Chapter 1

Introduction

At every waking moment, our visual system is bombarded with a wealth of information. Information from a world flooded with light and color, and filled with complex objects enters our eyes and is processed by the visual brain. Although this process intuitively appears to be effortless, the capacity of the visual system to handle all this information is limited. Consequently, only a small proportion of the information can be processed, while the rest is filtered out. Thus, an automatic, straightforward processing of the visual world does not proceed. Instead, a glance at a typical scene leads to competition between the multiple objects present in the scene for the processing resources of the visual system.

What is at stake in this competition is strikingly demonstrated by inattentional or change-blindness paradigms (Figure 1.1). In situations where multiple objects compete for the resources of the visual system, subjects can often fail to notice the sudden appearance of an unexpected stimulus, even at the center of gaze (Mack & Rock, 1998; Simons & Chabris, 1999). Similarly, observers are often completely unaware of the appearance or disappearance of large objects in a scene (such as the disappearance of an airplane propeller) (Rensink, O'Regan, & Clark, 1997; Simons & Chabris, 1999) and can even fail to notice changes made to the identity of the individual they are conversing with (Simons & Levin, 1998). These and other equally striking lapses of the visual system reveal its limited processing capacity–a limitation that places the perception of many aspects of visual scenes in jeopardy. Those features of the scene that lose out in the competition for this restricted processing resource can simply escape awareness resulting in inattentional or change-blindness.

It is obvious therefore that, unchecked, this drawback in visual processing would ultimately court disaster. For instance, failure to notice a car turning onto a busy street can be fatal. How is it then that despite these shortcomings in visual perception, we are fairly successful in noticing the relevant aspects of our environment, thus avoiding mishaps? It turns out, as a matter of fact, that much of visual perception relies on a critical resource: attention. Attention helps us “see” the world. Studies show that processing of particular objects can be favored by selectively attending to them. Thus, in the change-blindness situations described above, once attention is deployed to the location of change, observers are more likely to rapidly detect it. Conversely, failure to attend to the location or object of change leads to blindness to the change (Rensink, O'Regan et al., 1997; Kelley, Chun, & Chua, 2003). In other words, in the visual competition that takes place between all the stimuli in the field, attention has the ability to bias the struggle for processing in favor of attended objects, thereby rescuing them from oblivion, while filtering out the distracting information.

Given the importance of the attentional mechanism in daily vision, an understanding of its nature is therefore critical to our understanding of vision. However, it is equally important to also understand what type of processing can occur in the absence of attention. Attention is not always necessary for perception. Accordingly, a number of studies have demonstrated that a variety of objects can still reach awareness when attention is not available (Braun & Sagi, 1990; Braun & Julesz, 1998). In other words, some degree of processing occurs effortlessly without first having to deploy the attentional system. That this should occur is not wholly surprising or unexpected–it is crucial to our survival that certain stimuli in the visual environment be detected as rapidly as possible. Indeed, if perception were contingent on attention, it would be extremely disadvantageous because firstly, attention is a limited resource which can only be focused on a few objects at a time. Thus if all of vision were completely reliant on attention, all (and possibly threatening) details outside the focus of attention would be missed. Secondly, shifting attention from place to place is time-consuming (Anllo-Vento, Luck, & Hillyard, 1998; Hillyard & Anllo-Vento, 1998) and therefore precludes rapid processing of the visual world. Furthermore, before the attentional system can be engaged, some degree of pre-attentive processing must occur in order to direct the deployment of attention.

The work of this thesis is an endeavor to examine visual processing and attention from both ends: from rapid and effortless vision in the near-absence of spatial attention to vision that must rely on attention for correct processing. First I attempt to see how far the pre-attentive system can be pushed and to investigate the neuronal mechanisms that underlie this type of processing outside the focus of attention. Secondly, I examine the change-blindness situations described above where attention becomes necessary for vision, to see how the behavioral manifestations of an attentional deficit are represented at the neuronal level.

However, before proceeding with a description of this work, I will first review selective attention and its impact on neuronal processing.

1.2 What is attention?

Although we have an intuitive understanding of what it means to “pay attention,” or to have one’s attention “drawn” to something, vision scientists cannot agree upon a precise definition of attention nor a way to reliably measure it. Instead, it has been much easier to identify the effects of attention in different settings–its impact on performance and recognition, the detriments in vision when attention is not available, or the visual processing that can occur in its absence. Indeed attention reveals itself in such diverse ways that it has proved difficult to paint a single coherent picture of what it might be. Rather it is better understood in all its diversity, in examining its different manifestations.

1.3 The spotlight of attention

Attention has often been described as a spotlight that can be directed to a particular location in space (Posner, Snyder, & Davidson, 1980; Treisman & Gelade, 1980). Much like a searchlight that sequentially picks out specific regions, according to the ‘spotlight’ theory of attention, attention can be selectively shifted from place to place. The objects that lie within the spotlight of attention reap the benefit of having attention directed to them–information that lies within this region gets processed more quickly or efficiently compared to unattended areas of visual space (Posner & Petersen, 1990). Interestingly, a physiological correlate of the ‘spotlight’ has been observed—it has been shown that directing attention to specific locations in space results in enhanced hemodynamic activity in the corresponding areas of visual cortex (Tootell, Hadjikhani et al., 1998; Brefczynski & DeYoe, 1999; Martinez, Anllo-Vento et al., 1999; Somers, Dale, Seiffert, & Tootell, 1999). Similar enhancements at the single-cell level are also observed and are discussed in depth in section 1.4.

The exact shape of the attentional searchlight has been heavily debated. Eriksen and St. James (Eriksen & St James, 1986) have suggested that rather than being a fixed size, the spotlight is more like a zoom-lens, whose shape can be continuously adjusted. That is, in situations where one must search a region of space narrowly, (e.g., looking at a particular section of a map), the size of the focus of attention is scaled down accordingly. On the other hand, under different circumstances such as driving, the ‘spotlight’ of attention becomes more diffused to cover as much of the visual field as possible (Eriksen & Yeh, 1985; Muller, Bartelt, Donner, Villringer, & Brandt, 2003). Others have argued that the attentional focus is more like a ring with an inhibitory surround (Cutzu & Tsotsos, 2003). Alternatively, recent evidence suggests that rather than a single unitary spotlight, multiple spotlights can exist. Much like theater stage lights that can highlight locations in parallel across the stage, while ignoring intervening regions, attention has been shown to be able to select multiple, distinct regions in space (Pylyshyn & Storm, 1988; Awh & Pashler, 2000; Muller, Malinowski, Gruber, & Hillyard, 2003; McMains & Somers, 2004). And like theater stage lights, it has been argued, attention shifts in a non-continuous manner—instead of sweeping across space as it selects different areas, it is rather snuffed out at one location and turned on at the other (Sperling & Weichselgartner, 1990; Koch, 2004).

All in all, despite having been under investigation for decades, considerable disagreement exists regarding the details of the spotlight of attention. However, there is a general consensus as to its role in rendering visual processing rapid and more efficient.

1.3.1 Attention speeds up processing

Directing attention to a particular location facilitates the processing of objects at that location. For instance, it has been shown that objects that appear at an attended location are processed faster than those that appear at an unattended one. The Posner paradigm demonstrates this fact simply (Posner, Snyder et al., 1980). Consider a simple experiment in which subjects have to detect the presence of a vertical bar in an array of randomly oriented ones (e.g., Figure 1.2a). On each trial subjects are given a cue that signifies the possible location of the targeted vertical bar. On about 80% of the trials, the cue is valid—in other words, on these trials, the vertical bar appears at the cued location.

However, on the remaining trials, the cue is misleading because the target stimulus actually appears elsewhere. What Posner reported was that the time to detect the target was approximately 30 ms shorter when it appeared at a validly cued location compared to the invalid trials.

This phenomenon has been termed the ‘prior entry’ effect—the idea being that attended items have priority in visual processing and enter perception earlier than unattended ones—and is responsible for certain illusions. One of these is the ‘line motion illusion’ (Figure 1.2b) that effectively demonstrates the advantage bestowed by the spotlight of attention to the attended location (Cairney, 1975; Hikosaka, Miyauchi, & Shimojo, 1991, 1993). It occurs when a line is presented to the observer about 50 ms after a patch is flashed on. Instead of perceiving that the line appears all at once, observers believe it appears gradually, originating on the side of the flash. The explanation for this illusion is that the flash attracts attention automatically because of its salience, following which the parts of the line closest to the attentional focus enter perception first, giving rise to an illusion of motion.

Thus both these examples illustrate one role of attention in favoring the processing of particular locations with the result that attended objects and locations are perceived earlier than unattended ones.

1.3.2 Attention helps in object recognition

A popular paradigm that is used to study the effects of attention is called the visual search paradigm. In one version of this paradigm, made popular by Anne Treisman, subjects have to search for a target that appears on some proportion of trials, embedded within a display of distractors (Treisman & Gelade, 1980). Subjects are instructed to be as quick as possible, and the time taken to locate the target, the reaction time, is measured. On those detection tasks where the reaction time is independent of the number of distractors, search is said to be “parallel.” These tasks are also called ‘pop-out’ tasks because subjectively the target appears to ‘pop-out’ of the display (Figure 1.3). The tasks where reaction times increase linearly with the number of distracting elements are said to be serial. Treisman observed that simple feature-detection such as color or orientation discrimination were examples of parallel search. A red disk ‘pops-out’ from an array of green distractors, and a vertical bar is easily detected when surrounded by a number of horizontal ones. However, tasks that involved detecting conjunctions of features were not quite as easy and actually became more time-consuming as the number of distracting elements increased (Figure 1.3). These were the serial-search tasks.

