# ATTENTION

Wayne Wu

Routledge
Taylor & Francis Group
LONDON AND NEW YORK

# 1

# THE PSYCHOLOGY OF ATTENTION

## 1.1 Introduction

This chapter considers highlights of psychological research on attention since the 1950s, a period during which the conceptual scheme that frames contemporary theorizing about attention was firmly established. It examines central experimental paradigms used to probe attention, the initial questions and theories that drove early investigation, some conceptions of what attention is, and the concepts developed to characterize attention. Aside from a historical overview, there are two additional goals. First, with an eye to answering the metaphysical question, "What is attention?", I shall extract from experimental paradigms a link between attention and a subject's selecting information or targets to guide and control performance of a *task*. Specifically, I argue that a background assumption in experiments on attention is that such selecting for task is *sufficient* for attention. This condition provides the seed of an answer to the metaphysical question to be developed in subsequent chapters. Second, these experimental paradigms have informed the development of a theoretical vocabulary to characterize attention, and in particular, have led to descriptions of two basic kinds of attention: roughly, (a) attention that can be intentionally directed as when one looks for a missing object; and (b) attention that is captured as when a loud sound pulls one's focus to it. I will provide a rigorous analysis of the concepts used to characterize this division.

Section 1.2 begins with the common idea that attention is a form of selection but raises the question, "Selection *for what?*" Section 1.3 then examines an early debate about what stage of perceptual processing selection occurs at, in particular, whether it is at early or late stages of such processing. Here, attention was conceived of as a filter for information, selecting it for further processing. As the early versus late selection debate was never adequately resolved, section 1.4 discusses the proposal of Nilli Lavie's *Load Theory of Attention* that the conflicting data that drove the debate was a function of the different experimental tasks researchers used. The nature of the task makes a difference. In that vein, while early work focused on auditory attention, work on vision became prominent in the 1960s. Section 1.5 discusses the visual search paradigm and a resulting theory due to Anne Treisman: the *Feature Integration Theory*. While this theory is no longer at the center of current debates, it set the stage for how attention is conceptualized. Section 1.6 then discusses another paradigm for spatial attention, *spatial cueing*, and considers the contrast between *top-down* versus *bottom-up* attention. Are there different types of attention or different attentional mechanisms? Section 1.7 picks up on this theme and examines some central conceptual dichotomies used to characterize attention. I provide definitions of the central dichotomies. Finally, section 1.8 extracts from the standard experimental paradigms a sufficient condition for attention to an X: selection of X for a task. I argue that this is a shared assumption that can serve as an antidote to the widespread skepticism about an answer to the metaphysical question noted in the Introduction.

## 1.2 Attention as selection for what?

The Introduction presented five basic questions about attention:

***Metaphysical:*** What is attention?
***Function:*** What role does attention play?
***Properties:*** What are characteristic features of attention?
***Mechanism:*** How is attention implemented?
***Consciousness:*** What is the relation between attention and consciousness?

To begin the discussion of the psychology of attention, consider the function question. There is widespread agreement among cognitive scientists that attention is a process of selection. James's passage captures the selectivity commonly attributed to attention: "It is the taking possession by the mind, in clear and vivid form, of one out of what seem several simultaneously possible objects or trains of thought." Attention cannot, however, be merely selection. After all, there are many kinds of selection that do not count as attention. An object sorter can be highly selective yet does not attend to what it selects. As to be discussed in Chapter 2, a neuron can be highly selective in having a preferred stimulus, but it does not follow that the neuron thereby attends to its stimulus as opposed to its being part of a mechanism of attention. Indeed, there is something odd about the claim that a neuron, a part of a person, attends. The point is that if attention is selection, it is a specific kind. Psychologists often add that attention is selection *for further processing*, but this invites similar challenges: the object sorter and neuron can select for further processing, too. Further precision is needed in characterizing attentional selection.

One way to distinguish attentional selection from other forms of selection is to identify the type of thing that can attend. James speaks of the taking possession "by the mind" emphasizing that it is a psychological *subject* that pays attention, namely an entity that has a mind. The previously noted object sorter and selective neuron are not psychological subjects, so they cannot exemplify attentional selection even if they exhibit another kind of selection. One can then treat attention as a subject-level phenomenon or, as philosophers like to put it, a *personal*-level phenomenon. The relevant contrast is between the personal and the *subpersonal*. Although this distinction is widely invoked, it needs clarification. In the absence of a rigorous analysis of the distinction, I proceed with a simple division. One can think of personal-level states as those states that are attributable to a subject and not to the subject's parts, such as the brain or part of the brain. In contrast, subpersonal states are attributed to those parts but not to the subject. On this account, unconscious mental states count as personal in that they are attributed to the subject and not to the subject's parts. For example, certain parts of the brain might implement Freudian Oedipal desires, but while the subject might have such desires, that part of the brain does not. Similarly, attention is something that persons are capable of, not their parts. If one were to accept this division of the personal from the sub-personal, then one can discount selection exhibited by neurons and dumb machines as forms of attention.[1]

Still, the answer is not very informative, for while it suggests what kind of thing can be selective, it does not tell us much about selection. Might there be something in the nature of attentional selection that also divides it from other kinds of selection? Let's begin the historical overview of the psychology of attention while holding this question constantly in the background.

## 1.3 The debate over early versus late selection: capacity limitations

In the revival of modern attention research in the mid-twentieth century, attention theorists focused on the selection of *information*: psychological subjects are presented with a lot of information in experimental situations, and, to perform a task, they must select only relevant information. For example, a subject asked to selectively listen to one of two conversations selects information from that conversation. This emphasis on information, inspired by communication theory in the 1950s, led to the first major debate about attention: At what point in perceptual information processing does attentional selection occur? The answers to this question, usually divided between so-called *early selection* and *late selection* accounts, provide an early account of what attentional selection involves.

It is common among cognitive scientists to speak of both the mind and brain as processing information, but what is information? Claude Shannon (1953) provided a precise definition in his theory of communication in terms of what he called *mutual information*. The latter is defined in terms of *entropy*, which in information theory is a statistical measure of uncertainty. This concept of information is defined mathematically, but I eschew the technical details and make do with three points: (a) mutual information is tied to the reduction of *uncertainty* (a message about X is informative to the extent that it reduces uncertainty about X); (b) information can be precisely *quantified* (often measured in *bits*, derived from "binary digits"); and (c) it is *not identical to meaning*: the same meaningful sentence can carry different amounts of mutual information, while two sentences of different meaning can carry the same amount of mutual information (for more on information, see Appendix A). Meaning can be understood as a type of semantic information where "semantic information" identifies the content of a representation. Such content need not be linguistic such as the content (meaning) of a sentence, but can also be tied to representations of features or objects, such as the auditory system representing pitch or the visual system representing a ball. Given the distinction between semantic and mutual information, an ambiguity crops up in talk of processing and, later, of selecting information. Does "information" mean mutual information or semantic information? In fact, in psychological and philosophical theorizing, it is often the latter that is meant, but then what is the significance of mutual information?

In his book, *Perception and Communication* (1958), Donald Broadbent drew on Shannon's theory to propose a "fresh language" (35) and a "new set of descriptive terms" (36) for psychology. On Broadbent's view, the technical language of information allows for precise characterization of information processes that are *capacity limited*. These processes can only deal with a limited amount of information at a time. For example, Itti and Koch (2001) suggest that information can flow at $10^7$–$10^8$ bits per second along the optic nerve transmitting information from the eye. How can visual processing keep up with this vast input? Experience also suggests that there are capacity limits to perception. For example, there are a limited number of conversations you can listen to at once. It is natural to characterize this limit in terms of an informational bottleneck, although this is merely a metaphor. Given Shannon's work, Broadbent realized that psychology could go beyond metaphors to investigate capacity limits with mathematical precision. This is an important point that has been lost in recent years as psychologists and philosophers have focused on semantic information (meaning). Theorists have invoked capacity limits in theories of attention and, as we shall see, in theories of working memory in connection with phenomenal consciousness (see Chapter 6). They have suggested that such limits impose constraints on the nature of attention and consciousness, but in general, invocations of capacity often remain qualitative, rather than quantitative. These theories thus suffer the fate that Broadbent sought to avoid: metaphors rather than precision. Ultimately, serious talk of capacity limits must quantify these limits if the invocation is not to be merely a figure of speech. It will not be possible to invoke information theory in detail in our discussion. Rather, the point is to remember that where invocation of capacity limits becomes important, a theory must provide quantitative measures of the sort Broadbent drew from Shannon.

Capacity limits yield a plausible story of why attention is necessary. For, given a limited capacity to deal with an overabundance of information, a creature needs a capacity to select just what information is relevant for current goals on pain of information overload. That is, a capacity-limited creature needs attention. Capacity limits and selection provide the conceptual background for the debate over early versus late selection: Does attentional selection occur early or late in perceptual processing? The idea is that there are capacity limits on information processing, namely, the maximum amount of mutual information that can be processed at a time.

Plausibly, limits on processing of mutual information impose a limit on the amount of semantic information (representational content) that can be processed at a time. In what follows, we focus on limits in processing semantic information in light of a channel's limited capacity to process mutual information. Thus, we shall focus on the processing of representational contents, constrained by the (mutual) information capacity of the relevant processing channel. Talk of early and late *selection* concerns the different stages of perceptual processing of relevant representations. For example, in audition, an early stage of processing concerns the basic audible features of a sound (e.g., a voice)—say, its pitch or timbre—while a late stage of processing concerns the categorical features of the sound, say, the identity of the voice or the meaning it expresses. The question then is whether attention selects basic or categorical features.[2] Assuming that perceptual processing is capacity limited, an informational bottleneck must occur somewhere, and attention then serves to select information at the bottleneck. In light of this, Broadbent suggested that attention acts to *filter* information.

In early work on attention, pioneered by Colin Cherry (1953), the focus was on auditory processing of language. Cherry focused on *filtering tasks* where subjects are presented with multiple stimuli and asked to select some subset of them. In the *dichotic listening paradigm*, two streams of verbal inputs are presented, one to each ear treated as a separate information channel. Subjects then selectively "shadow," i.e., verbally repeat, only one of the sound streams. The basic finding was that when subjects attended to one stream, they did not pick up information from the other. When queried about what was said in the unattended stream, subjects were unable to provide accurate answers (notice that this experiment focuses on semantic information, i.e., what is heard, not mutual information). If perceptual processing was not capacity limited, psychologists initially reasoned, then subjects should be able to report the contents of both channels.

Jon Driver (2001) has noted that two questions were fundamental to theorists at that time: (a) What conditions allow people to effectively shadow the attended message?; and (b) What do people typically know about the unattended message? On the first question, it was ascertained that substantial differences in the physical properties of the sounds between the two channels facilitated performance. For example, shadowing improved when the two auditory streams were heard as if coming from distinct rather than the same locations. Shadowing was also aided by distinct acoustical properties such as presenting a low-pitch versus a high-pitch voice. In general, physical distinctness aided attentional selection.

Regarding the second question, many experiments suggested that subjects miss quite a lot from the unattended channel. Building on Cherry's work, Neville Moray (1959) observed that even when the unattended channel consisted of a small number of words repeated multiple times, subjects still failed to report accurately what those words were. In general, early observations suggested that while subjects could notice abrupt changes in lower-level physical properties of the unattended stream, higher-level perceptual properties like semantics (meaning) were typically missed. Consequently, Broadbent postulated that attentional filtering occurs after processing of basic physical features but prior to processing of categorical features. Thus, filtering occurs early in perceptual processing. The general picture entails a division between a preattentive and an attentive stage of processing, a distinction that remains to this day. On Broadbent's *early selection account*, preattentive processing concerns basic physical properties of the stimuli, with attention filtering relevant information about basic properties for higher-order, categorical processing. While talk of attention as a filter is metaphorical, Chapter 2 will consider one possible neural implementation of attentional filtering. For now, construe the metaphor as Broadbent's answer to the function question: attention filters (selects) information for the purpose of categorical processing. This provides a more concrete specification of attention as a type of selection.

Evidence quickly accumulated, however, showing that quite a bit of semantics in the unattended channel could get through. Drawing on the anecdote of the *cocktail party effect* where the mention of one's name in a nearby conversation is said to capture attention, Neville Moray (1959) did find that when the subject's name appeared in the unattended channel, the subject was more likely to notice it: subjects reported instructions expressed in the unattended channel when these were preceded by their name (33% of the time, Moray, op. cit. table IV, p. 58). Moray concluded that "[i]t is probably only material 'important' to the subject that will break through the [bottleneck] barrier" (op. cit., 56).

Subsequently, Anne Treisman (1960) argued that "the selective mechanism in attention acts on all [stimuli] not coming from one particular source by 'attenuating' rather than 'blocking' them" (246–7). In one experiment, Treisman instructed subjects to shadow a verbal stream presented to one ear, say the right ear, while ignoring a second stream presented to the other ear. Unbeknownst to her subjects, Treisman swapped the verbal streams midsentence, so a sentence that begins in the right ear, switches to the left, and vice versa. An example is given in Figure 1.1:

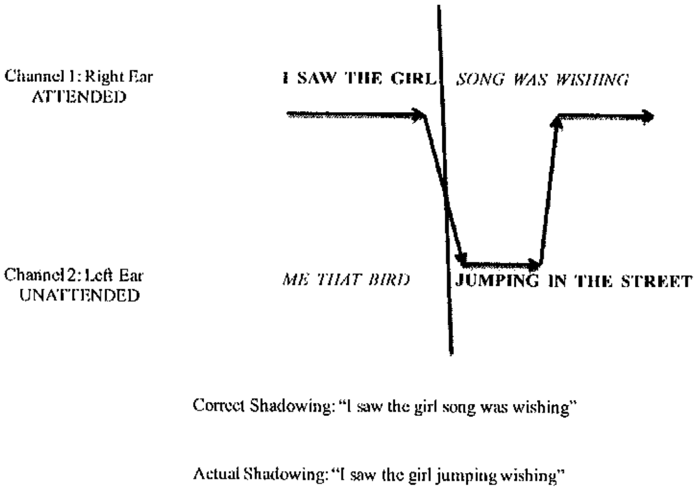


Figure 1.1 Treisman's experiment where two verbal streams are presented to each ear (one in italics, the other in bold). Normally, each stream is presented to a single ear, but Treisman changed channels mid-sentence so that the verbal streams switched ears at the point indicated by the vertical line. Thus, the sentence, "I saw the girl jumping in the street," begins in the left ear, the attended channel, but jumps to the right ear. Subjects were asked to shadow only one of the channels, so correct performance is just shadowing of the words in a single channel. The arrows indicate what the subjects actually shadowed, and, speculatively, they suggest that the subject's attention jumps between channels despite task instructions.

Here the sentence to be shadowed jumps from the right ear to the left, with the switch point indicated by the vertical line. To shadow correctly, however, the subject must continue to repeat the words on the right. In the example, the two sentences at issue are as follows: first, "I saw the girl jumping in the street," which begins in the right channel but switches to the left after "girl"; and, second, "me that bird song was wishing," which begins in the left channel but switches to the right after "bird". Correct shadowing would be the nonsensical "I saw the girl song was wishing." Surprisingly, the subjects shadowed "I saw the girl *jumping wishing*". It was as if attention jumped between the two ears despite task instructions. Indeed, subjects were *unaware* of doing this. Treisman reasoned that this intrusion of semantics from the unattended channel on the shadowing of the attended channel depends on contextual effects that continue to influence behavior, since the word "jumping" rather than "song" is more probable given the preceding "I saw the girl ... " It seems that unattended information remains available to influence behavior and is not completely filtered out. Given Treisman's plausible explanation of why the subject jumps between channels, it is natural to conclude that despite being unattended, the left channel must be analyzed to a higher linguistic level.[3] This means that information selected by the filter is not the only information being processed at higher stages. Since this unattended information is not being fully blocked, one might wonder if the filter isn't leaky: unattended information can get through the filter for higher level processing. Alternatively, perhaps the filter is operating at a *late* stage in perceptual processing. Indeed, the latter possibility began to gain wider acceptance.

By the mid-1980s, Daniel Kahneman and Treisman (1984) noted a shift from early to late selection theories (Deutsch and Deutsch 1963; Norman 1968). On late selection accounts, filtering occurs after all signals are perceptually processed up to a categorical level of representation (e.g., semantic). Thus, relevant capacity limits occur post-perceptually, and it is only at this late stage that attention is needed. As Driver puts it

> Late selectionists ... proposed that the limited awareness of unattended stimuli (as for the non-shadowed message in selective listening experiments) might have less to do with rejection from full perceptual processing, than with rejection from entry into memory or into the control of deliberate responses ... Thus, unattended stimuli might conceivably undergo full perceptual processing, yet without the person being able to base their deliberate responses upon this, and without the formation of explicit memories.
>
> (2001, 58)

The point of late selection accounts is that perceptual processing may not be limited at all; rather the bottleneck occurs when perception engages other systems.

