1 Attention—We All Know What It Is

But Do We Really?

The title of this chapter is adapted from the classic words of William James (1890), who wrote what has become perhaps the best-known plain language description of attention:

> Everyone knows what attention is. 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.

James specified two domains in which these objects occur: sensory and intellectual. He listed three physiologic processes that he believed played a role in the implementation of attention: the accommodation or adjustment of the sensory organs, the anticipatory preparation from within the ideational centers concerned with the object to which attention is paid, and an afflux of blood to the ideational center. With these processes, he set up a good deal of modern attention research including functional magnetic resonance imaging (fMRI) studies. However, since the time of James—and because of the myriad experimental findings exploring each of James' three processes—things have become less and less clear, and it is important to consider the many subsequent points of view.

A book on attention, computational or otherwise, needs to define what it means by attention. It would have been so convenient to end the introduction to attention with James' description. But it is not to be so. Many over a long period of time have written on how difficult it has seemed to pin down this domain of inquiry. Compare James' statement with that of Pillsbury (1908)

> [A]ttention is in disarray . . . .

or that of Groos, who wrote in 1896 that

> To the question, 'What is Attention,' there is not only no generally recognized answer, but the different attempts at a solution even diverge in the most disturbing manner.

Four decades later, it seemed that little had changed. Spearman (1937) commented on the diversity of meanings associated with the word:

> For the word attention quickly came to be associated . . . with a diversity of meanings that have the appearance of being more chaotic even than those of the term 'intelligence.'

Almost eleven decades after James, Sutherland (1998) suggested that:

> [A]fter many thousands of experiments, we know only marginally more about attention than about the interior of a black hole.

Taken together, these quotes make the situation seem bleak! The field is full of controversy, and it seems that a bit more care is required before moving on. A brief tour through some of the early thinking on the topic helps reveal sources of debate and key issues. A more detailed treatment can be found in Tsotsos, Itti, and Rees (2005). The first scientific reference to attention, even though its etymology is traced to ancient Rome, seems to be due to Descartes (1649), who connected attention to movements of the pineal body that acted on the animal spirit:

> Thus when one wishes to arrest one's attention so as to consider one object for a certain length of time, this volition keeps the gland tilted towards one side during that time.

Keeping with the idea that body organs are involved, Hobbes (1655) believed:

> While the sense organs are occupied with one object, they cannot simultaneously be moved by another so that an image of both arises. There cannot therefore be two images of two objects but one put together from the action of both.

Leibnitz (1765) first linked attention to consciousness, a possibility that has received much debate recently, and attributed this to inhibition from competing ideas:

> In order for the mind to become conscious of perceived objects, and therefore for the act of apperception, attention is required.

Hebart (1824) was the first to develop an elaborate algebraic model of attention using differential calculus and may be considered the first attention modeler. His general view on attention, however, was still rather simple:

> He is said to be attentive, whose mind is so disposed that it can receive an addition to its ideas: those who do not perceive obvious things are, on the other hand, lacking in attention.

Since the 1800s, much genius has gone into experimental methods that were hoped to shed some light on the phenomenon of attention. Helmholtz (1860) believed that nervous stimulations are perceived directly, never the objects themselves, and there are mental activities that enable us to form an idea as to the possible causes of the observed actions on the senses. These activities are instantaneous, unconscious, and

cannot be corrected by the perceiver by better knowledge—he called this **unconscious inference**, and thus he believed that attention is an unconscious phenomenon. On the other hand, Panum (1858) believed that attention is an activity entirely subservient to an observer's conscious will. Attention becomes difficult to hold once interest in an object fades. The greater the disparities between the intensities of two impressions, the harder it is to keep attention on the weaker one. Panum studied this in the specific context of binocular rivalry; but more generally, he observed that we are able to 'see' only a certain number of objects simultaneously. He therefore concluded that it makes sense that the field of view is first filled with the strongest objects. In studying an object, first attention, and then the eye, is directed to those contours that are seen by indirect vision.

Hamilton (1859) wondered about the span of attention:

> The doctrine that the mind can attend to, or be conscious of, only a single object at a time would in fact involve the conclusion that all comparison and discrimination are impossible. . . . Suppose that the mind is not limited to the simultaneous consideration of a single object, a question arises—how many objects can it embrace at once?