To explain these results, Treisman argued that individual feature detectors in the visual system could locate simple features in parallel across the visual field. Their conjunctions, however, could only be processed after first shifting attention to their location. Thus, for a complex object, attention was necessary to integrate its various features into a coherent representation. Since shifting attention is a time-consuming process, search times were expected to increase with the number of items that had to be serially searched through. This hypothesis is referred to as the ‘Feature Integration’ theory of attention or FIT (Treisman & Gelade, 1980).

The FIT proposes a solution to yet another problem referred to in the literature—the binding problem (von der Malsburg, 1981). The crux of this problem lies in the ambiguity that is present when different objects are presented together. For instance, when looking through a kitchen cabinet, a green round platter and a red long one could be confused such that the round plate could be thought to be red while the long one green. This problem is not as flippant or hypothetical as it sounds at first—it has been shown that when only a limited amount time is available for processing, features from different objects can indeed be mistakenly bound together giving rise to “illusory conjunctions” (Treisman & Schmidt, 1982). The solution that the FIT proposes to solve this problem is not an unreasonable one (Treisman, 1996, 1998, 1999). By focusing attention on an individual object at the expense of the rest, attention necessarily binds the different features of that object together, giving rise to an accurate representation of the visual scene.

1.4 Different forms of attention

Empirically, two different aspects of attention have been identified. One component is under our control—we decide to direct attention to a region of our choice and accordingly shift the ‘spotlight’ as it were, to that location. The second aspect of attention however is completely involuntary. Certain salient objects or events in the visual world automatically attract our attention, independent of our will. These two forms of attention are known as ‘Top-down’ and ‘Bottom-up’ attention, respectively, their names signifying the voluntary, directed nature of the one, and the stimulus-driven, automatic nature of the other. Additionally, attention can not only be directed to a location in space (‘spatial attention’) but can also focus on a particular feature of an object, or a specific object itself. For instance, we can choose to look for the color brown regardless of where it might appear, or look for an entire log of firewood itself. ‘Feature-based,’ ‘object-based,’ and ‘spatial’ attention are different manifestations of ‘top-down’ attention since to some extent they are under our voluntary control. Far from being merely pedagogical, these distinctions reveal themselves at both the behavioral and neuronal levels as is further discussed below.

1.4.1 Bottom-up attention

Bottom-up attention is driven by prominent stimuli in the visual scene—a black raven on a lush green lawn is automatically spotted because of its salience with respect to the rest of the scene. Generally speaking, objects that differ from the surrounding visual scene based on simple features such as contrast, color, orientation, and so on, bias visual processing in their favor and therefore can be effortlessly differentiated (Bergen & Julesz, 1983; Braun, 1994; Braun & Julesz, 1998). Such features, conspicuous in contrast to the rest of the display, are said to be 'salient.' We have already encountered this form of attention, for it is processing such as this that underlies the 'pop-out' mechanism that occurs during visual search—as we noted earlier, certain stimuli can be readily detected because of their salience in the visual search display (Treisman & Gelade, 1980).

Computational models of Bottom-up attention rely on the notion of 'saliency maps' (Koch & Ullman, 1985) that mark the locations of different salient features in a scene (Figure 1.4). The various features drive feature detectors that enter into competition with each other. The most salient features win the competition through a winner-take-all mechanism, and attention is directed to the corresponding location (Itti & Koch, 2001). To preclude the same object from attracting attention over and over, inhibition of return is modeled such that once a location is chosen as a result of the competition, it is prevented from doing so again for a certain period of time. This inhibition of processing at a previously attended location has been observed psychophysically as well. Behavioral studies show that processing at a particular location is hindered approximately 200 ms after attention is drawn to that location, resulting in slower reaction times (Posner & Cohen, 1984; Klein, 2000). This useful mechanism thus ensures that over time, visual processing can be directed towards novel locations, thereby facilitating an exploration of the environment.

1.4.2 Top-down attention

Vision does not always rely on objects catching our eye. Instead we spend a fair amount of time actively searching for objects of interest that cannot be detected easily. So for instance, a snowboarder must seek out his slope on a misty day through the snow and fog, or a shopper must scan the sea of faces at a crowded bazaar when looking for his favorite vendor. In situations such as these, the relevant objects are themselves not conspicuous enough and furthermore external noise and cluttered environments make the detection process more tedious. It is during these active searches that top-down attention is invoked. Top-down attention manifests itself in different forms–focal or spatial attention, feature-based attention, and object-based attention–each of which is described in more detail below.

1.4.2.1 Focal or spatial attention

In some circumstances when bottom-up salience is not adequate for detection, it becomes necessary to focus attention at different spatial locations to find the object of interest. For instance when stimulus features are present in a cluttered or noisy environment (such as a familiar face in a crowded market), or when their contrast is too low, they are not able to strongly activate the relevant saliency maps. In these cases, saliency maps cannot successfully resolve the location of objects, and the different stimuli in the scene compete against each other for representation. Top-down focal attention must then be directed to different locations to settle the competition in favor of the target stimulus. This form of top-down attention has been studied extensively over the last 20 years using electrophysiological and neurophysiological techniques, and thus its role in biasing competition in favor of target stimuli is relatively well understood.

Isolated stimuli

In order to understand the effects of attention at the single-cell level, much work has been done with behaving monkeys. Attention is usually manipulated by having the monkey either covertly attend (i.e., without moving its eyes) to a stimulus placed within a neuron's receptive field (RF) or by directing attention away from the RF. A number of studies have reported that for an isolated stimulus placed within the RF of a neuron, the neuron's response is enhanced when the animal's attention is drawn into the RF compared to when it attends outside the RF. The effect of attention on the stimulus is comparable to a multiplicative scaling (McAdams & Maunsell, 1999a): the response when attention is focused on the stimulus is a scaled factor of the response when attention is directed away (Figure 1.5a). This increased firing rate could be expected to improve stimulus discriminability by increasing the representation of the stimulus relative to the spontaneous background neuronal activity, a finding consistent with behavioral studies that report that detection is improved for attended stimuli (Posner, Snyder et al., 1980; Lee, Koch, & Braun, 1997; Lu & Dosher, 1998). Such response enhancements have been observed in V1 (Motter, 1993), V2 (Motter, 1993; Luck, Chelazzi, Hillyard, & Desimone, 1997), V4 (Spitzer, Desimone, & Moran, 1988; Connor, Preddie, Gallant, & Van Essen, 1997; McAdams & Maunsell, 1999b, 1999a), and also in dorsal areas MT/MST (Treue & Maunsell, 1996; Treue & Martinez Trujillo, 1999). However, not all scholars have observed increased firing to single attended stimuli in V4 and other areas (Moran & Desimone, 1985; Haenny, Maunsell, & Schiller, 1988; Maunsell, Sclar, Nealey,

& DePriest, 1991; Luck, Chelazzi et al., 1997). In these studies, the effect of attention was negligible. A possible explanation for these inconsistent observations has been suggested by Reynolds and Desimone who reported that increases in firing rates comparable to those observed when attention is drawn into the RF of a V4 neuron occur when the contrast of the stimulus is increased (Reynolds, Pasternak, & Desimone, 2000; Reynolds & Desimone, 2003). In other words, increasing the bottom-up salience of one of the stimuli and shifting the focus of top-down attention in favor of that stimulus had equivalent effects on neuronal responses. Furthermore, the effects of attention on neuronal responses depended on the contrast of the stimuli. For low and intermediate values of contrast, the largest increases in firing rates were observed. However, for high-contrast stimuli that saturated the neuronal response even in the absence of attention, the effect was almost negligible. Similar observations have been made in MT (Martinez-Trujillo & Treue, 2002). Thus, perhaps dissimilarities in the contrast of stimuli across different studies might help explain the inconsistent effects of attention that have been observed for isolated stimuli placed in the RF of neurons. Those studies that used well-contrasted stimuli would be expected to show negligible effects of attention, while studies using low-contrast stimuli should show larger attentional effects.

Multiple competing stimuli

The effects of attention in crowded environments, when multiple competing stimuli are placed inside the RF, are even more striking. In a seminal study, Moran and Desimone examined the neural correlates of spatially directed attention (Moran & Desimone, 1985). They presented monkeys with two stimuli while recording from neurons in visual area V4 and the infero-temporal cortex. One of these stimuli was the preferred stimulus of the neuron in that it enhanced the firing rate of the cell when presented alone within the receptive field (RF). The other “poor” stimulus elicited only a weak response when presented alone. When both stimuli were presented simultaneously in the RF, the response of the cell was a weighted-average of its responses to the two individual stimuli. Thus, the response to the pair was intermediate between the response to the preferred and poor stimuli. However, when the monkey was instructed to attend to either one of these stimuli, the cell’s response was strongly affected depending on which stimulus was attended. If the monkey attended the preferred stimulus, the neuron’s firing rate increased from the intermediate response to the strong response elicited by the preferred stimulus when presented alone. Conversely, when the monkey attended the poor stimulus, the response dropped to the low response elicited by this stimulus when seen alone. Thus, attention biased the response of the cell in favor of the attended stimulus. This result has since been replicated in V2 and V4 (Luck, Chelazzi et al., 1997; Reynolds, Chelazzi, & Desimone, 1999) (Figure 1.5b), and MT (Treue & Maunsell, 1996, 1999).