The debate about early versus late selection highlights some answers to the basic questions: attention is a type of selection, namely filtering; it occurs at specific moments in perceptual processing; and it functions to deal with capacity limitations. What is left hanging is whether attention operates early or late in perceptual processing.

## 1.4 Task demands and load: resolving early versus late selection

Kahneman and Treisman (1984) noted that the shift from early to late selection theories coincided with a shift in different types of experimental paradigms. Early work in audition involved filtering tasks where subjects were overloaded with task-irrelevant input. Later work focused less on filtering and more on target selection. This includes well-known paradigms like spatial cueing and visual search to be discussed in sections 1.5 and 1.6. Kahneman and Treisman argued that different experimental paradigms might tap into different mechanisms involving selection at different stages in processing. Thus, disparate results favoring early or late selection might merely reflect choice of experimental task.

Nilli Lavie and co-workers have proposed a *Load Theory of Attention* that builds on Kahneman and Treisman's observation (Lavie 2005). Begin with the idea of processing as *resource limited*, so that the amount of processing available for any task has an upper bound. Unlimited processing capacity is not available. What happens to the limited resource if current processing does not use all of it? Does the remainder lie dormant? Does another process tap into it? Treisman (1969) suggested that "we tend to use our perceptual capacity to the full on whatever sense data reach the receptors" (p. 296).[4] In line with this, Lavie and Tsal (1994) suggested that total processing resources are always deployed, and where the attended information channel does not exhaust available resources, remaining capacity is then apportioned to processing unattended channels. On their account, what is critical is the *perceptual load* of the attended channel, namely how much of available processing resources that channel consumes. This suggests an explanation of the conflicting data that drove the early versus late selection debate. Load Theory holds that both models are in a sense correct, for the observed effects adduced to support either early or late selection depend on task demands, namely what the subject is doing. The general prediction is that early selection effects will be seen in high perceptual load conditions where all available processing is consumed by heavy task demands. For example, auditory filtering tasks in early work in the 1950s might be high-load, involving a large amount of information to be sifted through. In contrast, late selection effects will be seen in low perceptual load conditions where the system is not overloaded with information and additional processing resources are available for processing unattended channels.

The crucial point is that the nature of the experimental task can effect how attention is deployed, which can give rise to either early or late selection effects. The character of these effects is *task dependent* because tasks determine the informational load that must be processed. Thus, a potential resolution of the early versus late selection debate is that both are in a sense correct, with their characteristic effects differentially occurring depending on the perceptual load in the task. More importantly, if early and late selection effects depend on the nature of the task, then it would be incorrect to tie the selectivity of attention down to a specific stage in processing (i.e., early or late). Rather, a more general possibility is beginning to emerge: sometimes, attention can be early in processing; sometimes it can be late. In either case, attention is dependent on the task.

## 1.5 Visual search and the Feature Integration Theory of attention

While early attention research focused on audition and verbal shadowing tasks, there was a gradual shift to vision and visual search tasks in the 1960s. Visual search is *looking for something*. Sometimes, search is difficult, as when you look for a friend in a crowded train station; sometimes it is easy, as when that friend is wearing a neon green shirt, jumping up and down in plain view. The attention that guides visual search can be understood as directed at objects and/or their features.

An influential model of visual search was Treisman's *Feature Integration Theory* of visual attention (FIT) (Treisman and Gelade 1980; for an informative assessment, see Quinlan 2003). In its initial version, FIT treats visual object recognition as a constructive process where basic visual features are first detected by dedicated receptors, e.g., those for color, shape, and motion. The visual system then *binds* these features to form representations of objects. As in Broadbent's filtering conception of attention, visual object recognition involves two stages, in this case a preattentive feature detection stage and an attentional binding stage. Treisman and Gelade (1980) wrote that in FIT:

> features are registered early, automatically, and in parallel across the visual field, while objects are identified separately and only at a later stage, which requires focused attention. The model assumes that the visual scene is initially coded along a number of separable dimensions, such as color, orientation, spatial frequency, brightness, direction of movement. In order to recombine these separate representations and to ensure the correct synthesis of features for each object in a complex display, stimulus locations are processed serially with focal attention. Any features which are present in the same central "fixation" of attention are combined to form a single object. Thus focal attention provides the "glue" which integrates the initially separable features into unitary objects. Once they have been correctly registered, the compound objects continue to be perceived and stored as such.
>
> (98)

Accordingly, in the preattentive stage, processing of features occurs in parallel and independent of focused attention. Focused attention then binds features into object representations. A critical aspect of the model is that the processing of features and objects is separate.

Standard visual search tasks require subjects to search for a target amid a set of distractor objects, the number of distractors constituting the *set size*. There are two experimental conditions, *target present* and *target absent*. The subject reports whether a target was present or absent (yes, in the first condition; no, in the second). Furthermore, there are two kinds of search: *feature search*, where a single basic feature is the target, and *conjunction search*, where targets are individuated by a combination of basic features.[5] For example, in feature search, where color and shape are basic features, one might look for a red **T** against a sea of green and brown **Ts** (i.e., red is the relevant feature). In this case, one looks for a difference in color to identify the target. In conjunction search combining both shape and color, one might look for a green **T** in a sea of brown **Ts** and green **Xs**. In this case, one looks for both a difference in color and shape (i.e. green and **T**-shape are the conjunction). For another example, consider searching for the black rectangle in Figure 1.2A versus searching for the same rectangle in Figure 1.2B.

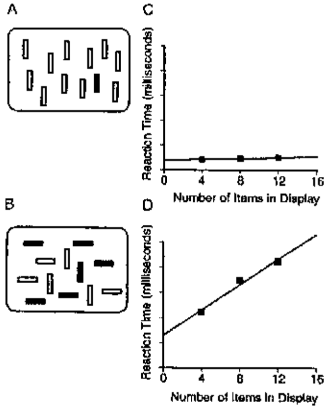


Figure 1.2 Visual search. In (A), the vertical dark rectangle pops-out because it is a feature singleton, the only one that differs from the distractors in terms of color. Pop-out is operationally defined as the relative independence of reaction time to set size (number of distractors) as given in (C). Notice the flat slope. In (B), we have a conjunction search where we must identify the vertical black rectangle. Here, visual search is harder and reaction time varies with set size as graphed in (D). Reprinted from S. P. Vecera and M. Rizzo (2003) "Spatial Attention: Normal Processes and their Breakdown." *Neurologic Clinics* 21: 575–607 with permission from Elsevier.

The relevant measure in these experiments is reaction time (RT), namely how long it takes the subject to report that the target is present or absent. Two basic findings are noteworthy. First, Treisman reported that in feature search, RT does not vary with set size (see Figure 1.2C). Here, the target seems to "pop out" from the display regardless of the number of distractors. Note that "pop-out" can describe the phenomenology (the object just seemed to pop out), but in the psychology of attention, it refers to the behavioral effect of constant reaction time despite increase in set size, as in Figure 1.2C. Second, in conjunction search, RT does vary with set size (see Figure 1.2D). These and other results led Treisman to propose a two-stage model. In feature search, processing of features happens concurrently or in *parallel* without capacity limitations. The target pops out because it is a *singleton* on one of the feature maps, namely a unique instance within that map (e.g., the color map might have a single red feature in a sea of green). In conjunction search, processing is non-parallel, a *serial* deployment of attention to one object at a time. To invoke a common metaphor, in conjunction search, focused attention operates like a moveable spotlight illuminating a subset of targets at a time.[6]

Treisman (1988) later modified FIT by postulating a master map of locations at an early stage in processing that serves as the target of focused (spatial) attention (see Figure 1.3).

Two points are worth highlighting. First, Treisman takes focused attention to have the function of binding features, and, in that way, construes attention as selection for object representation and thus for conscious awareness of objects.[7] This explicit connection to conscious awareness provides a distinctive conception of attention (see Chapters 4–6). Second, visual search tasks suggest the possibility that there are two types of attention, one involved in pop-out, the other more like a scanning spotlight. This possibility can also be seen in the final experimental paradigm to be discussed in this chapter, namely, the *Posner Spatial Cueing* paradigm.

## 1.6 The Posner spatial cueing paradigm

One aspect of attention that Helmholtz (Helmholtz 1896) demonstrated is that attention can be deployed *covertly* in a way that is sensitive to spatial location. To shift attention, one does not need to move a relevant sensory organ. In the visual domain, one can *overtly* attend to something by moving one's eyes to it. Such overt attention is plausibly present in other modalities: I can optimally orient my ears to a sound by moving my head; I can reach for an object pressing on my back; I can move closer to sniff

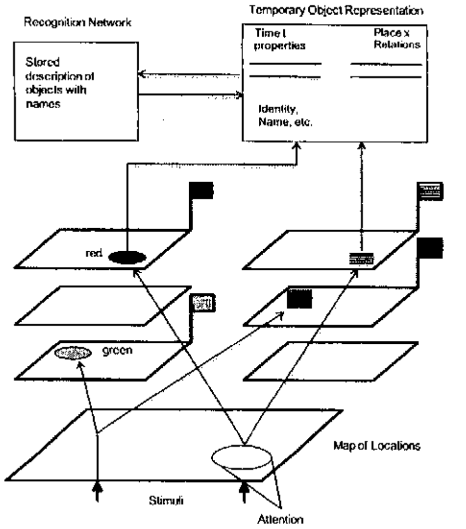


Figure 1.3 Later version of Feature Integration Theory. Redrawn from Treisman (1998). The spotlight of attention focuses on a spatial location in a spatial map that does not code features. Features, rather, are coded in separate feature maps, which give information that the feature is present (the 'flag' in the feature map) and information about the location of the feature. Attention to a location then selects certain features to be bound in an object representation, which can then be compared with stored information or used in other tasks.

something; and I can swish wine in my mouth. Nevertheless, I need not move a part of the body to shift attention, something vividly brought out in the cocktail party effect or looking out of the corner of the eye: I can surreptitiously listen to the more interesting conversation behind our group, even as I feign interest in our conversation, or I can keep my eyes on you while visually attending to something else. Should one then understand there to be two kinds of attention, overt and covert? I think the simplest position is to understand that the movement of a sensory organ is sometimes generated to serve attention, and where it is, attention is overt. There aren't, then, two kinds of attention but rather a single capacity that can involve movement.

In the visual domain, the *spatial cueing paradigm* developed by Michael Posner has become a standard test for the deployment of spatial (covert) attention, namely the selection of spatial location. The subject's task is to report the presence or the features of a visual target. The experiment begins with the subject looking at a screen on which a fixation point is presented and on which the subject must maintain fixation. The fixation point remains, say, for one second, at which point a cue is presented for 100 milliseconds (ms). After this, there is a temporal lag between the cue and the presentation of the target, the *cue-target onset asynchrony* (CTOA). Once the target appears, the subject makes a report.

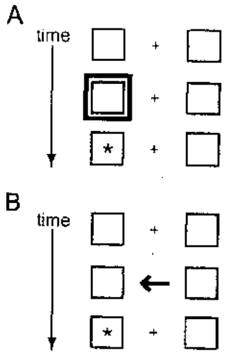


Figure 1.4 A depiction of the Posner spatial cueing paradigm. (A) shows a direct cue which occurs to the left of the fixation "+". In this case, the target, an "*", occurs at the cued location. The cue is a valid cue (an invalid cue would have occurred to the right of fixation, where the object appears on the left). (B) depicts indirect cueing with an arrow pointing to the left, and hence serving as a valid cue (an invalid cue would have pointed to the right with the object appearing on the left). In all cases, there is a temporal interval between cue and target. Reprinted from S. P. Vecera and M. Rizzo (2003) "Spatial Attention: Normal Processes and their Breakdown." *Neurologic Clinics* 21: 575–607 with permission from Elsevier.

There are two types of cues: a *direct* cue that appears at the target location, and an *indirect* or *symbolic* cue, such as an arrow, that indicates a distinct location. *Valid* cues correctly indicate target location, *invalid* cues incorrectly indicate target location, while *neutral* cues provide no information about target location. Within an experiment, cues are typically weighted more towards valid than invalid cues (e.g., about 80% valid-20% invalid, with some small amount of neutral cues, in Posner 1980). The relevant variable of interest can be either reaction time (RT) or response accuracy. The presence of the neutral cue allows comparison of valid versus invalid cueing.

The Posner paradigm has largely been applied to the visual domain, but it has also been deployed in audition (Spence and Driver 1994). What is consistently found is that there are advantages in reaction time and accuracy with valid cues over neutral cues: reaction times are faster and accuracy is higher. Similarly, there is a disadvantage when invalid cues are presented relative to neutral cues: reaction times are slower and accuracy is lower. The idea is that with a valid cue, the subject preemptively moves the

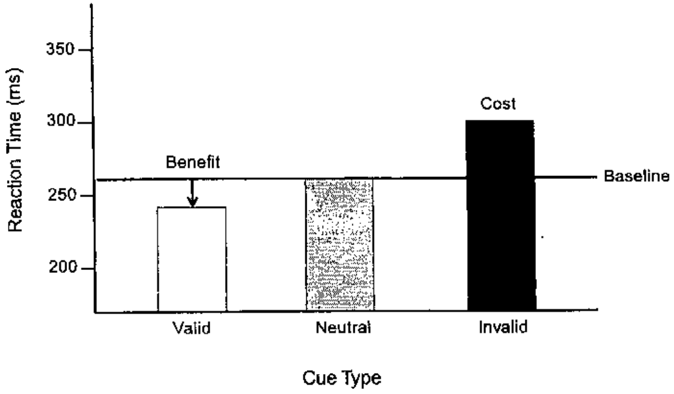


Figure 1.5 Standard effects in respect of reaction time in the Posner spatial cueing paradigm. Invalid cues are associated with a cost, namely, increased reaction time relative to neutral cues, while valid cues are associated with a benefit, namely, decreased reaction time relative to neutral cues. Adapted from Figure 2.6, Wright and Ward (2008), p. 20.

attentional spotlight to the target location with advantage in reaction time and accuracy whereas with the invalid cue, attention is misdirected and must move again to the actual target location, with concomitant cost in reaction time and accuracy.

The temporal differences between pop-out and serial search in the previous section and direct and indirect cueing in the current discussion might suggest two types of attention, or at least two different ways of deploying attention. In the empirical literature, the putative division in attention is often expressed as that between **top-down** versus **bottom-up** attention, although there are a plethora of dichotomies that are also used (see next section). Conjunction search and indirect (symbolic) cueing are often spoken of as top-down attention, while pop-out and direct cueing are often referred to as bottom-up attention (or in some other equivalent terms; this classification is not universally accepted as will be discussed in the next section).

How are top-down and bottom-up attention different? Consider some relevant effects from Posner's spatial cueing paradigm (Carrasco 2011 also gives a summary of relevant differences in Section 3.1). First, recall the cue-target onset asynchrony (CTOA), i.e., the time between onset of the spatial cue and appearance of the target. In Posner's paradigm, direct and indirect cues differ in their facilitation of reaction time (RT), in that direct cues yield a maximum facilitation on RT at a CTOA of 100 ms (i.e., target appears 100ms after the cue), while indirect cues yield the maximum facilitation at a CTOA of 300ms. Second, the benefits of cueing with direct cues are transient and decay fairly rapidly, while those of indirect cues are sustained. These observations suggest that there are different mechanisms underlying the effects of different cues. These differences are depicted in Figure 1.6:

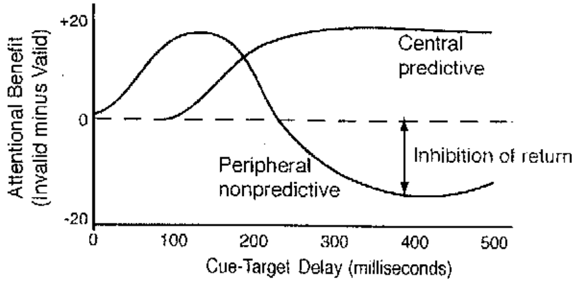


Figure 1.6 This graph shows the time course of the benefit in performance of direct (peripheral) versus indirect (central) cues. In this case, the direct cues were not predictive of target location (i.e., were equally likely to occur at the target location or not). Notice that the peak effect for direct, peripheral cues occurs earlier than that for indirect, central cues. For direct cues, there is also *inhibition of return*, as if attention is repelled for some time from the original cued location. Symbolic cues have a more sustained effect in terms of reaction-time benefit. Reprinted from S. P. Vecera and M. Rizzo (2003) "Spatial Attention: Normal Processes and their Breakdown." *Neurologic Clinics* 21: 575–607 with permission from Elsevier.

