Attention and the Processing of Natural Stimuli: Psychophysics, fMRI and Single Unit Recordings in the Human Brain.

Thesis by

Leila Reddy

In Partial Fulfillment of the Requirements
for the Degree of

Doctor of Philosophy

California Institute of Technology
Pasadena, California
2005
(Defended: May 17, 2005)

Acknowledgments

"Thou anointest my head with oil; my cup runneth over."

My cup doth indeed run over. As I write these lines, I am struck by how fortunate I am to even be in a position to write them, for the conclusion of my graduate work builds on the foundation of the solid, basic education I received as a child. Current estimates by UNICEF hold that 121 million children are denied a primary school education, 54% of them female. About 115 million of these children are from developing nations, such as the one I come from. Given this dismal backdrop, I do consider myself blessed.[1]

To my parents I owe a huge debt of gratitude—for their vision, constant encouragement, and unwavering confidence in me. For giving us many opportunities to learn, both by sending us to great schools and by surrounding us with wonderful books and a stimulating environment at home. For the incredible childhood they made possible—so incredible that I was sad on the eve of my 13th birthday at the thought of growing up! I am also deeply grateful to Lavi, for being a sister that sticketh closer than a friend, for her support, and especially for the hearty laughter. It did much good to my soul. And while I'm on the topic of family, I cannot fail to mention the Millers: Joanna, Josh, Susan, and Victor, who opened up their home to me and treated me like another Miller. The many blissful days I spent at their peaceful house, surrounded by plants, pets, good food, and fine companionship, were rest for mind and body.

The culmination of this thesis would not have been possible without the guidance of the past and present members of my thesis committee: John Allman, Itzhak Fried,

---

[1] I do not intend to imply that a primary school education is all that these children need. 29000 children are estimated to die *every day* as a result of malnutrition and the unavailability of basic medicines. Even as a relatively small portion of society makes progress and accrues more wealth, there are several urgent problems that need to be solved worldwide. We should be grateful for those who do give of themselves and what they have to try to alleviate some of these evils. And strive to help them.

David Kleinfeld, Pietro Perona, Thanos Siapas, Shin Shimojo, and Eric Winfree. Like all good committee members, they grilled me during my qualifying exam and made many useful suggestions during my candidacy examinations.

To my collaborators, I am grateful for valuable discussions and help with data collection: Farshad Moradi, Rodrigo Quian Quiroga, Lavi Reddy, and Patrick Wilken. Their specific and significant contributions to this thesis are detailed in the introductory chapter. My colleagues in the Koch laboratory have helped make my experience at Caltech a positive one: Alexander Kraskov, Lavi Reddy, Rufin VanRullen, Wei Ji Ma, Dirk Walther, Nao Tsuchiya, Fei Fei Li, and McKell Carter. To Farshad Moradi and Steve Flaherty, I am grateful for helping me get initiated with fMRI, and to Melissa Saenz for assistance with Brain Voyager. Candace Vavra, Shannon Moreen O’Dell, and Irene Wainwright have provided valuable help throughout my graduate work.

Alexander Kraskov has been a great office-mate and has helped me a lot in the short time I’ve known him—both in terms of the useful and illuminating discussions we’ve had, as well as collecting data for me at UCLA while I was out of town. The significance of this last good deed can only be fully appreciated by those who have spent long days and weeks at UCLA, waiting to record. David Kewley was a great source of cheerful support every time I ran out of disk space (because of the large quantities of data we collected at UCLA), as well as with all the other computer problems I’ve had. Emily Ho and Eve Isham have helped with all the human recordings described in this thesis, and Tony Fields has provided substantial technical support at UCLA.

Dr. Itzhak Fried, the neurosurgeon at UCLA, has been a constant source of motivation for the experiments that relied on recordings in human patients and has been involved in their design and implementation at each step. He has performed all the surgeries on which the neuronal data was based. His piercing questions and tireless enthusiasm were critical to the successful completion of my projects with him.

My advisor Christof Koch has been a wellspring of inspiration and energy these last five years. When I first joined his lab, my knowledge of neuroscience was minimal. Since then, his boundless enthusiasm and vast interests have enabled me to try my hand at psychophysics, fMRI, and human electrophysiology[2], and have also made the lab a place where diverse topics, from modeling the visual system to trying to model consciousness in mice, are pursued. Such diversity truly enables one to widen one's horizons and keep an open mind. His uncanny ability to ask just the right questions has been critical to my understanding of my own data. Working with him has been an exciting intellectual adventure for which I am sincerely grateful.

---

[2] Unfortunately, this is not to say that my current knowledge of neuroscience (or anything else for that matter) does not leave much to be desired.

Abstract

How does the visual system process stimuli it encounters frequently? To address this question, we investigated the attentional requirements associated with processing natural stimuli in isolation and in crowded environments. Using the dual-task paradigm, we tested observers on discriminating the gender of faces presented in isolation, in the near-absence of focal attention. We observed that performance on this task for poorly attended faces suffered only minimally. Furthermore, we also showed that finer discriminations, such as face identification (celebrities and unfamiliar faces), are unimpaired when attention is unavailable. From a computational perspective these results are surprising since observers are unable to categorize computationally simpler stimuli (e.g., a red-green color disk from its mirror image) under identical conditions. The brain mechanisms underlying the processing of face-gender were probed with fMRI. We found that the BOLD signal in the near-absence of attention was not significantly reduced, provided the faces were behaviorally relevant. This finding, that top-down expectations can be sufficient for high levels of the BOLD signal, is in contrast to current views that hold that the signal is reduced in the absence of focused attention. We took a closer look at attentional effects on neuronal activity by recording from individual neurons in the human brain while subjects performed a change detection paradigm (in which they reported whether they noticed changes made to (natural) stimuli presented in a crowded environment). Subjects were epileptic patients, implanted with depth electrodes in the medial temporal lobe (MTL) for identification of the seizure foci for potential surgical resection. We observed that neuronal responses when changes were correctly detected were significantly higher compared to incorrect trials. Under the common assumption that incorrect performance reflects the absence of attention, we show that MTL neuronal activity is reduced when attention is unavailable. For each cell, on a trial-by-trial basis, we were able to predict the occurrence of a change (67%) and the patients' behavior (58%), significantly above chance. Our results show that the brain can process isolated natural stimuli in the near-absence of attention, while in crowded environments, attention plays a key role, as we have observed at the neuronal level.

Contents

List of Figures

List of Tables

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.

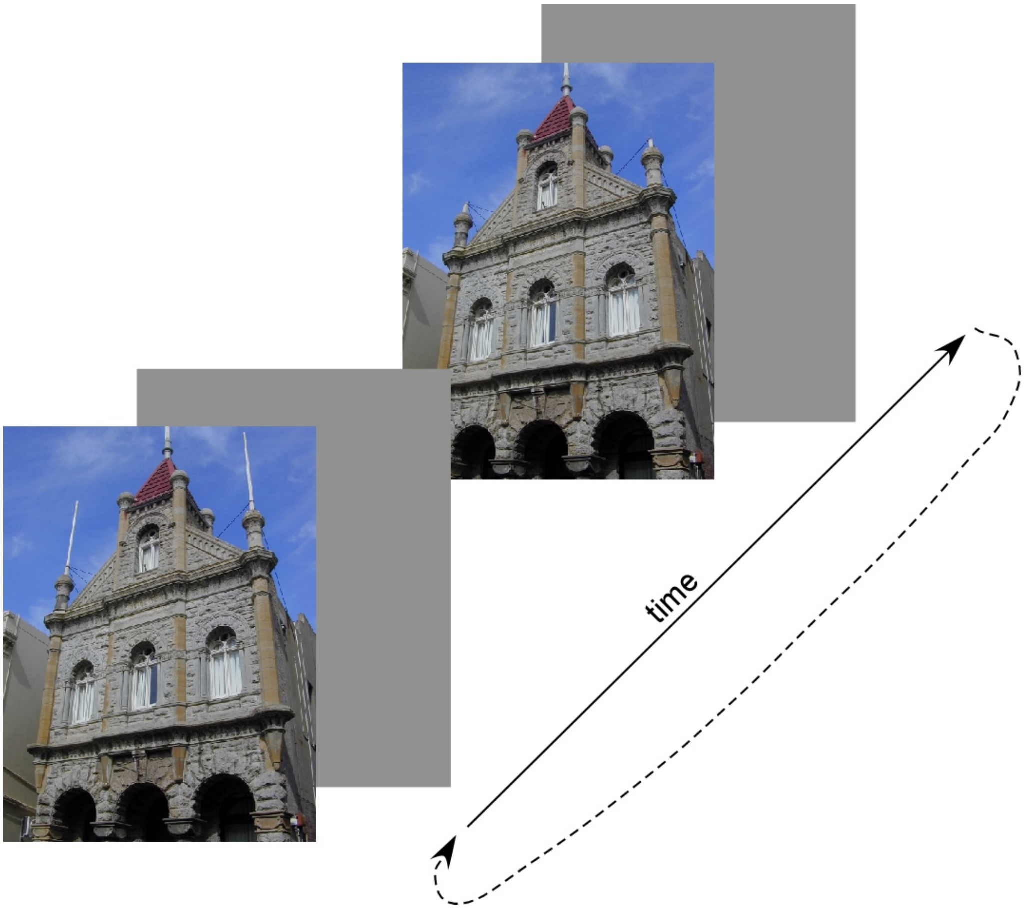


Figure 1.1 An example of a change-blindness paradigm. Two pictures differing in some respect (such as the color, size, position, or even presence of an object as shown here) are presented to observers, separated by a blank screen. It can take observers several repetitions of the cycle before they notice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

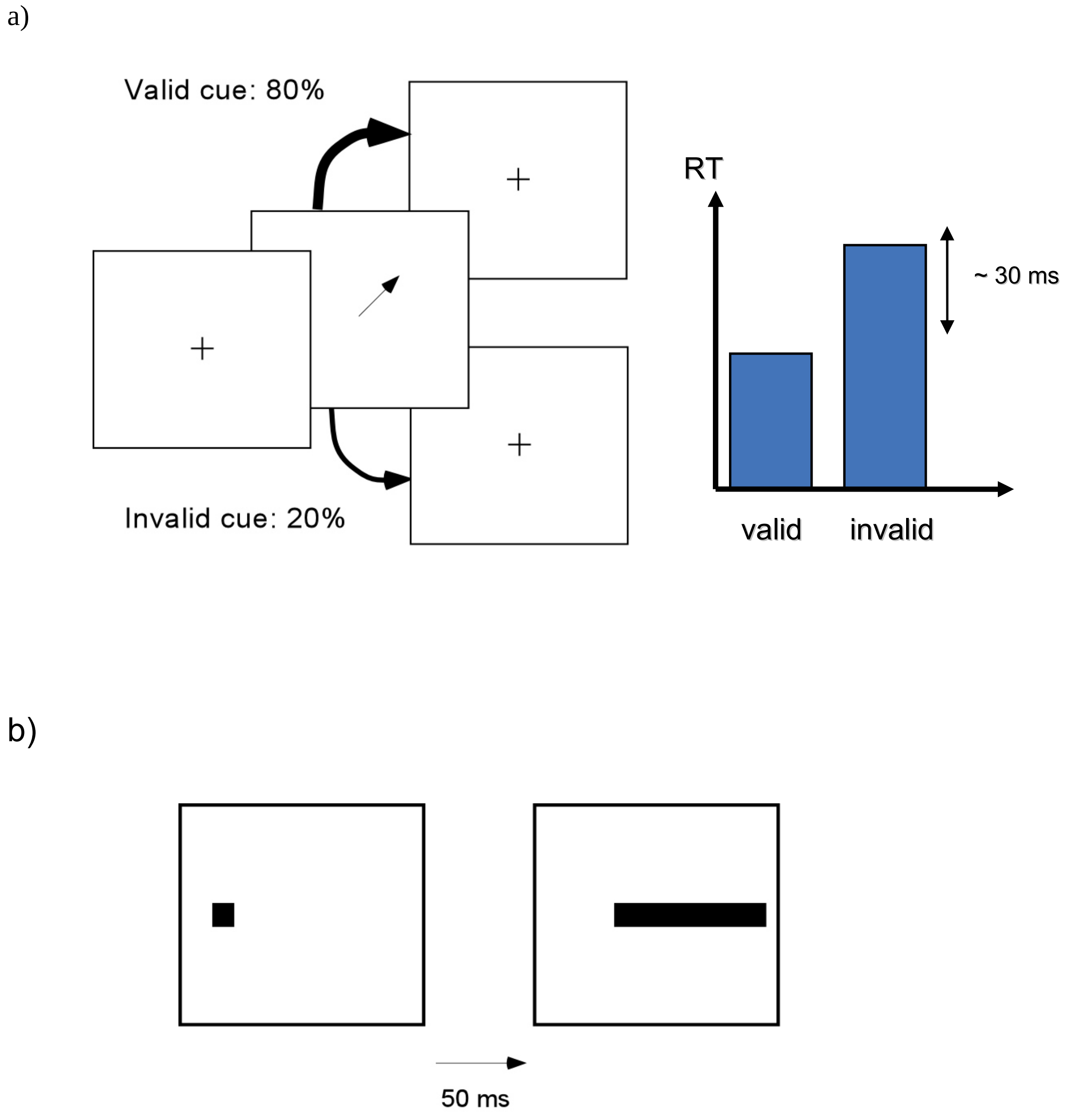


Figure 1.2: Attention speeds up processing. a) The Posner Paradigm. Subjects are shown a fixation cross at the beginning of the experiment, after which they are cued to attend to a particular location. On 80% of the trials, the cue is valid and correctly predicts the location of the target stimulus. 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.

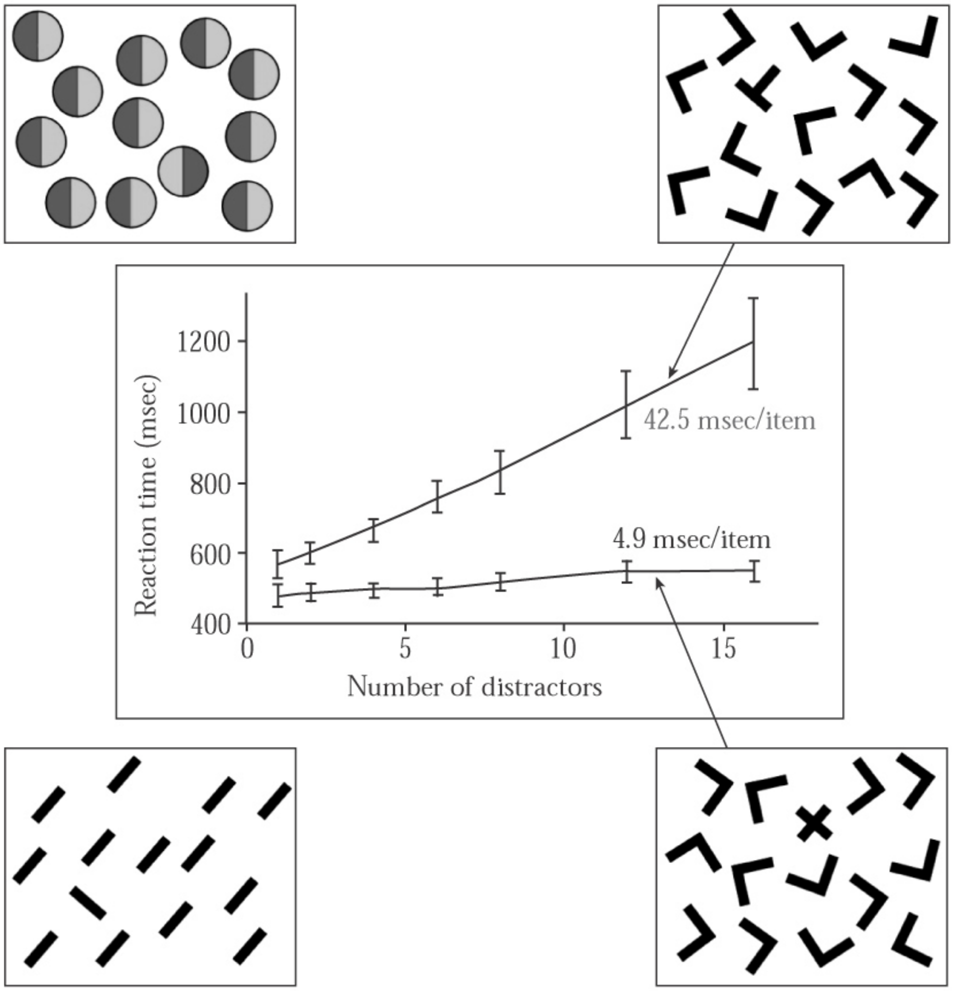


Figure 1.3: The visual search paradigm. Examples of parallel search (bottom panels) and serial search (top panels). In parallel search, the target (odd element) ‘pops-out’ of the display. The reaction time (RT) to detect the target is independent of the number of items in the search array. In serial 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).

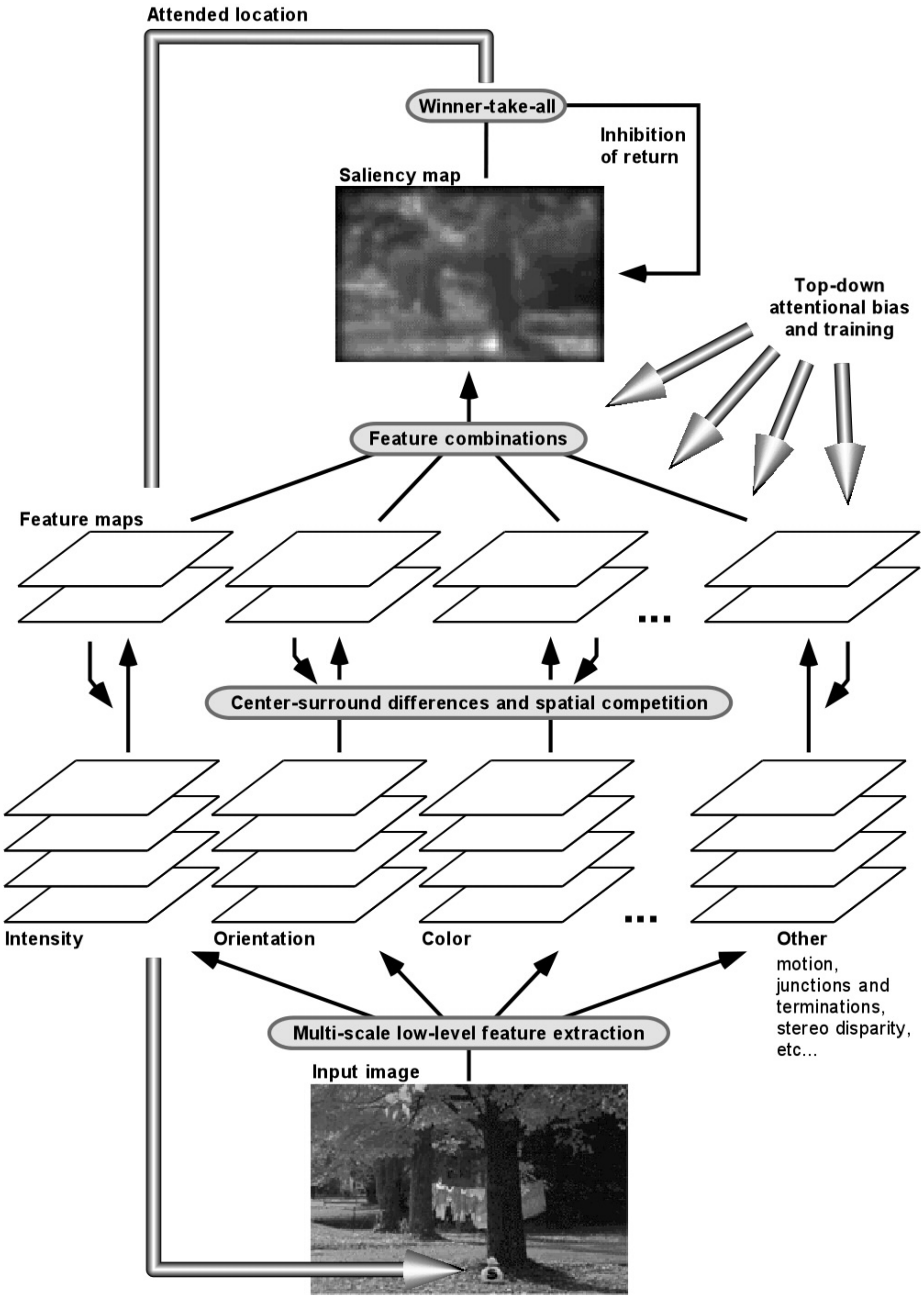


Figure 1.4: Saliency map model of attention (Itti & Koch, 2001). An input scene is broken up into its various feature dimensions and fed into feature detectors. Spatial competition occurs in each of the feature channels through center-surround differences. The resulting activities in each feature map 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.)

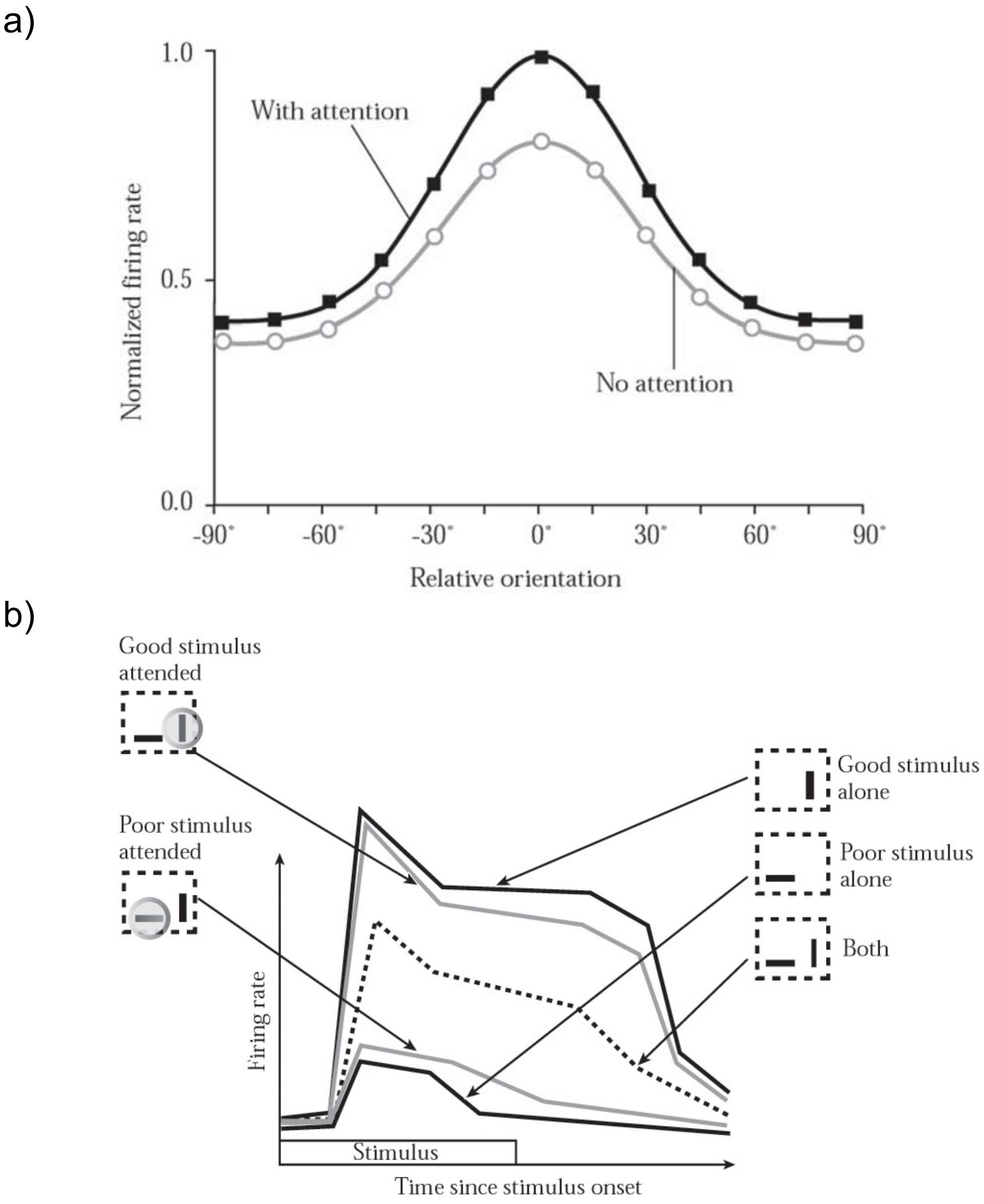


Figure 1.5: Attentional effects at the neuronal level. a) The effects of attention on an orientation tuning curve. The figure shows a typical tuning curve for a neuron in V4. The neuron has a preferred orientation that it responds to most strongly, and its response decreases with changes in orientation. 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)).