Tying it together with fMRI

For obvious reasons, studying the effects of attention on visual processing at the single cell level is very constrained in the human brain (although I will discuss my foray into this area in the last two chapters of this thesis). Instead, most of the work that has been done with the living human brain has relied on recent advances in fMRI techniques. Currently the relationship between the BOLD signal measured during fMRI studies and the action potentials measured in single-cell animal electrophysiology remains unclear. The BOLD signal is an indirect gauge of underlying neural activity, as it is based on the relative concentrations of oxygenated and deoxygenated blood. Increases in the measured signal are caused by a surplus in blood flow to different areas in the brain that overcompensates for the neural-activity dependent decrease in blood oxygenation levels. Currently the discrepancy between supply and demand for oxygenated blood is unclear. Logothetis and colleagues recently examined the relationship between BOLD and neural activity (Logothetis, Pauls, Augath, Trinath, & Oeltermann, 2001; Logothetis & Wandell, 2004). Using a 4.7T scanner, they were able to simultaneously acquire electrophysiological and fMRI data from monkeys. Based on their recordings, they concluded that the BOLD signal reflected changes in the field potential of the extracellular medium surrounding neurons. For a majority of their recordings, the BOLD signal was correlated strongly with the low frequency portion of the mean extracellular field potential (mEFP). This low frequency range, termed the local field potential (LFP), corresponds to slow electrical signals and sub-threshold activity that probably exerts a local influence on neurons. It is believed that although the LFP signal does not itself reflect the action potentials of the output neurons, it possibly carries the input signal for local processing. Thus the BOLD signal would correspond primarily to the input signals for local processing of neuronal information rather than the output action potentials that are transmitted to other brain areas and which are mainly recorded during single-cell recordings in monkeys.

Despite the differences between the two signals, fMRI has served as a valuable tool in probing the activity of the human brain. Pertinent to our current discussion, the effects of attention on visual processing have been studied in depth. Numerous investigators have explored the changes in the amplitude of the BOLD signal when the focus of attention is manipulated while subjects view different objects. Additionally, the role of attention in resolving competition between multiple objects present in the visual field has been probed. The upshot of all this work is that mechanisms similar to those observed in monkey single-cell experiments are observed in the human visual cortex.

Kastner and colleagues (Kastner, De Weerd, Desimone, & Ungerleider, 1998; Kastner, Pinsk, De Weerd, Desimone, & Ungerleider, 1999) tested the effects of attention in areas V1, V2, VP, V4, and TEO under two conditions where objects were presented either simultaneously (SIM) or sequentially (SEQ) with presentation times such that visual stimulation was equal in both conditions (Figure 1.6a). In the SEQ condition, a single isolated stimulus appeared at one of four locations in the visual field, while in the SIM condition, four stimuli appeared simultaneously. Attention was manipulated by instructing subjects to perform an attentionally demanding task at fixation or covertly pay attention to the stimulus closest to fixation. Thus, by manipulating attention in the SEQ and SIM conditions, Kastner et al. were able to investigate the changes in the BOLD signal for isolated and competing stimuli.

Based on monkey electrophysiology, one could predict that overall activity would be lower in the SIM condition because of competitive interactions between multiple objects in this condition. More specifically, this effect should be observed in higher brain areas such as V4 and TEO rather than area V1 because, while multiple stimuli can fit within the larger RFs of the higher areas, V1 RFs would be too small for competition to occur. A second prediction would be that activity in the attended condition should be higher than that evoked by the unattended stimuli.

Consistent with these predictions, activity was reduced in response to stimuli in the SIM condition in areas V4 and TEO, reflecting the suppressive effects due to competition between multiple objects in the RF. This effect was not observed in V1, consistent with the notion that V1 RFs are too small for competitive interactions to occur since it is difficult to encompass more than one stimulus in the RF. Mirroring the increase in firing rates observed at the single-cell level, covertly shifting attention to the stimuli increased the BOLD response. Furthermore, BOLD activity increased more to the simultaneously presented stimuli than to the sequentially presented ones, demonstrating that attention has a greater effect in cluttered environments compared to effects on isolated stimuli (Figure 1.6b). The magnitude of the attentional effect was also greater in the higher areas compared to V1—that is the effects of competition were reduced by paying attention, in proportion to the degree of competitive interactions in the different visual areas.

These findings in fMRI are thus in accordance with those observed in single cells in monkeys. Together this body of work supports the notion that attention plays a key role in visual processing by resolving the struggle among multiple objects competing for a neural representation. Furthermore, the results from the fMRI studies indicate that the attentional effects observed at the single-cell level are prevalent and dominant enough to be present in the populations of neurons comprising the voxels under study.

The role of attention in resolving competition among stimuli is well summarized by the ‘biased-competition’ framework of attention (Desimone & Duncan, 1995). According to this model, a cluttered visual environment activates populations of neurons that compete with each other. When observers direct attention towards a location in the visual field, top-down biasing signals are generated in favor of the stimuli at that location. Consequently, the neurons representing the favored stimuli remain active while the others are suppressed. In other words the attended stimuli dominate neuronal responses while the ignored stimuli are filtered out.

1.4.2.2 Feature based attention

Up till now, we have talked about spatial aspects of attention—when observers know where to look for the object of interest, they can spatially direct attention to those locations. Thus, if I know I left my keys either by the kitchen table, or the telephone, I can specifically look for them in those places. Often though, one does not know where to look—the item of interest could be anywhere in the visual scene. Rather, one has some information about what the item looks like. In this situation, attention is not paid to a position in space, but instead to some distinguishing feature of the object. Thus, for instance, at a race-track, if I know that my jockey's colors are green and white, to find my favorite horse, I can just look out for greens and whites.

This form of top-down attention, known as feature-based attention, has been shown to modulate the activity of neurons coding for the relevant features regardless of their spatial location (Motter, 1994b, 1994a; Chelazzi, 1995; Treue & Martinez Trujillo, 1999). In one study, Treue and Martinez Trujillo recorded from areas MT and MST in monkeys; two areas that are strongly associated with the processing of visual motion (Treue & Martinez Trujillo, 1999). On each trial, monkeys were required to fixate a central cross, while a random dot pattern moving in a given neuron's preferred direction was presented within the RF. Another random dot pattern that could either move in the same or opposite direction was presented outside the RF of the neuron (Figure 1.7). Thus with this setup, by instructing the monkey to always attend to the stimulus that appeared outside the RF, the attended direction of motion could be manipulated without changing the location of the attended stimulus. The results of this experiment showed that when the monkey attended to the preferred direction of motion outside the RF, neuronal responses were increased. On the other hand, when the monkey attended to a stimulus moving in the null direction, neuronal responses were suppressed. These attentional modulations could not be attributed to spatial attention since the attended location was constant in both conditions. Instead, these results provide evidence for the selection of stimulus features by attention. Similar feature-based attentional modulations have been observed in the human MT-complex as well (Saenz, Buracas, & Boynton, 2002). Thus, independent of the spatial location of stimuli, top-down signals appear to be able to selectively affect neuronal responses by targeting stimulus features.

Top-down influences set up as a result of actively expecting and processing certain features have also been observed in visual search experiments (Wolfe, 1994) where these signals appear to “guide the search” for relevant targets. For instance, in conjunction search tasks (e.g., Figure 1.3), if observers are looking for a red, vertical bar among red horizontal and green vertical bars, they can use top-down expectations about the target to restrict their search to only the red items. Within the set of red items, the ‘Guided Search’ theory would predict that a pop-out effect would occur for the vertical target.

These top-down expectations for certain features in the visual field will be of importance to us later on in this thesis. In Chapter 4, we will introduce the term ‘top-down behavioral relevance’ in referring to these signals, the name reflecting the relevance of actively expected features to the observer. As will be discussed later, we show that top-down signals can often be sufficient for the efficient processing of certain stimuli, even in the near-absence of spatial attention.

1.4.2.3 Object-based attention

Attention can also select out entire objects. For instance, when two objects are superimposed on each other, all features of one of the objects can be selected, while the other object is ignored. Remarkably, if these objects are novel, on later presentations, subjects fail to recognize the unattended object, even though it had previously occupied the same location as the attended one (Rock & Gutman, 1981). Similarly, subjects find it more difficult to distinguish two features if they belong to two different objects than if they are part of one object (Duncan, 1984; Baylis & Driver, 1993), suggesting that under some circumstances, entire objects are selected out with more ease than individual features.

Evidence for object-based attention at the physiological level comes from fMRI studies. In one such study, O’Craven and colleagues (O'Craven, Downing, & Kanwisher, 1999) demonstrated the neurophysiological effects of object-based attention in an fMRI study where they presented subjects with faces and buildings that were superimposed. On each trial, either the face or the building moved back and forth along one of four axes (Figure 1.8a). Depending on instructions, subjects had to report if the direction of motion, or the position of the stationary stimulus (which was jittered on each trial), was the same. BOLD activity was measured in the MT complex (which processes motion), the fusiform face area (FFA) located in the middle fusiform gyrus that has been shown to respond selectively to faces, and the parahippocampal place area (PPA), a region in the parahippocampal cortex that responds to buildings and spatial locations. Consistent with theories of feature-based attention, activity was higher in each area when the relevant feature (motion, faces, or buildings) was attended. Additionally, activity in each area was also increased when the corresponding object was irrelevant, as long as it was associated with the attended attribute rather than with the unattended one. For example, when subjects attended to the motion, and the face (house) stimulus had been moving, the BOLD signal was higher in the FFA (PPA), compared to those trials when the house (face) had been moving. On the other hand, when subjects attended to the position of the stationary stimulus, activity in the FFA (PPA) was higher when the face (house) was stationary (Figure 1.8b). This was true even though the faces and houses were never attended to directly. These results thus demonstrate that attention is able to select entire objects and increase the neuronal representation of the object associated with an attended feature.

1.5 Outside the spotlight

We have seen thus far that attention, in all its manifestations, selects stimuli or events for processing, making them available to perception. However, even when attention is intensely engaged in something, such as writing this section of the thesis or driving a car for the first time, we are still aware of our unattended surroundings—in other words, the world is not in darkness outside the spotlight of attention. In view of this then, what kinds of stimuli are detected in the absence of attention? What features are preferred by the visual system such that the processing of these objects occurs in a straightforward fashion? An understanding of this issue is essential to our understanding of vision because it should shed light on the design of the visual system. For certain objects to be processed rapidly and effortlessly, the visual system must be constructed in a way to support this processing. On the other hand, those stimuli that require attention reflect the constraints of the architecture of the system. Accordingly, this section addresses the processing that occurs in the absence of attention.