Memory load seems to have different effects on direct versus indirect cueing. When subjects are asked to do a task that requires keeping items in working memory (i.e., increased memory load), there are no significant effects on cueing facilitation with direct cues (e.g., in RT), while with indirect cues, the level of facilitation drops off with increased memory load. Perhaps the differences are not surprising. Symbols would seem to require, at a minimum, additional processing of the symbol in terms of its semantic significance. To respond to a symbolic cue like an arrow, one must understand its conventional meaning as an indicator.

This suggests the possibility of different mechanisms in direct and indirect cueing. The top-down and bottom-up distinction, defined at the psychological level, does seem to correspond to a division in underlying networks in the brain. The discussion of attentional networks gained much impetus with the publication of Michael Posner and Steven Petersen's, "The attention system of the human brain" (1990), a work cited over 3500 times in the intervening years and recently revisited by them (Petersen and Posner 2012). Posner and Petersen identified three networks associated with functions commonly attributed to attention: "(a) orienting to sensory events; (b) detecting signals for focal (conscious) processing, and (c) maintaining a vigilant or alert state" (1990, p. 26). In their original discussion, they emphasized that attention forms its own system separate from motor and sensory systems, that attention involves a network of anatomical areas in the brain and that these areas carry out distinct functions (ibid.).

In important imaging work, focusing on Petersen and Posner's proposed orienting network, Maurizio Corbetta and Gordon Shulman (Corbetta and Shulman 2002) later proposed

> that visual attention is controlled by two partially segregated neural systems. One system, which is centered on the dorsal posterior parietal and frontal cortex, is involved in the cognitive selection of sensory information and responses. The second system, which is largely lateralized to the right hemisphere and is centered on the temporoparietal and ventral frontal cortex, is recruited during the detection of behaviorally relevant sensory events, particularly when they are salient and unattended.
>
> (p. 201–2)

The relevant network is diagrammed in Figure 1.7:

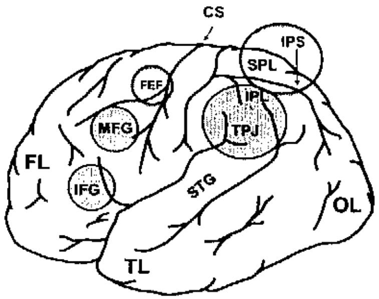


Figure 1.7 Rough localization of the regions of the two attentional networks: the dorsal frontoparietal network (open circles) and the ventral frontoparietal network (circles with gray shading). The former includes the intraparietal sulcus (IPS), the superior parietal lobule (SPL), and the frontal eye field (FEF); the latter includes the temporoparietal junction (TPJ) including the inferior parietal lobule (IPL) and superior temporal gyrus (STG), and the ventral frontal cortex (VFC), which includes the inferior frontal gyrus (IFG) and middle frontal gyrus (MFG). The IPS-FEF network plays a role in both top-down and bottom-up attentional processing; the TPJ-VFC network is involved in bottom-up attentional processing, including circuit-breaking in attentional capture. Frontal Lobe (FL); Occipital Lobe (OL); Temporal Lobe (TL); central sulcus (CS). This map of the attentional network is derived from Corbetta and Shulman (2002), figure 7. The figure is reprinted in adapted form from M. Behrmann, J. J. Geng, and S. Shomstein (2004) "Parietal cortex and attention." *Current opinion in neurobiology* 14: 212–217 with permission of Elsevier. Figure kindly provided by Marlene Behrmann.

The *dorsal frontoparietal network* is characterized as involved in the control of top-down attention. In Corbertta and Shulman's conceptualization, this top-down network generates and maintains an *attentional set*, namely "representations involved in the selection of task-relevant stimuli and responses" (202). It influences perceptual processing so as to serve current task demands, and in that way is sensitive to one's goals. On the other hand, the *ventral frontoparietal network* plays more of a role in bottom-up processing. Among its functions, this network serves as a circuit breaker. That is, certain salient stimuli, such as a loud sound, not only need to attract attention, but also stop other cognitive processes so that the subject can focus on the sudden stimulus. Note also that the two networks do not operate independently: while the dorsal network was recruited under all task conditions Shulman and Corbetta investigated, under bottom-up conditions, the ventral network was *also* recruited (Shulman and Corbetta 2012, 114). So, bottom-up and top-down attention seem to share some of the same neural substrates, but also differ in their neural substrates. The next chapter will return to the question of the neural implementation of attention, but the current task is to more critically scrutinize the conceptual contrasts that have been used to characterize attention.

## 1.7 Divisions of attention

This section considers some common ways of dividing attention:

- Top-down versus bottom-up
- Endogenous versus exogenous (cf. intrinsic versus extrinsic)
- Goal-directed versus stimulus-driven
- Controlled versus automatic
- Voluntary versus involuntary.

How should one understand these concepts so as to fruitfully invoke them in a theory of attention? Alan Allport has made the following observation:

> In general, despite the ingenuity and subtlety of much of the experimental literature that has been devoted to these two enduring controversies [early versus late selection, and the idea of automaticity and control in processing, to be discussed in this section], the key concepts (selection, automaticity, attention, capacity, etc) have remained hopelessly ill-defined

> and/or subject to divergent interpretations. Little wonder that these controversies have remained unresolved.
>
> (Allport 1993, 188)

For the concepts listed above, it is not hard to find papers where most of them are used, often in the same sentence. It is also not hard to find them being understood or applied in different ways between different papers. These notions are presumably technical terms but are never rigorously defined. No wonder Allport thinks there is muddle. Clarity requires definitions, and I shall provide definitions for what I think are the central notions: top-down versus bottom-up, and control versus automatic. Necessarily, the proposed definitions will involve some stipulation, but dissatisfied theorists are asked not to nay-say but to present concrete alternatives.

It is important to be clear that these terms apply to the subject. Thus, it is a psychological subject who exhibits top-down, endogenous, goal-directed, controlled, or voluntary attention. This leaves open other applications of these terms to the brain. For example, theorists speak of a brain region as exerting top-down influence on another region. This is a different use of "top-down" that can be perfectly appropriate, but it ascribes the relevant processing not to the psychological subject, but to a part of her. This recalls my earlier emphasis on the personal versus the subpersonal: some top-down effects are personal, as in attention; others are subpersonal, as in interactions between brain regions. It is no objection to the definitions to be given that they do not apply to interactions between brain regions. They are not intended to describe those interactions.

Let us begin with an initial proposal for *top-down versus bottom-up,* as much early work on attention divided mental processing into stages. Focusing on *perceptual* attention, if one thinks that perceptual processing forms the bottom of a processing hierarchy, then for S as subject and X as target:

> *S*'s attention is **top-down** if and only if *S*'s attention to *X* involves the influence of a non-perceptual psychological state/capacity for its occurrence.
>
> *S*'s attention to *X* is **bottom-up** if and only if *S*'s attention to *X* did not involve a non-perceptual psychological state/capacity for its occurrence.

An intuitive case of top-down attention is where a subject *intends* to pay attention in a certain way, say, to focus on a specific object. Thus, the subject's attending to that target occurs because the subject intends to attend to it. The selection at issue occurs because of the deployment of intention, a non-perceptual psychological capacity. Where perceptual attention happens without needing the influence of non-perceptual psychological capacities, attention is then bottom-up. This covers the intuitive cases when attention is captured by what one perceives, such as a loud bang. What this influence ultimately comes to, mechanistically speaking, is a matter for empirical research. Note that the definitions assume that one can divide the mind into systems, and in particular, between perceptual and non-perceptual systems. It is a good question whether one can adequately do this, an issue that must be set aside. In addition, any non-perceptual psychological system counts as part of the "top". So, motor influences on perceptual selection would count as top-down. Again, this is stipulative, but it allows for clarity.[8]

What of *exogenous* versus *endogenous* sources of attention (sometimes also *intrinsic* versus *extrinsic*)? It is not clear how this distinction differs from the previous. For example, Marisa Carrasco (2011) writes:

> The [endogenous system] is a voluntary system that corresponds to our ability to willfully monitor information at a given location; the [exogenous system] is an involuntary system that corresponds to an automatic orienting response to a location where sudden stimulation has occurred.
>
> (p. 1488)

Carrasco further points out that endogenous attention is sometimes spoken of as *sustained* attention while exogenous attention is spoken of as *transient* attention (recall the temporal properties of direct and indirect cueing discussed in the previous section and depicted in Figure 1.6). It is not clear, however, that the exogenous/endogenous dichotomy comes to anything more than the top-down, bottom-up contrast. For current purposes, I shall treat them as equivalent.

Bottom-up attention maps onto *stimulus-driven* attention, if one thinks of the stimulus as always first dealt with by perceptual systems. Stimulus-driven attention is often contrasted with *goal-directed* attention, but on any plausible account, goal-directed attention is only one type of top-down attention. Goals are, presumably, embodied in intentions or plans, but the account of top-down attention allows for all sorts of non-perceptual influences: memory, expectation, emotion, values, and habits.[9] Accordingly, the contrast between stimulus-driven and goal-directed attention is *not* exhaustive.

There are non-stimulus-driven forms of attention that are also not goal-directed, say my preference for chocolate over fruity candies that leads to my attending to chocolates in a candy store even if I am not intending to buy any candy. The stimulus-driven versus goal-directed contrast falls short of taxonomizing attention.

Things get murky with *control* versus *automatic* attention, on the one hand, and *voluntary* versus *involuntary* attention, on the other. The reason is that these notions point to *agency*. After all, one speaks of a person as being in control or doing something automatically, or of her doing something voluntarily or involuntarily. So, understanding these contrasts requires understanding action, a notion even more challenging than that of attention. I propose to focus on control versus automatic. The voluntary versus involuntary distinction is difficult for it either suggests a kind of agency, such as free agency or agency that involves the will in a specific way, or connotes a characteristic sort of consciousness, something that might be tied to felt effort or a sense of activity. Since the voluntary is tied up with further complex phenomena, it is not likely to help draw clear boundaries in attention.

One can, however, explicate automaticity and control more clearly using the notion of an intention, a goal-representational state. In psychology, the ideas of Richard Shiffrin and Walter Schneider (1977) greatly influenced subsequent discussions of the control-automaticity dichotomy. On automatic processes, they wrote:

> an automatic process can be defined ... as the activation of a sequence of nodes with the following properties: (a) The sequence of nodes (nearly) always becomes active in response to a particular input configuration, where the inputs may be externally or internally generated and include the general situational context. (b) The sequence is activated automatically without the necessity of active control or attention by the subject.
>
> (2)

This proposal connects automaticity to the absence of control (or attention) by the subject. What then is control on their conception? "A *controlled process* is a temporary sequence of nodes activated under control of, and through attention by, the subject" (ibid.). Ignoring the circularity in their definitions, one can take Shiffrin and Schneider as defining automaticity in terms of the absence of control, while control is tied to attention. In contrast, I propose to explicate the notion of control in terms of the role of intention. Here's the basic idea in a slogan: *control in attention is attending as you intend.*

Representations of a subject's goals are embodied in the subject's intentions, namely representations of a plan of action. These plans and their corresponding mental states can be expressed by reports such as *I intend to do X* or *I will do X*. Following Elizabeth Anscombe (1957), philosophers have noted that while actions can be described in many ways, only certain descriptions capture how agents conceive of their actions. That is, they are revealed as intentional only under certain descriptions. Thus, while Gavrilo Princip intended to assassinate Archduke Ferdinand, he did not intend to precipitate the First World War, even if the assassination was identical to the precipitation of war. Those descriptions describe the same action (Davidson 1980).

Control in attention is attention as one intends. Control also implies the absence of automaticity, or automaticity is the absence of control, as Shiffrin and Schneider emphasized. At the same time, if one looks at processes that are controlled, say deliberate actions, one also finds automaticity. You might intentionally throw a ball, but many aspects of your throwing such as its kinematics, the way your joints rotate, and the sequence of movements in your arm are automatic. You don't intend to throw with that speed, rotation or sequence of movements, but your intentional throwing wouldn't be what it is without them. So, despite the contrast between control and automaticity, intentional activities often involve both. How can this be?

Elsewhere, I have argued that one can define automaticity as the absence of control and allow for actions to be simultaneously controlled and automatic only if one relativizes automaticity and control to properties of the process. That is, one speaks of control of a process in respect of a specific feature of that process, and likewise for automaticity. Accordingly, automaticity entails the absence of control, yet a process can be both automatic and controlled *in respect of different properties*. I shall not give here a detailed version of the analyses of control and automaticity (see Wu 2013a), but the following biconditionals capture the essential idea and will suffice for current purposes.

> (C) *S*'s attention to *X* is **controlled** relative to its feature *F* iff *S*'s attention having *F* results from *S*'s intending it to have *F*.[10]

Following Shiffrin and Schneider in defining *automatic* negatively as the absence of control, one derives:

> (A) *S*'s attention to *X* is **automatic** relative to its feature *F* iff *S*'s attention having *F* is not due to control as per (C).

To see how this works, consider visual conjunction search tasks where the target is a red letter **E**. Where the subject attentionally selects a red **E**, her attention's having the feature of *selecting a red* **E** is controlled because it is precisely what the subject intends to do. Similarly, if the subject has her attention captured by a suddenly appearing stimulus, then attending *to that stimulus* is automatic because the subject did not intend to attend to that stimulus. In both cases, the relevant feature F is the subject's attention having the target that it has.

Here is an intuitive gloss of each definition. With top-down versus bottom-up, the key concern is how attention gets *initiated*, i.e., whether the subject is passive or active in that initiation. With top-down attention, the initiation of attention involves and is attributed to the subject due to some non-perceptual mental state including the subject's intentions. In bottom-up attention, by contrast, one can think of the stimulus as initiating attention, even if it disrupts a subject's current activities. On the other hand, think of the control versus automaticity distinction as concerned primarily with the shape of attention once it begins and with *how the features of that process unfold*: where attention is directed and in what sequence, how long it is sustained, to what specific features in the scene, and so on. Finally, it is worth pointing out that top-down and control are sometimes treated as equivalent; in our account, they are not.

Let us now relate the two central dichotomies.[11] Given the previous definitions, there are four categories:

(1) Top-down, controlled attention;
(2) Bottom-up, automatic attention;
(3) Top-down, automatic attention;
(4) Bottom-up, controlled attention.

The first two may not be that surprising, perhaps because one assumes that top-down implies control and bottom-up implies automaticity. In fact, this does not follow, for recall that the top-down/bottom-up distinction is tied to the occurrence of attention while the controlled/automatic distinction is tied to its features.

(1) and (2) are familiar categories. You tell me to follow the man in the fedora, and I attend to him. My attention to him is top-down and controlled. It is top-down because I initiate attention given my intention to follow your instructions, and I would not have done so otherwise. Further, it is controlled because my attention has the feature of being directed at that person as a result of my intention to keep my eyes on that person. In general, intentional forms of attention fit with (1). In a case of (2), a loud continuous sound pulls my attention to it. It thus looks like this capture of my attention occurs independently of any top-down influence, so it looks to be bottom-up. Moreover, given that I don't have any relevant intentions, many of the features of attention might be automatic although perhaps not for long. I hear the sound and subsequently intend to figure out where it is coming from, so attention thereby takes on a controlled aspect. It is sustained according to my intentions. The phenomenon of pop-out in visual search might also seem like a case of bottom-up, automatic attention, but this is controversial.

What of top-down, automatic attention? This seems an odd category, but consider the following experiment by Alfred Yarbus (1967). Yarbus presented his subjects with a painting of a homecoming scene and asked them to perform three tasks: (i) remember what the people in the picture are wearing; (ii) remember the location of people and objects; and (iii) estimate how long the visitor has been away. He then tracked their eye-movements (overt attention) while they carried out his instructions and noted the following patterns:

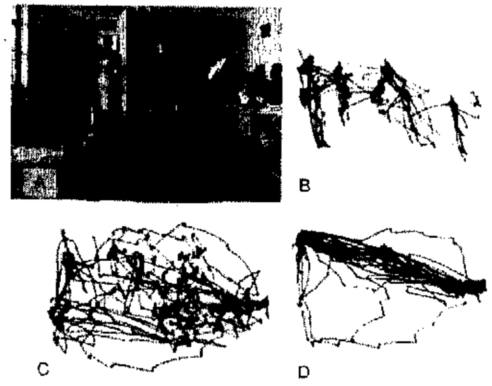


Figure 1.8 Yarbus asked subjects to perform a variety of tasks in relation to I. P. Repin's "Unexpected Visitor" (A). He monitored their eye movements as they visually interrogated the painting in order to perform his tasks. For example, panel (B) indicates the eye movements in response to the command to remember the clothes worn by the people; (C) to remember the position of people and objects in the room; and (D) to estimate how long the visitor has been away from the family. Material from Yarbus (1967), p. 174, figure 109 with kind permission from Springer Science+Business Media B.V. This figure reproduced from "Eye Movements and the Control of Action in Everyday Life" M. F. Land (2006) *Progress in Retinal and Eye Research* 25: 296–324 with permission from Elsevier.