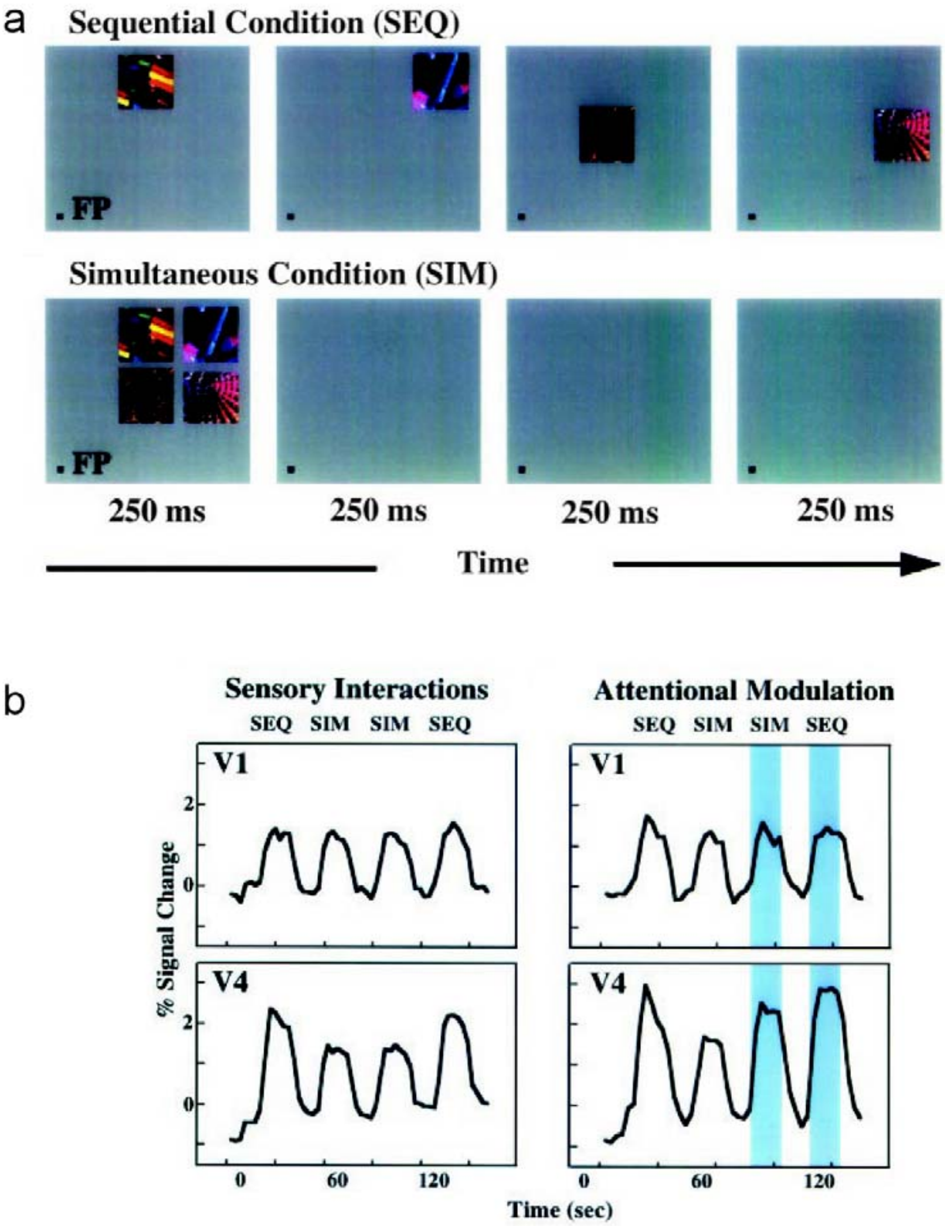


Figure 1.6: FMRI and attention in the human brain. a) Experimental design. Four images were presented at nearby locations to a fixation point (FP) in the upper right quadrant in 2 conditions: sequential (A) and simultaneous (B). Stimuli were presented for 250 ms, followed by a blank period for 750 ms, 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.)

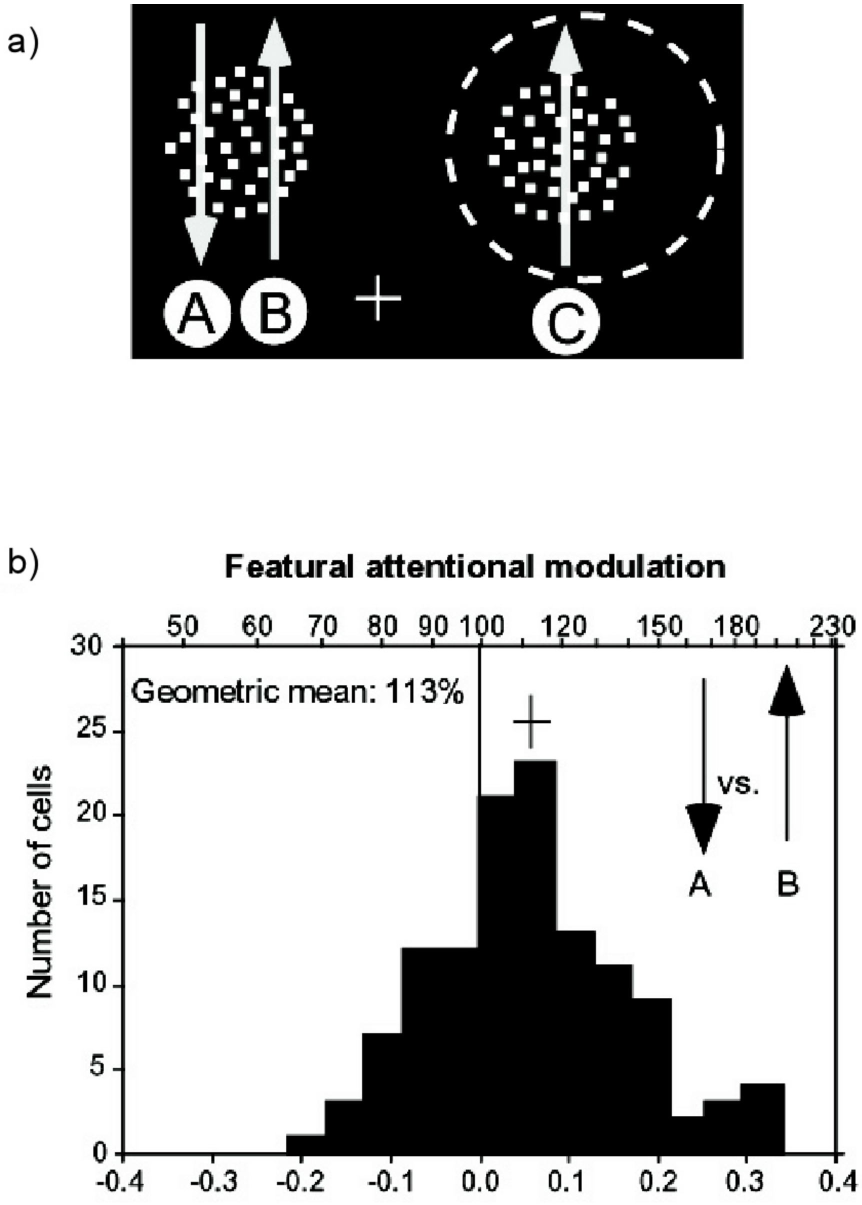


Figure 1.7. Feature-based attention. a) Experimental design used by Treue and Martinez Trujillo 1999. One random dot pattern (RDP) was presented within the RF of the neuron (dotted circle) while another was presented in the opposite hemifield. The stimulus inside the RF moved in the preferred direction 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.)

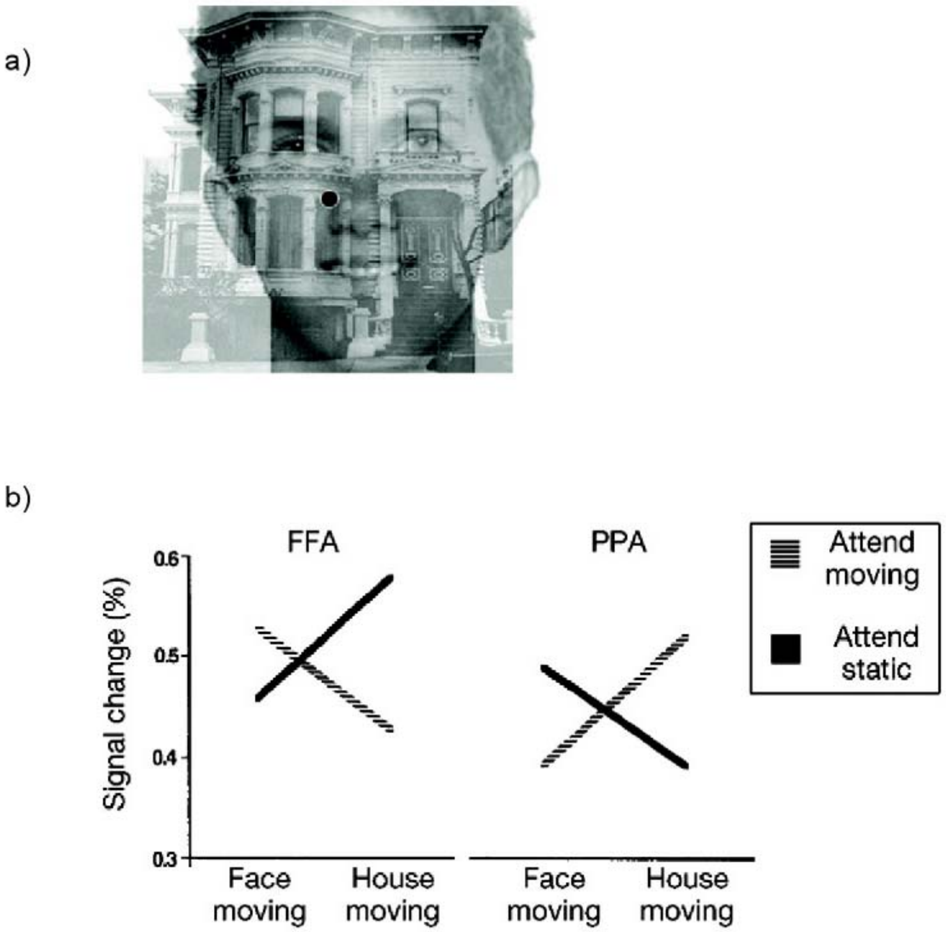


Figure 1.8. Object-based attention. a) Experimental design from O'Craven et al. (1999). Faces and houses superimposed on each other were presented to subjects. Either the face or house could move back and forth along one axis. Subjects had to report if the direction of motion of the moving object or 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.)

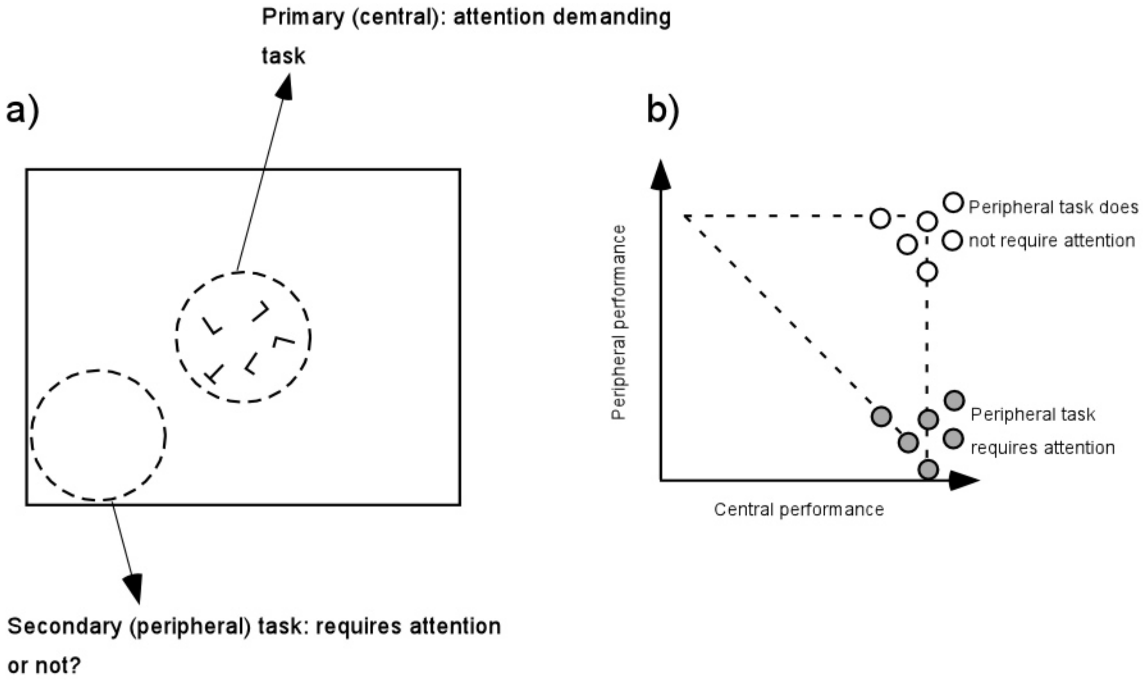


Figure 1.9: The dual-task paradigm. a) Subjects are presented with two tasks: the primary task is presented at the center and is a known attentionally demanding task (in this case a 5-letter discrimination task). The secondary task performed in the periphery is the one under investigation—does this task 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.

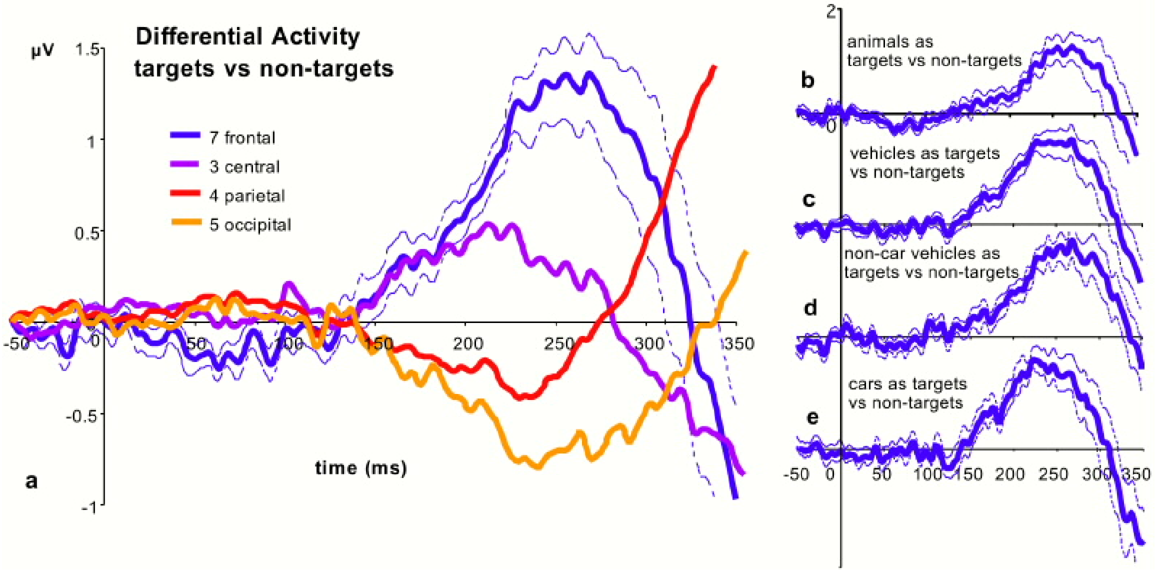


Figure 1.10: Rapid natural scene categorization. Each panel shows the difference in EEG activity following presentation of different categories of natural scenes on target versus non-target trials. a) Targets versus non-targets over all categories. The EEG time course begins to differ after approximately 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).

Chapter 2

Face-gender Discrimination in the Near-Absence of Attention

2.1 Introduction

Visual properties that can be extracted with little attention have long been under investigation. As discussed in the introduction to this thesis, performance on these pre-attentive tasks does not suffer, even when they are carried out in parallel with known attentionally demanding tasks (e.g., in the dual-task paradigm). Typically, these tasks involve simple feature discriminations, such as discriminating between two colors or two orientations. However, once tasks involve discriminating conjunctions of simple features, attention can become a necessary requirement. Thus, when subjects are asked to discriminate between a bisected color pattern and its mirror image, performance is strongly contingent on attention.

Even these “more complex” discrimination tasks, however are leagues away from the sophistication and complexity[3] involved in everyday vision where richly detailed scenes must be constantly inspected. The question of the attentional demands of the processing of these scenes is an interesting one, because on the one hand, classical theories of attention [4] would predict that the visual system must rely on attention to bind together the different features of the objects in the scenes. On the other hand, however, as we have discussed, evidence exists that natural scenes can be classified with

---

[3] We use the notion of “complexity” here and in the following chapter as a conceptual argument, in a computational sense. For instance, complexity can be defined loosely as the number of elementary computer operations it takes to solve a problem.

[4] e.g., the FIT discussed in Chapter 1.

remarkable speed (Thorpe, Fize et al., 1996). Furthermore, our everyday experience is not that of tunnel vision—rather, even when deeply engaged in some task, we can be aware of objects outside the focus of attention.

This issue was recently addressed by Li and colleagues who found that observers were able to discern the gist of natural scenes in the near-absence of attention (Li, VanRullen et al., 2002). Using a dual-task paradigm, they presented subjects with a variety of images of natural scenes and asked them to report the presence of an animal or vehicle (Figure 2.1a). Although each image was only presented for less than 30 ms, subjects were able to achieve surprisingly high levels of performance on these tasks, regardless of the availability of attention (Figure 2.1b). These data thus demonstrate that natural stimulus processing is not limited by the availability of the attentional resource.

From a computational point of view, the result that natural scene categorization can be performed in the near-absence of focal attention is surprising since such categorization tasks are substantially more “complex” than discriminating between two-color patterns. It therefore appears that the “complexity” of a visual discrimination task does not necessarily determine its attentional demands. Instead, the nature of the stimuli involved (natural stimuli versus artificial combinations of individual features) also seems to influence a task’s attentional requirements.

Where does this ability to process natural stimuli in the absence of focal attention break down? Would attention become necessary if the natural stimulus discrimination task became more “complex”—i.e. if targets and distractors are made more similar to each other and share the same features? To address this issue, we chose to determine the attentional demands of a face-gender discrimination task. Since both male and female faces share the same features (i.e., eyes, noses, mouths, etc.) and these features are similarly arranged in both sets of faces, this task could help us determine if the limits of pre-attentive processing of natural stimuli are determined merely by the complexity of the stimuli, or if even more complex natural stimuli can be processed pre-attentively.

Numerous experiments have explored the attentional demands of face processing. Although faces are believed to be of particular importance to the visual system (Farah, 1995; Kanwisher, McDermott, & Chun, 1997; Farah, Wilson, Drain, & Tanaka, 1998; Ro, Russell, & Lavie, 2001), most studies have failed to demonstrate a pop-out effect for faces or other natural objects in visual search (Nothdurft, 1993; Kuehn & Jolicoeur, 1994; Purcell, Stewart, & Skov, 1996; Brown, Huey, & Findlay, 1997; VanRullen, Reddy, & Koch, 2004). This suggests that face processing requires some form of attention. However, this result is still controversial (Hansen & Hansen, 1988; Suzuki & Cavanagh, 1995; Hochstein & Ahissar, 2002), and it is hoped that the present experiments will contribute to resolving this debate.

2.2 Methods

2.2.1 Participants

Six participants, including the author, were tested in Experiments 1, 2, and 3. Six additional participants were tested on Experiment 4 while another six were tested on Experiment 5. All participants (age range from 22 to 31 years) were undergraduate and graduate students or staff at the California Institute of Technology, who signed a consent form and were paid $13.50 per hour. By self-report they had normal or corrected-to-normal acuity.

2.2.2 Face database

The face stimuli used were obtained from the Max Planck Institute, Tübingen, Germany and contained 7 views of 100 male and 100 female faces (Vetter, 1998) (http://faces.kyb.tuebingen.mpg.de). This database of color photographs is well-matched for low-level features such as color, size, and illumination. All obvious gender cues, such as facial hair, have been removed from these images. Pilot experiments showed that the gender of some faces in the database was ambiguous with overall discrimination performance around 65%. Therefore, eight additional participants were asked to judge the gender of each face and rate their confidence on a 3-point Likert scale (this experiment was conducted by Patrick Wilken). The mean of these responses was converted into Z-scores. Each face was randomly presented 10 times for 1000 ms. The present gender-discrimination experiment used the 50 male and 50 female individuals that produced the highest mean male and female ratings. Examples of the faces used are shown in Figure 2.2b.

2.2.3 Apparatus

Participants were seated approximately 120 cm from a computer monitor connected to a Silicon Graphics (O2) computer for the dual-task experiments. The refresh rate of the monitor was 75 Hz. The face rating (described above) and face recognition experiments (Experiment 3) were performed on a Macintosh G4 computer; the refresh rate of the monitor was 75 Hz. The display was synchronized with the vertical retrace of the monitor.

2.2.4 Experiment 1: face-gender discrimination

The experiment consisted of two distinct tasks: an attentionally demanding, central letter discrimination task, and a peripheral, face-gender discrimination task. These tasks were performed in three conditions: blocks of the central or peripheral task alone, or a dual-task condition in which both central and peripheral tasks were performed concurrently. Subjects were instructed to be as accurate as possible, and no constraint was imposed on their reaction times. An auditory tone was provided as feedback on incorrect trials. The experimental timeline for one trial is illustrated in Figure 2.2a. In all three conditions the trials were arranged as shown in the figure and only the specific instructions to participants differed.