1.5.1 The dual-task paradigm

It is first necessary, however, to describe how attention is manipulated in the laboratory. Various paradigms are used to control attention, one of which has already been mentioned—the visual search paradigm. Another one, particularly relevant to this thesis, is the dual-task paradigm (Sperling & Melchner, 1978; Braun & Sagi, 1990; Braun, 1994; Braun & Julesz, 1998), which is discussed in depth in Chapter 2. For the present though, a brief description will suffice (Figure 1.9).

The idea behind the dual-task paradigm is fairly simple—subjects are presented with two tasks, which they are instructed to perform either separately (single-task conditions) or simultaneously (dual-task condition). The primary task is one that is known to engage attention, and the secondary task is the one the investigator is interested in probing. Performance on the secondary task is compared in the two conditions: when it is performed alone (and attention is available), and when it is performed in the dual-task condition (and attention is engaged by the primary task). If performance falls dramatically in the dual-task condition, then attention is necessary for the secondary task. However, if subjects perform equally well on the secondary task in the two conditions (i.e., performance is equivalent when attention is available and when it is removed by the primary task), then this task can be performed even when spatial attention is not fully available.

There is one caveat with this interpretation of the dual-task paradigm, which can also apply to other paradigms that attempt to measure the attentional requirements of different tasks. As was mentioned earlier, we do not have a reliable, agreed-upon measure of attention. Thus there is no way to quantify the amount of attention engaged by the primary task in the dual-task experiments. Based on control experiments (discussed in Chapter 2), we can be more or less sure that “most” of attention is made unavailable to the secondary task but we cannot conclude that the primary task engaged “all” the attentional resource. Hence, it is more accurate to demand what tasks can be done in the “near-absence” of attention, rather than discuss the processing that occurs in its complete absence. Conversely, we can avoid the problem of being at a loss to describe attention by using this particular property as a working definition: focal attention will be the resource engaged by the primary task. This is not an unreasonable definition because as we show in the following chapter, unavailability of this resource does sometimes result in severe performance deficits. Through much of this thesis, this is the definition of attention we use.

1.5.2 What can be done in the near-absence of attention?

Visual search experiments make a distinction between stimuli that ‘pop-out’ and those that require attention in order to be detected. Similar observations are made in one version of the dual-task paradigm when such search arrays are presented as the secondary task. Most ‘pop-out’ visual search tasks can be done in the dual-task condition with no decrement in performance, while performance on conjunctions of stimuli suffers (Braun & Sagi, 1990, 1991; Braun, 1994; Braun & Julesz, 1998).

In another version of this paradigm, single, isolated stimuli are presented in the secondary task, and on each trial, subjects are required to report whether the stimulus was the target or distracter. In this case, simple feature detection tasks, such as telling two colors apart, can be performed in the near-absence of attention. Other examples of tasks that do not suffer in the dual-task condition are orientation discriminations, simple shape discrimination (circle versus triangle), and simple motion discrimination (leftward versus rightward). However, once the task becomes slightly more complex, stimuli cannot be discriminated without attention (although occasionally after extensive training, these discriminations can be performed (Braun & Julesz, 1998)). For example, discriminating stimuli composed of two colors (e.g., a bisected red-green disk from its mirror image) or of two orientations (e.g., rotated letters ‘T’ versus ‘L’), or complex motion stimuli (e.g., clockwise versus counter-clockwise) are dependent on attention. The drop in performance on these tasks is dramatic (to chance levels) in the dual-task condition when the attentional resource is engaged by the primary task (Braun & Julesz, 1998; Lee, Koch, & Braun, 1999).

Based on such results from the dual-task and visual search paradigms, it was therefore concluded that simple, salient stimuli could be detected outside the focus of attention, but that performance on more complex stimuli was dependent on attention. This classical view of attention held sway in the literature for many decades.

1.5.3 Natural scene processing

Although a lot of work has gone into the formulation of this view of attention, much of it has been concerned with the processing of simple geometric shapes. While these stimuli serve us well in the laboratory, they are a far cry from the complexity of the natural visual world. In comparison to the artificial combinations of the few features that comprise geometric shapes, natural stimuli are composed of combinations of several orientations, colors, hues, contrasts, and spatial frequencies. According to the classical view of attention, processing of complex natural scenes should therefore necessarily require attention. However, in our day-to-day environment this constraint would be a severe encumbrance since for any object to enter perception, we would have to engage the attentional mechanism. Besides, evidence exists that natural scenes can be categorized very rapidly. Simon Thorpe and colleagues have used EEG to measure brain activity while subjects performed a natural scene categorization task (Thorpe, Fize, & Marlot, 1996). In those experiments, subjects were briefly presented with unmasked pictures (each for 20 ms) and asked to report as quickly as possible if they contained an animal or not. Differences in the EEG signal were observed as early as 150 ms between trials that contained an animal and those that did not. Similar results were observed when non-biologically significant stimuli such as vehicles, were the targets (Figure 1.10) (VanRullen & Thorpe, 2001). These results are inconsistent with the view that attention is necessary to process natural scenes because just shifting attention to a location can take up to a few hundred milliseconds.

Thus, does the visual system deal with these scenes differently, or do the findings from simple geometric shapes generalize to natural scenes? Recent work in fact indicates that natural scenes can be categorized even when top-down focal attention is not fully available. As we discuss in more detail in Chapter 2, using the dual-task paradigm, Li and colleagues presented subjects with natural photographs and asked them to report whether an animal or vehicle was present (Li, VanRullen, Koch, & Perona, 2002). It turned out that even when attention was focused away from these pictures, subjects were successfully able to detect the presence of the targets. Using ERPs, Rousselet et al. (Rousselet, Fabre-Thorpe, & Thorpe, 2002) have come to a similar conclusion. In a set-up related to the one used by Thorpe et al. they presented subjects with one or two pictures and asked them to detect animals. Differences in the ERP signal between target and distracter trials occurred at the same time, regardless of whether one or two pictures had to be processed. Thus, notwithstanding their "complexity," natural scenes can be easily dealt with by the visual system.

1.6. The specific questions and the organization of this thesis

It appears therefore that it is not only the "complexity" of a stimulus that determines its attentional requirements. The stimulus type (natural stimuli versus artificial, geometric shapes) must also be taken into account. Thus for artificial stimuli, with increased complexity, attention becomes necessary for successful processing—simple features can be processed outside the focus of attention but the more complex feature conjunctions rely on attention. Is a similar outcome observed with natural stimuli? Does attention become necessary when the categorization task is made more "complex"? This is the question that is addressed in Chapters 2 and 3.

In these two chapters, we focus on categorization tasks where targets and distractors are very similar to each other, in that they share the same features and only differ subtly in the spatial arrangement of these features. The dual-task paradigm is used to investigate the attentional requirements of this type of processing. Chapter 2 describes the attentional requirements of face-gender processing. This project was done in collaboration with Patrick Wilken who helped with the design of the experiment and also ran a group of subjects who rated the discriminability of the faces in the face database. Chapter 3 investigates an even finer discrimination, face-identification, in the near-absence of attention, a project done together with Lavanya Reddy who provided me with the large set of celebrity faces and also collected data for three of the subjects described in Experiment 1.

Chapter 4 examines the brain activity underlying the processing of face stimuli when attention is not entirely available. How does the activity elicited by these stimuli fare in the near-absence of attention? The FIT predicts that complex recognition tasks and their associated cortical areas must rely on attention for successful processing. But is this also true for natural stimuli? We use fMRI in this chapter to get a broad look at cortical responses over many thousands of face-responsive neurons as attention is manipulated. Farshad Moradi, who collaborated with me on this project, helped with the overall design of the experiment and provided me with the Matlab scripts for reading the Siemens images into Matlab.

The final two chapters address visual processing and attention in the human brain at an even greater resolution—at the single-cell level. In Chapter 5 we study the neuronal representation of natural stimuli in single neurons. I collaborated with Rodrigo Quian Quiroga for this project, in which we investigate how generally neurons in the medial temporal lobe encode particular stimuli. In other words, when a neuron responds to a picture of a particular person, for instance, will it also respond to other pictures of this person? What about non-pictorial representations, such as a letter string of the name of this person? For this project, both Rodrigo and I collected and clustered the data from the patients and did the preliminary analysis. The ROC and ANOVA analyses described in the chapter as well as some final clustering, were done by Rodrigo.

The goal in Chapter 6 is to study the effect of attention on natural stimuli at the single-cell level and to address an apparent contradiction that arises because of the discrepancy between the results from the dual-task paradigm and change-blindness experiments. For, on one hand, dual-task experiments provide evidence that natural stimuli can be processed outside the spotlight of attention. However, change-blindness occurs when attention is not available to process natural scenes. As will be discussed in the final chapter of the thesis, there is no real contradiction. Instead, it turns out that under some conditions, stimuli fail to be strongly represented at the neuronal level, and it is then that attention is called upon. This work was done together with Patrick Wilken and Rodrigo Quian Quiroga. Patrick Wilken helped with the design of the experiment. Rodrigo Quian Quiroga helped with data collection on some patients.