What is striking is that the patterns of eye movements make sense given the subjects' more abstract intentions to carry out Yarbus's instructions. For example, asked to remember the clothes, the subject intentionally looks at each figure. This intention need not be an intention to move one's eyes in any specific way, but the resulting pattern of eye movements in panel B is intelligible given the intention in question: the eyes gravitate around the people without spending time on the objects. Attention in the form of eye movements, *overt* attention, tracks the intention even if the intention is not to move the eyes in a specific pattern. The specific pattern of eye movements happens automatically and is not itself intended. Moreover, the pattern is a feature of overt attention, one that is not represented in the content of the intention. As the pattern is not controlled, it is automatic. At the same time, overt attention with this pattern would not have occurred without the subject having the requisite goal, so attention is top-down. Notice that when one toggles the subject's intention by presenting different tasks, the pattern of eye movement changes.[12] So, intentions are involved in the occurrence of overt attention with a characteristic pattern. While the idea of top-down automatic attention might seem contradictory, it is not. That we can categorize the phenomenon Yarbus observed suggests that the initial analysis is theoretically useful. This is a sign that the definitions are on the right track.

It seems likely that no process instantiates (4) since the causal processes imputed by each dichotomy operate at cross purposes: bottom-up attention requires a stimulus-driven initiation independent of any intentions, but control requires the influence of an intention. At best, attention might be bottom-up and automatic but quickly *becomes* controlled once intentions kick in to sustain attention to the stimulus. In any event, I want to conclude the discussion of the conceptual issues by returning to category (2): bottom-up, automatic attention when attention functions like a circuit breaker. This seems like an obvious, familiar category. Yet like (4), there are questions whether (2) is ever instantiated.

The previous section noted the difference between direct and indirect cueing. Richard Wright and Lawrence Ward (2008) suggest the following:

> Researchers can choose to study either voluntary or involuntary orienting, depending on whether they use symbolic or direct location cues ... Symbolic location cues initiate attention shifts in a fundamentally different way than direct location cues. The former are meaningfully associated with a particular location and therefore must be interpreted by an observer in order to be used. For this reason, the initiation of an attention shift by a symbolic cue is *goal-driven*. The observer processes the location information conveyed by the symbol and, on this basis, develops a computational goal for carrying out the task ... Direct cues, on the other hand, produce their effect by virtue of being physically close to the target location ... No cognitive interpretation of direct-cue meaning is required and, instead, attention is captured by the onset of the cue. For this reason, the initiation of an attention shift by a direct cue is *stimulus-driven*.
>
> (21–22)

It is natural to take the direct cue as capturing attention and in that way independent of goals. But is it goal-independent? Bradley Gibson and Erin Kelsey (1998) suggest that the influence of the direct cue is *goal-directed* (p. 699): "stimulus-driven attentional capture may be caused by goal-directed processes." How can this be?

In discussing Feature Integration Theory (FIT), I noted that feature singletons (i.e., features that are unique within a feature map such as a red shape in a sea of green shapes) seem to pop out. It would be natural to characterize pop-out as the capture of attention as occurs with auditory attention and loud noises.[13] John Jonides and Steven Yantis (1984; 1988) have argued, however, that most cases of pop-out in the attention literature are not genuinely bottom-up, automatic capture of attention but depend on the subject's goals. Hence, they are *top-down*! Consider the visual search tasks discussed above when the target seems to pop out. To undertake the task, you have to follow task instructions, say to locate a green **T**. Yet in intending to locate a green **T**, you've set yourself to complete a specific task. Locating that target is your explicit goal. The target *you intend to locate* is precisely what pops out. Again, it is top-down, and your locating it reflects attentional control.[14] Of course, there are automatic elements. What you don't control, and hence what is automatic, is *when* you locate the target. That you locate it in a way independent of set size reflects the automaticity with respect to when you locate it (reaction time is the same). Nevertheless, attending to the **T** is top-down and controlled. My definitions show how one can consistently and clearly apply the concepts of top-down, control, and automaticity to the same phenomenon.

It is striking that the original pop-out effects might in fact be top-down and controlled rather than bottom-up and automatic. Indeed, Charles Folk, Roger Remington, and colleagues (1992) claim that there are no pure cases of bottom-up, stimulus-driven attention (for a methodical review of the issues and experimental evidence, see Burnham 2007). One can pose the issue as a challenge: is attention ever independent of the goals of the perceiver?[15] The claim is that goals have a pervasive influence on attention. Still, it is hard to accept the claim that there is *never* attentional capture contrary to one's goals. Consider being engrossed in a performance of Beethoven's Ninth Symphony. Just before the climactic moment of the famous chorus, I pinch your shoulder. This is quite annoying, of course, since it breaks your concentration on the music, but it also does seem to be a compelling case of tactile attentional capture. Attention is devoted to *auditory* experience, as you listen to the music. Your intention for the past hour has been to listen, your attention has been focused on the music. There do not seem, then, to be any goals where tactile inputs are relevant. This is, of course, an anecdote, but a prima facie compelling one.

The dichotomies discussed in this section have been deployed for a long time in the study of attention, and they are well entrenched in psychological vocabulary. At the same time, there is something casual and slippery about their use that needs to be avoided once they are deployed in serious theory building. The proposals I have given provide concrete accounts of what these dichotomies come to. I suggest that barring any other concrete definitions (and there are none in the literature that I am aware of), theorists should start with the ones presented here.

## 1.8 A sufficient condition for attention: selection for task

There is a central idea towards which all the theories, paradigms, and conceptual dichotomies discussed thus far gravitate: the notion of a *task*. For example, the Load Theory of Attention argues for the task-dependence of where attention acts in perceptual processing. Further, the subject's goals pervasively influence attention, so much so that some theorists have questioned whether there is attention without the influence of goals. Finally, three specific experimental paradigms have been central to the psychological study of attention: dichotic listening, visual search, and spatial cueing. In each of these, a well-defined task structures the experiments. Given the centrality of tasks, might appeal to it provide a way to answer the growing skepticism to explaining what attention is?

A well-defined experimental task establishes conditions such that when they are fulfilled, the experimenter is confident that the subject has deployed the capacity the experimenter is studying. Specifically, where the subject *has followed task instructions and correctly performed the task*, the experimenter can be confident that the capacity in question has been deployed. Consider then studies of attention using verbal shadowing in dichotic listening paradigms. Where the subject correctly shadows the verbal stream assigned, the experimenter can be confident that the subject is attending to that stream, using the sounds in that stream to inform verbal response. Next, consider the use of reaction time to measure task performance in visual search and spatial cueing. The reaction at issue in both experiments is target detection, and reaction time reflects the temporal properties of attention in serving that task. Given that subjects perform that task, namely, producing a judgment about the target's presence or absence, this performance is a sign that the subjects have been attending to the relevant target, using it to render a judgment. This can also be discerned by looking at eye movements during the task. Obviously, when subjects are not doing the task, say when they twiddle their thumbs or continuously get things wrong, this is evidence that they are not appropriately selecting the relevant target and are being inattentive.

In the three experimental paradigms that I have discussed, it is clear that, for each, there is a well-defined target, reaction to which requires selection of that target to inform the response, whether tracking a conversation in verbal shadowing or examining targets in target detection. As these experiments are used to probe attention, there is a general assumption that all experimenters on attention hold in using these paradigms:

> **Empirical Sufficient Condition for Attention ($S_{emp}$):** Subject *S* perceptually attends to *X* if *S* perceptually selects *X* to guide performance of some experimental task *T*, i.e., selects *X* for that task.

Where the subject selects some target to guide their response in carrying out an instructed task, then the subject's selecting of that target is sufficient for the subject's attending to that target. Notice that the condition introduces a variable for the targets of attention and selection, targets that can be information, locations, features, or objects. Thus, dichotic listening and visual search involves the tracking of features and objects, say visible and audible entities and their properties, while Posner's spatial cueing paradigm tests, in part, for attention to locations. In what follows, the focus will be on locations, features (properties) and objects as targets of attention.

One might wonder why not just say that the subject's selecting *X* is just the subject's attending to X. This would, however, require that selecting *X* is a necessary condition as well, but that is controversial and more difficult to establish (I shall try to establish it in Chapter 3 by defending a selection for action account of attention). For current purposes, the sufficient condition provides an answer to the skepticism noted in the Introduction. For all their doubts concerning answering the metaphysical question, theorists of attention have done much interesting and important experimental work on attention. Furthermore, an assumption in their experimental practice, namely the empirical sufficient condition, provides a shared condition on attention that is relevant to the metaphysical question. Of course, not all sufficient conditions for a phenomenon are informative as to its nature. That someone wins the Electoral College in the U.S. presidential election is sufficient for their becoming U.S. president, but winning the Electoral College doesn't illuminate what a president is. The interest of the empirical sufficient condition is that it begins to flesh out talk of attention as selection by drawing on an assumption in experimental work on attention.

On reflection, this should not be surprising. Any experimentalist who wants a subject to direct attention knows how to do it, namely by having the subject perform specific tasks with respect to a target. That is, if the experimentalist wants to ensure that the subject attends to some X, then the experimenter designs a task where X is task-relevant and where X must be used to perform the task. To study attention, one needs to know how to manipulate it and to keep track of it. A well-designed experimental task is precisely one that creates conditions such that one can do so, and this is just manipulating the subject's task performance by manipulating what the subject must selectively respond to. The empirical sufficient condition then identifies a widely held assumption in empirical work on attention that can serve as an initial foothold in the face of skepticism about what attention is.

## 1.9 Summary

This chapter began by highlighting five basic questions, and in discussing the fruits of psychological research in the last 70 years, uncovered many answers, especially to the *properties question*. Of note, the properties of attention seem to point to two forms of attention. These forms have different temporal profiles regarding when they exert their greatest effect and how long they last, they have different dependencies on memory, and they seem to call on overlapping, but different, neural networks. As a result, it has been natural to divide attention, and this has led to two salient divisions when characterizing attention: top-down versus bottom-up and control versus automaticity. I have provided an analysis of the resulting dichotomies of attention, and highlighted an interesting category of attention, namely a top-down, automatic form illustrated in Yarbus's eye-tracking experiments. This work provides a more detailed picture of our capacity to attend.

The *function question* has also received some interesting answers that in turn suggest a possible answer to the *metaphysical question*. I have canvassed conceptions of attention as a filter and a spotlight. On the one hand, Broadbent emphasized filtering to explain the role of attention in perceptual processing, namely in selecting relevant information for further work-up. On the other hand, spotlighting suggests a phenomenal aspect to attention and has an echo in Treisman's talk of attention as selecting features to bind for conscious awareness of objects (she spoke of attention as "glue" for feature binding). I will examine the phenomenal conception of attention and attention's relation to consciousness in later chapters (Chapters 4–6), but the current discussion revealed an interesting property of attentional filtering, namely that the stage at which attention acts, namely, early versus late, seems to be task-dependent as hypothesized by the Load Theory. Rather than attention being tied to a specific stage in processing, perhaps it is tied to the task that the agent performs. Indeed, in the last section, I argued that an implicit assumption in experimental paradigms used to probe attention is that a subject's selecting an item to inform task performance is sufficient for the subject's attending to that item. This then points to another possible answer to the metaphysical question: might attention be selection for task, indeed, for action?

## Suggested reading

Mole (2011) and Hatfield (1998) discuss psychological work on attention pre-1950, and Tsotsos (2011, chap. 1) presents a nice overview as well. A succinct account of the psychology of attention from the 1950s onwards is provided for in Driver (2001). Pashler (1998) is a monograph discussion of similar terrain. Relevant recent review articles on the psychology of attention can be found in Posner (2011). Lavie and Tsal (1994) provide a discussion of the Load Theory of attention in light of the early versus late selection debate. On visual search and Feature Integration Theory, Treisman (1988) provides an overview, while Wolfe (1994) provides an update of his version of visual search. Wright and Ward (2008) provide an excellent overview of orienting and attention. Allport (1993) provides a well-known critique of 25 years of attention research, while Carrasco (2011) provides a

# 6

# ATTENTION AS THE GATEKEEPER FOR CONSCIOUSNESS: COGNITIVE ACCESS

## 6.1 Introduction

In Chapter 5, I contrasted the commonsense model with the gatekeeping model of the relation between attention and consciousness:

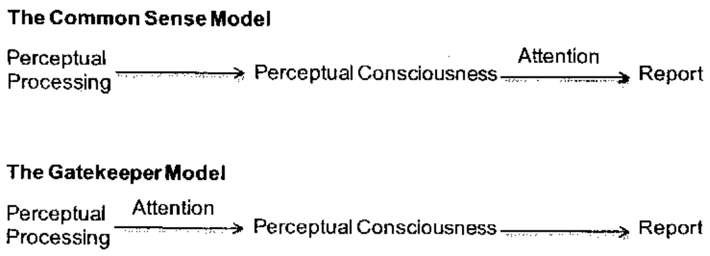


Figure 6.1 The Gatekeeper and Common Sense Models.

This chapter continues with a different elaboration of the two models that focuses on a connection between attention and *cognitive access*. Cognitive access to X – access to X by (some form of) cognition – implies attention. The canonical case of cognitive access to be discussed at length is access to X by working memory systems, where this access is mediated by attention. On some views to be discussed, attention is for working memory, and in

that way attention is for consciousness. This leads to the following elaboration of the previous models:

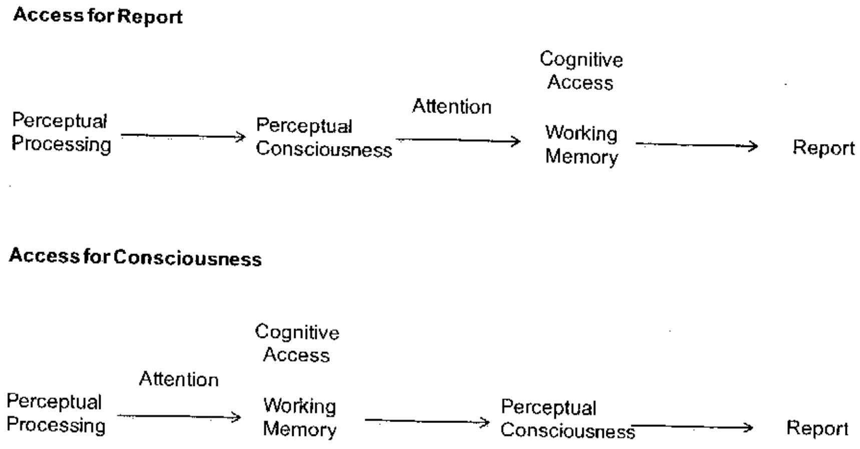


Figure 6.2 Cognitive Access and the Gatekeeper and Common Sense Models.

The crucial new element is the insertion of working memory as tied to cognitive access, and while there are complications, given how "access" is used, the simplest elaboration is to identify attention as for working memory and access as encoding in working memory. This results in a switch in emphasis in respect of the gatekeeping thesis. In the last chapter, the focus was on what attention is directed at in perception as determining the character of consciousness. In this chapter, the focus is on what attention delivers to working memory as determining the character of consciousness. Put another way, the shift is from focusing on attention in perception to attention for memory.

As the issues regarding memory and consciousness have been discussed in terms of access, Section 6.2 introduces Ned Block's notion of access consciousness, discusses different applications of the notions of access and accessibility, and ties access to a notion of attention for cognition. Section 6.3 examines two empirical theories of consciousness that take attention for working memory as a necessary condition for phenomenal consciousness. Then, Section 6.4 presents a famous experiment by George Sperling that has provided support for those who argue that attention does not limit phenomenology. Section 6.5 introduces Block's thesis that phenomenology *overflows* access, while section 6.6 discusses different responses to Block's thesis and considers the experimental evidence relevant to assessing that thesis. Finally, Section 6.7 briefly discusses a neurobiological argument for overflow due to Victor Lamme.