His last question is important even today. Brentano (1874) developed **act psychology,** where an act is a mental activity that affects percepts and images rather than objects. Examples include attending, picking out, laying stress on something, and similar actions. This was the first discussion of the possibility that a subject's actions play a dominant role in perception. Metzger (1974) lists aspects of action that contribute to perception: bringing stimuli to receptors, enlarging the 'accessible area,' **foveation** (the act of centering the central, highest-resolution part of the retina onto an object), optimization of the state of receptors, slowing down of fading and local adaptation, exploratory movement, and finally the search for principles of organization within visual stimuli.

Wundt (1874) further linked attention and consciousness, suggesting that attention, as an inner activity, causes ideas to be present in consciousness to differing degrees. The focus of attention can narrow or widen, reflecting these degrees of consciousness. For Titchener (1908), attention was an intensive attribute of a conscious experience equated with 'sensible clearness.' He compared attention to a wave, but with only one peak (corresponding with one's focus). He argued that the effect of attention is to increase clarity, whereas Kulpe (1902) suggested that attention enhanced not clarity but discriminability. Petermann (1929) argued against the subject being a passive perceiver of stimuli. He proposed an **attention-direction theor**y, based on actions, as the mechanism for an active attentive process. As will become apparent, this theme keeps reappearing. These and other ideas were never formalized in any way and remained conceptual, yet interesting, viewpoints on the issue.

Helmholtz (1896) introduced the idea of **covert attention**, independent of eye movements:

> The electrical discharge illuminated the printed page for a brief moment during which the image of the sheet became visible and persisted as a positive after-image for a short while. Hence, perception of the image was limited to the duration of the after-image. Eye movements of measurable size could not be performed during the duration of the flash and even those performed during the short persistence of the after-image could not shift its location on the retina. Nonetheless, I found myself able to choose in advance which part of the dark field off to the side of the constantly fixated pinhole I wanted to perceive by indirect vision. Consequently, during the electrical illumination, I in fact perceived several groups of letters in that region of the field. . . . The letters in most of the remaining part of the field, however, had not reached perception, not even those that were close to the point of fixation.

In other words, Helmholtz was able to attend to different portions of an image on his retina without eye movements. Such a demonstration is compelling and represents powerful evidence for the existence of attention independent of gaze change.

Even though experimental evidence supporting a variety of phenomena attributed to attention mounted, the field was not without its nonbelievers. The Gestalt school did not believe in attention. Köhler only barely mentions attention (Köhler, 1947). Gestaltists believed that the patterns of electrochemical activity in the brain are able to sort things out by themselves and to achieve an organization that best represents the visual world, reconciling any conflicts along the way. The resulting internal organization includes portions that seem more prominent than others. Attention, to them, was an emergent property and not a process in its own right. In this sense, Gestaltism was the precursor of the modern Emergent Attention theories that will be described in chapter 3. Figure-ground concerns loomed larger for them, the figure would dominate perceptions within a scene, thus emerging as the focus of attention rather than being explicitly computed as such. Berlyne (1974) tells us that Edgar Rubin, known for his vase/profile illusion of figure-ground perception, actually presented a paper at a meeting in Jena, Germany, in 1926 titled "On the Nonexistence of Attention." More recently, Marr basically discounted the importance of attention by not considering the time intervals of perception where attentive effects appear even though his goal was clearly to propose a theory for full vision. Describing grouping processes and the full primal sketch, he said:

> [O]ur approach requires that the discrimination be made quickly—to be safe, in less than 160 ms—and that a clear psychophysical boundary be present. (Marr, 1982, p. 96)

Attention has been viewed as **Early Selection** (Broadbent, 1958), using **Attenuator Theory** (Treisman, 1964), as a **Late Selection** process (Deutsch & Deutsch, 1963; MacKay, 1973; Moray, 1969; Norman, 1968), as a two-part process, **preattentive fol-**

lowed by attentive processing (Neisser, 1967), as a result of **neural synchrony** (Milner, 1974), using the metaphor of a **spotlight** (Shulman, Remington, & McLean, 1979), within **Feature Integration Theory** (Treisman & Gelade, 1980), as an **object-based** phenomenon (Duncan, 1984), as a **shrink-wrap** process (Moran & Desimone, 1985), using the **Zoom Lens** metaphor (Eriksen & St. James, 1986), as a **Premotor Theory** subserving eye movements (Rizzolatti, Riggio, Dascola, & Umilta, 1987), as **Guided Search** (Wolfe, Cave, & Franzel, 1989), as **Biased Competition** (an extension of the shrink-wrap interpretation; Desimone & Duncan, 1995), as **Feature Similarity Gain** (Treue & Martinez-Trujillo, 1999), and more. These are all defined and discussed in later chapters, and they are listed here to show the diversity of opinion on the nature of attention. The field is rich with ideas, but can they all be right?