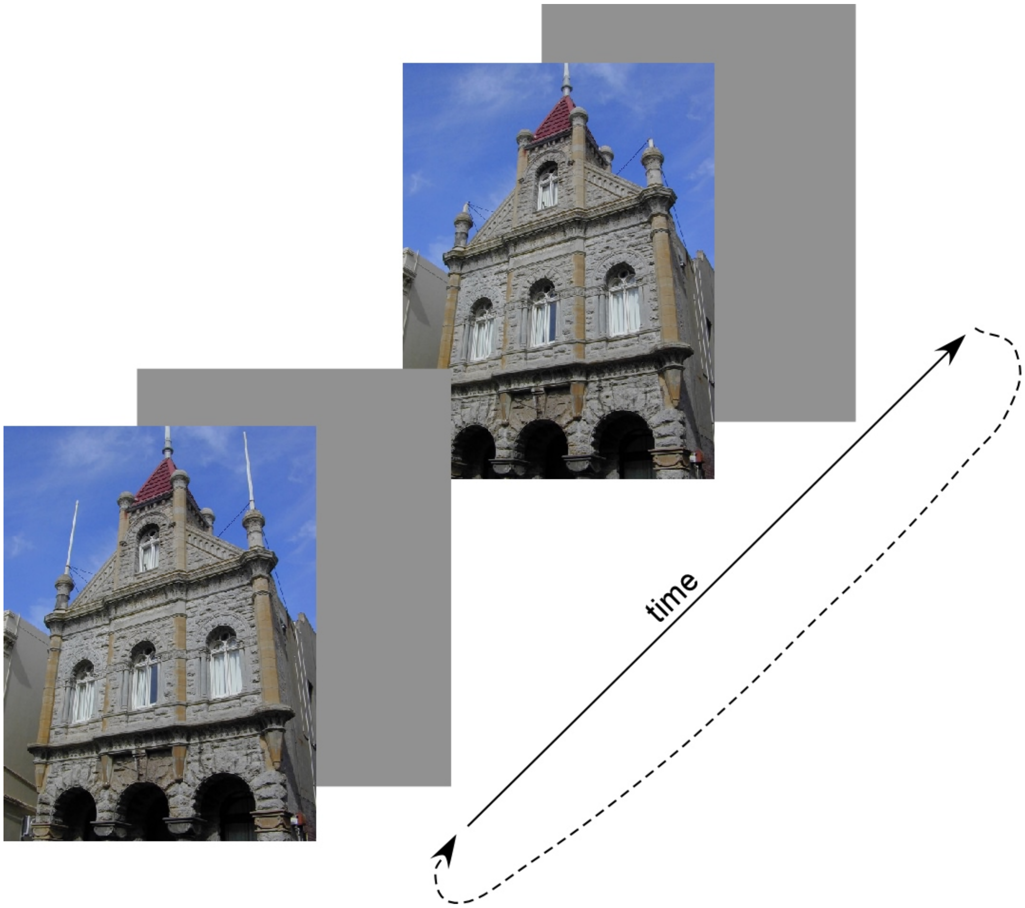

Figure 1.1 An example of a change-blindness paradigm. Two pictures differing in some respect (such as the color, size, position, or even presence of an object as shown here) are presented to observers, separated by a blank screen. It can take observers several repetitions of the cycle before they notice the change. (Modified from www.cognitrn.psych.indiana.edu)

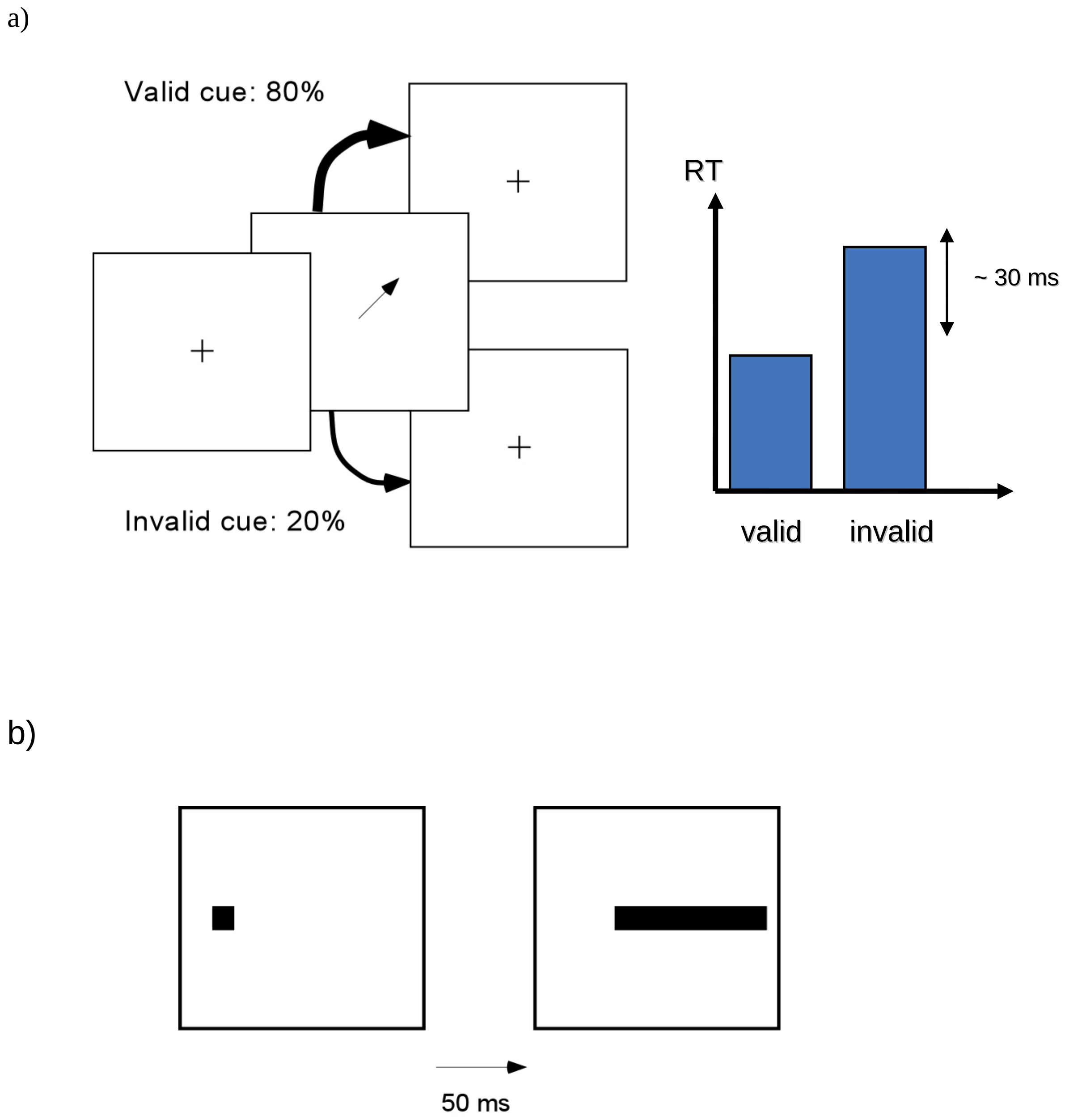

Figure 1.2: Attention speeds up processing. a) The Posner Paradigm. Subjects are shown a fixation cross at the beginning of the experiment, after which they are cued to attend to a particular location. On 80% of the trials, the cue is valid and correctly predicts the location of the target stimulus. On the remaining trials, the cue is invalid. Reaction times (RT) in detecting the target are measured, and it is observed that valid locations are processed faster (by about 30 ms) than invalidly cued ones. b) The 'Line-Motion' illusion. A patch is flashed on the screen (left) about 50 ms before a line appears on the right). Observers believe that the line appeared from theleft to right (i.e., it originated on the side of the flash), rather than appearing all at once, thus giving rise to an illusion of motion.

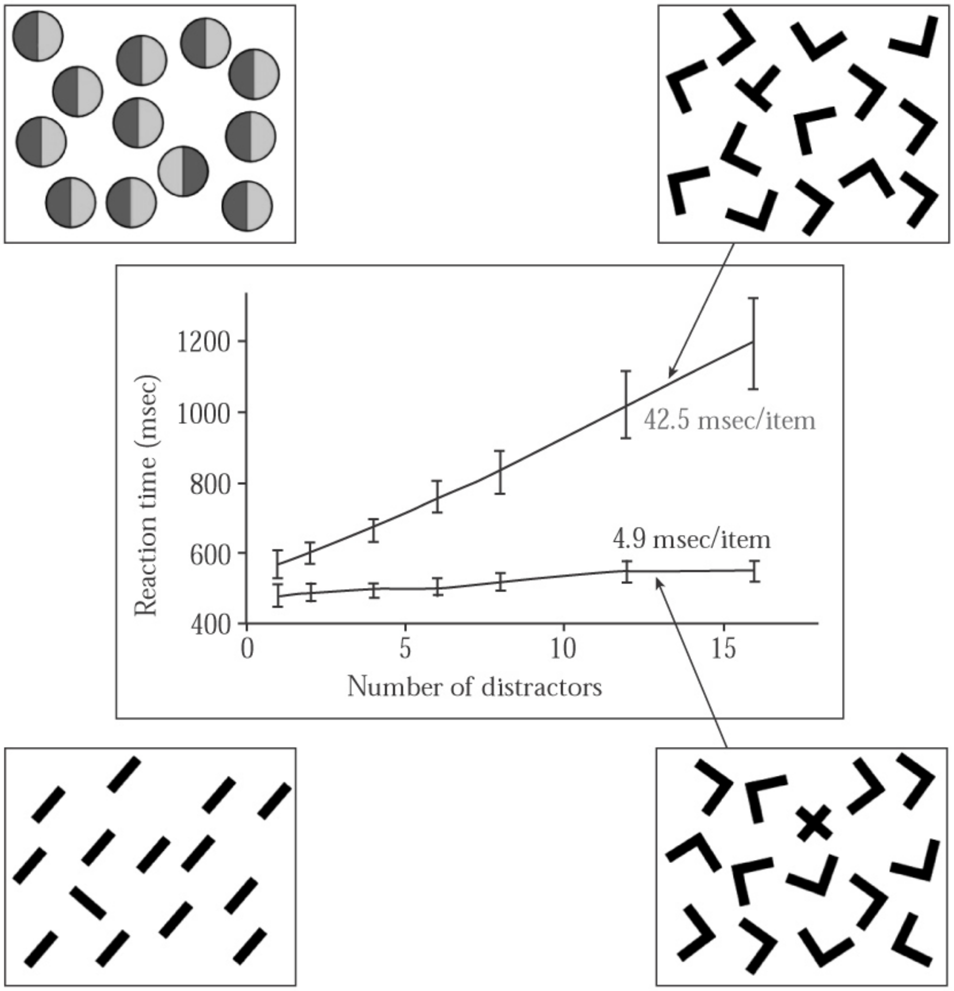

Figure 1.3: The visual search paradigm. Examples of parallel search (bottom panels) and serial search (top panels). In parallel search, the target (odd element) ‘pops-out’ of the display. The reaction time (RT) to detect the target is independent of the number of items in the search array. In serial search, detecting the target item requires one to search the entire display. The RT increases linearly with the size of the search array (VanRullen & Koch, 2003). (With permission from Koch,2004).