Central letter discrimination task

The attentionally demanding central task consisted of letter discrimination. Each trial started with a fixation cross presented 300 ± 100 ms before the onset of the first stimulus. At 0 ms, five randomly rotated letters (Ts and Ls, either all the same or one different from the other four) were presented at the center of the display at nine possible locations within 1.2° of fixation. Participants were required to report whether all five letters were identical or not by pressing one of two keys on the keyboard. The letters were individually masked by an "F," rotated by an angle corresponding to the "T" or "L" it replaced. The SOA (“Stimulus Onset Asynchrony”) was determined individually for each participant (see “Training” below). This task has been shown to be effective in engaging spatial attention at the center of the display (Braun & Julesz, 1998; Lee, Koch et al., 1999). In particular, Braun has shown that reducing the SOAs for each subject by as little as 30 ms can result in substantial drops in performance on this task (Braun & Julesz, 1998); see also (Li, VanRullen et al., 2002). Thus, the SOAs obtained for each subject by using the training procedure described below ensure that subjects do not have time to shift attention to and from the letters during the task, without impairment in performance.

Peripheral face-gender discrimination task

A face subtending approximately 2.5° of visual angle was presented peripherally 26 ms following the onset of the central stimulus. The face appeared at a random location centered on an edge of an imaginary rectangle subtending 8° x 10° of visual angle. Each face was masked by a pattern mask composed of scrambled faces; the peripheral stimulus was always masked before the central stimulus. The peripheral SOA was determined individually for each participant (see “Training” below). Participants were required to report the gender of the face using two keys on the keyboard.

Dual-task

In the dual-task condition, participants were asked to respond to first the central task (with the left hand) and then the peripheral one (with the right hand), and fixate at the center.

Training

At the beginning of training, the letters were displayed for 500 ms and the faces for 160 ms before the mask appeared (Figure 2.2, (see also Movie 1))[5]. Through the course of training, “letter” and “face” SOAs were decreased (when mean performance in a 48-trial block exceeded 90%). In order to limit the possibility of eye movements, the letter SOA was decreased to below 250 ms for all subjects. Thus, training was complete when participants’ letter SOA had stabilized below 250 ms for a one-hour session. After training, over the group of participants, the “face” SOA varied between 133 ms-160 ms and the “letter” SOA between 173 ms–240 ms. This procedure, coupled with the high motor demands of the dual-task paradigm, meant that participants required extensive training (between 6 and 12 hours per participant). For three of the six participants, one

---

[5] See http://journalofvision.org/4/2/4/article.aspx for the movie

set of 350 randomly selected faces (7 views of 50 individuals) was used as stimuli, while the other three participants were trained on a different set of 350 faces. Participants received the same amount of training in all tasks.

Data collection

Once training was complete, the letter and face SOAs were fixed for each subject and data was collected over five one-hour sessions. Each session consisted of four blocks of 48 trials in each single-task condition and six blocks of 48 trials of the dual-task condition. A session was considered valid if dual-task letter performance was not significantly lower (t test, $p > .05$) than single-task letter performance. This served to ensure that participants were effectively focusing attention on the central letter task. Over the six participants, only two sessions were rejected as a result of this criterion.

2.2.5 Experiment 2

In a separate dual-task session, all six participants from Experiment 1 were asked to perform gender discrimination on a set of novel stimuli (7 views of the 50 individuals they had not seen in Experiment 1) using the same method as previously used, but with no further training. Participants performed only one session of this type.

2.2.6 Experiment 3

Experiment 3 was performed on the same day as Experiment 2, with participants (except the author) from Experiments 1 and 2. In two separate sessions of this experiment, participants were presented with the faces they had viewed during Experiment 1 (the “familiar” images) and Experiment 2 (“control” faces), respectively, along with an equal number of faces they had never seen before. Each face was shown centrally for 1000 ms. Participants reported whether they recognized the face or not using two keys on the keyboard. The first session was run before, and the second after, Experiment 2.

2.2.7 Experiment 4

In a separate set of experiments, six participants who had been trained on a different dual-task experiment (Li, VanRullen et al.) were tested in our paradigm for one hour per day for two consecutive days. These participants had been trained on the same central letter discrimination task, but a different peripheral task (animal vs. non-animal and vehicle vs. non-vehicle discrimination). In the paradigm they had been trained on, these participants responded to the peripheral stimulus by releasing the mouse button. Thus, instead of reporting whether the face presented was male or female with different keys on the keyboard, three of these participants were asked to release the mouse button if the face was male while the other three released the button if the face was female. During the experiment, the participants viewed a different image set each day.

2.2.8 Experiment 5

In Experiment 5, six new participants were tested. They were trained on three different peripheral tasks: upright face-gender discrimination (i.e., the same task as in Experiments 1 and 2), inverted face-gender discrimination (i.e., where each face was rotated by 180°), and a discrimination between two color patterns—a vertically bisected disk with red and green halves or such a disk rotated by 180° (i.e., they had to discriminate a red-green disk from its mirror image, a green-red disk). In individual dual-task blocks, participants performed the central letter discrimination task together with one of the three peripheral tasks. Each session consisted of four blocks of the single central-letter task, two blocks of each single peripheral-task, and three blocks of each dual-task. The faces were masked by a pattern mask composed of scrambled faces (as before) while the disks were masked by a disk divided into four red and green alternating quadrants. The tasks were matched for difficulty such that single-task performance for all three peripheral tasks was on average 75%. Participants received an equal amount of training on the three peripheral tasks. The same face set and training and data collection methods were used as in Experiment 1. The SOA for the disks was 98.1 ± 4.1 ms.

2.2.9 Data analysis

A 1-way ANOVA and paired t-tests were computed for each experiment to compare single and dual-task performance. An alpha value of .05 was used for all statistical tests. Normalized performances in the dual-task experiment were calculated by a simple linear scaling of the mean value of each participant's performance. The scaling mapped the mean single-task performance to 100% leaving chance at 50%: Normalized performance = 1/2 + 1/2. $[(P_2 – 1/2) / (P_1 – 1/2)]$, where P_2 and P_1 refer to performance in the dual-task and single-task conditions, respectively.

2.2.10 Eye movement control

In all our dual-task sessions, the presentation times of the central and peripheral stimuli were less that 250 ms, and the peripheral stimuli were always presented at random locations. These constraints were imposed in order to limit the possibility of eye movements, and it is indeed unlikely that eye movements contributed significantly to the observed behavioral performance of our subjects. However, to directly confirm this, in a separate control experiment 3 subjects performed the face-gender dual task experiment while their eye movements were monitored. They were seated 75 cm from a computer monitor on which the stimuli were presented. The monitor measured 26º x 19º of visual angle. Subjects rested their chins on a chinrest in order to minimize head movements. An infrared (IR) eye-tracking system obtained from ISCAN, Inc. was used to measure and record eye positions. The camera and IR beam generator were placed approximately 60 cm away from the subject, and the right eye was illuminated with invisible IR light (~850 nm). The camera sampled the image of the eye at 120 Hz, and the image was processed in real time to obtain the position of two features of the eye: the IR-dark spot at the center of the pupil and the IR-bright spot on the cornea (the corneal reflection). The difference of these two values, in principle, gives a measure of the position of the eye independent of head position.

During calibration trials that were run prior to the start of the experiment, and after every two blocks, subjects fixated 9 calibration points for 2 s each. The calibration points were presented on a 3 x 3 grid on the monitor. These calibration trials gave us the value of the eye position in camera coordinates. Using these values and the position of the fixation points in the stimulus display, a correspondence between camera coordinates and the stimulus display coordinates could be established. This correspondence was then used to ascertain eye positions during the dual-task experiment in stimulus display coordinates. The mean error of the scanner over all subjects during the calibration runs was estimated to be about 0.40 º.

In the dual-task experiment, the eye positions were measured during each trial (from the appearance of the first fixation cross up until when the letter masks appeared) at 120 Hz. The median and mean of these values over all trials in each of the three experimental conditions were computed. This data is discussed in the results section below and is presented in Figure 2.7.

2.3 Results

The effects of attentional manipulation on face-gender discrimination were studied with a dual-task paradigm in which participants performed a central, attentionally demanding, letter discrimination task as well as a second, peripheral face-gender discrimination task either concurrently or separately. The role of attention on gender discrimination was measured by comparing performance on the peripheral task, when this task was performed alone (single-task condition), with performance under dual-task conditions. If gender discrimination requires little or no attentional resources, peripheral performance will suffer minimally in the dual-task condition compared to the single-task condition. If, however, the peripheral task does require attention, performance should be severely impaired under the dual-task condition (Sperling & Melchner, 1978; Braun & Sagi, 1990; Braun & Julesz, 1998).

In Experiment 1, six participants were tested on this paradigm (Figure 2.3). Their performance on the central letter discrimination task when performed alone was on average 83.1% ± 4.1% (mean ± s.d.). This value can be compared with performance on this task in the dual-task condition (83.4%± 5.6%): if a participant's attention is engaged by the central letter task, performance in the dual-task condition should be equivalent to performance in the single-task condition; otherwise there should be a significant decrease in performance levels. For our participants, there was no significant difference in performance on this task between the single and dual-task conditions (*t*-test, $p > .05$). When participants performed the face-gender discrimination task alone, performance was on average 77.6% ± 3.8%. This comparatively lower value reflects the short stimulus exposure and the fact that obvious gender cues, such as the presence of facial hair, were removed from the images. Performance on this task in the dual-task condition (74.9% ± 4.0%) was also not significantly different (*F* (1, 10) =1.52, *p* = .2) from performance in the single-task condition over the group of six participants (Figure 2.3a). For five of the six participants, individual *t*-tests revealed no significant difference in performance (*p*>.05) between these two conditions. Figure 2.3b summarizes these results: in the face-gender discrimination task, performance for all six participants in the dual-task condition was above 90% of their performance in the single-task condition (normalized plot, see Methods). These results indicate that, although there may be a small decrement in the dual-task condition, face-gender discrimination can still be performed efficiently with little or no spatial attention resources available and constitutes the main finding of this study.

In order to limit the possibility of eye movements, the central SOA was maintained below 250 ms for all participants, and the peripheral stimulus could appear anywhere at one of eight peripheral locations. This constraint, together with the high motor demands of the dual-task procedure, meant that participants required extensive training (between 6 and 12 hours per participant) with the same set of male and female images (referenced hereafter as the “familiar” face set). Consequently, it could be argued that instead of performing gender discrimination as required, participants were actually using a strategy akin to face recognition. To control for this potentially confounding effect, the same participants were tested on a set of novel faces (“control” faces) in Experiment 2 (Figure 2.4a). Despite the novelty of the control face set, over the group of participants, the difference in performance on gender discrimination between single and dual-task conditions was not significant (*F*(1,10)=1.43, *p*=.3). Individually, for five of the six participants, performance was not significantly different between these two conditions (79.1% ± 4.8% and 75.6% ± 5.1%, respectively; paired *t*-test, *p* > .05).

Although there was a modest decrement in the dual-task condition, face-gender discrimination performance for all six participants was above 85% of their original single-task performance (normalized plot, Figure 2.4a). Note that the central task performance in the dual-task condition was not significantly lower than performance in the single-task condition for each participant (*t*-test, *p*>.05), indicating again that attention was effectively engaged at the center in the dual-task condition. From this control experiment, it appears that familiarity with the face set is not critical to the observed performance.

In fact, results from an additional control experiment (Experiment 3), indicate that participants had not gained any appreciable familiarity with either of the face sets they had viewed during Experiments 1 or 2. In separate sessions, participants were presented with the faces viewed extensively during the training and data collection phases in Experiment 1 (“familiar” faces) or faces viewed in Experiment 2 (“control” faces), as well as an equal number of completely novel faces. (The “familiar” faces had been viewed between 18 and 30 times, while the “control” faces had each been viewed twice during the course of experiments 1 and 2, respectively. Each presentation of the face had lasted between 143 ms and 160 ms, depending on the participants’ SOA; see Methods). Participants were asked to report for each face whether they had seen it at least once during Experiment 1 or 2. Surprisingly, for both the “familiar” and the “control” sets of images, participants’ performance on this recognition task (52.1 ± 3.4% for the familiar face set and 51.1 ± 2.2 % for the control face set, Figure 2.5) was not significantly different from chance levels (*p* = .2, *p* = .4, respectively, paired *t*-test). Thus, it appears that despite having viewed some of the faces several dozen times, participants were unable to differentiate the stimuli in either face set. These results confirm that the pattern of performance observed in Experiments 1 and 2 cannot be accounted for by familiarity with the stimuli used.

Since participants had been extensively trained on the face-gender discrimination task, it could still be argued that they had learned low-level features in the image set, which would contribute significantly to the observed performance. To control for this, six new participants were tested on our gender-discrimination task (Experiment 4). They had been trained on a completely different dual-task experiment (natural scene categorization: animal vs. non-animal or vehicle vs. non-vehicle) (Li, VanRullen et al., 2002). Data was collected over two days with a new set of stimuli on each day. Despite the novelty of the peripheral task, participants performed comparably well in the dual-task and single-task conditions (Figure 2.4b). While performance on the gender-discrimination task was significantly lower (*F* (1, 10) =5.4, *p*=.04) in the dual-task (69.7 ± 5.6%) vs. single-task (75.91 ± 6.2%) condition over the group of participants, there was, individually, no significant difference in performance for four of the six participants (p>.05, paired t-tests). The normalized results shown in Figure 2.4b indicate that despite the novelty of the task, performance in the dual-task condition was above 80% of performance on the single-task condition for all six participants. We conclude therefore that there was no strong or consistent confounding effect of training in our gender discrimination task.

Thus, whether involving highly familiar or completely novel faces, or even a completely novel discrimination task, there is only a modest decrement in performance on face-gender discrimination in the near-absence of focal attention.

Finally, in order to rule out the possibility that low-level cues in the face dataset could account for the observed results, we tested six additional participants in Experiment 5. In this experiment, participants were required to perform face-gender discrimination on both upright and inverted faces, using the same method as Experiment 1. Inverted faces provide a suitable control for basic low-level characteristics (e.g., contrast, luminance, spatial frequency, etc.), which might aid gender discrimination. If the observed results were due to low-level statistical properties of male and female faces, equally high levels of performance would be observed in both the upright and inverted face-gender discrimination tasks.

Participants received the same amount of training in both the upright and inverted face-gender discrimination tasks, and the level of difficulty was matched so that the mean single-task performance was about 75% for both tasks. Consistent with the results of Experiments 1 and 2, participants again achieved a high level of performance on upright face-gender discrimination in the dual-task condition compared to the single-task condition (Figure 2.6a; 71.3% ± 3.4%, 75.5% ± 4.0%, respectively). Over the group of six participants, a 1-way ANOVA revealed no significant difference in performance in the dual and single-task conditions (F (1, 10) =3.62, p = .09). Individually, there was no significant difference between these two conditions for four of the six participants (t-test, p>.05), and all six participants performed above 85% of their original single-task performance. In contrast, based on a 1-way ANOVA, the six participants showed a significant decrease in performance (F(1,10) = 25.7, p < .001) in the inverted face-gender discrimination task when attention was unavailable compared to the single-task condition (59.7% ± 4.7%, 71.7% ± 3.3%, respectively), and individual tests for each participant revealed a significant decrease in performance in this dual-task condition for all six participants (t-test, p<.05; Figure 2.6b). Further, for each of the six participants, performance in the inverted dual-task condition was significantly lower than performance in the upright dual-task condition (p <.05, t-test). We conclude that the observed performance in upright face-gender discrimination cannot be accounted for by the low-level statistical properties of the stimulus set.

The interpretation of the results reported here relies on the assumption that the central letter task efficiently engages attention in the dual-task condition and that performance on attentionally-demanding tasks should suffer dramatically in the dual-task condition. As a further control, we verified that performance on a known attentionally-demanding task would indeed be severely impaired in the dual-task condition (Braun & Julesz, 1998; Li, VanRullen et al., 2002). We had the same six participants discriminate between a masked color disk and its mirror image in the dual-task condition. In our experiment, participants received the same amount of training in all three discrimination tasks (upright face-gender, inverted face-gender, and colored-disk discrimination), and task difficulty was matched so that single-task performance was about 75% for all three tasks. In contrast to the results observed for upright face-gender discrimination and consistent with previous studies (Braun & Julesz, 1998; Lee, Koch et al., 1999; Li, VanRullen et al., 2002), we observed (Figure 2.6c) a dramatic decrease in performance over the group of six participants when the colored-disk discrimination task was performed in the dual-task versus single-task condition (51.8% ± 3.4%, 76.1% ± 8.5% respectively, $F(1,10) = 42.0$, $p < 10^{-4}$). As shown in Figure 2.6c, normalized performance values were between 45% and 60% of single-task levels, and for five of the six participants these values were not significantly different from chance levels of performance (paired *t*-test, $p > 0.1$). These results confirm that under our experimental conditions, the attentional requirements of the central task can result in a clear decrease in dual-task performance.

To determine whether eye movements contributed significantly to the observed performance, three subjects performed the dual-task experiment while their eye positions were monitored. For each of the three experimental conditions (central, peripheral, and dual), the median and maximum eye positions were calculated on each trial for each subject. The data for one subject is presented in Figure 2.7a. In this figure (drawn to scale), the outermost gray rectangle depicts the extent of the computer monitor on which the stimuli were presented. The yellow hashed area represents the region occupied by the face stimuli. It is important to note that no part of the face stimuli extended beyond this yellow area. The innermost rectangle, delineated by a solid black line, represents the 1.2º window in which the letter stimuli were presented. The red crosses mark the median and maximum eye positions on individual trials in each of the three conditions. As the figure illustrates, eye positions in each of the three conditions were limited to the area occupied by the letters and never entered the area occupied by the face stimuli. Thus, this subject did not make large eye movements in the direction of the peripheral stimuli in any of the three experimental conditions. Comparable results were obtained for other subjects in this experiment.

Of course, this eye movement control is only meaningful if subjects are performing the task with comparable performance levels as those obtained in our main experiment. The behavioral data for the three subjects tested with the eye-tracker are shown in Figure 2.7b. The first panel presents the normalized performance on face-gender discrimination over all trials in the experiment. In the second and third panels, we have imposed an upper bound on the permissible eye movements. Thus in the second panel, all trials on which the maximum eye position during the trial exceeded 2º were rejected. Behavioral performance over the remaining trials was calculated. In the third panel, all trials on which the median eye position exceeded 1º were rejected. After imposing these limits on eye positions, the behavioral performance observed on the face-gender discrimination task (the second and third panels) is not significantly different from performance shown in the first panel. Thus, this data confirms that eye movements do not make a significant contribution to the behavioral performance of subjects on this task.

2.4 Discussion

Our findings demonstrate that telling a very briefly flashed male from female face, a fine discrimination task, can be performed remarkably well when spatial attention is engaged elsewhere. We have shown that participants can achieve a high level of performance in the presence of little or no focal attention when they are tested on a set of completely unfamiliar faces, even when they are unfamiliar with the task itself. In control experiments, we have shown that these results cannot be explained by eye movements. Further, when participants perform the same face-gender discrimination task in the near-absence of spatial attention on a set of inverted faces, performance is significantly impaired compared to performance on this task with upright faces. These results demonstrate that the observed findings cannot be attributed to low-level characteristics of the image set. Previous psychophysical studies have shown that face recognition is impaired when the faces are inverted rather than upright (Yin, 1969; Valentine, 1988; Valentine & Bruce, 1988; Brown, Huey et al., 1997). Additionally, even though functional imaging studies have suggested that inverted face processing recruits additional brain areas compared to upright face processing (Haxby, Ungerleider et al., 1999), electrophysiology in monkeys has revealed that face-specific cells maintain responses to inverted faces, but that these responses are weaker and longer in latency compared to those evoked by upright faces (Perrett, Mistlin et al., 1988). Our results suggest a differential requirement of spatial attention by these two tasks: the absence of spatial attention has a pronounced effect on the processing of inverted faces, but not upright faces.

It should be noted that while 19 of the 24 datasets we obtained did not demonstrate any significant decrease in performance in the dual-task conditions, the remaining five did show some decrement. However, some decrement in performance is expected to occur when participants perform two demanding tasks concurrently, compared to when the tasks are performed alone. These performance decrements do not necessarily imply competition for an attentional resource, but could be attributed to other factors such as having to maintain two sets of task goals, or having to encode and produce two sets of motor responses (Allport, 1980; Duncan, 1980a; Pashler, 1984, 1994). In addition to comparing single and dual-task performance, it is revealing to compare the dual-task performance of our participants in the face-gender discrimination task with dual-task performance on tasks that are known to require attention (Braun & Julesz, 1998; Li, VanRullen et al., 2002). As we have shown, performance on a known attentionally demanding task (discriminating a red-green from a green-red disk) drops to chance levels when the available spatial attention is severely reduced. In contrast, performance on our face-gender discrimination task remains consistently above 80% of single-task performance when attention is engaged elsewhere. Indeed, a statistical comparison of all 24 datasets we collected indicates that all our participants perform face-gender discrimination in the dual-task condition significantly above chance (*t*-test, $p < 10^{-16}$).

From a computational perspective, we designed our peripheral task to be challenging: this task did not merely involve the discrimination of targets and distractors at a basic level of categorization, but required a fine discrimination within a category level, between male and female faces that share the same overall structure and lack hair and other obvious gender cues. In essence, this meant a fine discrimination of the spatial arrangement of highly similar features present in both targets and distractors. Our results indicate that such discrimination can be carried out in the presence of a primary task highly effective in requiring attentional resources (Braun & Julesz, 1998; Lee, Koch et al., 1999; Li, VanRullen et al., 2002). This supports the notion that the "complexity" of a task as measured by its computational demands does not necessarily determine its attentional requirements. Classical views of selective, visual attention have suggested that while simple salient stimuli can be detected outside the focus of attention, attention plays a key role in the recognition of more complex stimuli. In other words, it has been proposed that attention is necessary to combine the different low-level features of a stimulus into a coherent representation of the object (Treisman & Gelade, 1980). Access to this representation is supposed to be necessary for object recognition and behavior. Our findings argue that face-gender discrimination is possible in the near-absence of attention. Although this conclusion cannot be generally extended to other subordinate—level categorization tasks involving natural stimuli, our approach shows that attention is not always necessary for such tasks. The possibility that faces hold a special status for the visual system is still under debate (Farah, 1995; Gauthier & Tarr, 1997; Kanwisher, McDermott et al., 1997; Farah, Wilson et al., 1998; Tovee, 1998; Gauthier, Skudlarski, Gore, & Anderson, 2000; Ro, Russell et al., 2001; Bogen & Berker, 2002). It would thus be interesting to test the role of attention in other “complex” discrimination tasks and to determine whether expertise in other areas yields similar results.