## 6.2 Phenomenology and access

Ned Block (1995) introduced a distinction between *phenomenal* consciousness (P-consciousness) and *access* consciousness (A-consciousness). As discussed in previous chapters, phenomenal consciousness is what it is like for the subject. Block characterized A-consciousness as follows:

> A state is A-conscious if it is poised for direct control of thought and action. To add more detail, a representation is A-conscious if it is poised for free use in reasoning and for direct "rational" control of action and speech. (The "rational" is meant to rule out the kind of control that obtains in blindsight.)
>
> (In the version printed in Block 2007b, 168)[1]

A-conscious representations are poised for access in the sense of being *accessible* for use by action systems (with "action" broadly construed). When such representations are in fact used, then they are *accessed*. So, the central notions are access and accessibility. These notions, however, must be deployed with care. One of Block's early examples of P-consciousness without A-consciousness is the following:

> Suppose that you are engaged in intense conversation when suddenly at noon you realize that right outside your window, there is – and has been for some time – a pneumatic drill digging up the street. You were aware of the noise all along, one might say, but only at noon are you consciously aware of it. That is, you were P-conscious of the noise all along, but at noon you are both P-conscious and A-conscious of it.
>
> (Block 2007b, p. 174)

How does the access/accessibility distinction apply in this context? Certainly, before one notices the drilling, one doesn't have access to it in the sense that it does not guide or prompt a report of the sound. Were one to make a report, then one accesses that information to guide behavior. Block also points out that you might realize that the drilling has been going on for some time. This realization calls upon information from memory, but until the memory is recalled, it is only accessible to and not yet accessed by report.[2] Yet there is a further dimension, for imagine that as the buzzing occurs during your blissful unawareness of it, your perceptual system registers the sound. There might be a further step between registering the sound in perception to being accessible to encoding in memory, so it is possible to have perceptual information about the drill without this being accessible to memory. This leads to two further distinctions: (a) perceptual representations of the drill that are accessible to working memory, but short of being actually accessed by (encoded in) working memory; and (b) perceptual representations that are not even accessible to memory. The challenge is that the notions of access and accessibility can be used to describe different points in processing. This means that talk of cognitive access or accessibility might refer to different things, and that in discussing the elaboration of the two models, one must be explicit about which meaning is intended lest confusion ensue.[3]

To regiment the use of the notions of access and accessibility, consider the following flow of information:

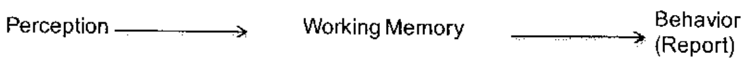


Figure 6.3 The Flow of Information via Working Memory.

Given discussion in the literature on cognitive access, four stages are salient:

1. Perceptually encoded, currently inaccessible, but potentially accessible to memory
2. Perceptually encoded, accessible to, but not yet accessed by, memory
3. Encoded in memory, accessible to, but not yet accessed by, behavior
4. Accessed from memory to guide behavior such as report.

To keep things orderly, I will use the access/accessibility distinction to describe the flow of information from perception to action where one always speaks of X's *accessibility-to-Y* or X's *being accessed-by-Y* (i.e., *Y's accessing X*). In these locutions, Y identifies a system to which information from X is sent (e.g., working memory, reporting systems). So, in (1), perception is only potentially accessible to working memory; in (2), perception is actually accessible to working memory; in (3), perception is in fact accessed by working memory, but only accessible to behavior systems; in (4), working memory is then accessed by behavior systems. The required regimentation is to always be clear on what the arguments for X and Y are. The focus in what follows is largely on (2) and (3) with ultimate emphasis on (3).[4]

Is A-consciousness necessary for P-consciousness? Note that there are two interpretations of "A", namely, "access" or "accessibility". To disambiguate, I will drop talk of A-consciousness and focus on the difference between access and accessibility. This leads to the following claims:

(A1) Subject *S* is P-conscious of *X* only if *X* is accessed by *S*.

(A2) Subject *S* is P-conscious of *X* only if *X* is accessible to *S*.

For these claims, that X is accessed or accessible to S implies that X is accessed by, or accessible to, respectively, S's working memory. If access or accessibility is tied to attention, then we have a version of the gatekeeper view. It is *prima facie* plausible that access is tied to attention, for we access an item X for some T by selecting it for T (e.g. let "T" stand for task, action, or phenomenal consciousness). For the relevant T, say a task, selection of X for T suffices for attention to X for T. Since accessibility is defined as potential access, both notions are then tied to attention. (A1) and (A2) can then be used to derive the following gatekeeping (GK) theses:

($\text{GK}_{\text{A1}}$) Subject *S* is P-conscious of *X* only if *S* attends to *X* for working memory.

($\text{GK}_{\text{A2}}$) Subject *S* is P-conscious of *X* only if *S* could attend to *X* for working memory.[5]

If one were inclined to think that *actual* report of a stimulus (verbal or some relevant behavior) is a necessary condition for P-consciousness, then one would endorse (A1) and hold that P-consciousness arises only when processing reaches stage (4). It is not clear that anyone holds this view except, perhaps, a hard-headed behaviorist. Instead, most hold that reports or relevant behavior provide evidence for phenomenal consciousness, but that actual reports are not necessary for consciousness. Those who endorse (A1) will require access in terms of stage (3) as necessary for phenomenal consciousness (see Global Workspace Theory in Section 6.3.1). Encoding in working memory determines the content of consciousness. Those who endorse (A2) will only require accessibility in terms of (2) as necessary for phenomenal consciousness (see Attended Intermediate Representations Theory in Section 6.3.2). Finally, those who deny that P-consciousness implies access or accessibility will claim that one can have perceptual consciousness without reaching any of stages (2)-(4). To return to (1), there can be perception that is not (currently) accessible to working memory, and yet is conscious. It is not clear that anyone holds this view.[6] I will focus on (A1) since all the parties in the debate endorse some version of (A2).

## 6.3 Two empirical theories of consciousness

This section examines two empirical theories of consciousness that provide accounts of access and its relation to phenomenal consciousness: the Global Workspace Theory and the Attended Intermediate Representations (AIR) Theory. The first theory was initially presented by Bernard Baars (1988), although I shall focus on recent elaborations by Stanislas Dehaene and Lionel Naccache (2001); the second is defended by Jesse Prinz (2012). Both theories share an assumption about the relation between information carried by neurons and the contents of consciousness, namely, the *content realization principle* (CRP):

(CRP) There is a necessary correlation between the content of consciousness and the information carried by the neural realizers of consciousness.

CRP implies that conscious content correlates or covaries with neural information. Thus: let neural population N realize conscious state C with content P. Then the information I in N realizes P such that, where there is a change in the content of C, there is also a change in information in N, and *vice versa*. CRP leads to the following question: Why does some information rise to the level of conscious content while other information does not?

### 6.3.1 The Global Workspace Theory

In the prologue to his book, *In the Theater of Consciousness*, Bernard Baars writes: "Consciousness seems to be the publicity organ of the brain. It is a facility for *accessing, disseminating, and exchanging information, and for exercising global coordination and control*" (1997, 7). This functional conception that originated with Baars (1988) has been more explicitly linked by Dehaene and Naccache to the organization of the brain:

> the human brain also comprises a distributed neural system or "workspace" with long-distance connectivity that can potentially interconnect

> multiple specialized brain areas in a coordinated, though variable manner ... The global workspace thus provides a common "communication protocol" through which a particularly large potential for the combination of multiple input, output, and internal systems becomes available.
>
> (2001, 13)

We shall focus on Dehaene and Naccache's account.[7] The general picture can be depicted as follows:

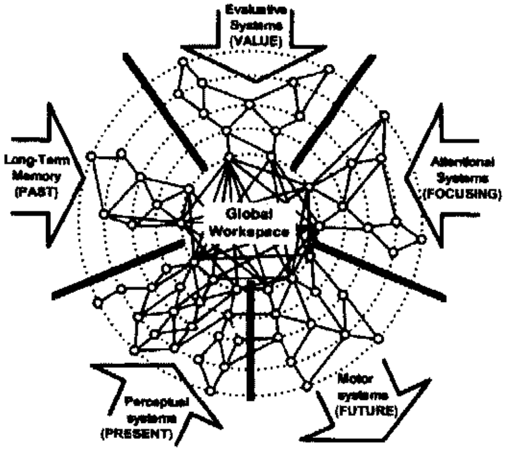


*Figure 6.4* Model of the neural global workspace from S. Dehaene, M. Kerszberg and J.-P. Changeaux (1998) "A neuronal model of a global workspace in effortful cognitive tasks." *Proceedings of the National Academy*, USA 95: 14529–34. Copyright (1998) National Academy of Sciences, U.S.A. Figure courtesy of Stanislas Dehaene.

This "neural" version of Global Workspace Theory focuses on the structure of specific networks in the brain. It is important to note, however, that there is no single brain structure that constitutes the global workspace, though neural workspace theorists tend to emphasize the frontal and parietal lobes (the *fronto-parietal* network). Rather, the issue concerns the activity of regions of the brain that have the requisite connectivity. Thus there is an

> absence of a sharp anatomical delineation of the workspace system. In time, the contours of the workspace fluctuate as different brain circuits are temporarily mobilized, then demobilized. It would therefore be incorrect to identify the workspace, and therefore consciousness, with a fixed set of brain areas. Rather, many brain areas contain workspace neurons with the appropriate long-distance and widespread connectivity, and at any given time only a fraction of these neurons constitute the mobilized workspace.
>
> (op. cit., p. 14)

The workspace is then not just an anatomical notion but a functional characterization of a widely distributed, dynamic neural network. It is a network that makes available information to multiple systems. This requires that the network be realized in circuits that have broad and long-range connections to other parts of the brain. Information that is in the workspace can then be *broadcast* to (accessed by) other systems.

To understand how this provides a theory of consciousness, one must understand what Dehaene and Naccache take consciousness to be. They account for a *transitive* notion of consciousness, as when one speaks of the consciousness *of* color (recall CRP). Further, this notion of consciousness is necessarily tied to reportability (indeed, they call transitive consciousness "access to conscious report" (Dehaene et al. 2006).[8] On their view, the idea of consciousness that is not reportable, i.e. accessible, is empirically empty (Naccache and Dehaene 2007). So, consciousness of X requires that information regarding X be encoded in the workspace so as to be accessible to guide report and other behaviors. This implicates working memory. How then is information from perception that is available to working memory *encoded* in working memory? Dehaene and Naccache attribute this role to *top-down attention*. It is attentional selection that determines which accessible perceptual representations become encoded in, and thus accessed by, working memory. In terms of the neural global workspace, the upshot of attention is that a larger part of the workspace becomes active, spanning the parietal and frontal regions. The *global* workspace is thereby engaged. Recalling CRP, one can say that, on the Dehaene/Naccache theory, subjects are in conscious states with content P when relevant information is modulated by attention such that it is encoded in the global workspace and accessible to behavior. Attention, then, is the gatekeeper for consciousness by serving as the gatekeeper for working memory (recall Figure 6.2).

### 6.3.2 Attended Intermediate Representations (AIR) Theory

Jesse Prinz (2012) has argued for a theory of consciousness that endorses the following:

> (AIR) Consciousness arises when and only when intermediate-level representations are modulated by attention.
>
> (89)

While AIR (Attended Intermediate Representations) applies to all forms of consciousness, Prinz largely focuses on visual consciousness. Accordingly, AIR applied to vision holds that attention to intermediate visual representations is necessary and sufficient for consciousness. In earlier chapters, I argued against sufficiency and responded to Prinz's defense, but, in this chapter, it is the necessary condition that matters.

Prinz begins with ideas originally presented by Ray Jackendoff (1987) who himself drew on David Marr's (1982) seminal book, *Vision*, where visual processing is divided into distinct stages. For discussion purposes, understand the division of visual processing as follows:

> ***Low-level vision***: where basic features such as edges are processed;
>
> ***Intermediate-level vision***: that represents the world in a viewpoint-dependent way capturing object boundaries, textures, and depth;
>
> ***High-level vision***: that abstracts away from viewpoint and involves categorical representations of objects and properties.

Thus, if the visual stimulus is Bill Clinton's face, then low-level vision encodes basic visual properties like the *boundaries* of the face; intermediate level vision encodes a viewpoint, say, a *lateral profile* of the face, if one is looking at Clinton from the side, and high-level vision encodes its being *Clinton's* face, a representation that might be activated whatever view one has of his face (e.g., head-on versus from the side). Where does visual information become conscious? Prinz follows Jackendoff in emphasizing that consciousness arises at the intermediate-level, for the representational content of visual experience correlates best with intermediate rather than low- or high-level vision (again, recall CRP). That is, visual experience is tied to a viewpoint and in some sense presents objects as relative to the location that the perceiver occupies. Objects look the way they do given that viewpoint (cognitive scientists speak of *egocentric* representations which are, presumably, at least a subset of the intermediate representations Prinz has in mind). Prinz makes further claims about the neural realization of intermediate representations, suggesting that they involve specific parts of the visual system such as visual areas V2, V3, V4, and V5 (also known as MT, the middle temporal area), among other areas (2012, p. 52).

Of more direct concern is Prinz's conception of attention. His strategy is to look for a common mechanism which is found in all cases of attention and which might then serve as the referent of the term "attention" (2012, p. 91). A good candidate is *change in information flow*. Specifically, a stimulus that is attended

> becomes available for processes that are controlled and deliberative. For example, we can *report* the stimulus that we consciously perceive, we can reason about it, we can keep it in our minds for a while, and we can willfully choose to examine it further.
>
> (92)

Given the discussion in earlier chapters, this passage might make one think of change in information flow as selection for *action*. Prinz's focus, however, is more specific, for he sees a connection to *working memory*: "attention can be identified with the processes that allow information to be encoded in working memory" (93). He characterizes working memory as "a short-term storage capacity that allows for 'executive control'" (92). One might wonder whether this leaves out a simple form of attention, e.g., when one is directly acting on an object currently perceived. Here, attention might seem to serve action, not working memory.[9] In any case, the link between attention and working memory provides Prinz a functional analysis of the folk concept of attention (95) and leads to an unpacking of AIR:

> (AIR) Consciousness arises when and only when intermediate-level representations undergo changes that allow them to become available to working memory.
>
> (97)

Note that both AIR and Global Workspace Theory acknowledge a role for attention and working memory, in that it is attention for working memory that explains conscious content (hence, attention for cognition). AIR theory differs from the Global Workspace theory in that, while the latter ties conscious content to information *encoded* in working memory, AIR ties it to information *available* to working memory. Put in terms of access and accessibility, Global Workspace theory takes P-consciousness to depend on access, in that information must be *accessed* by working memory (hence, encoded; thesis (A1) Section 6.2); AIR theory takes P-consciousness to depend on accessibility in that information must be *accessible* to working memory (thesis (A2), section 6.2). In part, Prinz favors AIR due to some evidence suggesting that the elimination of working memory, and thus working memory encoding, does not eliminate consciousness (2012, Chapter 3).[10] What one can say is that both AIR and Global Workspace theory agree that attention, in the sense of selection that is tied in some way to working memory, is necessary for phenomenal consciousness. Thus, both theories entail gatekeeping.

### 6.3.3 *Attention, A- and P-consciousness: the issues*

There are two theses about the dependency of P-consciousness on access/accessibility.

(A1) Subject *S* is P-conscious of *X* only if *X* is accessed by *S*.

(A2) Subject *S* is P-conscious of *X* only if *X* is accessible to *S*.

Attention is relevant because it provides a route to cognitive access/accessibility in respect of selection for working memory. Thus, on AIR theory, attention renders intermediate perceptual representations accessible to working memory while in Global Workspace theory (GWT), attention allows working memory to access perceptual representations.

(GWT) *X* is accessed by *S* only if *S* attends to *X*.

(AIR) *X* is accessible to *S* only if *S* attends to *X*.[11]

In this way, attention, by being tied to access or accessibility, serves as a gatekeeper for phenomenal consciousness by being a gatekeeper for working memory. For (A1) conjoined with (GWT), and (A2) conjoined with (AIR), imply a familiar gatekeeping conditional: Subject S is P-conscious of X only if S attends to X, i.e., only if S selects X in some way for working memory. Let us simplify matters by focusing on GWT and (A1). (A1) implies (A2) in that if S accesses X, then X is accessible to S.[12] This then makes the challenge to the gatekeeping view very specific: Can it be demonstrated that there is phenomenal consciousness outside of what is encoded in working memory?