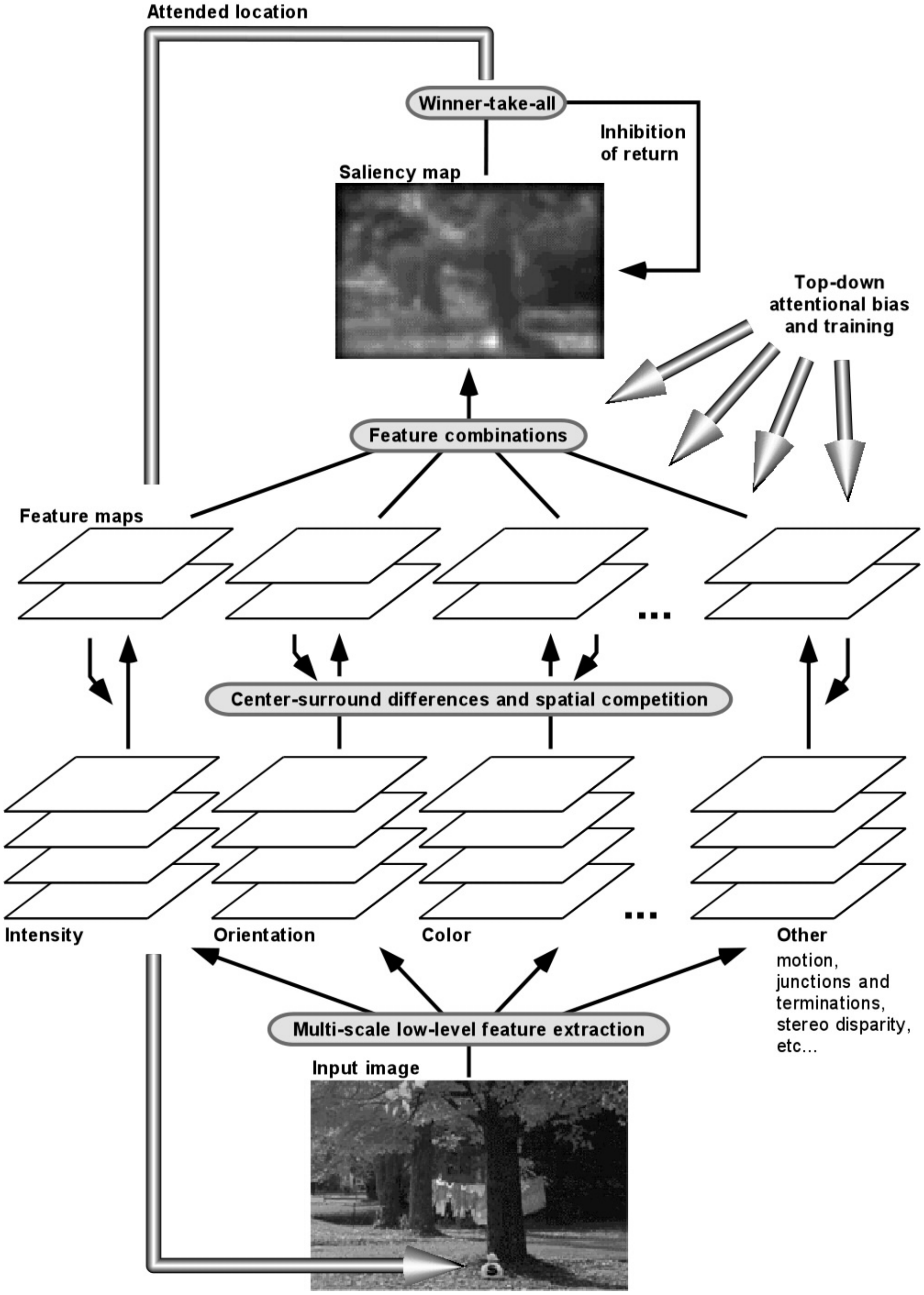

Figure 1.4: Saliency map model of attention (Itti & Koch, 2001). An input scene is broken up into its various feature dimensions and fed into feature detectors. Spatial competition occurs in each of the feature channels through center-surround differences. The resulting activities in each feature map are combined together and fed into the saliency map. A winner-take-all mechanism then determines which are the winning locations, and this information is transmitted to areas responsible for shifting attention. A process called ‘inhibition of return’ ensures that the same winning location does not win over and over and thus allows attention to visit other areas. (With permission from (VanRullen, 2003). Copyright Elsevier, 2003.)

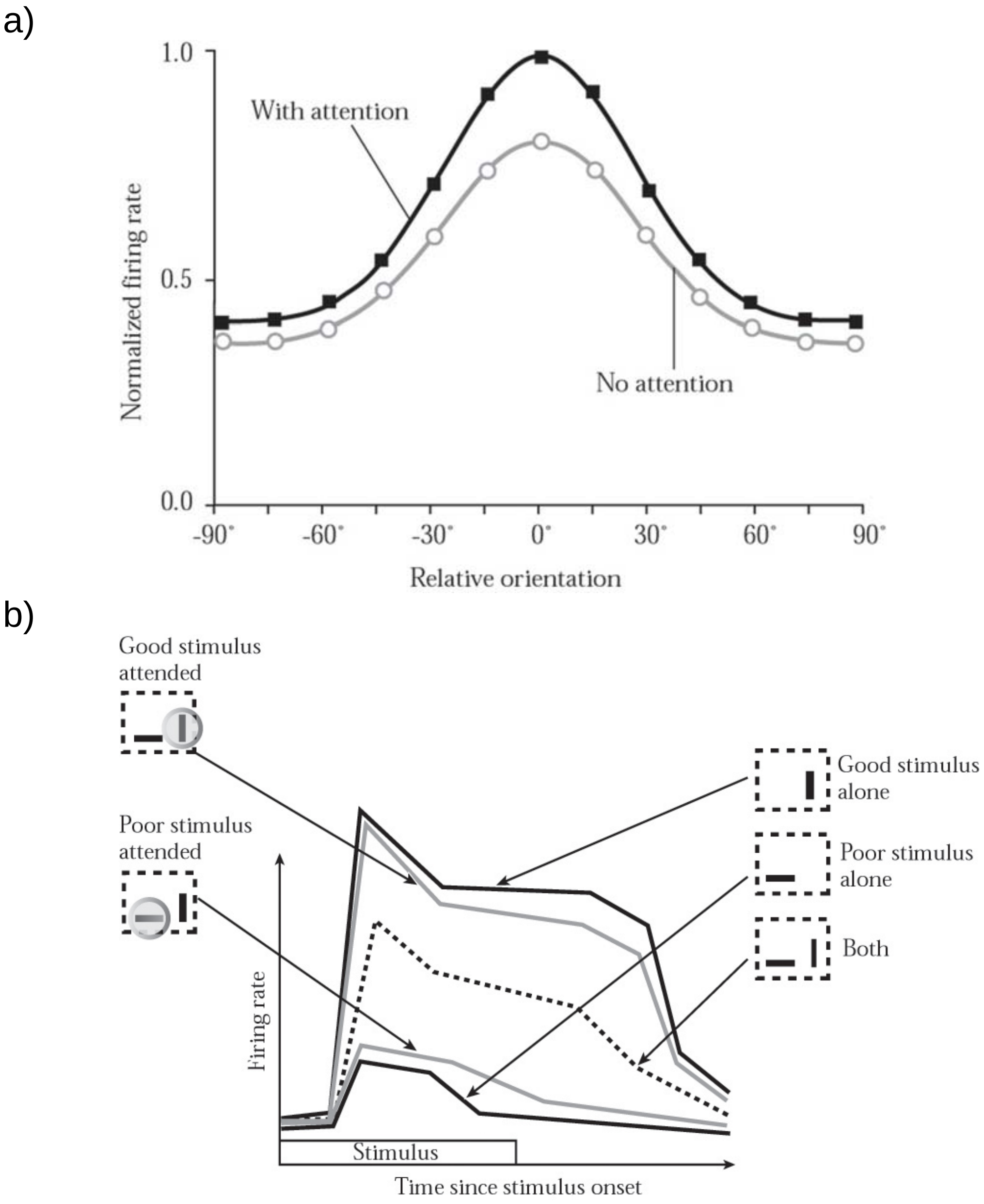

Figure 1.5: Attentional effects at the neuronal level. a) The effects of attention on an orientation tuning curve. The figure shows a typical tuning curve for a neuron in V4. The neuron has a preferred orientation that it responds to most strongly, and its response decreases with changes in orientation. Attention results in a multiplicative increase in firing rates (filled circles) at all orientations (e.g., McAdams and Maunsell, 1999). b) The effects of attention in a crowded environment. This hypothetical neuron prefers vertically oriented bars (‘good stimulus’) and responds only weakly to a horizontal bar (‘poor stimulus’). When both bars are placed in the RF of the neuron (dashed square), the cell’s response is a weighted average of the response to the individual stimuli. However, when attention is directed to the good stimulus (circled stimulus), the response increases and approaches the response elicited by this stimulus alone. Similarly, attention to the poor stimulus causes a drop in response. Thus, attention biases processing in favor of one of the stimuli. (With permission from (Koch, 2004)).