We have seen how Helmholtz provided a convincing demonstration for the existence of covert attention. Yet eye movements are the most obvious external manifestation of a change of visual attention. Yarbus' classic work (Yarbus, 1979) showed how task requirements affected fixation scan paths for an image. Given the same picture of a family in a Victorian living room scene, Yarbus asked subjects to either freely view the picture or to answer one of the following six questions about the people and situation depicted in the picture:

1. What is their economic status?
2. What were they doing before the visitor arrived?
3. What clothes are they wearing?
4. Where are they?
5. How long is it since the visitor has seen the family?
6. How long has the unexpected visitor been away from the family?

He recorded subject's eye movements while freely viewing and for the period of time before subjects provided a reply to a question. Each recording lasted 3 minutes. The surprise was the large differences among the summary scan paths demonstrating that the reason for looking at a picture plays a strong role in determining what was looked at. In fact, this was a nice extension of the basic Posner cueing paradigm that has played such a large role in experimental work (Posner, Nissen, & Ogden, 1978). Instead of providing a spatial cue that directed attention, Yarbus' questions directed attention. Posner (1980) suggested how overt and covert attentional fixations may be related by proposing that attention had three major functions: (1) providing the ability to process high-priority signals or alerting; (2) permitting orienting and overt foveation of a stimulus; and (3) allowing search to detect targets in cluttered scenes. This is the **Sequential Attention Model:** Eye movements are necessarily preceded by covert attentional fixations. Other views have also appeared. Klein put forth another hypothesis (Klein, 1980), advocating

the **Oculomotor Readiness Hypothesis:** Covert and overt attention are independent and co-occur because they are driven by the same visual input. Finally, the aforementioned Premotor Theory of Attention also has an opinion: Covert attention is the result of activity of the motor system that prepares eye saccades, and thus attention is a by-product of the motor system (Rizzollati et al., 1987). However, as Klein more recently writes (Klein, 2004), the evidence points to three conclusions: that overt orienting is preceded by covert orienting; that overt and covert orienting are exogenously (by external stimuli) activated by similar stimulus conditions; and that endogenous (due to internal activity) covert orienting of attention is not mediated by endogenously generated saccadic programming.

What role do stimuli themselves play in attentional behavior? What is the role of the salience of the visual stimuli observed (see Wolfe, 1998a)? Just about everything someone may have studied can be considered a feature or can capture attention. Wolfe presents the kinds of features that humans can detect efficiently and thus might be considered salient within an image: color, orientation, curvature, texture, scale, vernier offset, size, spatial frequency, motion, shape, onset/offset, pictorial depth cues, and stereoscopic depth. For most, subjects can select features or feature values to attend in advance. Saliency has played a key role in many models of attention, most prominently those of Koch and Ullman (1985) and Itti, Koch, and Niebur (1998).

Modern techniques in neurophysiology and brain imaging have led to major advances in the understanding of brain mechanisms of attention through experiments in awake, behaving animals and in humans. It is not possible to do justice to the large and impressive body of such research here (but see Itti, Rees, & Tsotsos, 2005). Suffice it to say that evidence abounds for how attention changes perception, and it seems manifested as both enhancement as well as suppression of signals. We also have a better idea about where attentional computations may be taking place in the brain.

How can it be that so many different and sometimes opposing views can be held all for the same "we all know what it is" phenomenon? One possibility is that the nature of a purely experimental discipline lends itself to fragmented theories. Most of the theories and models described earlier are constructed so that they provide explanations for some set of experimental observations with a focus being on the experiments conducted by each researcher. To be sure, each attempts to be as consistent with past work as possible so to build upon the past and not to continually reinvent. However, the explanations are almost always stated in natural language, using the ambiguous terminology of attention. In other words, there is no quantitative or formal statement of the theory such that it is unambiguous and not open to different interpretations. For many of the main theories of attention, it is easy to find subsequent interpretations that seem rather unjustified. As a result, a large part

of the controversy in the field may have two main sources: a vocabulary that has never been defined unambiguously and a theoretical framework that is not formal in a mathematical sense and thus open to interpretation.