In the last chapter, I argued that experiments aimed at teasing apart different models of attention's role in consciousness falter because the conditions of inattention needed to show inattentional blindness suffice to undercut the ability to report the stimulus. At the same time, since report is how one gains access to consciousness and report implicates attention, then it looks like the primary evidence for consciousness cannot also provide evidence for consciousness in the absence of attention. The consciousness one attests to in a report is also consciousness to which one is attentive. This raises what Ned Block (Block 2007b) has called a *Methodological Puzzle:* How can one experimentally address the issues given the limitations just noted?

Block's solution to the puzzle is to deploy *inference to the best explanation*. That is, he advocates choosing the model that best explains the relevant data. His argument against the gatekeeping view can be reconstructed as follows:

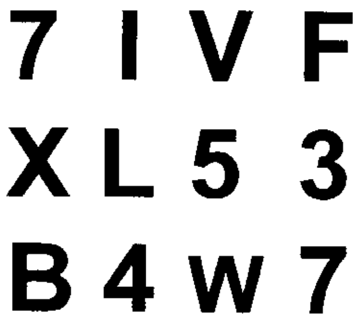
1. Visual working memory (the workspace) has a limited capacity.
2. Overflow: phenomenology has a higher capacity than working memory
3. "The control of working memory is in the front of the head" (496).
4. Arguably, the "core neural basis of visual phenomenology is in the back of the head" (ibid.).
5. If one assumes that the machinery controlling working memory is necessary for visual phenomenology, then one cannot explain overflow.
6. If one assumes that the machinery controlling working memory is not necessary for visual phenomenology, than one can explain overflow.

The idea is that the best explanation of overflow is that the machinery of phenomenology is distinct from the machinery of working memory. Overflow implies that phenomenal consciousness is not limited by attention for working memory, for the capacity of phenomenology is greater than the capacity of working memory. But why accept overflow?

## 6.4 Sperling, partial reports, and iconic memory

In 1960, George Sperling published a paper titled, "The information available in brief visual presentations" (Sperling 1960). Sperling's question was: How much does one see in a glance? To answer this, he presented visual stimuli to subjects for very brief durations, an experimental paradigm with a very long history dating back to the late nineteenth century. In those earlier studies, subjects were asked to report what they saw of briefly presented stimuli, and thus they had to draw on memory of the stimuli.[13] As Sperling noted, a repeated early finding was that subjects could only report a subset of what was presented to them. At the same time, *subjects typically claimed to see more than they could report*. Sperling's advance was to take this sense of seeing more than can be reported as a basis for asking a further question: Does one see more than can be remembered? An answer to this question is directly relevant to the gatekeeper view.

Sperling's goal was to determine the informational capacity of what is seen and whether this is tied to the capacity of memory. He recognized, however, that if memory for report is limited, then attempts to report everything that was seen (*total report*) can never exceed the capacity of memory for report (what is now called working memory). Accordingly, he opted for a *partial report* paradigm: the subject reports only on part of what was seen, as determined by task instructions. Sperling's ingenious approach was to use partial reports to circumvent the limits of working memory as revealed in total reports.



7 I V F

X L 5 3

B 4 W 7

Figure 6.5 Letter Array in the Sperling Partial Report Paradigm.

Sperling presented subjects with stimuli containing a number of numerals and letters (from 3–12) in various configurations. A sample 12-figure configuration is reproduced here:

When his subjects were asked to give a total report of the identity of the letters flashed, they were able to report on average 4.3 letters (experiment 1, p. 6). This estimate was stable across changes in stimulation durations from 0.015 to 0.5 seconds (experiment 2, p. 6). In his third experiment, Sperling shifted to partial report where subjects were required to report no more than four letters from a stimulus display. Consider then a presentation of 12 letters in three lines arranged top to bottom, with four letters per line (Figure 6.5). Sperling used a tone after stimulus presentation to indicate randomly which line subjects were to report. He assumed that if subjects used the tone to tap into a specific part of a memory representation of the array, namely, that corresponding to the cued line, then by taking the number of letters reported in partial reports and multiplying it by the number of lines in the array, he could obtain an estimate of the total number of letters that were seen. Doing this, Sperling found the number of letters reported to be on average 9.1, about three of four letters in each line. In other words, using partial reports, what was perceptually available—and presumably seen—was measured to be about nine letters; using total reports, what was remembered was measured to be about four letters. So, visual capacity exceeds working memory capacity. The effect is called *the partial report advantage* and lasts for about 300ms after the stimulus is removed, the *stimulus offset*. The work has become one of the classic experiments in modern psychology.

What seems to be largely uncontroversial is that Sperling showed that: (a) what is seen, in the specific sense of information processed by the visual system, can persist after stimulus offset; (b) that it can be accessed in report as in the partial report paradigm; and (c) the content of what is seen exceeds the content of what Sperling spoke of as immediate memory (i.e., working memory). It is a further question how to use Sperling's results to adjudicate questions about the gatekeeper view.

The persistence of what is seen after stimulus offset is *visual persistence*. Max Coltheart (1980) suggested, however, that "visual persistence" is ambiguous between *neural, visible,* and *informational persistence*. By "neural persistence" Coltheart referred to the persistent activity of visual neurons after the stimulus is removed; by "visible persistence" he meant the continued visibility of the stimulus after offset, such as in an afterimage; finally, by "informational persistence", Coltheart intended the continued accessibility of the stimulus after offset, referring to this as *iconic memory* (the term in this context was introduced by Ulric Neisser 1967). The crucial next question is whether iconic memory, what Sperling uncovered in his partial report paradigm, reflects conscious or unconscious perception. If it reflects conscious perception, then the capacity of phenomenology in iconic memory exceeds the capacity of working memory. This then would provide a counterexample to the gatekeeper view.

## 6.5 Assessing the phenomenology of overflow

One of the central issues in the debate concerns how to characterize the different forms of visual short-term memory (VSTM) elicited by Sperling's paradigm and others inspired by it. Theorists speak of *iconic* memory, and recent work by Victor Lamme and collaborators suggests that there is a second form of VSTM, what they call *fragile* VSTM (Landman, Spekreijse, and Lamme 2003). The final section will briefly discuss fragile VSTM, but I will focus here on three positions regarding the content of any relevant form of VSTM (and thus Sperling's iconic memory) in respect of arrays like those in Sperling's experiments:

1. *Unconscious:* The information is specific and unconscious (or reflects unconsciousness), but can be brought to consciousness, say by attention.
2. *Nonspecific:* The information is conscious (or reflects consciousness), but in some way it is nonspecific, though it can be rendered more specific due to attention.
3. *Specific:* The information is conscious (or reflects consciousness) and highly specific.

"Specific" indicates that the information regarding each letter identity in memory is sufficient to support report of the identity of each letter when appropriately cued (Sperling's result). For example, **A**s are represented as **A**s. Where information is nonspecific, then identity information is in some way not present or degraded, although this idea needs elaboration. Roughly, **A**s are not represented as **A**s.[14] Talk of "reflects" acknowledges that the memory system itself might not be conscious, though it is a trace of a conscious or unconscious state. For letters in Sperling's array, (1) holds that the relevant representation is unconscious; (2) maintains that the subject consciously perceives the letters, but not necessarily as letters, but rather (perhaps) as symbols, shapes, or even as a jumble of features (recall inattentional agnosia, discussed in the previous chapter); (3) asserts that subjects consciously see the letters as the letters they are, though they cannot report on all of them.

Those who endorse cognitive access, and hence attention for cognition, as necessary for phenomenology often endorse (1). Block's version of overflow endorses a version of (3): what one sees exceeds what one can report, and in a way that allows for rich detail. But what of (2)? On first glance, one might take (2) to be inconsistent with gatekeeping views, but in fact it is consistent with those views. If this is correct, then (3) is the only viable (or at least clear) anti-gatekeeping position. To see why, consider how Sperling's paradigm differs from the inattentional blindness paradigms discussed in Chapter 5. The aim of the latter paradigms is to ensure that subjects deploy their attention in a specific, focused manner *away from a target stimulus* (a dancing gorilla, a large scale change in a scene). Yet in Sperling's paradigm, subjects cast perceptual attention in the broadest manner possible, namely to the entirety of the letter array. There is no issue of distraction here. Accordingly, the specific issues regarding gatekeeping that are raised by Sperling's paradigm and similar experiments concern the limits of attention on the output side, irrespective of how much attention might be deployed relative to its targets.

Let me report my own phenomenology when experiencing one of Sperling's displays (12 letters, three lines of four letters).[15] Let's call the cued letters subjects are to report the *reported letters* and the remaining letters, the *unreported letters* (this is a bit rough, but it will do). The letters I can report (reported letters) visually appear to me as the letters they are. That's why I can report them! Yet it also seems to me that the letters I cannot report (the unreported letters) are nevertheless visible to me. I see something at those positions although they don't appear to be specific letters. Rather, the unreported letters seem to be a smudge, as if blurrily seen, perhaps not even symbol-like. I am grasping for an adequate description, but I would venture to say that what it is like for me to see the unreported letters is similar to what it is like to see the letters at the edge of this page when I look at the middle of the page (I admit, I worry that this description is theory-ladened). Among undergraduates I have taught who have been presented with Sperling's stimulus, they have spontaneously suggested something more like Sid Kouider's (Kouider et al. 2010) contention that the figures appear as fragments.[16] So, the phenomenology I and some of my students report is consistent with Nonspecific. Block reports that his phenomenology is more in line with Specific.

Conflicts in introspection are often hard to adjudicate, but proponents of Nonspecific and Specific can allow that subjects do see more than the specific letters they report or remember. This was Sperling's starting point, and it identifies a crucial difference between Sperling's paradigm and inattentional blindness paradigms. For unlike the latter, the crucial stimuli in the former are *in some sense* reported. That is, it is not like the case of the gorilla where subjects make no reports at all regarding it. Rather, subjects notice and make reports about all the letters in the array. The difference is in the *specificity* of the report. The point, then, is that subjects do have access to all the letters, but possibly in different degrees (see also Stazicker 2011, Section 3). This is reflected in their reports that rely on working memory. Thus, Nonspecific is *prima facie* consistent with gatekeeping views. It is not the case that there are any conscious elements that outstrip cognitive access.

Subjects report on what they are conscious of, and this is more than the four specific letters they can name. If this is correct, then it is *Specific* that is needed to refute gatekeeping views. Hence, proponents of overflow must endorse Specific. The core of the overflow hypothesis is that the content of experience outruns the capacity of access in this way: phenomenology is specific in its content in a way that working memory is not.

What is clear is that the invocation of *capacity* needs to be made more precise. For Sperling, the task was to name the identity of the letters, requiring the coding of specific information regarding identity (an **A** or a **3**). Here, capacity is measured in terms of letter identity, and the consistent result is a limit of about four. Yet subjects also *report* that there are more letters visible than the four they identify, so this information about the other letters is also cognitively accessible. Subjects thus recall more information than merely four letter identities, and, in another sense, working memory capacity is greater than four. Not greater than four letter identities, of course, but greater in terms of a different notion of information, say, the resolution of uncertainty. Subjects have information not only about letter identity but also about the array. For example, they can accurately report that there are more items than four letters in the array. Perhaps this additional information concerns gist, or perhaps it is more specific. It is, however, additional information about the array over and above letter identity. There is, then, a counting question regarding measuring capacity. This is a fairly technical matter that will have to be set aside, but more work needs to be done here if proponents either of Nonspecific or of Specific are to make clear talk of capacity. Remember, it was the tools of information theory, in allowing for precise quantification of informational capacity, that Broadbent thought to be a big step forward for psychology, a new language (see Chapter 1 and Appendix).

Before focusing on relevant experiments and differing interpretations of them, let's consider two reasons Block has emphasized in favor of Specific. In his (2007b), Block notes:

1. Subjects in experiments attest to drawing on specific phenomenology in making their partial reports.
2. Denying specific phenomenology suggests that, when subjects have specific phenomenology restricted to the specific letters they report, then there is a shift from unconsciousness or generic phenomenology to specific phenomenology. Subjects should notice a change, but they do not.

The first point Block mentions is one that defenders of overflow often raise (see also Burge 2007), yet it is unclear how much weight one should give to it. The claim is largely anecdotal. For example, Block (2011, 570) notes Bernard Baars' observation that "subjects – and experimenters serving as subjects – continue to insist that they are momentarily conscious of all the elements in the array." Yet Baars seems to be reporting Sperling's own observations here, not independent studies that provide clear empirical support for the first point. Further, someone inclined to endorse Nonspecific can insist accurately that "they are momentarily conscious of all the elements in the array." The difference is whether one is aware of them in a specific or nonspecific way. Thus, what is needed, but currently lacking, is a systematic study of subjects' reports about their phenomenology in partial report paradigms.[17] It is worth emphasizing that subjects in the experiment *know* that the stimulus array presents letters, or at least are told so. Sperling's original subjects were told what they would see (letters) and underwent many trials with the same kind of letter stimuli. They knew or expected that the other positions they were unable to report contained letters. Thus, even if they were to report seeing each specific letter stimulus as the specific letter stimulus it is, their judgment might be affected by their expectation, rather than being an accurate readout of what perception gives them in each trial. Certainly, this is a potentially confounding factor. De Gardelle et al. (2009) showed that when pseudoletters are substituted in a Sperling letter array, subjects still think they are seeing only letters. They suggest that subjects' confidence in being aware of all the letters is a cognitive illusion (for a response, see Block 2011).

Let us turn to Block's point that if attention were needed for specific phenomenology, then one would notice a shift from unconscious/nonspecific to specific phenomenology in respect of the letters attended to. But would subjects notice such change? After all, change blindness studies show that even when a subject focuses attention, they miss substantial changes in a visual scene. Of course, if one focuses attention on the location of the change, then the change is easily seen. Yet this might explain why one would never notice the shift from unconscious/generic to specific phenomenology. In the partial report paradigm, the proposed change is induced by attention's moving to a location, rather than the change as occurring in a location where attention is already present. It is not clear that under such conditions, a change of the sort Block considers would be obvious.[18] Block's main argument, however, is to draw on interpretations of the experiments that provide the best explanation of the data, an inference to the best explanation. His claim is that Specific provides the overall best explanation of a diverse set of results. Let us then pursue alternative explanations in light of the distinction between Nonspecific and Specific.

### 6.5.1 Postdiction

Sperling arrived at his estimate of the capacity of iconic memory by summing each partial report across the total number of rows. One might wonder, however, whether summation is appropriate. It would be appropriate if the iconic memory representation is unaffected by any further processing induced by the cue. In particular, the representation must not be affected by attention as induced by the cue. If so, the subject could "read off" the data from a stable iconic memory representation. Ian Phillips (2011a), however, has questioned this *independence* assumption, i.e., the assumption "that a subject's experience of the stimulus in a [partial report] condition is independent of which report is cued because the cue comes only *after* display offset" (386). To show this, Phillips draws on the phenomenon of *postdiction*.

Consider the sensory processing of two stimuli, *A* at time $t_1$ and B at time $t_2$ where $t_1$ is prior to $t_2$ (this formulation allows that *A* and B can be processed in different sensory modalities). The counterintuitive idea of postdiction is that sensory processing of B can affect one's experience of *A*. This idea is counterintuitive if one assumes that sensory experience is more atomistic: one first experiences *A*, and then experiences B, where later experiences, or at least processing of later stimuli, cannot affect earlier experiences. An alternative is that sensory experience is a more complicated function of sensory processing over time. In particular, conscious experience of *A* might result from the unconscious sensory processing of *A and* B. Accordingly, sensory experience might result from sensory processing that spans significantly more than an instant. Let us call such effects *postdictive*.

Phillips provides an overview of various postdictive effects, many involving temporal offsets between the relevant stimuli (e.g., array and cue) similar to those found in Sperling's paradigm. Some of these are multimodal, involving two senses. Consider one striking postdictive effect found in the *sound-induced visual bounce* where two circles ("balls") are depicted on a screen as moving towards each other. When these two circles intersect and then continue to move, there are two experiences subjects report: the circles pass through each other, or the circles bounce off each other. When a

sound suggesting collision is played at, or around, the intersection of the circles, the intersection is more likely to be experienced as a bounce. Interestingly, this effect can occur even when the sound comes 200 milliseconds (ms) *after* the initial intersection (intuitively, one expects the experience of collision to work only if the sound comes right at the initial intersection, as if the circles were real balls).

Recall that the independence assumption leads us to infer that there is a uniform representation of the letter array that subjects tap into in different ways, depending on which line is cued. If this representation is or reflects phenomenal visual states, then the capacity of visual consciousness exceeds working memory capacity. Since any of the letters that the subject can report are represented in specific detail, the iconic memory representation represents each in specific detail, as per Specific. Thus, there is a rich phenomenal representation that exceeds cognitive access.