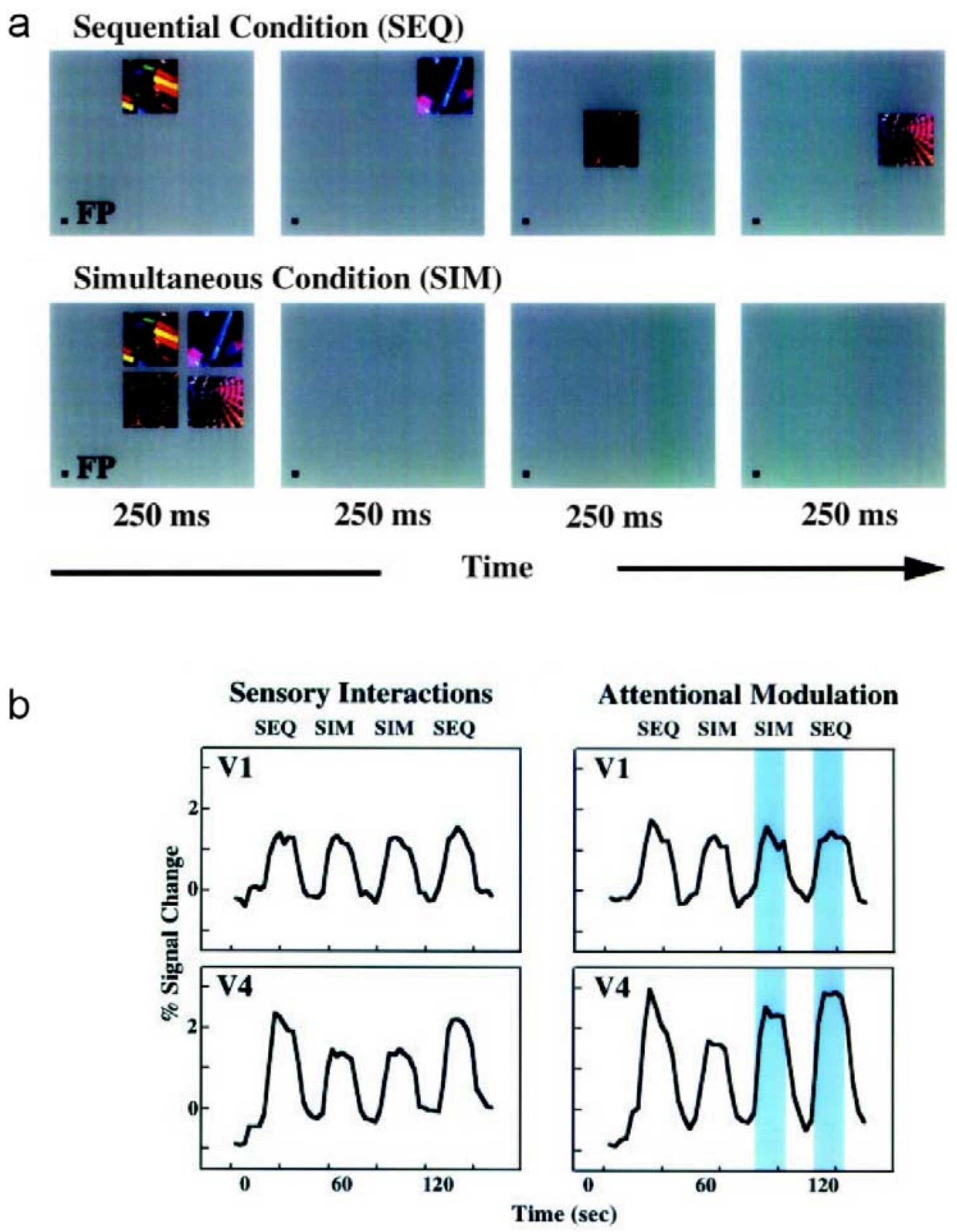

Figure 1.6: FMRI and attention in the human brain. a) Experimental design. Four images were presented at nearby locations to a fixation point (FP) in the upper right quadrant in 2 conditions: sequential (A) and simultaneous (B). Stimuli were presented for 250 ms, followed by a blank period for 750 ms, in each location. This figure shows a stimulation period of 1 s in both conditions, which was repeated in blocks of 18 s. Over time, the physical stimulation was identical in each location. b) Sensory interactions and attentional modulation in V1 and V4. As shown in the time-courses in the left panel, simultaneous presentations evoked less activity than sequential stimulation in V4 but not in V1. This result suggests that competitive interactions were scaled to the size of the receptive fields in these two areas. In the right panel, the effects of attention are shown. In the attended conditions (blue shaded areas) activity was increased more to the simultaneously presented stimuli than to the sequentially presented ones in V4, relative to the unattended condition (unshaded area). (Adapted from (Kastner & Ungerleider, 2000), copyright Annual Reviews of Neuroscience, 2000.)

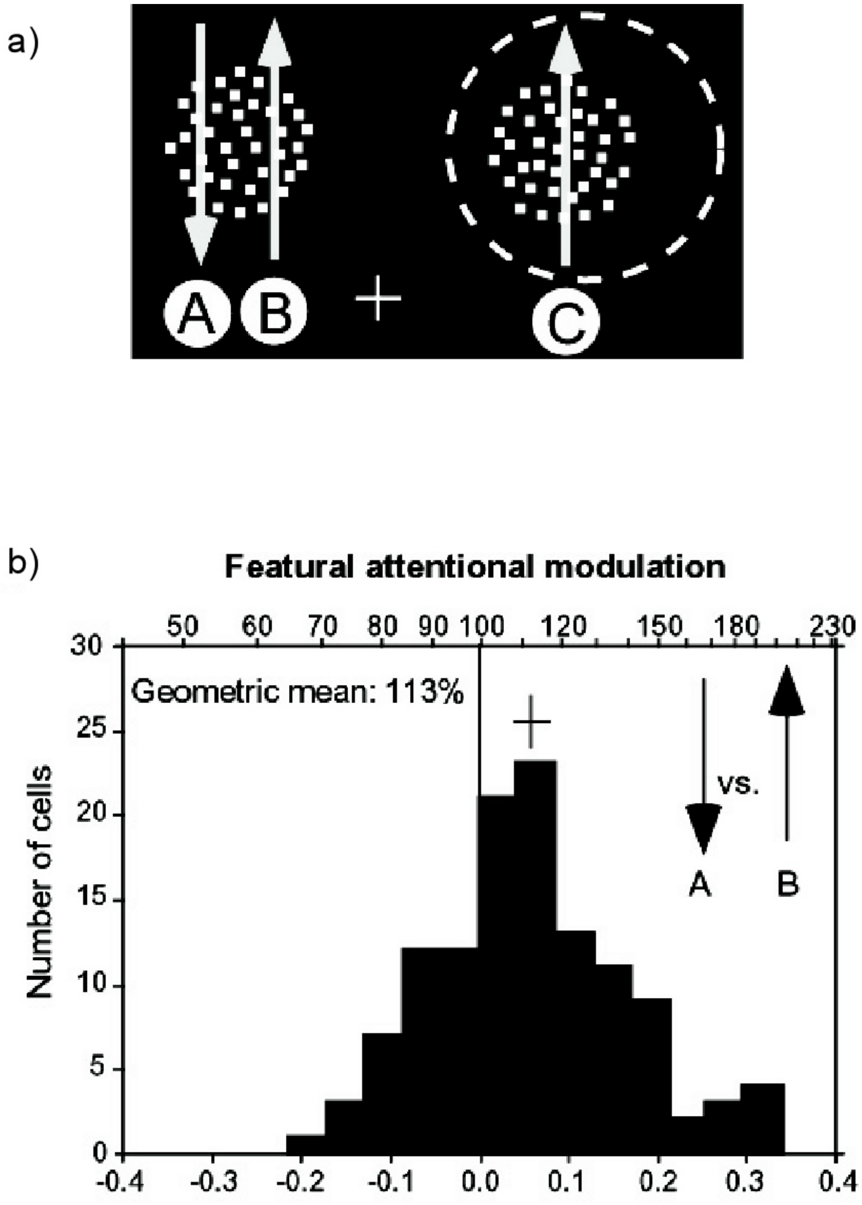

Figure 1.7. Feature-based attention. a) Experimental design used by Treue and Martinez Trujillo 1999. One random dot pattern (RDP) was presented within the RF of the neuron (dotted circle) while another was presented in the opposite hemifield. The stimulus inside the RF moved in the preferred direction of the cell (upward arrow). The other stimulus moved in either the same (B) or the opposite direction (A). The monkey was instructed to attend to one or the other direction of motion, outside the RF. Thus the attended direction could be switched without changing the attended location. b). The histogram shows the attentional modulation of the responses of 131 neurons when the animal attended the preferred (B) versus the anti-preferred (A) directions of motion outside the receptive field. The attentional modulation index AI = $(X_B - X_A)/ (X_B+X_A)$, where X is the width of the tuning curve is plotted on the y-axis. The top scale shows the ratios—the mean ratio is 113%. The rightward-shift of the histogram indicates an increased response when the stimulus moved in the preferred direction of the cell. (With permission from Treue and Martinez Trujillo, 1999. Copyright, Nature Publishing Group, 1999.)