If a failure to pop-out during a search task is taken to indicate the necessity of focal attention for recognition, then our results appear to contradict a number of studies that have shown that facial information does not "pop-out" in a visual search situation (Nothdurft, 1993; Kuehn & Jolicoeur, 1994; Purcell, Stewart et al., 1996; Brown, Huey et al., 1997). However, it is worth noting that earlier studies had suggested that faces can be processed in parallel (Hansen & Hansen, 1988), and this issue is still controversial and open to debate (Hochstein & Ahissar, 2002). Furthermore, the correspondence between dual-task and visual search results has recently been called into question (VanRullen, Reddy et al., 2004). More supportive evidence for the pre-attentive processing of faces comes from clinical reports of patients with visual neglect (Vuilleumier, 2000; Vuilleumier, Sagiv et al., 2001). For these patients, extinction was less likely to occur for faces presented in the neglected hemifield than other objects (e.g., meaningless shapes). In other words, faces can attract attention more efficiently and thus probably have a competitive advantage at the pre-attentive level. Such observations are compatible with ERP and magneto-encephalography (MEG) investigations of the latency of face or face-gender selective responses, which was found to be on the order of 100 ms–150 ms (Schendan, Ganis, & Kutas, 1998; Yamamoto & Kashikura, 1999; Mouchetant-Rostaing, Giard, Bentin, Aguera, & Pernier, 2000; Liu, Harris, & Kanwisher, 2002). Given this remarkable speed, one wonders whether such processing can depend critically on visual attention.

In neural terms, several electrophysiological investigations have found single neurons responsive to faces in the infero-temporal cortex of monkeys, the "end-point" of the ventral visual hierarchy (Gross, Rocha-Miranda, & Bender, 1972; Bruce, 1982; Perrett, Rolls, & Caan, 1982; Desimone, Albright, Gross, & Bruce, 1984; Perrett, Smith et al., 1984; Rolls, 1984; Perrett, 1987). Similar observations have been made in humans in medial temporal lobe structures ((Kreiman, Koch, & Fried, 2000a; Quian Quiroga, Reddy, Kreiman, Koch, & Fried, 2005); see also Chapter 5). Several neuro-imaging studies have shown the existence of higher-level brain regions (such as the Fusiform Face Area, FFA) that selectively process facial information (Sergent, Ohta, & MacDonald, 1992; Haxby, Horwitz et al., 1994; Kanwisher, McDermott et al., 1997; Kanwisher, Stanley, & Harris, 1999) although some models of face recognition have conjectured that gender discrimination could occur in more posterior temporal areas (Bruce & Young, 1986). Consequently, it is not unreasonable to suppose that our stimuli differentially activate neurons in such high-level areas and that gender discrimination can rely on the selectivity of these neurons. Some evidence shows that these areas can be modulated by attention (Wojciulik, Kanwisher, & Driver, 1998; O'Craven, Downing et al., 1999; Pessoa, McKenna, Gutierrez, & Ungerleider, 2002), but the present results indicate that the residual activity in the near absence of attention is sufficient for the efficient processing of faces. Our findings, together with those of Li *et al*. (Li, VanRullen et al., 2002) and Rousselet *et al*. (Rousselet, Fabre-Thorpe et al., 2002), suggest that the activation of such high-level neuronal populations can take place in the near-absence of attention.

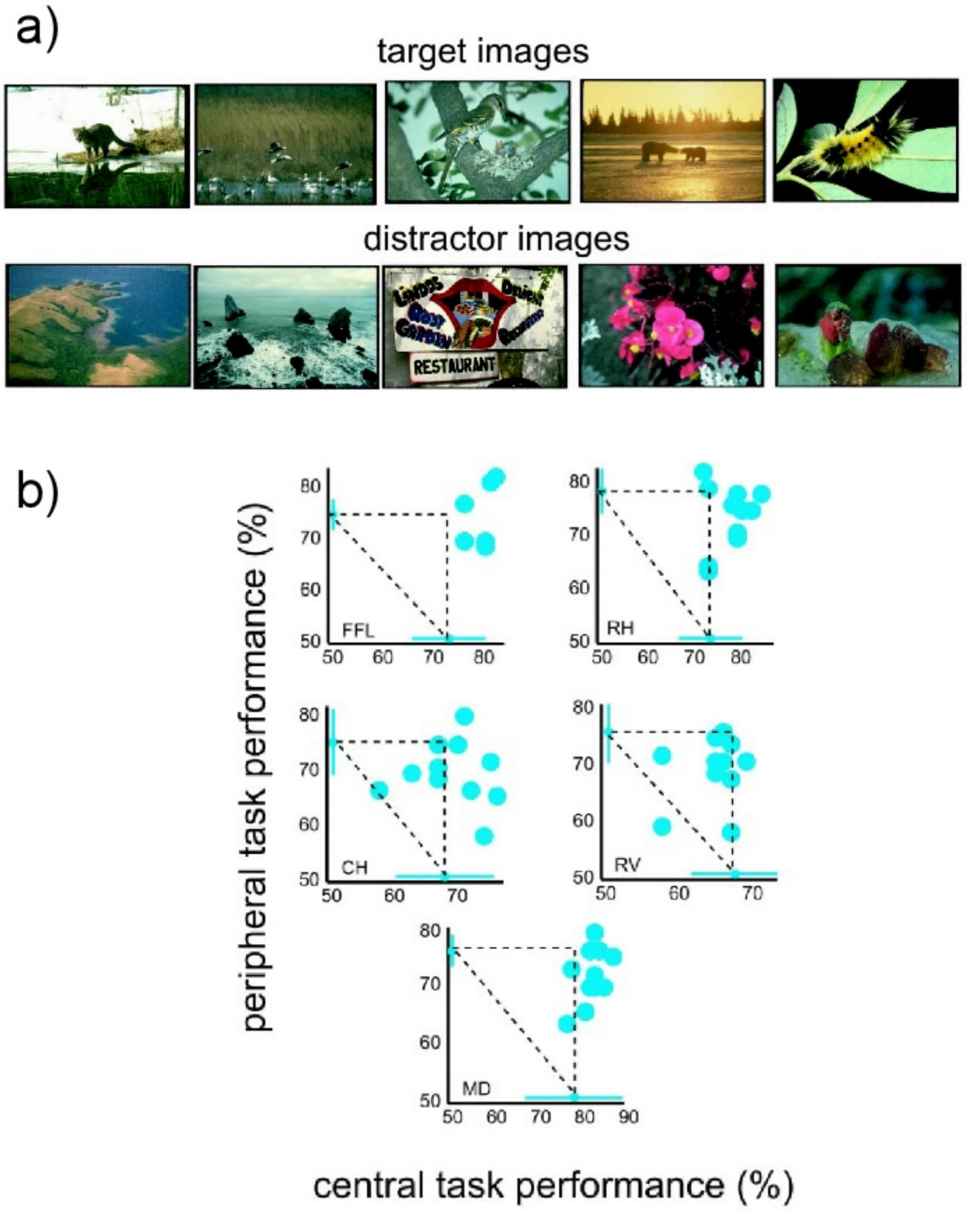


Figure 2.1: Natural scene categorization in the near-absence of attention (Adapted from Li et al., 2002). a) Examples of target (scenes containing animals) and distracter scenes (scenes with no animals) used in the experiment. b) Performance of 5 subjects on the natural scene categorization task. Performance on the peripheral task (natural scene categorization) is plotted against performance on the central letter discrimination task for each subject. The filled circles represent performance on individual blocks in the dual-task condition. Performance on the two tasks in the single task condition is plotted as a line along the respective axes, where the length of the line represents the standard deviation in performance. For each subject the difference in performance on both tasks was not significantly different ($p > .05$) between single and dual-task conditions indicating that the animal detection task is performed well in the near-absence of attention. Copyright 2002, National Academy of Sciences, USA.

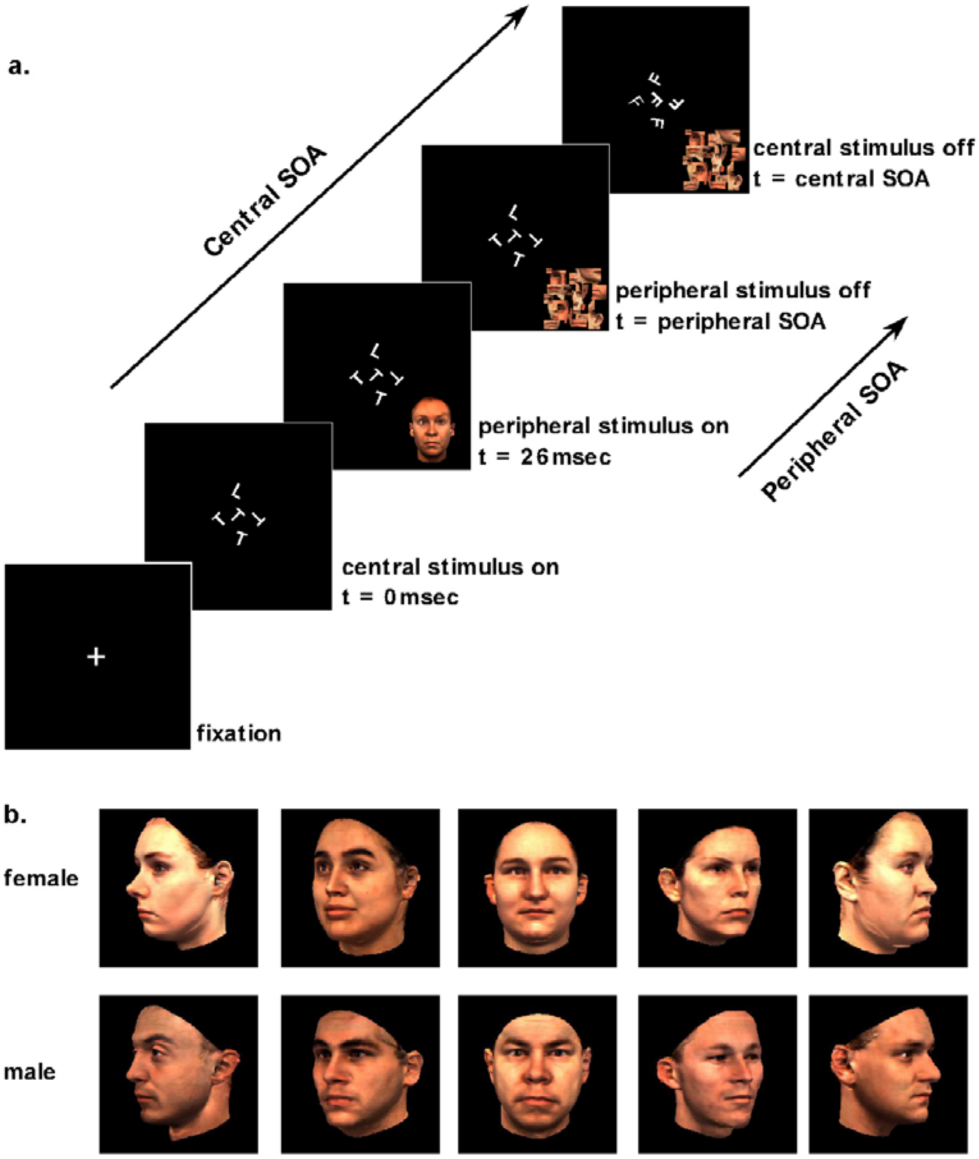
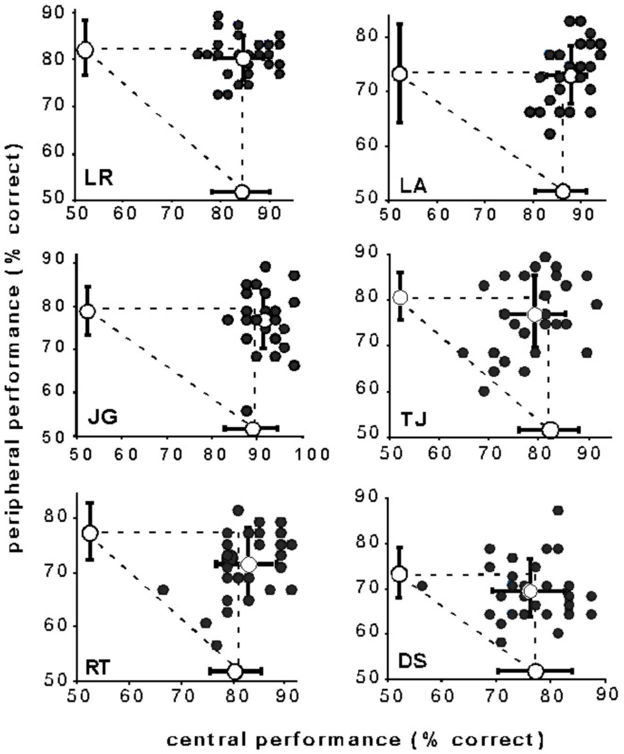


Figure 2.2: Face-gender discrimination dual-task experiment. (a) Schematic timeline for one trial in the dual-task experiment. At the end of a trial, participants are required to report the gender of the face presented and/or whether the 5 central letters were the same (either 5 Ts or 5 Ls) or different (4 Ts and 1 L or 4 Ls and 1 T). All trials are arranged similarly, independent of the specific instructions. Both letters and faces were masked individually. Central SOA (~200 ms) and peripheral SOA (~145 ms) indicate the presentation time for letters and faces, respectively. (b) Exemplars of male and female faces used in the experiment. (From (Reddy, Wilken, & Koch, 2004); copyright, Journal of Vision, 2004)).

a. Individual results



b. Normalized results

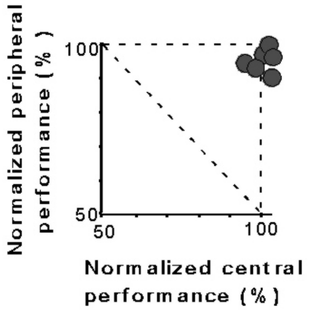


Figure 2.3. Results from six participants in the dual-task paradigm. *(a)* The horizontal axis represents performance on the attentionally demanding central letter task. The vertical axis represents performance on the peripheral gender discrimination task. Each filled circle is the participant's mean performance in the dual-task in one block of 48 trials, while an open circle represents mean performance in the three experimental conditions: single-central task, single-peripheral task, and the dual-task. By default, performance of the "to-be-ignored" task is assumed to be at chance level (50%) in the single-task condition. Error bars represent standard deviation. For all participants except one, (RT), face-gender discrimination performance in the dual-task condition is not significantly worse (*t*-test, *p*>0.05) than performance in the single-task condition indicating that face-gender discrimination suffers only minimally when performed concurrently with an attentionally-demanding task. *(b)* Normalized average performance for each participant in the dual-task paradigm. Each point represents a participant's performance in the dual-task normalized to their single-task performance. Normalized values are obtained by a linear scaling that maps the average single-task performance to 100% leaving chance at 50%(Methods). Normalized gender-discrimination performance values lie above 90% of single-task performance, suggesting that participants can perform face-gender discrimination remarkably well in the near-absence of attention. (From (Reddy, Wilken et al., 2004); copyright, Journal of Vision, 2004)).

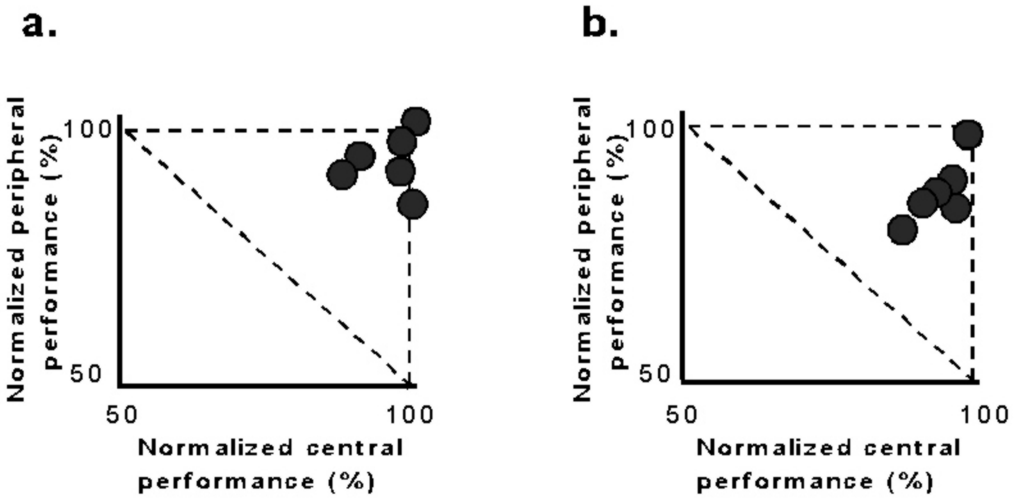


Figure 2.4. Control experiments for familiarity with images and training. *(a)* Normalized average performance for six participants in the dual-task paradigm using a completely novel set of faces. (Notation as in Figure 3b). Normalized dual-task performance lies above 85% of single-task performance for all participants, indicating that even with a novel set of faces, gender discrimination is performed well under the dual-task condition. *(b)* Normalized average performance for six participants who had been trained on a completely different dual-task paradigm. Normalized dual-task performance lies above 80% for all participants. This suggests that in spite of unfamiliarity with the gender discrimination task, performance was only marginally impaired in the dual-task condition. (From (Reddy, Wilken et al., 2004); copyright, Journal of Vision, 2004)).

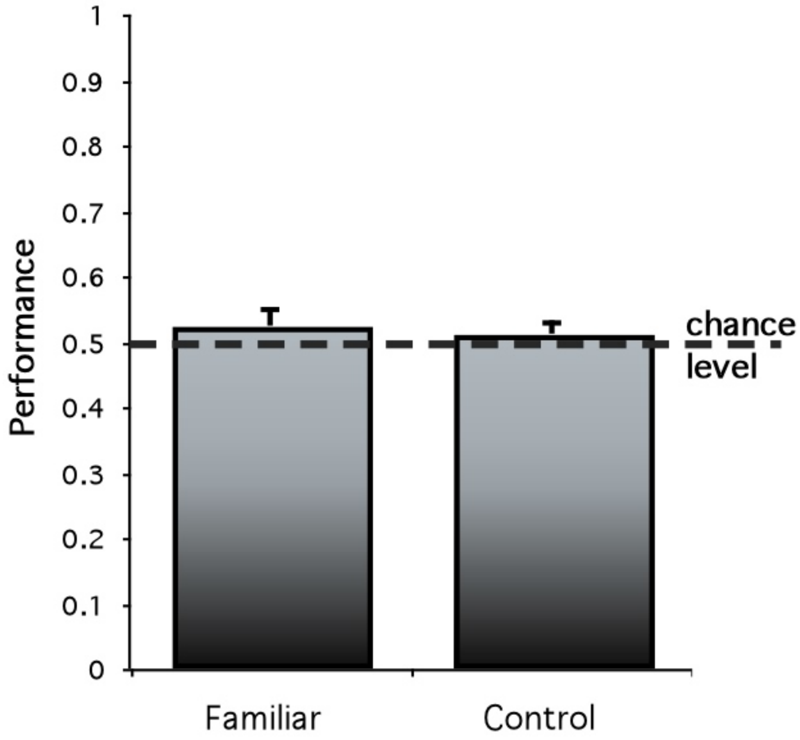


Figure 2.5. Face recognition experiment. Participants (the same as those shown in Figures 3 and 4a) were presented with faces they had viewed during the study and an equal number of novel faces and asked to report whether they recognized the face or not. The “familiar” image set is the one participants were trained on whereas the “control” faces had been viewed only twice each. In both cases, participants are at chance level at discriminating previously seen faces from novel faces, indicating that they had formed no explicit representation of the face sets. Error bars represent standard deviation. (From (Reddy, Wilken et al., 2004); copyright, Journal of Vision, 2004)).

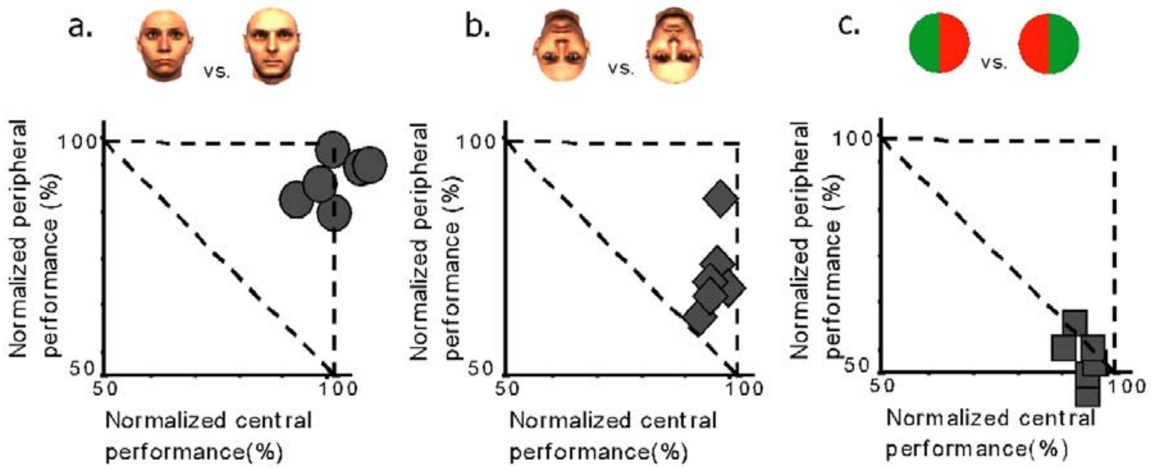


Figure 2.6. Control experiments for low-level visual cues and to test the efficacy of the central task in withdrawing attention. Normalized dual-task results of six participants in three tasks. *(a)* Upright face-gender discrimination task. Normalized dual-task performance values are on average 92% of single-task performance levels for upright face-gender discrimination, as expected from results shown in Figure 3b. *(b)* On the other hand, in the inverted face-gender discrimination task, normalized dual-task performance values are on average 72% of single-task performance levels, demonstrating that in the near-absence of attention, performance is impaired. In addition, for each participant, there is a significant decrease in performance when the task involves inverted faces compared to upright faces. Thus low-level visual cues cannot account for the pattern of results obtained in the upright face-gender discrimination task. *(c)* Color pattern discrimination task. Normalized dual-task values are on average 53%, demonstrating that attention is effectively withdrawn by the central letter task in dual-task conditions. (From (Reddy, Wilken et al., 2004); copyright, Journal of Vision, 2004)).