An alternative interpretation invokes postdiction. The content and structure of the underlying representation depends on which line is cued and, hence, on attention. Attention, on this picture, alters the underlying representation that serves task demands. Where the cue directs subjects to the top line, the underlying representation is brought by attention to be in a format that best serves reporting the top line; where the cue directs subjects to the middle line, the representation is brought by attention to be in a format that best serves reporting the middle line, etc. Attention can either bring the targeted line into consciousness from an unconscious representation or it can sharpen nonspecific representations into specific ones. Either way, there is no uniform phenomenal representation underlying reports across conditions. Rather, the nature of the iconic memory representation varies with attention in light of a subsequent cue. Given postdiction, Nonspecific or Unconscious might be the correct account of iconic memory.

### 6.5.2 Generic representations and the determinable/determinacy distinction

The cogency of Nonspecific depends on how one understands nonspecific representations of letters. Rick Grush (2007), in his commentary on Block (2007a), suggested that the relevant visual representations are as of generic letters. His example concerns how experience represents text on a page at the periphery of the visual field. See now for yourself. Focus on a word at the center of this page, and covertly attend to words at the periphery. Grush is certainly correct that the way they appear differs from the way the words at fixation appear in the glory of their specificity.[19] Yet how to understand a visual representation of a generic letter is not completely clear. Given my example about objects in the periphery, we might understand the proposal in terms of visual spatial resolution. James Stazicker (2011) has recently developed a line of response to Block, emphasizing limits on visual spatial resolution. Stazicker puts this in terms of the determinate/determinable relationship.[20] A standard example of this relation involves colors, say the determinable red and its determinates, crimson and burgundy. Determinates are ways of instantiating the determinable. As Stazicker comments: "To represent something indeterminately ... is to represent it as instantiating a determinable property, without commitment as to which determination of that determinable it instantiates. Roughly, property A determines property B where to have A is to have B in a specific way" (170). So, if a visual representation represents an object as (say) crimson, then it represents the object as being red in a specific way, namely, as crimson.

So, one can appeal to the spatial resolution of the visual system as a constraint on the determinacy of its spatial representations. Since visually representing shapes is a form of spatial representation, the spatial resolution of vision will provide constraints on the visual representation of shape. This idea can be spelled out by understanding, at least in a sketchy way, spatial processing in vision. Think of the retina as containing spatial filters that are sensitive to specific spatial frequencies. Consider a band of alternative black and white lines as in Figure 6.6:

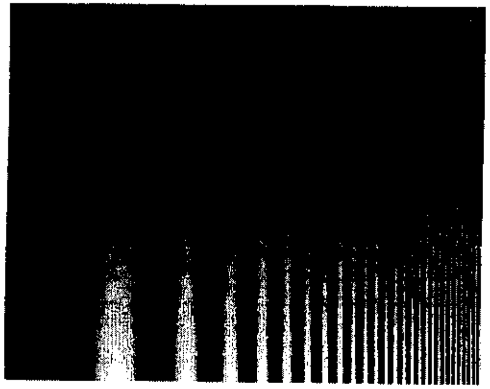


Figure 6.6 Figure showing increasing spatial frequency from left to right. Note that contrast increases from top to bottom, and higher contrasts are needed to adequately see higher spatial frequencies. This is why the lines appear to be taller as one proceeds to the right. Reprinted from G. M. Boynton (2005) "Contrast Gain in the Brain." *Neuron* (47): 476–77 with permission from Elsevier.

As you can see, the frequency of alternation of the lines increases per unit distance as one goes from left to right. In principle, this spatial frequency can be represented as a sinusoidal wave (represented as cycles per visual degree (cpd); your thumb held at arms length covers about two visual degrees from left to right). Where the relevant visual spatial filters can detect high spatial frequency, they can more finely resolve spatial properties such as the gap between two lines. For high-resolution spatial filters, two lines with a small gap separating them can be distinguished; for low-resolution spatial filters, the two lines cannot be distinguished, and the visual system will fail to detect the gap. Spatial resolution is greatest at the fovea and falls off rapidly. Stazicker's emphasis on spatial resolution is an important addition to the debate, but what should one say about nonspecific representations as postulated by Nonspecific?

A natural thought is that the phenomenal upshot of degrees of spatial resolution is degrees of sharpness in visual representation. One way visual experience can be less sharp is for experience to involve blurriness. Let us understand this in respect of the visual experience of the boundaries of a line with a sharp edge (so there is, objectively, no blurring at the edge). An ideal visual system not limited by spatial resolution can represent the edge of the line as at a determinate location, say at $y$ (how one specifies $y$ does not matter beyond it involving a magnitude reflecting position in some appropriate spatial coordinate system). A less determinate representation of the location of the edge might place it within a range, say between $x$ and $z$, where $x < y < z$. One can understand this difference in terms of the uncertainty tied to visual information in respect of where the edge is located. Recalling Shannon information theory (see Appendix), one can say that visual information leaves more uncertainty in the second case regarding the location of the line, but resolves it in the first case. This characterization of differences in spatial resolution allows us to speak of determinates and determinables if one wishes: being at $y$ is a way of being between $x$ and $z$. Moreover, both representations of the location of the edge can be veridical.

But how does this help with specifying a less determinate representation of a letter? Notice that the previous point was a way of spelling out blurriness in terms of the representation of edge location. Now, a letter such as **E** consists of lines (edges), and to the extent that the letter is experienced blurrily, then the visual information one gains about the structure of the letter, including its edges, carries a high level of uncertainty, first in terms of the location of edges, but correspondingly in terms of the figure constituted by the edges. It might then be useful to think of spatial resolution in terms of uncertainty. This is a description of spatial content, though it leaves open how best to characterize the corresponding phenomenology. Still, it gives us a handle on how experience should be characterized in *Nonspecific*: the experience of letters is nonspecific in that it is tied to a high degree of uncertainty, indeed not only about spatial information, but also information about other features.

Emphasizing spatial resolution as characterizing Nonspecific, we can raise a question for its proponents. Why can't proponents of *Specific*, such as Block, acknowledge that the visual system faces limits of spatial resolution, but plausibly insist that these limits are not at issue in the Sperling experiments? All parties agree that in order to explain Sperling's results, the specific identity of the letters must be visually represented *somewhere* in the cognitive system. Accordingly, the spatial resolution of the retinal locations stimulated by the unreported letters must be sufficiently sensitive to allow for determinate short term memory representations of letter identity, for any of these letters can be accessed for report when cued in Sperling's partial report paradigm.

Now the question is this: If one agrees that spatial resolution of the relevant letters is enough to perform the task, then why does the machinery of phenomenology seem to *blur* those letters in order to generate nonspecific representations as required by Nonspecific? This might seem like a pointless step, like purposely defocusing modern auto-focusing cameras while taking a picture. In terms of information, the idea is that the machinery of phenomenology adds noise to the system, increasing uncertainty. But why not just maintain, at the level of consciousness, the spatial resolution that is already present in iconic memory? If this is correct, why not then take the blurring to be an unnecessary further step such that explanatory parsimony pushes us to accept Block's alternative instead, namely, that the phenomenology is as Specific claims? One might respond by saying that moving from unconscious to conscious representations will inevitably involve a loss of information since it is an extra step in transmission of information, and this loss of information can precisely result in a phenomenology more like Nonspecific.[21]

Perhaps one way to settle this debate is to understand how much information is lost as it moves from step to step in visual processing (say from iconic memory to what comes next in the processing hierarchy). For example, given Sperling's result with cueing, there is good evidence that the identity of many letters is registered by the visual system where this

iconic memory exceeds working memory capacity (nine versus four letters in Sperling's estimation). It also seems that subjects don't just visually experience four letters. Rather, they see more letters but, minimally, only four of those letters *as the letters they are*. The issue then concerns their experience of the additional letters. In principle, one question that might be raised concerns the decay of information as it is processed and transmitted during visual processing. The idea of decay is that there is an increase in uncertainty about the layout of the array. Information is thereby lost. If that decay is rapid, then one could make the argument that by the time processing occurs that is necessary for visual consciousness, there is insufficient information content to support the detailed phenomenology that proponents of Specific claim there to be. Attention counteracts this loss by helping to maintain some subset of the information content about the letters from decaying when the subject is appropriately cued. Attention thereby preserves information by selecting it for memory, and this forms the basis of the subject's reports. On this view, those letters not selected for working memory cannot be seen in detail because information regarding them is quickly lost. If so, one can question whether consciousness can reflect the detail proponents of Specific aver, and instead argue that the information content present could only support Nonspecific phenomenology. On the other hand, there might be a rate of decay of information regarding the letters that (a) both explains the specific performance Sperling observed, but (b) also allows that the information regarding the identity of more than four letters is preserved at later stages of processing, even if this information is not funneled into working memory. If informational detail remains at later stages of processing, one might have the basis of an argument for Specific. This proposal is admittedly sketchy, but the point is that more detailed models are needed to connect with the behavioral data that has largely driven this debate. We need an alternative approach to pry the two models at issue apart, one that returns to the concrete specification of capacity limits that information theory can provide.

There seems to be a general sense among theorists in this area that proponents of Specific face a steep uphill battle, but let me raise a question for those who endorse gatekeeping and, specifically, the idea that consciousness is limited by cognitive access and attention: What does it mean to say that consciousness is limited by working memory capacity? When discussing Sperling, I spoke of working memory capacity as about four letters plus perhaps gist, but that is not a theoretically useful way to measure information. To explore the issue further, imagine looking at the ocean from a boat, marveling at the blue expanse that extends to the horizon. Gatekeeping claims that what you experience, a seemingly vast colored expanse, is in some way limited by cognitive access. But how is phenomenal consciousness of a large spatial area limited by working memory? Is working memory essential to one's *online* experience of the ocean blue? Might the phenomenology of experience of a colored space outstrip working memory? If not, why not? It is hard to understand the awareness of the blue expanse as constrained by working memory. The point is that gatekeeper theorists can't sit at the sidelines, enjoying the spectacle of their opponents climbing a steep hill. Gatekeeper theorists also have a difficult job to do, namely, to provide a concrete explanation of precisely what it means to say that conscious experience is limited by the capacity of cognitive access. As I noted earlier, this talk of capacity must be made more concrete, and until it is, *gatekeeping remains a vague thesis.* It does not allow us to say concretely in relevant cases what it means for consciousness to be limited in this way. But being concrete is a way to allow for an adequate assessment of the thesis.

## 6.6 Fragile visual short-term memory

I now briefly consider a neuroscientific argument for the Overflow thesis and Specific by Victor Lamme. Lamme and his coworkers have empirically isolated a different form of visual short-term memory (VSTM), what they call *fragile visual short-term memory.* This is a form of short-term memory that is intermediate in capacity between iconic memory probed in Sperling's work and working memory. Lamme has used these results in an argument in support of a version of Specific and to leverage the formulation of new explanatory concepts in this area.

Lamme's work on VSTM is important, extending Sperling's original findings. Adapting a change blindness paradigm, Landman, Spekreijse and Lamme (2003) presented subjects with an array of eight rectangles around a fixation point, with each rectangle oriented either vertically or horizontally (see also Sligte, Scholte, and Lamme 2008). This was followed by a presentation of a second array where the orientation of only one of the rectangles was changed. The time interval between the arrays varied from nearly 0.5 seconds to about 1.5 seconds in different experiments. Subjects were also provided a cue either (a) in the first array; (b) during the interstimulus interval; or (c) in the second array. Not surprisingly, when the subject is cued in the first array, they are highly accurate in detecting whether the cued rectangle changes its orientation in the second array, presumably because the cue allows the subject to attend to that object. Perhaps not surprisingly, subjects are also fairly poor at detecting changes when cued in the second array. The striking result is seen when the cue is presented in the interstimulus interval, because now performance accuracy is surprisingly high, even two seconds after the offset of the first array. Here, the cue seems to enhance performance even after stimulus offset, something Sperling observed as well.

Based on these and other studies, Lamme has argued that there are three forms of VSTM:

1. Iconic VSTM
2. Fragile VSTM
3. Working VSTM.

Lamme sees iconic and fragile VSTM as tied to the phenomena that Sperling characterized, and indeed, Lamme speaks of (1) as *retinal* iconic memory and of (2) as *cortical* iconic VSTM to emphasize the areas of the visual system that he takes to subserve each. For example, Lamme takes retinal iconic memory to essentially be the afterimage of the display, something that disappears quickly. Nevertheless, the informational content of iconic memory can survive the end of the afterimage, at which point it becomes cortical or fragile VSTM, fragile because it is easily disrupted by new retinal information. In either case, the capacity of (1) and (2) exceeds that of (3).

These are important extensions of Sperling's work, but how does this help provide a distinctive argument for Specific? Again, the central question is what iconic memory reflects: specific conscious or unconscious information (i.e., Specific or Unconscious). Lamme's argument appears to be as follows:

1. There is a high capacity VSTM distinct from working VSTM, namely, *iconic* (cortical/fragile) VSTM;
2. Representations in iconic VSTM exhibit many facets of *perceptual organization;*
3. Conscious representations exhibit perceptual organization;
4. Many unconscious representations do not exhibit perceptual organization;
5. The most *parsimonious* explanation is to take iconic VSTM to reflect conscious, and not unconscious, representations.

(reconstructed from Lamme 2010, 210–11)

By perceptual organization, Lamme means features like feature binding, figure-ground segregation, grouping, and organization that allows for illusions. The first two premises are derived from Lamme's own work on VSTM. Lamme notes that the empirical evidence for (2) is an ongoing project, but that many facets of perceptual organization have been observed for iconic representations. Premise (3) is in a way derived from introspection and cognitive access to experience, while (4) is empirical, derived from what is known about early visual processing, which many deem to be unconscious. The central question, then, is why (5) is correct in taking Specific to be the most parsimonious representation?[22] Premises (3) and (4) suggest that certain features of perceptual organization tend to track conscious versus nonconscious processing, but it is not clear that to then associate iconic VSTM with conscious processing amounts to an explanatory parsimonious inference. This move does echo Block's strategy of providing an account that makes best sense of all the data, in response to the methodological puzzle, but it is unclear what parsimony comes to here. Until that is clarified, it is not clear that we have solid grounds to endorse Specific from a neuroscientific perspective.

## 6.7 Summary

What is it with attention and consciousness? Why is it seemingly so obvious and yet so elusive? I have examined whether attention itself entails phenomenal consciousness, though I argued that it does not (Chapter 4). The past two chapters have considered whether attention has a specific role as gatekeeper for consciousness. This is, as we have seen, a difficult question in that it is hard to find a clear way to empirically engage the issues so as to help us decide between alternative models. Let me summarize some lessons from this and the previous chapter:

1. The central contrast is between the common-sense model, where consciousness is not limited by attention, and the gatekeeping model, where consciousness is so limited;
2. Theorists must provide clearer formulations of which gatekeeper thesis they are defending;
3. Inattentional blindness paradigms, where attention is purposely pulled from a stimulus, cannot provide evidence to settle the issue regarding which model is correct;
4. Alternative experimental paradigms or approaches must then be found to test gatekeeping;

5. The issue of gatekeeping can be emphasized either from focusing on attention's inputs (as in Chapter 5) or on attention's outputs as discussed here, namely, as being for working memory;
6. Sperling's Paradigm and similar approaches provide an alternative approach but the experiments are subject to divergent interpretations;
7. The first step to a way out is to develop clearer models about information processing and capacity that can lead to predictions about what experience of unreported targets in a briefly flashed array should be like.

Again, as I noted in the last chapter, these are areas where conceptual clarity and new approaches are needed. In light of the discussion over the past three chapters, there is no doubt that attention plays some important role in consciousness. The question, nevertheless, remains: What is its precise role?

## Suggested reading

For an overview of the neural Global Workspace Theory, see Dehaene and Naccache (2001); for an overview of the Attended Intermediate Representation Theory (AIR) see Prinz (2012). Lamme (2010) makes a detailed case for an empirical basis for endorsing the overflow thesis. Block provides an extended presentation of the overflow thesis from an empirical perspective in his (2007b) and a more philosophical perspective in (2008). His (2011) provides a summary of recent work. Phillips (2011b) provides a discussion of Sperling type experiments with emphasis on attention.

## Notes

[1] In his (2007b), Block opts to characterize access in terms of broadcasting in the global workspace (see below on the Global Workspace Theory).

[2] I am indebted to distinctions drawn by Jesse Prinz (2012). See also Dehaene and Naccache (2001).