Moving Toward a Computational Viewpoint

Although attention is a human ability we all intuitively think we understand, the computational foundations for attentive processes in the brain or in computer systems are not quite as obvious. Notions such as those of capacity limits pervade the attention literature but remain vague. Within all of the different viewpoints and considerations of the previous section, the only real constant—something that everyone seems to believe and thus the only logical substitute for James' original statement—is that attentional phenomena seem to be due to inherent limits in processing capacity in the brain. But if we seek an explanation of attentional processing, even this does not constrain the possible solutions. Even if we all agree that there is a processing limit, what is its nature? How does it lead to the mechanisms in the brain that produce the phenomena observed experimentally?

This presentation, focusing on vision and sensory perception mostly, attempts to make these more concrete and formal. Through mathematical proofs, it is possible to derive the necessity of attentive processes, and through algorithmic approximations and processing optimizations it is possible to discover realizations given either biological or computational resource constraints. Perhaps the most important conclusion is that the brain is not solving some generic perception problem and, by extension, a generic cognition problem. Rather, the generic problem is reshaped—changed—through approximations so that it becomes solvable by the amount of processing power available in the brain.

The human cognitive ability to attend has been widely researched in cognitive and perceptual psychology, neurophysiology, and in computational systems. Regardless of discipline, the core issue has been identified to be **information reduction**. Humans, and many other animals as well, are faced with immense amounts of sensory input, and the size of the brain limits its ability to process all of this input. This is the qualitative statement that has appeared many times in the literature. It is not simply that there is too much input; the problem is that each component of each stimulus can be matched to many different objects and scenes in memory resulting in a combinatorial explosion of potential interpretations, as is caricatured in figure 1.1.

Perhaps the bulk of all perceptual research has focused on how the brain decomposes the visual signal into manageable components. Individual neurons are selective for oriented bars, for binocular disparity, for speed of translational motion, for color opponency, and so on. We know that individual neurons also exist that are

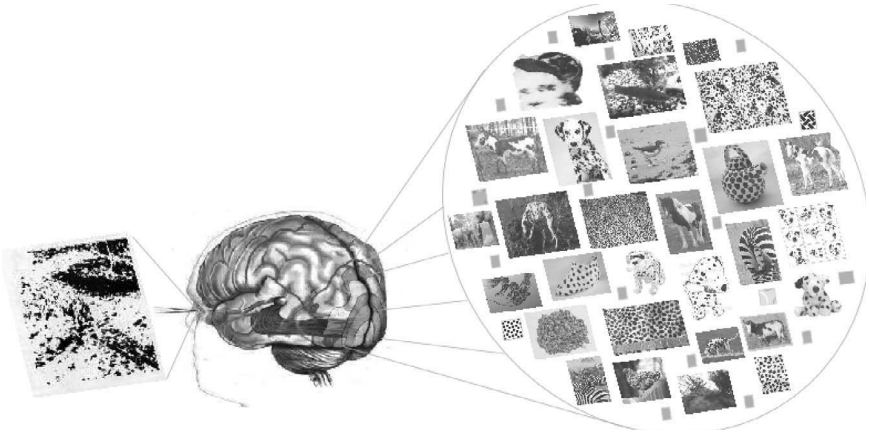

Figure 1.1
The classic "Dalmatian sniffing at leaves" picture (attributed to Richard Gregory) is sufficiently complex to activate an enormous number of possible interpretations. Each small piece of it has similarities (some strong, some weaker) to many other possible objects and scenes. The combinatorial explosion of possibilities that results is what any system—brain or machine—must effectively deal with to perceive successfully and act on the world.

tuned to particular faces or other known objects. But how can we deal with unknown scenes and objects? The neural decomposition of a visual scene gives the brain many, many pieces of information about a scene. It is in effect a *Humpty-Dumpty*-like problem—we know how the visual image may be decomposed, but how is it reassembled into percepts that we can use to guide our day-to-day lives? It is here where the combinatorial explosion has greatest impact.

This combinatorial view is the one that is central to the theory presented in this book. However, it is not the only view. For example, Tishby, Pereira, and Bialek (1999), using information theory, view the relevant information in a signal as being the information that it provides about some other signal. They formalize this problem as that of finding a short code that preserves the maximum information about the other signal, squeezing information through a 'bottleneck' formed by a limited set of code words (the **information bottleneck method**). Clearly, they address information reduction and do it in a principled and well-defined manner. Although an interesting and important perspective, it seems difficult to understand how it may relate to brain processing because it does not address what sort of process may be responsible for determining what those code words may be; Tishby et al.'s major concern is the amount of information not its content or how it is processed. The issues cannot be separated if one wishes to develop a theory of human attention.

The basic idea that humans can be viewed as limited-capacity information processing systems was first proposed by Broadbent (Broadbent, 1958). In computa-