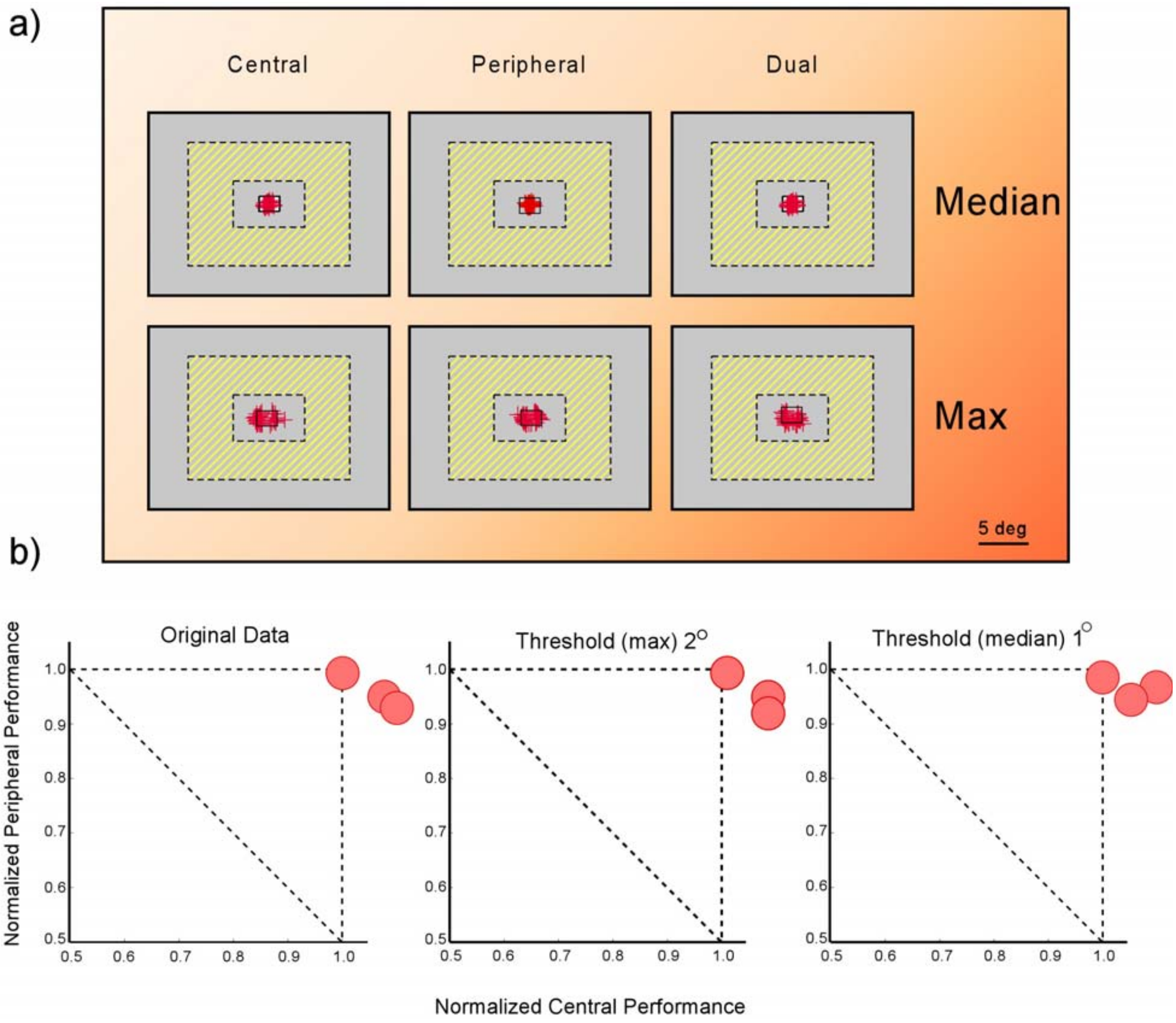


Figure 2.7 Eye movement control for 3 subjects. a) The median and maximum eye position during individual trials in all three conditions is presented for one subject. The figure is to scale. The grey area represents the extent of the computer monitor on which stimuli were displayed. The yellow-hashed area marks the inner and outer limits of the region occupied by the faces. Thus no part of the face stimuli was presented outside this area. The innermost rectangle, delineated by a solid black line, marks the extent of the letter stimuli (1.2°). Each red cross is the median and maximum eye position on individual trials. The data presented here shows that the subject's eye movements never entered the region occupied by the face stimuli but were instead limited to the area occupied by the letters. b) Behavioral performance for the three subjects. In the first panel, performance over all trials is shown. In the second panel, trials on which the maximum eye position was larger than 2° were rejected, and performance over the remaining trials was calculated. In the third panel, trials on which the median eye position was larger than 1° were rejected. The behavioral performance obtained after imposing these restrictions is not significantly different from performance shown in panel 1. Thus, the data indicate that eye movements do not contribute significantly to the high levels of behavioral performance we have observed.

Chapter 3

Face Identification in the Near-Absence of Attention

3.1 Introduction

The processing of naturalistic stimuli has recently come under a fair amount of attention (Li, VanRullen et al., 2002; Rousselet, Fabre-Thorpe et al., 2002; Braun, 2003; Kayser, Kording, & Konig, 2004). In particular, Li and colleagues showed that the visual system can categorize natural scenes more efficiently than artificial geometric shapes (e.g. a red-green bisected disk from its mirror image) (Li, VanRullen et al., 2002). They therefore concluded that the attentional demands of a task are not determined by the complexity of the stimuli used, but by their nature—natural scenes versus artificial stimuli.

As we discussed in Chapter 2, we previously asked what the limits of this type of pre-attentive processing of natural stimuli might be. We were interested in determining whether attention would become necessary if the natural targets and distractors were made more similar to each other, in a more constrained stimulus space. Indeed, in contrast to the bisected disk task, where the stimulus space was restricted and the targets and distractors differed from each other along well-defined feature dimensions, in the study by Li et al. the set of images were very diverse, and both the target and distracter ensembles probably populated a high dimensional space. Thus it remained possible that natural scene processing in the near-absence of attention would break down for more constrained stimulus spaces (and thus presumably more complex discriminations).

To address this issue, we had used a face-gender discrimination task (Reddy, Wilken et al., 2004). The feature dimensions involved in this discrimination are well characterized: eye brows, eyes, jaws, noses, and mouths, in order of decreasing relevance (Brown & Perrett, 1993; Bruce, Burton et al., 1993; Yamaguchi, Hirukawa, & Kanazawa, 1995). Additionally, male and female faces share many common features, in particular their global structure, which makes the input space more constrained than other natural scene categorization tasks. Remarkably, our results demonstrated that subjects could still identify the gender of a face even when spatial attention was not fully available. Furthermore, this effect was not due to low-level feature discrimination since the male and female faces in our database were well-matched for low-level features. Additionally, we also observed an inversion effect (Yin, 1969; Valentine, 1988; Valentine & Bruce, 1988; Brown, Huey et al., 1997), which confirms that the result depends on holistic properties of faces.

Unfortunately, these results also constituted a failure in the sense that we did not succeed in finding the limit of pre-attentive natural scene processing. Instead it appears that for such pre-attentive processing to break down, even more subtle discriminations will be necessary.

In the present study we thus probe the attentional requirements of face identification. Several lines of evidence indicate that identifying a particular face is a finer and more complex discrimination task than telling male and female faces apart. For instance, current models of face recognition posit that, although gender discrimination and face recognition could proceed in distinct modules (Bruce & Young, 1986) (see however (Haxby, Hoffman, & Gobbini, 2000)), gender discrimination occurs prior to face recognition (Ellis, 1986). Psychophysical data lends credence to this hypothesis since it has been demonstrated that gender categorization is performed faster than face recognition (Sergent, 1986; Bruyer, Galvez, & Prairial, 1993). Accordingly, it has been shown that the two systems interact with each other (Rossion, 2002) and that determining the gender of a face can influence its subsequent recognition (Baudouin & Tiberghien, 2002). Additionally, face identification generally exploits higher spatial frequency information than gender discrimination. Consequently, while global descriptors appear to be sufficient for gender processing, face identification might instead be based on the finer details of a face (Sergent, 1986). Indeed, the relevant features for discriminating two female faces (for example) are necessarily more specific than those needed to categorize female and male faces (e.g. (Campbell, Benson, Wallace, Doesbergh, & Coleman, 1999).

Thus while face gender discrimination failed to achieve sufficient complexity, face identification might constitute a better candidate to probe the limits of pre-attentive processing of natural stimuli.

3.2 Methods

3.2.1 Participants

Five subjects (including the author) were tested in Experiment 1. Four of these subjects and two new subjects were tested in Experiment 2. Four subjects from Experiment 2 were tested in Experiments 3–5. All participants were undergraduate and graduate students at the California Institute of Technology. They were paid $13.50 per hour for participation in the experiment and gave informed consent. All subjects reported that they had normal or corrected-to-normal visual acuity. For the experiments, subjects were seated approximately 120 cm in front of a Macintosh G4 computer.

3.2.2 Face database

The face database used in Experiment 1 was provided by Lavanya Reddy and consisted of pictures of male and female Hollywood celebrities obtained from the web. These were usually high resolution, color shots of the faces of actors and actresses in “natural” settings. Six female and 5 male celebrities were the target individuals, and there were 24 views of each of these targets. Several views of 43 different celebrities (150 images in total), obtained under comparable conditions, were the distracter images.

For Experiments 2–4, the face database was obtained from the Max-Planck Institute in Tübingen, Germany, (http://faces.kyb.tuebingen.mpg.de) and contained seven color views of 100 male and 100 female individuals, unknown to our subjects. For these experiments five viewpoints were used (frontal view, and left and right profiles at 30˚ and 45˚). For all experiments, individuals were chosen randomly to be the targets and distractors.

3.2.3 Training

All of our subjects had previously been trained in the dual-task paradigm prior to participating in this set of experiments. However, none of them had been trained on the face-identification task. Instead, three of our subjects had been trained on the animal versus non-animal discrimination task (Li, VanRullen et al., 2002), while the rest had been trained on the face-gender discrimination task described in Chapter 2 (Reddy, Wilken et al., 2004). Details of the training procedures used in these two paradigms are similar and are described in Chapter 2 and the corresponding papers.

3.2.4 Experiment 1: face identification with famous faces

The dual-task paradigm was used to test the effects of attentional manipulation on face identification performance. The experiment consisted of two separate tasks: an attentionally demanding central letter discrimination task and a peripheral face identification task. Subjects performed these tasks in three conditions: blocks of the letter discrimination task or face identification tasks alone, or blocks of both tasks together in the dual-task condition. Each block consisted of 48 trials, with 24 target trials, and 24 distracter trials. An auditory tone was provided as feedback on incorrect trials. A typical trial is shown in Figure 3.1.

Central letter discrimination task

The letter discrimination task has been described in detail in Chapter 2. The average presentation time for the letter stimuli for this experiment was 197.0 ± 14.3 ms.

Peripheral face identification task

A face subtending approximately 2.5˚ of visual angle was presented at a random location on the edge of an imaginary rectangle subtending 8˚ x 10˚ of visual angle. The faces were backward-masked by a pattern mask composed of scrambled faces. The face-mask always appeared before the letter-masks. The average presentation time of the faces was 167.8 ± 14.4 ms.

In Experiment 1, the faces presented to subjects were of Hollywood celebrities. Before each block of this task, subjects were given the name of one of the set of target individuals. All 11 target celebrities were known to all subjects. On 24 of the 48 trials in the block, different images of this target celebrity were presented to the subject, while on the other 24 trials other celebrities of the same gender as the target were presented as distractors. The order of the trials was randomized. Subjects reported if the face was the target face or not by pressing two keys on the keyboard.

Dual-task condition

In the dual-task condition, subjects had to perform both the central letter discrimination task and the peripheral face identification task together while fixating at the center. In this experiment, subjects performed at least 7 blocks of the dual-task condition, and 3 blocks each of the central and peripheral tasks.

3.2.5 Experiment 2: face identification with non-famous faces

In a separate dual-task experiment, subjects performed a face-identification task as before, but this time with a set of non-famous faces. Except for the following details, the organization of the experiment was the same as Experiment 1. In this experiment, a set of 16 individuals from our non-famous set of faces (see above) were randomly chosen as targets. There were 24 distracter individuals. As mentioned before, we used five different views of each of these faces. Each block of trials started with a "familiarization phase" during which subjects were shown all five views of a particular target individual. They were instructed to familiarize themselves with this individual for subsequent identification in the ensuing block. By self-report, on average, subjects took 30 seconds to look at the faces before they started the block by pressing the space bar. In the block that followed, the target individual was presented on 24 of the 48 trials; while distractors of the same gender as the target were presented on the remaining trials. The targets and distracter trials were randomized, and subjects reported on each trial if the face was a target or not by pressing one of two keys on the keyboard. The average presentation time was 179.8 ± 15.8 ms for the letters and 162.3 ± 17.1 ms for the faces. These SOAs were not significantly different from those obtained for Experiment 1.

Six subjects performed 4 one-hour sessions of this experiment on 4 consecutive days. On each day, they performed 6 dual-task blocks and 4 blocks of the two single tasks. The target face was never repeated in any blocks of the experiment during a one-hour session. However, since the same target face could be presented on different days, we tested subjects in Experiment 3 on a set of unrepeated target faces.

3.2.6 Experiment 3

The layout of this experiment was identical to Experiment 2. The only difference was that on each block of trials, a new target individual was presented to subjects. Thus, targets were never repeated across blocks. Four subjects performed at least 11 blocks of the dual-task condition, and 5 blocks of the two single-task conditions in this experiment.

3.2.7 Experiment 4

In Experiment 4, 4 subjects were tested on face identification but with a set of inverted faces. This experiment was only performed with the non-famous faces. Subjects performed at least 12 blocks of the dual-task condition and 8 blocks of the two single-task conditions. In all other respects, the design of the experiment was identical to that of Experiment 2.

3.2.8 Experiment 5

In this experiment, 4 of the subjects performed a disk discrimination task in the periphery. Two color patterns—a vertically bisected disk with red and green halves and such a disk rotated by 180°—were presented to subjects. On each trial, the disk was masked by a disk divided into four red and green alternating quadrants. At least 12 blocks of the dual-task condition and 8 blocks of the two single-task conditions were collected for each subject. The average presentation time for the disks was 78.9 ± 19.4 ms.

3.3 Results

The dual-task paradigm was used to determine the effects of manipulating spatial attention on face-identification performance. One each trial, subjects were presented with different faces and were asked to report if the face was of the individual who had been designated as the target at the beginning of the block (Figure 3.1). In each block, the distractors were always of the same gender as the target. In the dual-task condition, subjects performed on both tasks, while prioritizing the central letter discrimination task.

In Experiment 1, five subjects were tested on face identification with faces of well-known Hollywood actors and actresses. A list of their names appears in Table 3.1. Subjects' performances on this task are shown in Figure 3.2a. The average performance on the letter discrimination task in the single- and dual-task conditions were 80.3 ± 5.4% and 78.2 ± 5.5%, respectively. These values are not significantly different from each other for each subject (t test, p>.05), indicating that in the dual-task condition the focus of attention was engaged by the letter discrimination task. Average performance on the face-identification task was 83.2 ± 5.3% when it was performed alone, and 81.1 ± 4.9% in the dual-task condition. For each of the five subjects, performance on this task in the dual-task and single-task conditions was not significantly different (t test, p>.05.). These results are summarized in Figure 3.2b, in which each participant's performance in the dual-task condition is plotted relative to the performance they achieved in the single-task conditions[6]. As the data shows, for each of the five subjects, dual-task performance was above 90% of their performance in the single-task condition. Eye movements would not have played a major role in achieving such performance since the peripheral faces were only presented briefly on each trial (see methods). Additionally, control experiments with an eye tracker have allowed us to verify that eye movements do not contribute to the performance achieved in the dual-task condition (Figure 2.7). These results thus indicate that subjects are able to efficiently identify famous individuals even when spatial attention is not fully available for the task.

Given these results with familiar faces, it is interesting to ask whether this performance extends to lesser known faces as well. It is possible that the subjects' ability to identify individuals in the near-absence of spatial attention is limited to a small group of famous or highly familiar people, and that identifying relative strangers would require closer attention. Accordingly, in Experiment 2 we repeated the face-identification experiment, but this time with a set of non-famous faces. As mentioned in the methods section, this face set contained five views of several unknown individuals, and on different blocks, a particular individual was chosen as the target. Before each block began, all five views of the target were presented to subjects who were instructed to acquaint themselves with that individual for subsequent identification.

Performance of six subjects on this task is shown in Figure 3.3a. The average performance on the letter discrimination task was comparable in the single- and dual-task conditions signifying that subjects were paying attention to this task in the dual-task condition (82.7 ± 3.4% and 80.3 ± 4.7%, $p>.05$, two-tailed t-test). On the peripheral task, average performance was 86.4 ± 3.0% and 82.1 ± 3.3% in the single- and dual-task conditions, respectively. The difference in performance on the face identification task

---

[6] The exact formula used was: normalized performance = $1/2 + 1/2.[(P_2 - 1/2) / (P_1 - 1/2)]$ where P_2 and P_1 refer to performance in the dual-task and single-task conditions, respectively.

between the single- and dual-task conditions was significant for 2 of the 6 subjects. Figure 3.3b shows the performance of each subject in the dual-task condition normalized to their performance in the single-task conditions. From this figure we see that, on average, face-identification performance in the dual-task condition was above 90% for all subjects. Thus these results demonstrate that although there is a small decrement in performance in the dual-task condition, identifying relatively unfamiliar faces is possible in the near-absence of attention. Additionally, the data also allows us to confirm that the results we observed with famous faces were really due to a face-identification process and were not a result of artifacts introduced by the image set. The images used in Experiment 1 were obtained from the web and were therefore not very well controlled for low-level or other cues specific to a particular celebrity. For instance, stars are well known for their distinctive hairstyles and facial expressions, and subjects could have based their decisions in the face-identification task on these cues. Thus, even though the image set allowed us to use a broad range of photographs of people in natural everyday environments, the face-identification results could have been disputed. The results of Experiment 2, however, confirm that subjects are able to identify individuals in the near-absence of attention, even in the absence of any obvious cues.

Over the course of Experiment 2, although a particular individual was never the target on more than one block in each session, he or she could have been re-assigned to be the target in another session. On average, for a particular target this would have occurred no more than two times, but it could still be argued that the results we have observed this far were biased by the familiarity gained with particular targets. For this reason, in Experiment 3 we re-tested four of our subjects on face identification but this time with a novel, unknown target individual on each block. Their average performance on the face-identification task was 80.4 ± 5.1% and 79.7 ± 3.1% in the single- and dual-task conditions, respectively. This difference was not significant for any of these subjects (p>.05). The normalized data for these subjects shown in Figure 3.4a demonstrates that even under these conditions, subjects achieve a high level of performance on this task. Over the group of subjects the normalized face-identification performance was greater that 95%.

In Experiment 1, at the beginning of each block, subjects were told who the target was by simply presenting the corresponding celebrity's name. However, this was not possible in Experiments 2 and 3 because all the targets were unknown people. Instead, before each block in these experiments, subjects were presented with all five views of the target so that they could begin to recognize them (typically subjects spent 30 seconds in examining these faces before proceeding with the block). It is possible that during this familiarization phase of the experiment, subjects memorized the overall layout of each view of the target, and in the block that followed, rather than identifying the face, they just used a template matching strategy by comparing the presented face on each trial with the templates they had in memory.

To control for this possibility we tested subjects on face identification with (unfamiliar) inverted faces in Experiment 4. As in the previous two experiments, in this experiment, subjects were presented with all five views of the target face at the beginning of each block, but this time all the faces were inverted. If subjects had indeed relied on a template-matching strategy in the previous experiments, then they would have been able to use a similar strategy with the inverted face task. This would have led to performance similar to that observed in Experiments 2 and 3. Four subjects from Experiments 2 and 3 were tested with inverted faces in Experiment 4 (Figure 3.4b). Their average performance on the face-identification task was 79.8 ± 4.2% and 67.1 ± 3.0% in the single- and dual-task conditions, respectively. This difference was significant for all of these subjects (p<.01). In comparing performance on upright and inverted faces, for each subject, we also observed a significant drop in face-identification performance (p<.01). Thus these results demonstrate that a template-matching strategy would not have been sufficient to achieve the high levels of performance observed in the previous experiments. Additionally, the results with inverted faces also confirm that subjects did not use a strategy based on low-level differences in the image set to perform the task in Experiments 2 and 3 (since otherwise similar performance levels would have been observed with upright and inverted faces).

These results thus demonstrate that subjects are able to perform subtle discriminations about the identity of individuals in the near-absence of attention. However, in all these experiments, an obvious concern arises about the efficacy of the central letter discrimination task in effectively engaging the focus of attention away from the periphery. This concern can be addressed by verifying that for some tasks that are known to require attention, performance suffers in the dual-task condition. We ensured this was the case in our dual-task condition by testing 4 of the subjects in the dual-task condition on a bisected disk discrimination task, which has been shown previously to require attention (Braun & Sagi, 1990; Braun & Julesz, 1998; Li, VanRullen et al., 2002; Reddy, Wilken et al., 2004). As is shown in Figure 3.4c, in contrast to their performance on face identification (Figure 3.4a), subjects' performance on this task was severely impaired (p<.001). Over the group of subjects, the average performance on this task was 53.5 ± 2.6% in the dual-task condition, compared to 80.2 ± 5.7% observed in the single-task condition. These results confirm that under our dual-task condition, the central letter discrimination task does remove some attentional resource from the periphery, resulting in performance decrements in tasks known to require attention.

3.4 Discussion

The results of these experiments extend previous findings (Li, VanRullen et al., 2002; Rousselet, Fabre-Thorpe et al., 2002; Reddy, Wilken et al., 2004) on the processing of natural stimuli in the near-absence of spatial attention by demonstrating that faceidentification is also possible when spatial attention is not fully available. Given that finer and more complex levels of discrimination are required for face identification (compared to the natural tasks previously used as has been discussed in the introduction), it is surprising that processing does not break down completely when the attentional focus is shifted away from the faces. As we mentioned earlier, the goal of the experiments described in this chapter was to ascertain how far the ability to process natural stimuli in the near-absence of spatial attention could extend. We had speculated that the face-identification task would reveal the limits of pre-attentive natural stimuli processing. Surprisingly, our results indicate that the visual system is not overwhelmed by discriminations of this caliber. While earlier work had indicated that super-ordinate levels of categorization (Mervis & Rosch, 1981) for natural stimuli (e.g. animal versus non-animal) are unimpaired in the near-absence of attention (Li, VanRullen et al., 2002), we now show that this finding extends to categorization at the individual level. In contrast to processing with natural stimuli, however, the near-absence of attention is a severe limitation to performance on discrimination of artificial geometric shapes, even though they are computationally simpler tasks.

In our data, we observed a modest drop in performance in identifying faces in the dual-task condition compared to the single-task condition (which was significant for only two subjects in Experiment 2 with the relatively unfamiliar face set, and did not occur at all with the celebrity faces in Experiment 1), but it should be remembered that some decrement in performance is expected to occur when subjects perform two tasks simultaneously. These decrements do not necessarily reflect competition for an attentional resource but could be ascribed to other factors such as having to remember two sets of instructions or produce two motor responses instead of one (Allport, 1980; Duncan, 1980b; Pashler, 1984, 1994).

Does this ability to make fine discriminations on natural stimuli also extend to other classes of natural objects, or is it only specific to faces? For instance, could one discriminate between two similar breeds of dogs in the near-absence of attention? It could be contended that the results we observe here would not generalize to other natural categories since various studies have claimed that dedicated areas in temporal cortex exist that preferentially process face stimuli, and that faces are thus of special importance to the visual system (Kanwisher, McDermott et al., 1997; Kanwisher, 2000; Grill-Spector, Knouf, & Kanwisher, 2004; Yovel & Kanwisher, 2004); (see however (Ishai, Ungerleider, Martin, Schouten, & Haxby, 1999; Ishai, Ungerleider, Martin, & Haxby, 2000; Haxby, Gobbini et al., 2001) for evidence for more distributed representations of objects in the temporal and lateral-occipital regions). However, it has also been suggested that faces are so well represented in the brain primarily because human beings are experts on face processing and that similar representations should also be observed for other well-known categories. In support of this claim, Gauthier and colleagues have shown that the areas underlying face processing also participate in the processing of other objects of expertise (such as cars for car experts) (Gauthier, Tarr, Anderson, Skudlarski, & Gore, 1999; Gauthier & Logothetis, 2000; Gauthier, Skudlarski et al., 2000; Tarr & Gauthier, 2000; Gauthier, Curran, Curby, & Collins, 2003). Thus it is possible that these regions would similarly facilitate the processing of highly familiar natural categories and that complex discriminations in the near-absence of attention would be possible for these objects as well.