[3] For example, Ned Block focuses on the relation between phenomenology and cognitive *accessibility* in his (2007b) which is geared towards the empirical community, while in his (2008), which is geared towards the philosophical community, he switches to talking about phenomenology as overflowing cognitive *access*. To keep things clear, it is then imperative to be explicit about access/accessibility *to what* (system). I am not saying

that Block is confused about this. Rather, for readers to keep track of the meanings behind invocation of access and accessibility, regimentation is required.

4 In his (2007a) discussion of Dehaene and Naccache (2001), Block comments on their division between ($I_1$) "permanently" inaccessible states, ($I_2$) states that are *accessible* in that were they to be attended to, they would be accessed by working memory (the global workspace, see Section 6.3.1) and ($I_3$) states that are accessed by working memory. Block points out two notions of cognitive accessibility, a broad sense that covers ($I_2$) and ($I_3$), and a narrow sense that covers ($I_3$). It is cognitive accessibility in the narrow sense that is the focus of Block's discussion. Notice that cognitive access in his terminology is access by systems subsequent to working memory. We are using "cognitive" in a different sense, namely, where it refers in the first instance to working memory.

5 ($GK_{A2}$) has to be rewritten slightly to accommodate Jesse Prinz's AIR view to be discussed in later sections, but the current version will do for now.

6 Block is sometimes read as endorsing this extreme view, and I suspect the reason is due to the slipperiness of talk of phenomenology outside of cognitive accessibility. Such talk reasonably suggests to a reader that one means that consciousness is tied to stage (1) and thus not even accessible to working memory. Many find such a view barely coherent. As Block emphasizes in later writings, that is not his claim. The previous conceptual regimentation discussed in the text is crucial for clarity.

7 Shanahan and Baars (2007) emphasize that their account of the global workspace does not identify it with working memory, but, rather, as something that gives access to working memory.

8 Given the previous regimentation, what they should have said is *accessibility* to conscious report.

9 Prinz would presumably respond that even here, attention makes the action-guiding representations accessible to working memory. Fair enough, though that is an empirical question that requires a more concrete specification of what it means to make a representation accessible. It might turn out to be false. Nevertheless, wouldn't the emphasis on working memory lose the forest for the trees in the case imagined, where attention's role seems to be to support action?

10 For a critical discussion of Prinz's theory, see (Wu, 2013c). See also (Mole 2013).

11 There is a slight complication here that makes terminology fraught with potential peril. It is in fact natural to talk about Prinz as emphasizing perceptual attention, i.e., attention that influences perceptual representations, while Dehaene and Naccache emphasize cognitive attention, i.e., attention for cognition. The relevant modulations here are all "pointing" towards working memory, even if they occur at different points in processing. Accordingly, I group them together as emphasizing attention that is for cognition.

12 This sets aside the central issue that Prinz raises, namely that accessibility is what matters for phenomenal consciousness. This is unfortunate, but the issues will otherwise get overly complicated. Here's why. Ned Block can largely agree with Prinz that phenomenology is always accessible. Where both will disagree is whether accessibility entails attention. Prinz says yes; Block says no (Block 2007b). The debate then centers on whether the property of a representation that renders it accessible is one that is brought about or not by attention. This is an interesting question, but difficult to get a clear handle on. One question we can raise to Prinz is the following: presumably, access also requires attention, but then it looks like attention is involved in two steps, namely, making a visual representation accessible and then, when needed, accessing the visual representation for working memory. But you might wonder if attention ever operates like that. Perhaps attention always just enables access which, of course, implies accessibility. There is no stopping point between accessibility and access once attention gets involved. Still, a fuller discussion is warranted, something that space constraints prevent us from pursuing.

13 For a discussion of some of this earlier work and an interesting analysis of the issues, see (Phillips 2011a).

14 Again, I find it more helpful to think about information in terms of decreasing uncertainty, so degraded information increases uncertainty. Specific holds that the information content in memory regarding the letter is much less uncertain than what is imputed by Nonspecific. That is, information content resolves uncertainty about letter identity (see Appendix A).

15 You can see a version of the Sperling stimulus in an online Ted$^X$ talk by Ian Phillips titled "Swimming against the stream of consciousness" which can be obtained by searching on the internet. The stimulus is presented about 1:30 seconds into the video, which also gives a brief summary of Phillips' account of the experiment, something we discuss in a later section.

16 Kouider defends a picture of awareness where the letters that are not identified are given to the subject in fragmentary form. This account is tied to certain assumptions about perceptual processing and our access to it. Specifically, Kouider et al. (2010) hold that perceptual processing involves multiple levels, from basic features to higher order categories

(including gist) *such that each level can be independently accessed*. Among the perceptual levels are those processing representations of fragments or perhaps parts of the objects present in the visual field. On Kouider's account, we can sometimes grasp the gist of the scene without grasping much detail concerning basic features or objects, or we can focus on a feature and not grasp the gist. This allows for the possibility of what he calls *partial awareness*, awareness that is restricted to some subset of visual processing levels.

On the Kouider view, phenomenal consciousness depends on access, but access can involve all, some, or none of the levels of visual processing. In the case of Sperling's experiments, when subjects claim to see more than they can remember, they are responding to partial awareness, where they have access to low-level representations of the stimuli, namely, letter fragments. Some of these letters, e.g., in cued rows, might be accessed at higher levels, namely, those that present the identity of the letters, while others, say in uncued rows, are accessed as fragments. In this way, the account explains subjects' sense that they see more than can be remembered. Alternatively, when subjects claim to see the specific identities of all the letters, they are under a *cognitive* illusion.

In response, Block (2011) points out that the fragments hypothesis nevertheless suggests that consciousness is rich in content, going beyond the letters that the subjects can explicitly report. He notes that if there is disagreement, it is on "*how* degraded the specific phenomenology is" (2007, p. 532). So, Kouider and Block can agree that either (2) or (3) is correct, as against (1). Consciousness is not limited to the letters that are reported. Nevertheless, Kouider et al. would emphasize that phenomenology nevertheless does not exceed access. For, on their model, the fragments are *in fact accessed*, and that is why subjects in Sperling's experiment report that they see more than they could report. Indeed, the sense of seeing more than can be (specifically) reported, demonstrates a kind of access and is thus consistent with gatekeeper views. What remains accessible, even after the capacity to remember the specific identity of particular objects is saturated, is the gist, here the presence of fragmentary forms. Thus, to the extent that Kouider and Block agree, it is that experience is rich in a way inconsistent with (1). They nevertheless continue to disagree about whether (2) or (3) is the correct view. This does suggest the independence between the rich/sparse content distinction and the claim of overflow. Overflow implies that content is rich in the sense that it exceeds cognitive access or accessibility. But deniers of overflow can

also claim that content is in a sense rich, in that it exceeds the specific letters that are reported by subjects in Sperling's experiment.

17 Block makes passing reference to an observation of Rogier Landman "that the extent to which subjects evince specific phenomenology may be correlated with how well they do in the experiments [such as those reported in Landman, Spekreijse, and Lamme 2003]" (Block 2007b, 531). This is the sort of evidence that would help buttress overflow, but as far as I know, this observation has never been verified or published by Landman.

18 See also (Stazicker 2011, 175–76). For a different response to this issue, see (Phillips 2011b, 215).

19 Block characterizes generic phenomenology in terms of existentially quantified content, namely the visual system's representing that there is an array of letters, as opposed to the representation of the identity of each specific letter.

20 An influential application of the determinable/determinate distinction to the case of the visual experience of blurriness occurs in Tye (2003).

21 Block emphasizes that opponents of Specific endorse the unconscious representation of highly specific visual information, an unconscious icon, but points out that there is no evidence for unconscious iconic memory of the requisite specificity. See his (2011) for a brief discussion.

22 Lamme provides a second argument that appeals to *recurrent processing*. Here is a description of the flow of information after a visual signal reaches the brain, one consisting of four stages (see Lamme 2010):

> Stage 1: A *superficial* feed forward sweep (FFS) of the signal up the visual hierarchy that does not travel deep into the visual system.
>
> Stage 2: Deep processing of the FFS, where the signal travels the entire sensory hierarchy to motor and prefrontal areas.
>
> Stage 3: *Superficial* recurrent processing involving horizontal and feed*back* connections, of a more local nature.
>
> Stage 4: *Widespread* recurrent processing across the hierarchy (cf. the global workspace).

When the stimulus is removed, iconic memory is associated with Stage 3, while working memory is associated with Stage 4 processing. Thus, encoding in the global workspace is tied to Stage 4. For Lamme's *Recurrent Processing Theory* of consciousness, however, the crucial stage for consciousness begins at Stage 3. Since this is prior to the activation of the global workspace at stage 4, Lamme disagrees that *cognitive access* is necessary for consciousness (he might yet agree with Prinz that *cognitive accessibility* is necessary).

Why does Lamme think that consciousness is tied to Stage 3 as well? In short, it is because recurrent processing looks to be a good neural correlate for consciousness, in part because it is looks to be a good neural correlate for perceptual organization, a critical feature of phenomenal consciousness (Lamme allows that there are open empirical questions here; he is offering a hypothesis). This assumes, as does the argument in the text, that there is some important connection between phenomenal consciousness and perceptual organization. It seems that for Lamme's argument to be compelling, there should be a necessary relation between perceptual organization and phenomenal consciousness. But is there?

# CONCLUSION

## WHAT ATTENTION IS AND WHY IT IS CENTRAL

It is perhaps appropriate here to return once more to James and his answer to the metaphysical question:

> [Attention] is the taking possession by the mind, in clear and vivid form, of one out of what seem several simultaneously possible objects or trains of thought. Focalization, concentration, of consciousness are of its essence. It implies withdrawal from some things in order to deal effectively with others, and is a condition which has a real opposite in the confused, dazed, scatterbrained state which in French is called distraction, and Zerstreutheit in German.
>
> (James, 1890, p.403)

Given the discussion of the past eight chapters, I think James was in many ways right! Recall the five basic questions regarding attention. Here are brief responses to each in light of the discussion of this book. To underscore certain themes, I shall state the claims more boldly than perhaps the evidence and arguments warrant.

### Metaphysical: What is attention?

Attention is by most accounts a selective psychological capacity. As noted in Chapter 1, there are many forms of selection that do not count as

attention. A natural specification of attentional selection, however, ties it to tasks. This link to task was uncovered in the empirical sufficient condition – selection of *X* for task *T* suffices for attention to *X* for *T* – a condition that is a shared assumption in experimental practice within the science of attention. As such, the sufficient condition provides a basic starting point for the analysis of attention. It uncovers a commonality among theorists who have lamented the possibility of defining attention. To them, one can say, "a definition is (nearly) in hand in what you already assume." That is, if theorists want to leave the mosh pit of attempts to state what attention is and not simply surrender, then the best option is to start with what everyone already knows in the experimental practice of attention: attention is, at least sometimes, selection for task.

As we have noted, the task-centered conception of attention, couched in the empirical sufficient condition, constrains interpretation of data in the neuroscience of attention. So, even in the search for basic mechanisms of attention, nothing is achieved without the empirical sufficient condition, a condition that gives neuroscientists grounds for concluding that the mechanisms and circuits that they uncover are "attentional". This points to the empirical centrality of a task-centered account of attention, one that can be expanded conceptually to a selection for action account. Of course, there are alternatives such as attention as selection for consciousness or memory, but it is likely that selection for (working) memory is subsumed by the selection for action account (selection for working memory being something necessary for much selection for action), and selection for consciousness founders on unconscious attention. The claim then is that the action-centered account provides the best current answer to the metaphysical question: attention is selection for action.

### Function: What role does attention play?

One of the lessons that David Marr conveyed, in his posthumously published book, *Vision*, is that to understand capacities like vision, one must understand what that capacity is for. Without such an understanding, the cognitive science of psychological capacities can get nowhere. A computational theory, as Marr put it, is a necessary foothold for discovering mechanisms for vision at different levels of analysis. The same lesson is true of attention. What, then, is the functional role of attention? As James' quote suggests, the functional role of attention centers on its selectivity, and the usual suspects emerge: attention for action, for consciousness, for memory.

One can, of course, study attentional functions in a more task-dependent and fine-grained way, say, by focusing on attention in reasoning, in perception, or in imagination, but it is likely that these more fine-grained investigations will simply uncover instantiations of the more general functions noted. Still, several of these more fine-grained functional contributions of attention are of great philosophical significance: attention's role in fixing demonstrative thought, in enabling agency, in determining and affecting the character of consciousness, in making justification possible, and in fixing introspective thought. Attention is not merely pervasive. It is fundamental to central aspects of the mind. One can put matters this way: without attention, agency, justification, certain forms of external and internal thought, and certain features of consciousness would not be possible. These are strong claims, and they merit further sustained reflection.

## Properties: What are characteristic features of attention?

A slew of experimental paradigms have uncovered different features of attention: its targets, its temporal profile, its duration, its processing demands, and its interaction with other systems. Some features are notable, such as how attention is affected by the nature of the task and how it responds in different ways to different types of stimuli (e.g., direct versus symbolic cues, or feature singletons in pop-out and conjunctions in visual search). The science of attention will continue to uncover interesting features of this central psychological capacity, and as philosophers continue to explore the philosophical significance of attention, they will need to keep abreast of these developments. Still, there does seem to be a way to carve attention at its joints, namely, in the divisions between top-down and bottom-up attention and between controlled and automatic attention. These types of attention seem to involve different neural realizations, but also reflect the different sources of attention: a reliance on intentions and higher-order nonperceptual states on the one hand, and a reliance on the world on the other. Attention in that way can be both active and passive.

## Mechanism: How is attention implemented?

The question of implementation can be pursued at different levels of abstraction, from abstract computational descriptions to the concrete neural realizations of attention. We have discussed, among others, Broadbent's conception of attention as a filter for selecting information for further processing, Treisman's Feature Integration Theory, with its focus on attention as binding features for object representation and awareness, Desimone and Duncane's biased competition model where attention emerges from neural competition for limited resources, Rizzolatti's Premotor Theory that takes spatial attention to result from the activation of action representations for eye movement, and a plethora of effects at the level of the activity of single neurons. The challenge for these mechanistic accounts of how attention works or is implemented will be not just whether they are adequate to the phenomenon, but also how well they can be integrated with each other. A central task for cognitive science, in which philosophers will play a critical role, is not in defining attention, for we have the basis of such a definition, but in using that definition to integrate and bridge these disparate levels of analysis. Too often within cognitive science, work at different levels fails to be bridged in illuminating ways. That is hard work, and in the case of attention, work which presents interesting challenges and fertile ground for new approaches.

## Consciousness: What is the relation between attention and consciousness?

Finally, attention seems to be closely tied to consciousness. James' quote suggests that attention is essentially connected to consciousness, but some forms of attention are unconscious. So, on one reading, James was wrong: consciousness is not of attention's essence. On another reading, he was right: attention has an essential role to play in allowing us to respond to the deliverances of consciousness, for attention has a necessary role to play in our capacities to respond in general. This is not to deny that attention has a phenomenal upshot, but the precise nature of the phenomenology of attention remains a difficult question. One concrete proposal is that attention is not a conscious state with its own characteristic phenomenology, but is a state that affects consciousness. Some of the phenomenal effects of attention are quite disparate. Whether there is something more uniform in the phenomenology of attention, some way that attention makes things phenomenally salient, is a difficult issue on which our intuitions might simply clash at rock bottom.

Attention has another potential connection to consciousness, namely, attention as the gatekeeper of consciousness. On the one hand, the basic notion of gatekeeping can be simply expressed: one is phenomenally conscious of X only if one attends to X. On the other hand, it is not clear exactly what this thesis comes to in detail, and it is not clear that we have strong evidence in favor of it. Some of the evidence, namely, work in the realm of inattentional blindness, is not adequate to settle the issue. At the same time, the contrary thesis—that there is phenomenology outside of attention—might seem impossible to establish, since the evidence that we have for phenomenology, namely, some form of report, relies on attention. If there is to be evidence for such phenomenal overflow, it will require ingenuity to establish. Still, gatekeeper theorists have to make more concrete exactly how to understand the limits on consciousness beyond talk of capacity limits. Many years ago, Donald Broadbent drew inspiration for psychology from the precise tools that Claude Shannon provided him in information theory. Capacity limits could be precisely quantified. We need to return to that inspiration in understanding the role of attention in consciousness and how associated capacity limits provide concrete boundaries to the character of consciousness.

So, James was right. We do know what attention is and this knowledge puts us in a position to investigate and discover what attention does, what it is good for, and how it works. The science of attention has now been established for over a century. We can, I think, look forward to a healthy philosophy of attention as well, and a positive synergy between the two approaches.