the literature. Some examples are the differences between structural (bottleneck) constraints and processing (resource) constraints (see Figure 8.4); differences between the various ways in which information is selected or discarded, and the problems that exist in trying to do more than one thing at once. From Pashler's (1998) work it seems that both assumptions about structural and processing constraints are necessary. However, it may well turn out that even such hybrid theories are still too simplistic and that much more imaginative thought is needed on the ways in which the human information processing system may be constrained. Our theories need to take account of both structural and processing constraints but are we really thinking about these in the most appropriate fashion?

It may well be that our conceptions of the human information processing system are misguided and that the mind is much more complex than the parallel-then-serial framework as suggest by Pashler and others. Indeed, some very interesting and quite different speculations about human information processing have been put forward by Navon (1989a, 1989b). He speculated that the mind may actually comprise many different processing modules, each specialised for some particular purpose: sometimes the modules co-operate, sometimes they compete. By this view the mind is understood as an *anarchic intelligence system*. The modules are said to operate autonomously in attempting to fulfil their own purposes such as 'collecting information . . . , making interpretations, attempting prognoses, and initiating operations pertinent to the system's subsistence' (p. 195). Attention within this framework is 'assumed to regulate only the communication among modules' (p. 191). Within this framework the idea that the filter operates early or late is seen to be far too simplistic – each module has its own constraints and these may come into play at different stages throughout the system.

These ideas have yet to be fleshed out fully and the critical experiments have yet to be carried out. Nevertheless, such an alternative framework reveals that, despite what might seem to be an emerging consensus on how best to construe constraints in the human information processing system, interesting avenues remain to be explored.

CHAPTER SUMMARY

- The world can be an overwhelming place in terms of the sensory information it provides. The aim of attention therefore is to allow us to focus on those things that are critical in our environment and filter out those aspects that are less relevant.
- Broadbent's (1958) filter theory proposed an early structural bottleneck, which acts as a gateway between early sensory analysis and later processes of stimulus identification. Although early processes handle vast amounts of information concurrently – in parallel – the more central processes operate less slowly in a serial one-at-a-time mode. Items that are deemed relevant are selected for further processing but such selection can only operate on the physical characteristics of the stimulus as represented in the early sensory store.

  Support for Broadbent's theory was provided by split-span experiments in which participants reported information split across two channels, and shadowing experiments in which participants were asked to report information from a single channel while ignoring a second channel.

▶

- Treisman's (1964) attenuated filter theory model revised Broadbent's (1958) idea and argued that sometimes information on unattended channels could be processed to a semantic level, such that meaning was derived. The evidence seemed to suggest that occasionally the meanings of items were recovered despite attention being directed elsewhere. One implication was that the division of attention between attended and unattended channels was not all-or-nothing.
- The late selection theory of Deutsch and Deutsch (1963) argued for a late structural bottleneck in that attentional selection takes place at a post-categorical stage, wherein all items are identified and selection is now carried out on the basis of meaning. This has the implication that attended and unattended information are processed to a semantic level. Further evidence for late selection comes from the experiments of Duncan (1980) who showed that limitations in processing only occurred when more than one target item competed for attention.
- Alternative accounts argued for processing rather than structural constraints. This is the difference between the speed of a lift (processing constraint) and the size of the lift (structural constraint). Theorists have debated the relative merits of single resource models in which all concurrent tasks draw on a single limited pool of attentional resources (Kahneman, 1973), and, alternatively, models in which multiple resources each relate to different input channels (Treisman & Davies, 1973).
- Pashler's (1998) controlled parallel model argued for the combined existence of structural and processing constraints in attention, with an initial filtering mechanism selecting items, and such items going on to compete for resources involved in semantic processing.

  Lavie (2000) used the concept of different sorts of mental load to argue for different kinds of processing limitations. From her experiments, the ability to ignore irrelevant information was *enhanced* when the perceptual load of the task was increased. In contrast, the ability to ignore irrelevant information was *diminished* when memory load was increased.

ANSWERS TO PINPOINT QUESTIONS

8.1 Item decay can be avoided by rehearsal. This allows information that has already exited the S system to be recirculated via the P system back into the S system.

8.2 Participants performed better at the binaural condition when reporting the content of one ear followed by the content of the other ear, and when the rate of presentation was slowed.

8.3 Wood and Cowan (1995b) challenged Broadbent's (1958) model by arguing that unattended information (such as your own name) can be semantically processed.

8.4 B D G Q share curved properties with C and O whereas T Z X Y do not (Rabbitt, 1964). This is an example of categorising (Broadbent, 1958) in which learned categories can determine which items are selected and which are essentially filtered out or ignored.

8.5 Duncan (1980) provided evidence in support of the Deutsch and Deutsch (1963) model in that non-targets failed to impact upon responding while multiple targets did. This seems to indicate that that the nature of the item (target or non-target) is made explicit prior to selection. Selection seems therefore to take place at a post-categorical level.

8.6 A dual-task decrement should occur in that performance in the two tasks should not be as good as either task performed in isolation. According to Norman and Bobrow (1976), complementarity should also obtain in that the more resources allocated to task A, the worse participants should be at task B (or vice versa).

8.7 Neutral words did not impact upon responding in the Navon and Miller (1987) study because they were not relevant to either task.

8.8 Two tasks are less likely to interfere with one another if they (a) place demands on different stages of processing, (b) involve different modalities, and (c) do not share internal forms of coding.

8.9 The idea of serial exhaustive processing (Sternberg, 1967) is supported by the facts that (i) the increase in reaction time was linear with increases in memory set size, and (ii) identical functions obtained for both present and absent responses.

8.10 The ability to ignore irrelevant information was *enhanced* when the perceptual load of the task was increased. In contrast, the ability to ignore irrelevant information was *diminished* when memory load was increased.