In fact, the disparity between performance on natural object and artificial geometric shape discrimination in the dual-task condition could be directly linked to the difference in how familiar the visual system is with these different stimuli. Indeed in contrast to natural scenes and faces, bisected red-green disks are rarely encountered in everyday life. The different degrees of familiarity for different object categories might be mirrored by a continuum of “attentional requirements” in which faces would be one extreme. This could explain why very subtle discriminations among face stimuli (such as recognition of individuals) can be done pre-attentively, whereas, to answer the question posed above, if subjects do not routinely engage in discriminating breeds of dogs, then the corresponding task might require attention. This speculation is not entirely unreasonable since if the brain is optimally adapted for everyday stimuli, the processing of oft-encountered natural stimuli should be favored over unusual geometric shapes (Kayser, Kording et al., 2004; VanRullen, Reddy et al., 2004).

fMRI studies have shown that visual perception and imagery of famous or well-known faces activate face-selective voxels in the fusiform gyrus and other temporal areas (Dubois, Rossion et al., 1999; Leveroni, Seidenberg et al., 2000; Gorno-Tempini & Price, 2001; Rossion, Schiltz, Robaye, Pirenne, & Crommelinck, 2001; Ishai, Haxby, & Ungerleider, 2002). Furthermore, evidence for single neurons in the human temporal lobe that are selectively responsive to different views of familiar natural objects has been reported recently. These neurons have been shown to respond to famous buildings and celebrities, or to entire categories of natural objects (e.g. animals or cars; see Chapter 5 and (Kreiman, Koch et al., 2000a; Kreiman, 2002; Quian Quiroga, Reddy et al., 2005)). The hardwired selectivity of these temporal neurons, established through repeated exposures with a particular stimulus category, could underlie dual-task performance for our celebrity identification task (Experiment 1) or the animal versus non-animal categorization task (Li, VanRullen et al., 2002), respectively. On each trial, the activity of the neurons explicitly encoding the relevant famous person or natural category might be sufficient to give rise to a perception of the target (even in the near-absence of attention), resulting in the high levels of behavioral performance we observe (VanRullen, Reddy et al., 2004).

However, in the present study we have also reported that similar levels of performance are observed for relatively unfamiliar faces (Experiment 2 and 3). What are the neural candidates that could underlie the processing of unfamiliar faces in the near-absence of attention? A coding strategy based on explicitly encoding every possible face individually at the neural level is clearly neither efficient nor feasible. Furthermore, it is also unlikely that explicit representations for the target faces (such as those mentioned above for famous faces) could have been created during the approximately 30-second-long familiarization phase at the beginning of the block when subjects were presented with these faces. At the neuronal level, changes in synaptic connectivity (LTP, LTD) are known to involve much longer time scales (Abbott & Nelson, 2000; Bi & Poo, 2001). Instead, our results suggest that face processing mechanisms can be flexible and generalize their abilities to faces that have not been encountered previously, by setting up representations (not necessarily explicit) online over a few seconds. Furthermore, these representations are probably of a temporary nature since it would be wasteful to maintain permanent representations for faces one is not likely to come across again. Computational models of face processing have suggested that upon encountering a face, recognition units that describe the face are established (Bruce & Young, 1986). Undoubtedly, based on neurophysiological evidence (Hasselmo, Rolls, & Baylis, 1989; Sergent, Ohta et al., 1992; George, Dolan et al., 1999; Hoffman & Haxby, 2000), areas in the temporal lobe are likely to be implicated in this processing (Haxby, Hoffman et al., 2000; Haxby, Hoffman, & Gobbini, 2002); however, the manner in which these online representations are implemented at the neural level is still unclear and open to future research.

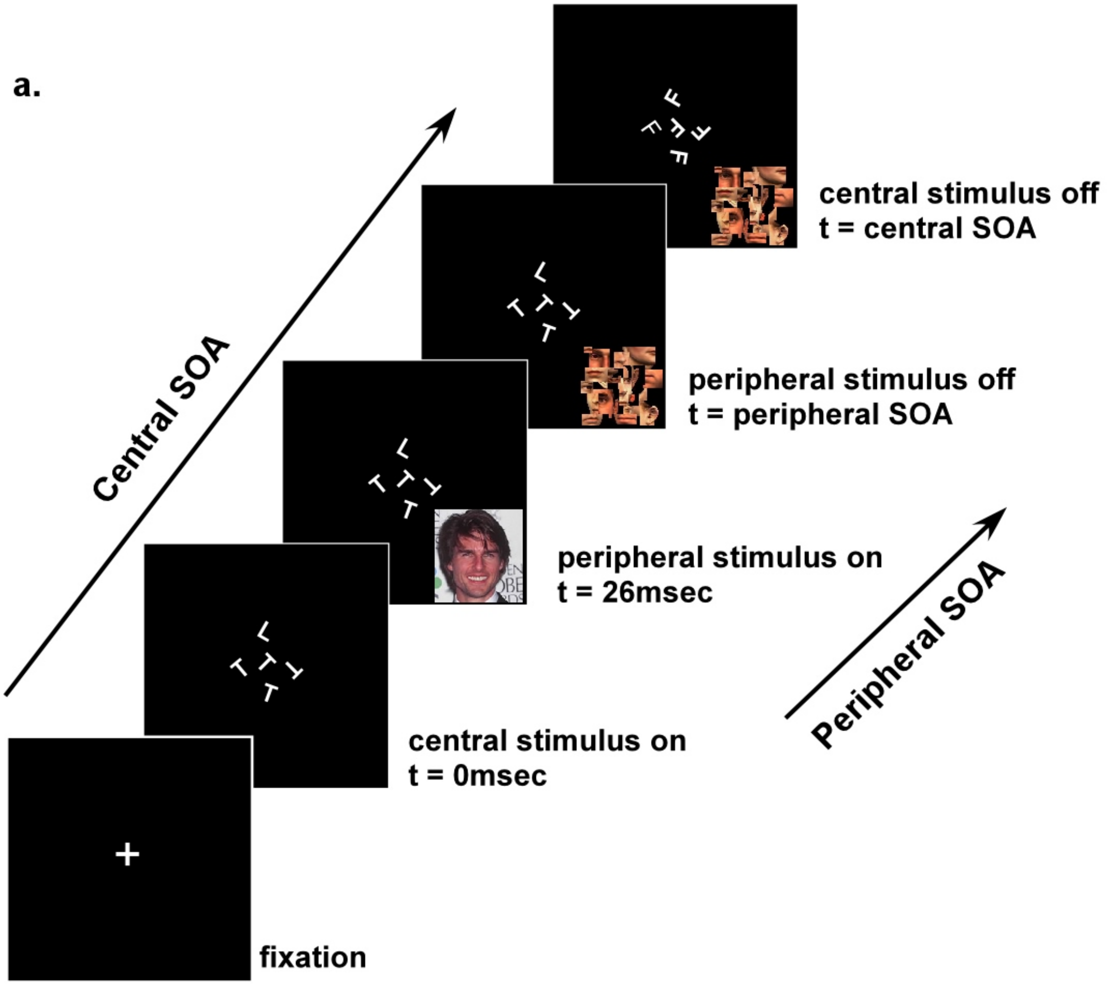


Figure 3.1: Face-identification discrimination dual-task experiment. (a) Schematic timeline for one trial in the dual-task experiment. At the end of a trial, participants are required to report whether or not the face was that of the target celebrity and/or whether the 5 central letters were the same (either 5 Ts or 5 Ls) or different (4 Ts and 1L or 4 Ls and 1T). All trials are arranged similarly, independent of the specific instructions. Both letters and faces were masked individually. Central SOA (~200 ms) and peripheral SOA (~167 ms) indicate the presentation time for letters and celebrity faces respectively.

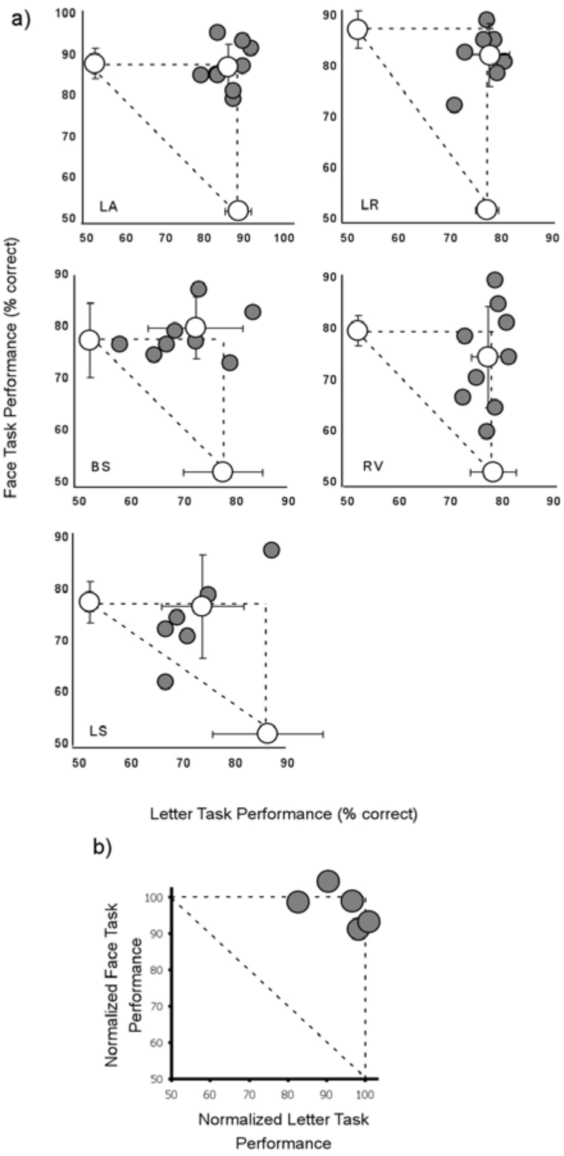


Figure 3.2. Results of the famous face-identification task (Experiment 1). *(a)* Results of 5 subjects. The horizontal axis represents performance on the attentionally demanding central letter task. The vertical axis represents performance on the peripheral celebrity identification task. Each filled circle is the participant's mean performance in the dual task in one block of 48 trials, while an open circle represents mean performance over all blocks in the three experimental conditions: single-central task, single-peripheral task, and the dual task. By default, performance of the "to-be-ignored" task is assumed to be at chance level (50%) in the single-task condition. Error bars represent standard deviation. For all participants face-identification performance in the dual-task condition is not significantly worse (*t*-test, *p*>0.05) than performance in the single-task condition indicating that face-identification suffers only minimally when performed concurrently with an attentionally demanding task. *(b)* Normalized average performance for each participant in the dual-task paradigm. Each point represents a participants' performance in the dual-task normalized to their single-task performance. Normalized values are obtained by a linear scaling, which maps the average single-task performance to 100% leaving chance at 50%. Normalized face identification performance values lie above 90% of single-task performance, suggesting that participants can perform this task remarkably well in the near-absence of attention.

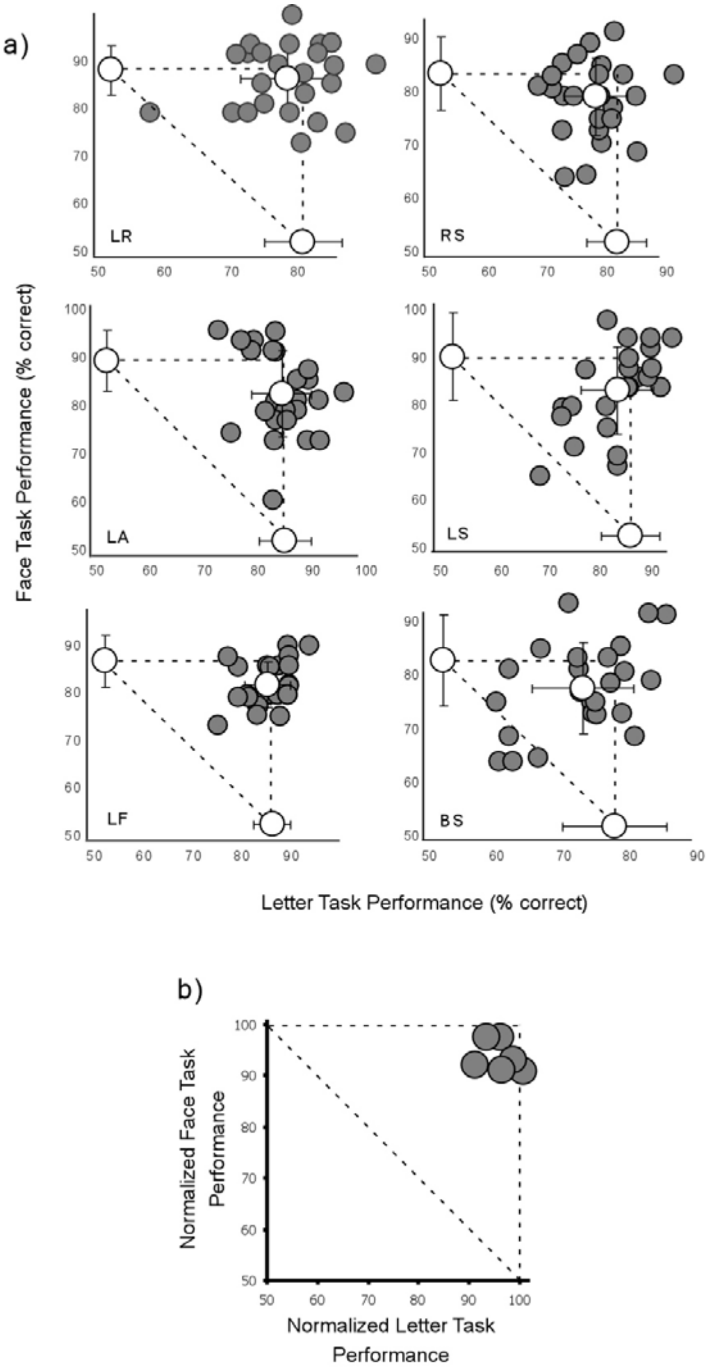


Figure 3.3. Identifying unfamiliar faces. Performance of six participants in the dual-task paradigm with the unfamiliar faces (Experiment 2). The format of this figure is the same as Figure 3.2. Normalized performance values shown in b) are >90% for all subjects. Thus, even with unfamiliar faces, subjects are able to identify faces in the near-absence of attention.

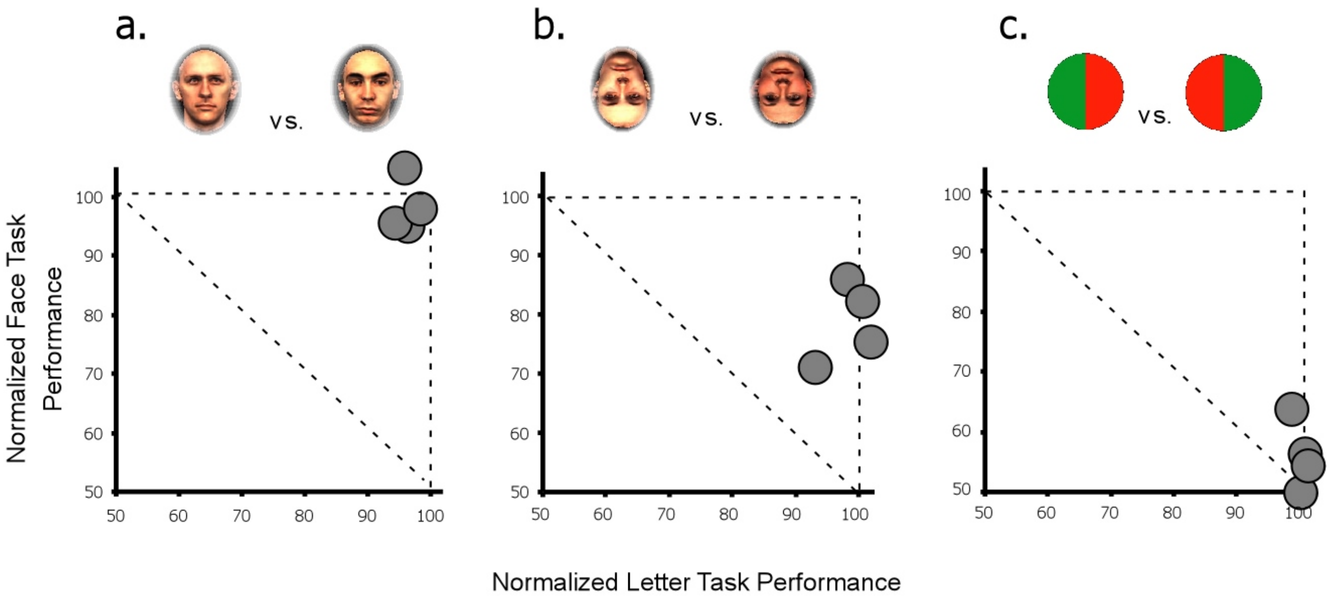


Figure 3.4. Face-identification control experiments. Normalized performance values for 4 subjects in three dual-task experiments (Experiments 3–5). a) Performance values on upright face identification with a new target face on each block of the experiment. The normalized performance is greater than 95% for each subject. b) Performance values on face-identification with inverted faces. The average normalized performance is 78.4 ± 6.6% for these subjects. A significant drop in performance is observed for each subject with inverted faces compared to upright face identification. The results indicate that the performance observed with upright faces is not due to a template matching strategy (see text), nor a strategy based on low-level differences between images. c) Performance values on a disk-discrimination task in the periphery fall to chance levels in the dual-task condition. This indicates that the central letter task does withdraw some attentional resource away from the periphery resulting in a sharp drop in performance on certain tasks.

Table 3. 1 Celebrity target and distracter images shown for face identification

Targets:

Catherine Zeta Jones
Courtney Cox
George Clooney
Harrison Ford
Jennifer Aniston
Jennifer Lopez
Julia Roberts
Nicole Kidman
Tom Cruise
Tom Hanks

Distracters:

Al Pacino
Andy Garcia
Antonio Banderas
Arnold Schwarzenegger
Ben Affleck
Billy Boyd
Cameron Diaz
Colin Farrell
Colin Firth
Cuba Gooding Jr.
Danny DeVito
Denzel Washington
Drew Barrymore
Elijah Wood
Ethan Hawke
Hugh Grant
Hugo Weaving
Ian McKellen
Jennifer Garner
Jim Carrey
Julia Stiles
Keanu Reeves
Liv Tyler
Lucy Liu
Madonna
Matthew McConaughey
Meg Ryan
Merryl Streep
Miranda Otto
Nicholas Cage
Orlando Bloom
Penelope Cruz
Rene Zellwegger
Robert DiNiro
Robin Williams
Russell Crowe
Salma Hayek
Sarah Jessica Parker
Sean Astin
Sean Connery
Tim Robbins
Viggo Mortensen
Winona Ryder

Chapter 4

When does Visual Attention Modulate fMRI Activity in High-level Visual Areas?

4.1 Introduction

The work presented thus far demonstrates that the visual system is well equipped to handle the stimuli it encounters on a daily basis. As we have argued, rather than relying on laborious attention-dependent processing, the underlying architecture of the visual system capably supports straightforward computations of natural objects and scenes. It has been suggested that a possible basis for these computations could be the neuronal selectivity observed in higher-level brain areas for natural stimuli (VanRullen, Reddy et al., 2004). Neurons in the anterior part of the infero-temporal cortex in monkeys and the medial temporal lobe in humans have been observed to be highly selective to natural categories such as animals, faces, foodstuffs, buildings, and so on ((Gross, Rocha-Miranda et al., 1972; Perrett, Rolls et al., 1982; Desimone, Albright et al., 1984; Kreiman, Koch et al., 2000a; Quian Quiroga, Reddy et al., 2005); see also Chapter 5 of this thesis), and these built-in preferences for different types of natural stimuli could ensure that the corresponding stimuli reach perception in the near-absence of attention, resulting in high levels of performance on the discrimination tasks.

However, several fMRI studies have shown that BOLD activity for unattended objects is decreased or even abolished in high-level areas specifically encoding these stimuli (Wojciulik, Kanwisher et al., 1998; O'Craven, Downing et al., 1999; Vuilleumier, Armony, Driver, & Dolan, 2001; Pessoa, McKenna et al., 2002; Marois, Yi, & Chun, 2004; Yi, Woodman, Widders, Marois, & Chun, 2004). In these studies, the BOLD signal measured in the fusiform face area (FFA) or the parahippocampal place area (PPA), (areas that typically show enhanced activation to images of faces and places, respectively) was shown to be significantly reduced in the absence of attention.[7] These results appear to contradict the observation that subjects are efficiently able to categorize natural stimuli when attention is not fully available. For if the cortical areas that supposedly support the processing of these stimuli show significantly lower levels of activity in the near-absence of attention, then how is it that the corresponding behavioral performance remains unaffected?

In this chapter, in an attempt to reconcile these two lines of evidence, we studied the effects of attentional modulation on face processing in the FFA, a region in the fusiform gyrus that responds strongly to images of faces compared to other non-face stimuli (Puce, Allison, Gore, & McCarthy, 1995; Kanwisher, McDermott et al., 1997; Kanwisher, Stanley et al., 1999). We employed the dual-task paradigm in a 3.0T fMRI scanner while subjects performed a face-gender discrimination task. We must note here that an animal or vehicle discrimination task could have served just as well in principle. However, in practice, faces turn out to be a very convenient set of stimuli to use in fMRI studies since the FFA shows very specific responses to faces, and can be readily localized (Kanwisher, McDermott et al., 1997; Kanwisher, 2000).

Previously, fMRI studies have investigated the effects of attentional modulation on face processing as paradigmatic for object processing in general (Wojciulik, Kanwisher et al., 1998; O'Craven, Downing et al., 1999; Vuilleumier, Armony et al., 2001; Pessoa, McKenna et al., 2002). Most of these studies compared activity in the FFA in two conditions: when the faces were fully attended and task relevant, and when

---

[7] Attention was manipulated using various methods: e.g. the attentional blink paradigm, face/house matching tasks, or reporting on the orientation of peripherally presented bars, which were displayed along with the faces.

the faces were fully ignored (and task irrelevant). A significant drop in BOLD activity has generally been observed between these two conditions, and the difference was attributed to attentional modulation. However, because subjects were instructed to make a behavioral report on the faces in the first, but not in the second condition, it is unclear whether the observed drop in activity is solely due to attentional modulation, or if it can also be explained by a change in the behavioral relevance of the faces.

An advantage of the dual-task paradigm is that, in addition to the two conditions used in previous studies (faces attended vs. faces ignored), it includes a third condition in which attention is spatially focused away from the faces (on an attentionally demanding letter discrimination task), yet they remain task relevant. Thus we are in a position to separately examine the effects of focal attention and task relevance on activity in the FFA. In agreement with previous reports, we observed a significant drop in activity when the faces were unattended and behaviorally irrelevant. However, when the spatial focus of attention was not on the faces but they remained behaviorally relevant, activity was left unaffected. This high level of activity in the near-absence of focal attention might constitute the neural basis of the surprising behavioral performance obtained in the dual task condition.