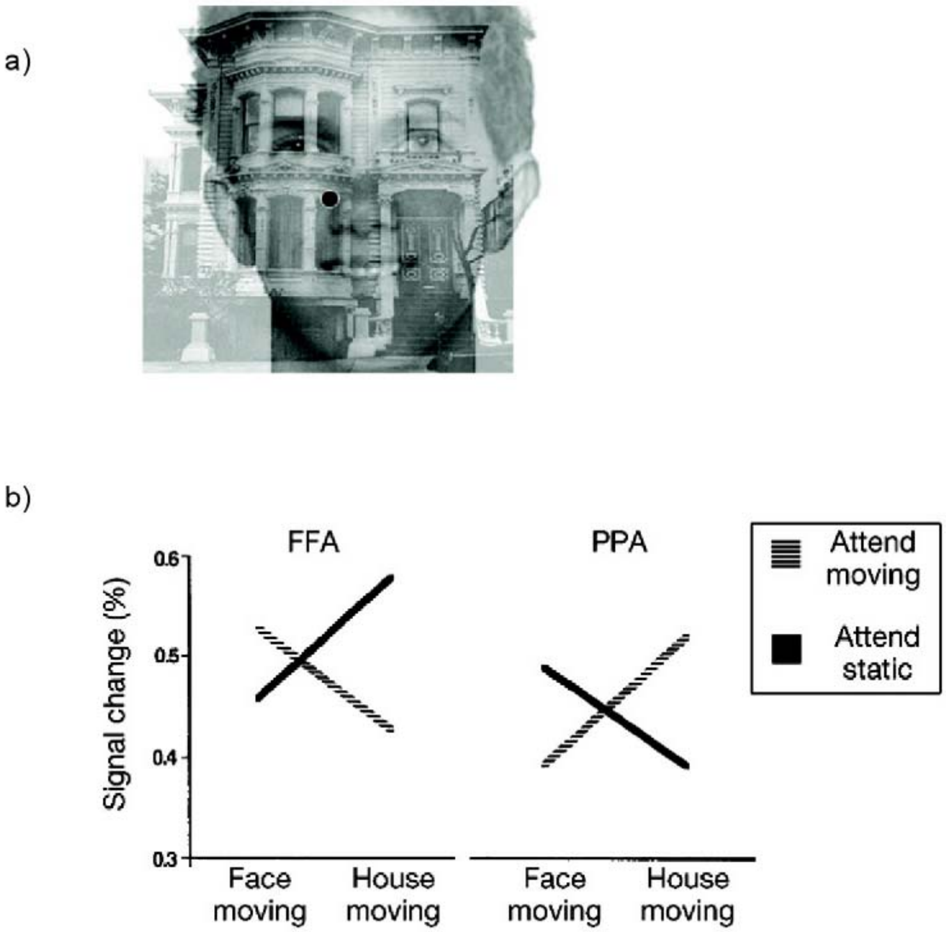

Figure 1.8. Object-based attention. a) Experimental design from O’Craven et al. (1999). Faces and houses superimposed on each other were presented to subjects. Either the face or house could move back and forth along one axis. Subjects had to report if the direction of motion of the moving object or the position of the stationary one (which was slightly jittered in each stimulus) was repeated across trials. b) Mean peak response evoked in the FFA and PPA during the fMRI experiment. In the motion task, the response in the FFA was greater if it was the face that moved, while in the PPA the response was greater if the moving object was a house. In the stationary task, the response in the FFA was greater when the house moved (and the face was stationary) while the converse was true in the PPA. Thus, although neither face, nor house was attended to, their corresponding neuronal representations were enhanced depending on if one of their attributes (motion or position) was attended to. (With permission from O’Craven et al, 1999; copyright, Nature Publishing Group, 1999.)

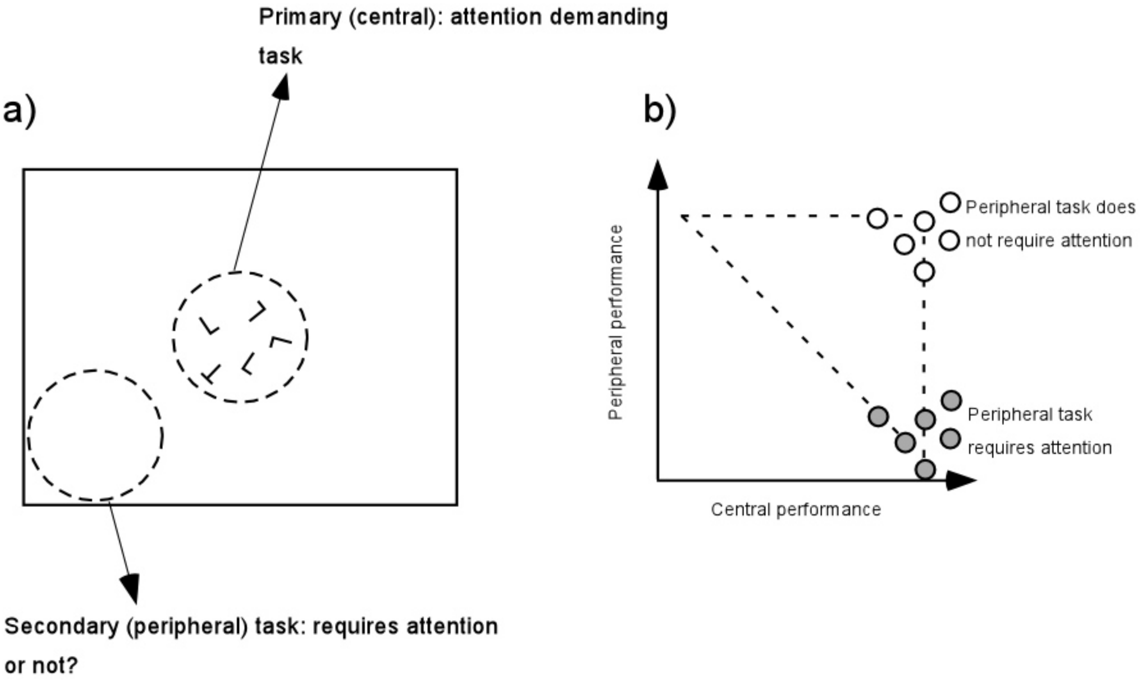

Figure 1.9: The dual-task paradigm. a) Subjects are presented with two tasks: the primary task is presented at the center and is a known attentionally demanding task (in this case a 5-letter discrimination task). The secondary task performed in the periphery is the one under investigation—does this task require attention or not? b) Results from a dual-task paradigm are plotted on an attention-operating characteristic (AOC) plot. Peripheral task performance is plotted against central task performance. If performance in the dual-task condition is comparable to performance when the tasks are done alone, then we conclude that attention is not required by this task. Otherwise, performance on the task is dependent on attention.

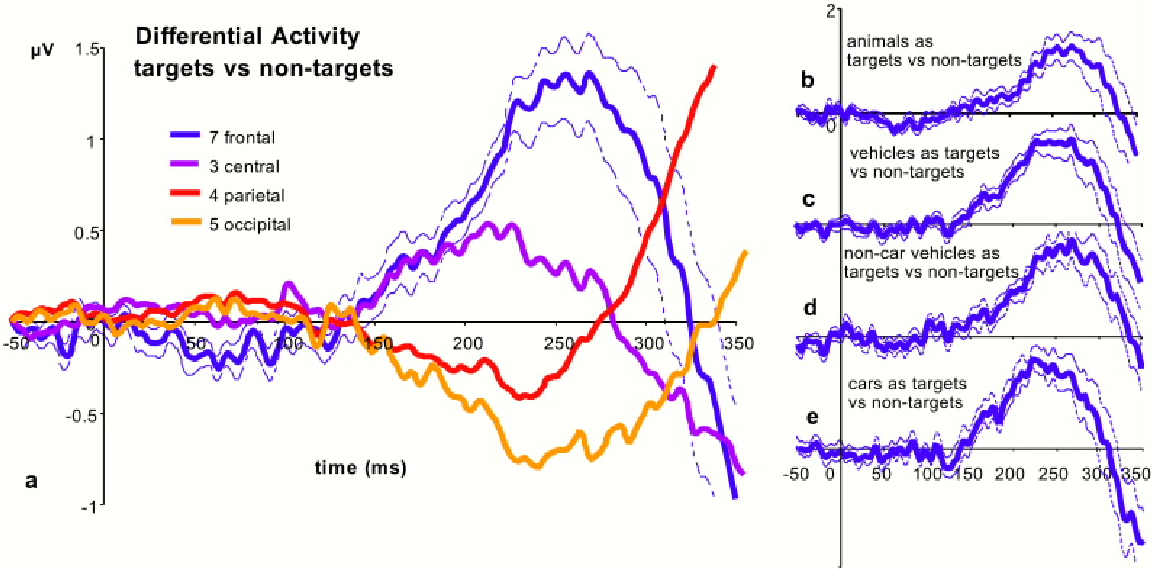

Figure 1.10: Rapid natural scene categorization. Each panel shows the difference in EEG activity following presentation of different categories of natural scenes on target versus non-target trials. a) Targets versus non-targets over all categories. The EEG time course begins to differ after approximately 150 ms over all electrodes reflecting that the visual system can differentiate between targets and distractors at this early time. This discrimination cannot be based on low-level features since targets and distractors are actually the same pictures (viewed under different task instructions). (b-e). Differences between target and non-target trials for different categories of natural stimuli. With permission from R. VanRullen (modified from VanRullen and Thorpe, 2001).