4.2 Methods

4.2.1 Behavioral training

Twelve paid volunteers, aged 20–30 participated in dual-task face-gender discrimination experiments. All experiments were conducted according to the guidelines of the Institute’s committee for protection of human subjects. The details of the paradigm are described in Chapter 2. For the current set of experiments, six “trained” subjects were trained on the two single-task conditions (letter and face-gender discrimination) as well as the dual-task condition, while six “untrained” subjects were trained on only the two single-task conditions. The entire training procedure typically lasted between 5 and 10 hours on consecutive days for each subject and was carried out outside the scanner on a Silicon Graphics computer. The details of the training procedure were identical to those described in Chapter 2. Over the group of 12 subjects, the SOAs varied between 133–160 ms and 173–240 ms on the face and letter tasks, respectively.

The “trained” subjects also performed a control experiment outside the scanner. In this experiment, the peripheral task involved discriminating a vertically bisected disk with red and green halves (equated for gray values) from their mirror image. Each disk was masked by a disk divided into four red and green alternating quadrants. Subjects received equal amounts of training on this task (central, peripheral, and dual blocks) as on the face-gender discrimination task; the tasks were matched for difficulty such that average single-task performance was around 80%.

The behavioral performance of subjects on the dual-task experiment is reported below both in terms of percent correct, as well as the discriminability index, d’ (a measure based on signal detection theory, which gives a response bias-free measure of performance).

Normalized performance values reported in Figure 4.2 are calculated by a simple linear scaling of the mean value of each participant’s performance. The scaling mapped the mean single-task performance to 100%, leaving chance at 50%:

Normalized performance = $0.5 + 0.5.[(P_2-0.5)/(P_1-0.5)]$, where P_2 and P_1 refer to performance in the dual-task and single-task conditions respectively.

d’ was calculated for the letter and gender discrimination tasks by estimating the proportion of hits and false alarms on each task. Note that d’ is a measure of the distance between the noise and signal+noise distributions and assumes that the noise follows a Gaussian distribution with a fixed variance.

4.2.2 fMRI sessions

The twelve subjects performed the face-gender, dual-task experiment in a 3.0 T whole-body Siemens scanner at the California Institute of Technology. Stimuli were presented on a Macintosh laptop and viewed through optical goggles (Resonance Technologies, Northridge, CA, VisuaStim XGA. 800 x 600 resolution at 60 Hz). The size and eccentricity of the stimuli in the goggles was the same as that during training. T2* weighted Echo-planar images (TR=2s, TE=30ms, FA=90°, FOV=210mm, 31X3 mm interleaved axial slices) were acquired using the whole-head coil and an in-line motion correction sequence (Thesen, Heid, Mueller, & Schad, 2000). The slices were positioned to cover the temporal lobe. Each functional run consisted of 186 volumes; the first two volumes were discarded. A 12-minute high resolution (1x1x1mm) whole-head T1-weighted MPRAGE sequence (T1=1.5s, TI = 0.8s, TE=3.05ms, 176 sagital slices) was also acquired for each subject.

Localizer task

FFA and PPA

Subjects were first presented with sequences of images in order to localize the FFA and PPA (Kanwisher, McDermott et al., 1997). The design consisted of alternating 30 s blocks of color faces and outdoor scenes (approximately 9 x 13.5° of visual angle), obtained from the Vision Lab at Caltech (http://www.vision.caltech.edu/html-files/archive.html) presented for 1 s each at fixation. The faces were different from the ones used in the dual-task experiment. Each run lasted for 4 minutes, and subjects participated in 3 runs. The FFA and PPA ROIs were localized in each individual by contrasting average brain activity in face versus scene blocks. Due to different signal-to-noise ratios, there was some variation in the threshold used for defining the ROIs across subjects. The average t value that defined the ROI was $t(6) = 4.1 \pm 0.4$; $p < .005$. The mean Talairach coordinates of the FFA and PPA (right FFA: $x = 38 \pm 1$mm, $y = -50 \pm 2$ mm, $z = -13 \pm 1$mm; left FFA: $x = -40 \pm 1$mm, $y = -51 \pm 2$mm, $z = -15 \pm 2$mm; right PPA: $x = 24 \pm 2$mm, $y = -42 \pm 2$mm, $z = -6 \pm 1$mm; left PPA: $x = -28 \pm 1$mm, $y = -46 \pm 4$mm, $z = -9 \pm 2$mm) are consistent with previous reports (Kanwisher, McDermott et al., 1997; Epstein, Graham, & Downing, 2003). The mean number of voxels over all subjects was 28 ± 3 and 67 ± 29 in the FFA and PPA, respectively. The FFA and PPA localized for one subject are shown in Figure 4.1. In a separate localizer session, 4 of the 6 subjects were presented with a rapid sequence of faces and scenes in the periphery at 5Hz. The faces and scenes were presented at the same locations as during the dual-task experiment (see below). At each peripheral location, the sequences of faces and scenes were presented for a total of 2.5 s, with each face or scene being shown for 200 ms.

LOC

The Lateral Occipital Complex (LOC) in both hemispheres was localized for the 6 untrained subjects by contrasting images of grayscale objects obtained from (Murray & Wojciulik, 2004) with scrambled versions of these images. The design and analysis of these localizer runs were otherwise identical to those for identifying the FFA and PPA as described above.

Dual-task paradigm

The fMRI experiment was based on a block-design. Subjects performed 10 runs of the dual-task paradigm in the scanner using the parameters described in Chapter 2 except for the following modifications. In a single run, 2 blocks each of the central, peripheral- and dual-task conditions were presented in a randomized counterbalanced order, with each block consisting of 16 trials. On average each block lasted approximately 40 s and was followed by a 20 s blank interval. Subjects made their behavioral report using a 4 button response box. As in Chapter 2, the peripherally presented face subtending approximately 2.5 degrees of visual angle appeared at a random location on one of the four edges of an imaginary rectangle. This imaginary rectangle measured 8 degrees by 10 degrees of visual angle. The eccentricity of the faces was thus 4 to 5 degrees, (minimum = 4, maximum = 6.4 degrees (on the diagonal)).

Data analysis

For each subject, images were co-registered between runs using a Linear Image Registration Tool (FLIRT) (Jenkinson, Bannister, Brady, & Smith, 2002). Analyses were conducted using custom Matlab software (Mathworks). For each subject, the time-course in the 3-task conditions was calculated by averaging the BOLD signal change in all voxels of the relevant ROI over all repetitions of each condition (after removing linear drifts over the course of each run). These time-courses were collapsed across hemispheres and subjects. This is the data shown in Figures 4.4a, c and 4.6a, c and 4.8a, c. The time-courses for each subject were then integrated over the duration of the 40s block (shifted by 7s to account for the delay in the hemodynamic response) and normalized by dividing by the mean activation over all 3 conditions (group data shown in Figures 4.4b, 4.6b, and 4.8b. Statistical analyses (ANOVA and paired *t*-tests corrected for multiple comparisons using Scheffe’s method) were performed on these values.

4.3 Results

4.3.1 Behavioral performance

We used the dual-task paradigm described in Chapter 2 to examine the effects of attentional manipulation on brain activity in the FFA and the PPA. A typical trial from the experiment is shown in Figure 4.2a. In all trials, both the face and letters were displayed as shown, and depending on instructions, subjects could perform one of three tasks : (1) In the peripheral face-gender task condition, subjects had to report whether the face was male or female. Since attention was focused on the faces, and subjects made a behavioral report on them, this condition corresponds to the "attended" condition of most previous studies. (2) In the central letter task condition, subjects reported whether the letters were all the same, or if one differed from the other four. In this condition, attention is focused on the letters and away from the faces, which are task irrelevant. Thus this task is similar to the "unattended" condition in previous studies. (3) In the critical dual-task condition, subjects had to perform on both tasks simultaneously. As we have shown previously ((Reddy, Wilken et al., 2004); Chapter 2) and discuss below, in this condition, focal attention must be focused away from the faces although subjects are required to make a behavioral report on them. Thus, this condition allows us to dissociate the effects of attentional manipulation and task relevance (Fig 4.2b).

Six subjects who had been previously trained on the face-gender, dual-task paradigm, performed 20 blocks each of the central, peripheral, and dual-task conditions in the scanner. Each 16-trial block lasted 40 seconds and was followed by a blank display for 20 seconds. The behavioral results for these subjects on this task are shown in Figure 4.3a. As reported previously (Chapter 2 and (Reddy, Wilken et al., 2004)), participants performance on face-gender discrimination in the dual-task condition was comparable to that obtained in the single-task condition (mean performance ± s.e.m. normalized to single-task performance : 92.5 ± 2.0%; difference in performance between single- and dual-task conditions was significant for 3 of the subjects ($p<.05$)). The performance of these subjects was comparable to the performance of the six subjects reported in the first experiment of Chapter 2 (performance on central letter task, $p=0.79$; performance on face-gender task, $p=0.69$). Thus, a high level of performance is achieved even when the spatial focus of attention is shifted away from the faces. This conclusion also holds if we use d' as a response bias-free measure of performance: d' in the single- and dual-task conditions were on average 1.8 ± 0.5 and 1.4 ± 0.3 (mean ± s.d), respectively, and were not significantly different ($p=.15$). Thus a change in strategy between these two conditions is unlikely to explain the good performance we observe in the dual-task condition. Similarly d' values were not significantly different between single and dual-task conditions of the central task ($p=0.1$). Eye movements do not play a major role in achieving this performance since the peripheral faces are presented only briefly and at random locations. Furthermore, control experiments with an eye-tracker have allowed us to verify the absence of eye movement contribution to this performance (Chapter 2).

An obvious concern that arises, however, is whether the central letter discrimination task is sufficiently demanding to engage the focus of attention away from the periphery. This concern can be addressed by verifying that performance on known, attentionally demanding tasks suffers under the same dual-task conditions. As shown in Figure 4.3b, performance for five of the subjects (the remaining subject was not available for testing) on a bisected disk discrimination task in the periphery falls dramatically when performed concurrently with the central letter task, even though their performance on this task was comparable to face-gender performance when both tasks were performed alone. The difference in d' values on this task between peripheral (1.9 ± 0.2) and dual-task (0.25 ± 0.1) conditions was significant ($p<0.00005$) thus demonstrating that the central letter task did effectively engage the focus of attention away from the periphery. As before, we thus define focal attention operationally as that resource that, when engaged by the central T/L discrimination task, is unavailable to the periphery, consequently impairing performance on some concurrent tasks (such as bisected disk discrimination but not face-gender discrimination or natural scene categorization. (See also (Braun & Julesz, 1998; Lee, Koch et al., 1999; Li, VanRullen et al., 2002)).

4.3.2 fMRI activity: main result

The FFA and PPA (Epstein & Kanwisher, 1998) of each subject were isolated in separate localizer scans run before the dual-task sessions by contrasting brain activity in blocked presentations of faces and scenes presented centrally. The BOLD activity during the dual-task experiment was analyzed separately in the isolated FFAs from the left and right hemispheres. Since similar results were observed in both hemispheres, in the data presented here, the FFAs were collapsed across hemispheres. Figure 4.4a shows the raw time course of BOLD activity in the FFA during the three experimental conditions. Consistent with previous reports, we observed a significant decrease in activity when the spatial focus of attention was removed from the faces and they were ignored (peripheral task condition vs. central task condition, [$F(2,15) = 11.89$, $p = 0.0008$], Figure 4.4b). Remarkably, in the dual-task condition, when attention was not focused on the face, but a gender-specific response was still required, activity was as high as in the peripheral-task condition. Thus, we observed a dissociation between the effects of attentional manipulation *per se* and task relevance.

To rule out the possibility that BOLD activity was saturated in the peripheral and dual-task conditions (thereby accounting for the similar levels of activity in these conditions), we compared peripheral-task activity with that evoked by the face stimuli during a localizer experiment in the FFA when faces were presented peripherally at 2Hz, in both hemispheres of our subjects. We observed significantly lower levels of activity in the former condition compared to the latter (mean ± s.e.m.: 1.0 ± 0.1 vs. 1.9 ± 0.2 % signal change, $t(22) = 2.9$, $p < 0.01$). Note that this result is not unexpected: it is known that peripherally presented faces (peripheral-task condition) activate the face regions significantly less than centrally presented ones (localizer runs) (Levy, Hasson, Avidan, Hendler, & Malach, 2001). The frequency of presentation of faces (once per trial) was also considerably less than in the localizer scans. On the other hand we cannot rule out the possibility that a subset of neurons within these voxels, coding specifically for our peripheral face stimuli, could have been saturated in the dual-task experiment. To control for this possibility, we measured the activation in the FFA to peripherally presented faces and scenes (see Methods for details of the second localizer experiment). The faces were presented at 5Hz, and we observed significantly higher activity to these rapidly presented peripheral faces (2.1 ± 0.4 % signal change; $p < .05$) compared to activity in the peripheral-task condition. Thus the similar levels of activity observed between the dual and peripheral-task conditions truly mirror the experimental manipulation rather than saturation in the BOLD signal.

In contrast to a recent study that reported total absence of activation to unattended and ignored faces (Pessoa, McKenna et al., 2002), we observed decreased but significant levels of BOLD activity in the FFA in the central-task condition, presumably in response to the peripheral and task-irrelevant faces (Figure 4.4 a,b). We estimated the amount of activation elicited by the faces in this condition by examining the effect of removing the faces altogether. In a control experiment, three of our subjects performed the letter discrimination task exactly as in the central-task condition with the exception that the faces were not presented in the periphery. We observed a large (67%) and significant ($p < .05$) drop in FFA BOLD activity when the faces were not presented compared to the central-task condition when the faces were present although unattended and fully ignored (Figure 4.5b). Thus, we find that even unattended and ignored faces produce significant ($p < .05$) levels of FFA activation, in agreement with previous reports (Wojciulik, Kanwisher et al., 1998; Vuilleumier, Armony et al., 2001). In contrast to this relatively large effect of eliminating the influence of face stimuli, removing the letter stimuli from the peripheral task condition resulted in a much smaller (27%) drop in FFA activity (Figure 4.5c).

Undoubtedly, the dual task condition is more demanding than either of the two single tasks, since subjects have to perform two tasks simultaneously. Therefore, subjects might compensate for this perceived increase in difficulty by a more aroused state resulting in a widespread increase in activity in the dual-task condition. To determine whether this effect accounts for the high level of activity obtained in the dual-task condition, we looked at activity in the PPA region that had been identified in the localizer runs. This region is more strongly activated by spatial layouts than by faces (Epstein & Kanwisher, 1998). As shown in Figure 4.4c, no differential effect was present in the PPA ($p > .05$), thus ruling out the possibility that non-specific arousal effects are responsible for the observed results.

4.3.3 fMRI activity: the role of training

Even for simple tasks such as color or orientation discrimination, achieving good performance in the dual-task paradigm is contingent on substantial training (Braun & Julesz, 1998; Li, VanRullen et al., 2002). We thus wondered whether the high level of activity in the dual-task condition was merely a consequence of training. In other words, can everyone profit from this ability to process faces outside the focus of attention, or is it only acquired after extensive training? Furthermore, since untrained subjects usually show a drop in behavioral performance in the dual-task condition, we were also interested in determining if FFA activity would mirror this decrease. Therefore, in a second experiment, six subjects, untrained in the dual-task condition performed the face-gender dual-task experiment in the scanner.

Six subjects who had never performed in any dual-task condition before (but had been exposed to both single tasks) were tested in the scanner under exactly the same conditions as in the previous experiment. As expected, their behavioral performance on face-gender discrimination dropped considerably when the focus of attention was shifted away from the faces: for this group of subjects, average performance in the dual-task condition was 73.7 ± 3.0% of performance in the single-task condition (difference in performance between single- and dual-task conditions on the peripheral task was significant for all six subjects, $p<0.009$; Figure 4.3c). On average, the corresponding d' dropped significantly ($p=.008$) from 1.3 ± 0.4 in the single-task to 0.6 ± 0.3 in the dual-task conditions.

Figure 4.6 shows the activity in the FFA for these six subjects. As in the data for Figure 4.4, levels of BOLD activity in the dual-task and peripheral-task conditions were similar, and a significant reduction was only observed in the central-task condition [$F(2,15) = 9.46$, $p = 0.002$]. Thus, despite the substantial drop in behavioral performance in the dual-task condition, the corresponding BOLD activity did not decrease compared to the single-task level. This implies that training is not a necessary condition for maintaining high levels of FFA activity in the absence of focal attention.

4.4 Discussion

Our results demonstrate a dissociation between the effects of focal, spatial attention and behavioral relevance. Manipulating the spatial focus of attention alone did not significantly affect the BOLD signal in the FFA as long as the face stimuli remained relevant to the task at hand. At first glance, this result may appear to be at odds with a number of other studies that have reported a significant decrease in FFA BOLD activity as a result of attentional manipulation of faces (Wojciulik, Kanwisher et al., 1998; O'Craven, Downing et al., 1999; Vuilleumier, Armony et al., 2001; Pessoa, McKenna et al., 2002). In fact, the two experimental conditions measured in most of these studies roughly correspond to our peripheral and central task conditions: subjects either perform a task involving the faces or ignore them completely. And, in agreement with previous findings, we observed significantly distinct levels of BOLD activity between these two conditions. However, by introducing a third condition in our paradigm (the dual-task condition), we were able to separate the effects of two factors (spatial attention and task relevance), which were confounded in previous reports. Our data shows that these factors indeed have distinct effects on the BOLD signal in the FFA. Thus, rather than being at variance with previous reports, these results, together with earlier findings, reveal a more complete picture of how focal attention and behavioral relevance interact in higher levels of visual cortex.

4.4.1 Focal attention and behavioral Relevance

Our interpretation here relies on a clear distinction between focal, spatial attention and top-down behavioral relevance. But how exactly do we define behavioral relevance? It is obvious from our data that there are two factors at play in our paradigm. We have operationally defined focal attention as the resource that is needed to solve the central task. Similarly we use the term “behavioral relevance” as an operational definition for the distinction that arises in our paradigm as a result of giving subjects different face-based task instructions in the three conditions. It could be argued, that in fact, what we term top-down behavioral relevance is a form of top-down and/or feature-based attention. Indeed, previous reports have shown that paying attention to a particular feature increases feature-related activity throughout the visual field (Treue & Martinez Trujillo, 1999; Saenz, Buracas et al., 2002). We believe that our interpretation is not incompatible with such alternative definitions, provided one assumes that face-gender constitutes a "feature" for the visual system.

Although we observed that the FFA BOLD activity was primarily dependent on behavioral relevance, irrespective of the current focus of attention, we do not wish to imply that behavioral relevance alone is the dominant factor affecting neuronal responses. Our paradigm does not allow us to examine the reverse interaction, namely how focal attention would affect processing when the stimuli are not task relevant. It is entirely possible that focusing attention on task irrelevant stimuli enhances their representation in cortex in a manner similar to when poorly attended stimuli are made behaviorally relevant (central vs. dual-task conditions). Although this issue has not been addressed explicitly, indirect evidence comes from studies in which manipulating the attentional load results in a significant modulation of BOLD activity in response to task-irrelevant stimuli in several brain areas (Rees, Frith, & Lavie, 1997; O'Connor, Fukui, Pinsk, & Kastner, 2002; Pinsk, Doniger, & Kastner, 2004; Yi, Woodman et al., 2004). Thus, focal attention and behavioral relevance might very well have interchangeable effects on brain activity, and it remains to be seen how these effects add up and/or interact in the visual cortex (see for example, (Treue & Martinez Trujillo, 1999)).

4.4.2 Effects of training in the FFA

What is the specific role of training in our paradigm? Our data shows that, within each group of subjects, FFA activity obtained in the dual and peripheral-task conditions was comparable (Figures 4.4 and 4.6 a, b). Additionally, between the two groups of subjects, the BOLD activity in the central, peripheral, and dual-task conditions did not differ significantly (p=0.3, p=0.3, and p=0.7, respectively). In terms of behavioral performance, for both groups of subjects, the information provided by this level of activity was sufficient to yield good performance, at least in the peripheral-task condition (presumably FFA activity can constitute a basis for behavioral decisions regarding face stimuli). However, in the dual-task condition, the same level of FFA activity only seemed to benefit the trained subjects. Thus, it appears that although information was similarly available to both sets of subjects, only the trained subjects were able to use it effectively. Although we cannot discount the (unlikely) possibility that the FFA is not necessary for face-gender discrimination tasks, it is possible that the BOLD signal is too crude a measure to show a difference in activation between trained and untrained subjects. Alternatively, rather than increasing BOLD activity in the FFA, training could facilitate a more efficient use of the relevant activity, possibly by minimizing competitive bottlenecks that arise (at stages later than the FFA) from processing two tasks simultaneously (Allport, 1980; Duncan, 1980b; Pashler, 1984, 1994).

4.4.3 The fMRI signal and behavioral performance in the FFA

It has been previously reported that the fMRI signal correlates with behavioral performance during object recognition tasks (Grill-Spector, Kushnir, Hendler, & Malach, 2000; Bar, Tootell et al., 2001). In the study by Grill Spector et al., various objects were presented at very short SOAs (20-500ms) and subjects' performance at correctly recognizing these objects was correlated with changes in the BOLD signal. Bar and colleagues presented masked objects for 67 ms and correlated the BOLD signal with subjects' confidence in reporting the identity of the objects. Both studies report that higher levels of behavioral performance were correlated with larger increases in the BOLD signal.

However our data reveal that for our subjects there is no apparent correlation between behavioral performance and BOLD activity in the FFA. Whereas we observed significantly lower behavioral performance for our untrained subjects compared to the trained subjects, both groups had comparable levels of FFA BOLD activity in the dual-task condition, as discussed above. Over the groups of trained and untrained subjects, the correlation coefficients (r^2) between the average FFA BOLD activity of a subject and their average face-task performance (over all face-task blocks) were 0.04 and 0.08, respectively. At the level of individual subjects, additional analysis revealed that on each face-task block, there was no correlation between BOLD activity in the FFA and the corresponding behavioral performance ($r^2 = 0.001$ on average, and $r^2 < 0.13$ in all cases; Figure 4.7 shows this data for the 'trained' subjects). Thus, while FFA activity might be necessary for good behavioral performance in our gender discrimination task, it is not sufficient. One possible difference from earlier studies (Grill-Spector, Kushnir et al., 2000; Bar, Tootell et al., 2001) lies in the procedure used to modulate performance. These studies used backward masking with short SOAs (20–500 ms, compared with ~160 ms for our paradigm) to limit stimulus visibility. It is possible that the attentional manipulation in our paradigm limited discriminability (and thus behavioral performance) without affecting the visibility of the faces (although we did not directly test for stimulus visibility). When male faces are confused with female faces at the same level of visibility (and vice-versa), no net difference in FFA activation is expected.

4.4.4 Visual responses in the FFA

As we mentioned previously, faces are a convenient set of stimuli to use in fMRI studies because the regions (Kanwisher, McDermott et al., 1997; Halgren, Dale et al., 1999; Haxby, Hoffman et al., 2000; Hoffman & Haxby, 2000) responding to face stimuli can be easily identified. However, although it has consistently been shown that these regions respond more to faces than to other objects, whether or not these areas are specialized for face processing is a heavily debated matter. In particular, fMRI studies have observed significant responses to other types of natural stimuli, including animals in the FFA ((Chao, Martin, & Haxby, 1999; Ishai, Ungerleider et al., 1999), but see also (Kanwisher, Stanley et al., 1999)). Besides the stimulus category, other factors such as the subjects' expertise with the relevant stimuli are thought to play a role in determining the visual properties of these areas as well (Gauthier, Tarr et al., 1999; Gauthier & Logothetis, 2000; Gauthier, Skudlarski et al., 2000; Gauthier & Nelson, 2001). In this context, our results suggest that the behavioral relevance of the stimuli must also be taken into account when considering the visual response of the FFA.

Another factor that is thought to influence processing in the FFA (and other higher level cortical areas) is related to the eccentricity at which the stimuli are presented. Specifically, it has been suggested that the FFA is associated with tasks that require a great degree of visual acuity (such as processing facial information or reading), and consequently it is an area with a central visual field bias (Levy, Hasson et al., 2001; Malach, Levy, & Hasson, 2002; Hasson, Harel, Levy, & Malach, 2003). Our results however indicate that peripheral faces can also significantly activate the FFA, even when they are unattended (Figure 4.5c).

4.4.5 Saturating the BOLD signal—some thoughts

The main result of this study is that top-down behavioral relevance is sufficient for evoking high levels of BOLD activity in the FFA, even when the face stimuli are outside the focus of attention. However as we mentioned earlier, a possible concern with this interpretation could be that if the signal during the dual-task condition was saturated, we would necessarily not have observed any additional benefit from shifting attention onto the faces (i.e. the peripheral condition). We have already presented data (Section 4.3.2) that confirms that the voxels in our ROIs were not saturated, since for each of 4 subjects significantly higher levels of activity were observed during the localizer tasks with peripherally presented faces, compared to the dual-task experiment. However, these results do not necessarily preclude the possibility that the underlying neuronal populations were not saturated during the peripheral-task condition.

The primary argument against saturation in our analysis hinges on the supposition that presenting faces for longer durations would cause the underlying neuronal population to be more active, thereby increasing the overall signal measured in the relevant voxels. Indeed it is known that masked stimuli activate neuronal populations much less than unmasked stimuli presented for longer durations (Rolls & Tovee, 1994; Rolls, Tovee, & Panzeri, 1999; Rolls, 2004). This observation, however, is generally only true when neuronal activity is considered over the longer intervals of time during which the unmasked stimuli are present. At any particular moment in time when the stimulus is displayed, the firing rates for unmasked stimuli are not necessarily higher compared to those of masked stimuli. In other words, the optimal stimuli of the cell under study, whether masked or unmasked can saturate the neuronal response during the time they are presented. The reason the fMRI signal could possibly show differences in the levels of activity elicited by masked versus unmasked stimuli is that it integrates the underlying neuronal signal over a longer time scale, during which interval the effects of presenting stimuli for longer periods become apparent in the average neuronal activity.

For our purposes, from a neuronal point of view, the face stimuli could very well have saturated the relevant neural signal during the dual-task condition, at least during the time they were presented (before the onset of the mask). Thus at the level of the neurons responding to our peripheral stimuli, the additional effects of shifting the focus of attention onto the faces in the peripheral-task condition would not have been detected.

Does this possibility then undermine our main conclusions? Not really! Our conclusion is that, in the absence of focused attention, top-down behavioral relevance can produce high levels of neuronal activation. Whether saturation occurs or not this conclusion still holds—we see high levels of activity in the dual-task condition when focal attention is not on the faces. In fact, if saturation does occur, our results would imply that neuronal activities can reach their maximal level of response in the absence of focused attention! This result is indeed intriguing, given that current views hold that neuronal activity must be decreased in the absence of attention.

It must be noted that saturation could become a problematic issue if we wish to demonstrate that focal attention has absolutely no effect on BOLD activity—beyond that induced by behavioral relevance. But this is not our stance. Although no significant effect of focal attention on BOLD activity was observed in the particular conditions of our experiment, we are not opposed to the idea that spatial attention can increase the neuronal response under some conditions. Indeed when the stimulus is presented in a cluttered environment, or is degraded, several monkey electrophysiology and human fMRI studies show the pronounced effect of spatial attention, as we discuss below. However, our results suggest that, at least under some conditions, removing spatial attention does not necessarily result in lower levels of cortical activity. This finding, which holds independent of whether saturation occurs or not, contrasts with the previous understanding of face-related BOLD activity and focal attention.

4.4.6 Activity in other brain regions

This far, we have primarily focused on activity in the FFA because this region responds specifically to faces. However, it is of interest to also consider activity in other brain areas during the different task conditions. We localized the parahippocampal place area and the lateral occipital complex for our subjects, and our findings in these areas are discussed below.

4.4.6.1 The parahippocampal place area (PPA)

The parahippocampal place area is a region in the parahippocampal cortex that responds strongly to spatial layouts compared to faces and other objects (Epstein & Kanwisher, 1998). In our localizer task, we identified this area for all 12 of our subjects, and as expected from its known visual properties, the face and letter stimuli failed to elicit large levels of activation in the PPA (compared to the FFA; Figures 4.4c and 4.6c). In addition, we did not observe differential effects of focal attention and task relevance in these voxels. This is not entirely unexpected given the lack of a visual response in the first place. This finding is reassuring since, as discussed in section 4.3.2 it confirms that our observations are specific to the voxels we targeted and is not a whole-brain effect.

4.4.6.2 The lateral occipital complex (LOC)

The LOC, a cortical area known to be important in the processing of object shapes (Malach, Reppas et al., 1995; Kanwisher, Woods, Iacoboni, & Mazziotta, 1997; Grill-Spector, Kushnir, Edelman, Itzchak, & Malach, 1998; Kourtzi & Kanwisher, 2000, 2001), was localized for the six untrained subjects. Over these six subjects, we observed a significant increase in the LOC BOLD signal compared to baseline for both our central and peripheral tasks. This result is compatible with previous reports showing that letters and faces activate regions in the lateral occipital region (Puce, Allison, Asgari, Gore, & McCarthy, 1996). The activity elicited by the attended letter stimuli was significantly higher than that elicited by the attended faces (p=.02; Figure 4.8) suggesting that the ROI we localized in the lateral occipital cortex may exhibit a preference for letters over faces. However, we did not observe a significant effect of focal attention and behavioral relevance on the BOLD signal in the LOC in the different task conditions (p=.14, dual-task vs. peripheral-task condition).

4.4.6.3 Retinotopic visual areas

Although we did not localize the boundaries of the early visual areas in our subjects, it is worth speculating about what we would have observed in retinotopically organized regions. Several fMRI studies have reported that focal attention modulates activity in early areas including V1 (Tootell, Hadjikhani et al., 1998; Watanabe, Harner et al., 1998; Sengpiel & Hubener, 1999; Somers, Dale et al., 1999), and thus it is likely that different patterns of activity would have been observed in response to the peripheral and central stimuli in the different task conditions. In particular, it is probable that activity at foveal representations would have been higher in the central and dual-task conditions (compared to the peripheral-task conditions) since focal attention would have been directed to the foveal stimuli. Conversely, at the peripheral representations, activity would have been expected to be lower during the central task condition compared to the peripheral task condition. But it is interesting to consider what would happen at peripheral locations during the dual-task condition, when focal attention is not available to the periphery, but the peripheral stimuli are behaviorally relevant. Because we observe a modulation of activity due to top-down behavioral relevance in higher areas (FFA), it is legitimate to ask whether the same effect arises in earlier retinotopic areas. This is unlikely however since these retinotopic areas do not explicitly encode our peripheral stimuli (faces). The fact that these effects are not observed in the PPA, which also gets input from the early visual cortex, probably implies that the effect is not present in such retinotopic areas. This does not rule out the possibility however that top-down behavioral relevance could affect activity in these areas when appropriate stimuli (such as sinusoidal gratings) are used.

4.4.7 In comparison with monkey electrophysiology

With reference to electrophysiology studies, our results showing the absence of a significant effect of focal attention on BOLD activity might seem to contradict previous reports of strong attentional modulation at the neuronal level (Moran & Desimone, 1985; Spitzer, Desimone et al., 1988; Reynolds, Chelazzi et al., 1999; Reynolds, Pasternak et al., 2000; Reynolds & Desimone, 2003). However, it is important to note that these observations have generally been made when two competing stimuli (a preferred and a non-preferred stimulus) are placed within the receptive field (RF) of a recorded neuron. When a single stimulus is placed within the RF (in the absence of competition), the effect of shifting attention in and out of the RF is much less pronounced, at least for well-contrasted stimuli (Moran & Desimone, 1985; Haenny, Maunsell et al., 1988; Maunsell, Sclar et al., 1991; Motter, 1993; Luck, Chelazzi et al., 1997; Reynolds, Chelazzi et al., 1999). Attentional effects compatible with these single-cell observations have also been reported in fMRI studies (Kastner, De Weerd et al., 1998). Thus, far from being inconsistent, electrophysiological accounts are compatible with the data we report here—namely that, for well-isolated and contrasted stimuli (such as our face stimuli), the effect of removing the focus of attention alone, while keeping task demands comparable, does not result in a substantial drop in neuronal activity.

4.5 Conclusions

In conclusion, our results might help explain a puzzling discrepancy observed at the behavioral level: in some cases, visual processing can fail dramatically in the absence of attention (e.g. change-blindness or inattentional blindness (Rock, Linnett, Grant, & Mack, 1992; Simons & Levin, 1997)) while in other cases (dual-task), perception outside the focus of attention can be quite successful. When visual stimuli or changes are totally unexpected, absence of top-down behavioral relevance might decrease their associated neuronal representations, as shown here. On the other hand, as soon as the range of possible relevant events can be constrained by top-down influence (as in dual-task), neural activities need not suffer even in the absence of focal attention.

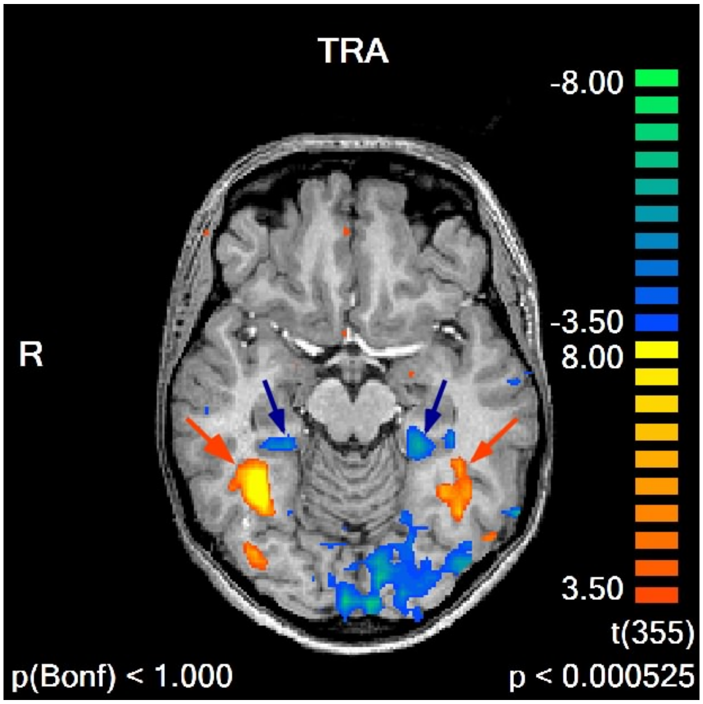


Figure 4.1: Localizing the FFA and PPA. The FFA (orange arrows) and PPA (blue arrows) localized for one of our subjects are shown in this transverse slice. The FFA and PPA were localized by contrasting BOLD activity during blocked presentation of color faces against blocked presentation of outdoor scenes. The mean Talairach coordinates of the FFA over all subjects were: Right FFA: x = 38 ± 1mm, y = -50 ± 2 mm, z = -13 ± 1mm; Left FFA: x = -40 ± 1mm, y = -51 ± 2mm, z = -15 ± 2mm. The mean Talairach coordinates of the PPA over all subjects were: right PPA: x = 24 ± 2mm, y = -42 ± 2mm, z = -6 ± 1mm; left PPA: x = -28 ± 1mm, y = -46 ± 4mm, z = -9 ± 2mm.

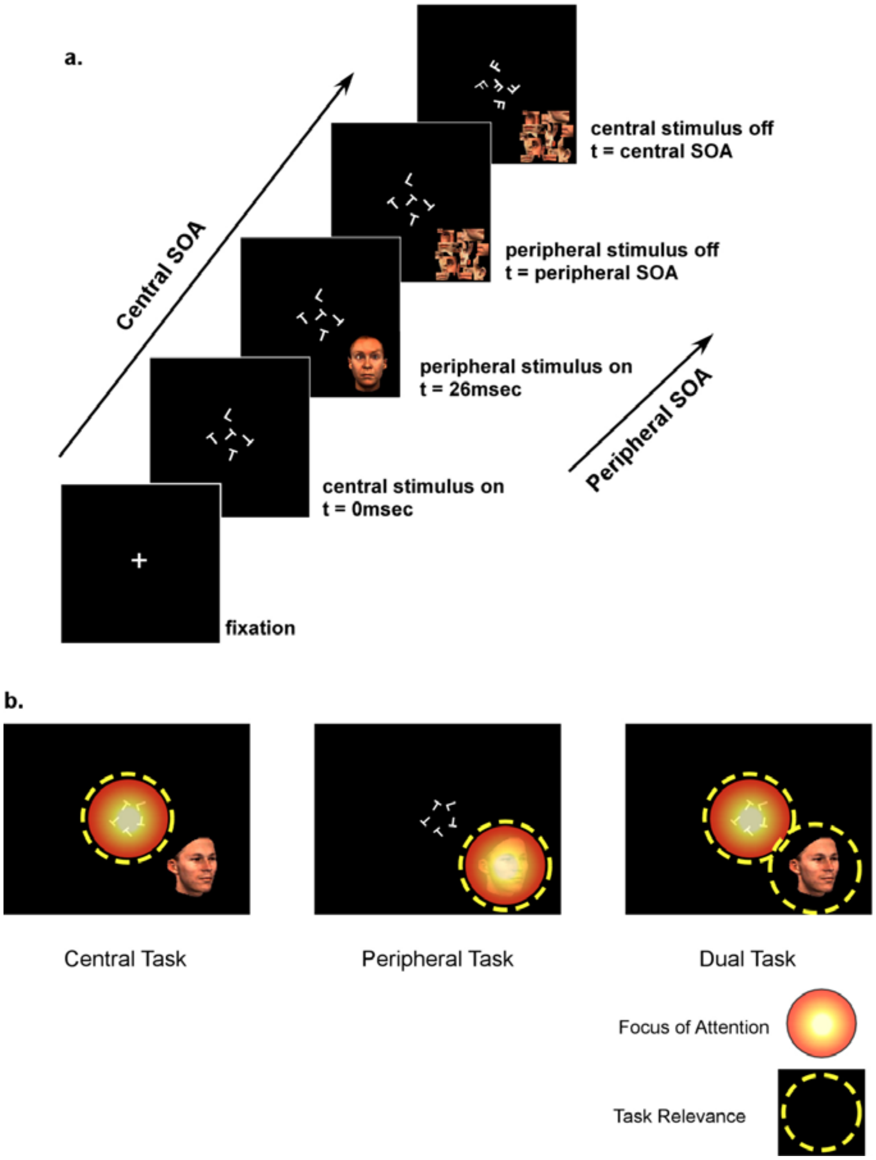


Figure 4.2: Manipulating attention and behavioral relevance in the dual-task experiment. a) Schematic timeline for a typical trial in the dual-task. At the end of the trial, subjects are required to report the gender of the face presented and/or whether the letters were all the same or different. The layout was the same for all trials, independent of specific instructions given to subjects at the beginning of each block. Central SOA(~200 ms) and peripheral SOA (~160 ms) indicate the presentation time for letters and faces, respectively. b) A simplified explanation of the manipulation of the focus of attention and behavioral relevance in the 3 conditions of the paradigm. In the central-task condition, the focus of attention is on the letters, and since subjects report on the letters, the faces are behaviorally irrelevant. In the peripheral-task condition, the focus of attention is available to the faces, and the faces are also behaviorally relevant. However, in the dual-task condition, the focus of attention is away from the faces, but they are still behaviorally relevant since subjects make a report on them. Thus, in this paradigm, we have conditions in which the faces are relevant and attended, a condition in which we remove only the focus of attention, and finally a condition in which we also make the faces behaviorally irrelevant. The three conditions enable us to tease apart the effects of these two factors.

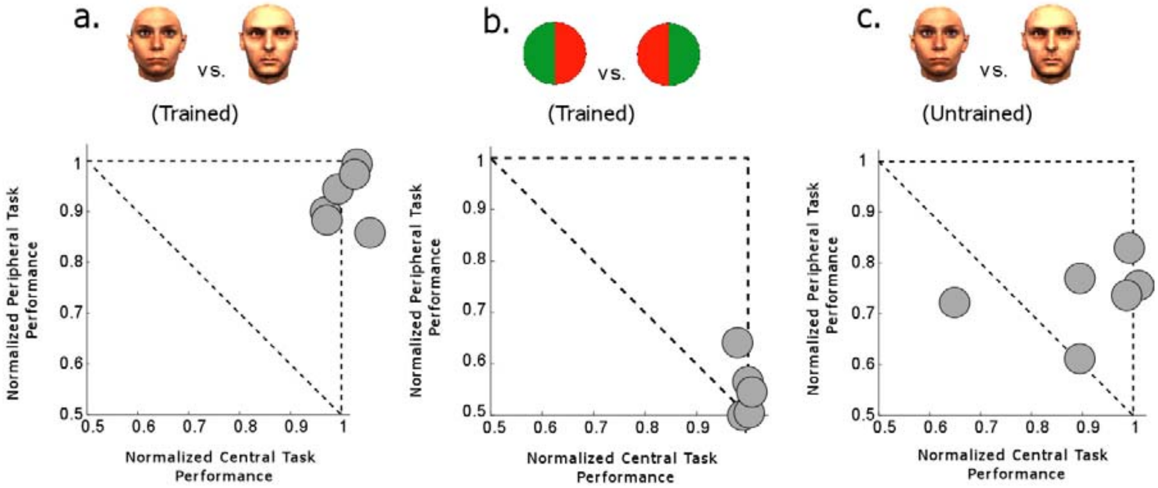


Figure 4.3: Behavioral performance. Dual-task performance values relative to performance achieved in the single-task conditions. a) Face-gender discrimination performance for 6 trained subjects obtained in the scanner. Each point represents a subject's dual-task performance on the central and peripheral tasks relative to their single-task performance. This group of subjects, achieves a high level of performance (average performance in the dual-task condition: 92.5% of performance in the single-task condition). The performance of these subjects and those reported in Chapter 2 were comparable (performance on letter discrimination task, p=0.79; performance on face-gender task, p=0.69). b) Performance for the same 6 subjects on a known attentionally demanding disk discrimination task (outside the scanner) falls to chance levels in the dual-task condition. Thus, the central task is effective in withdrawing focal attention away from the periphery. c) Face-gender performance for 6 new untrained subjects (in the scanner). As expected, we observe a significant drop (p<.009) in performance for these subjects in the dual-task condition.

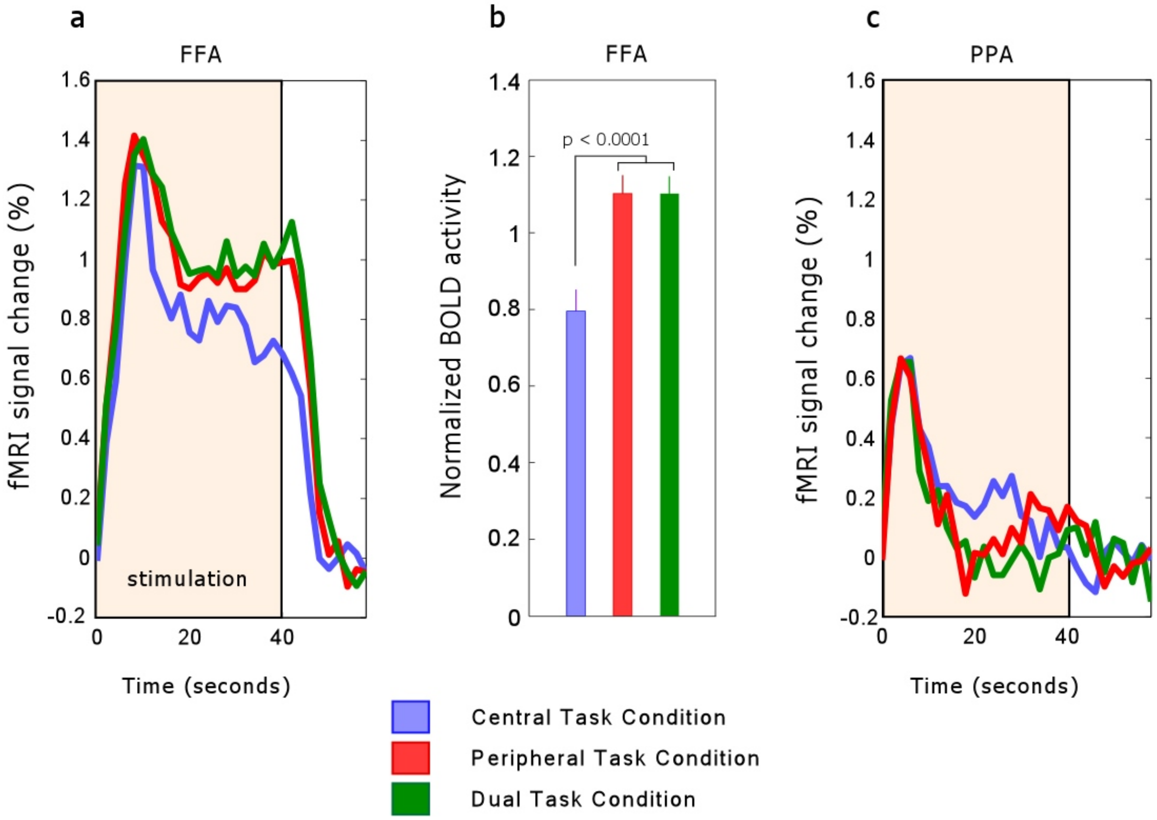


Figure 4.4: BOLD activity for 6 trained subjects in the dual-task condition. a) Time course of FFA activity. The data shows a significant drop in activity in the central task condition compared to the peripheral task condition. However, there is no significant drop in activity in the dual-task condition. Thus, removing the focus of attention alone does not affect activity. Making the faces task-irrelevant results in a significant decrease in the signal. b) The same data shown in a different format for the 3 conditions. Each bar corresponds to the area under the 3 curves shown in a) over the 40 s stimulation period (shifted by 7 s to take into account the delay in the hemodynamic response). A value of 1 corresponds to the average activation in the ROIs over all stimulation periods of the entire experiment. c) Activity in the PPA. The trend observed in the FFA is not observed in other brain areas, thus ruling out the possibility that the observed effects are due to general arousal.

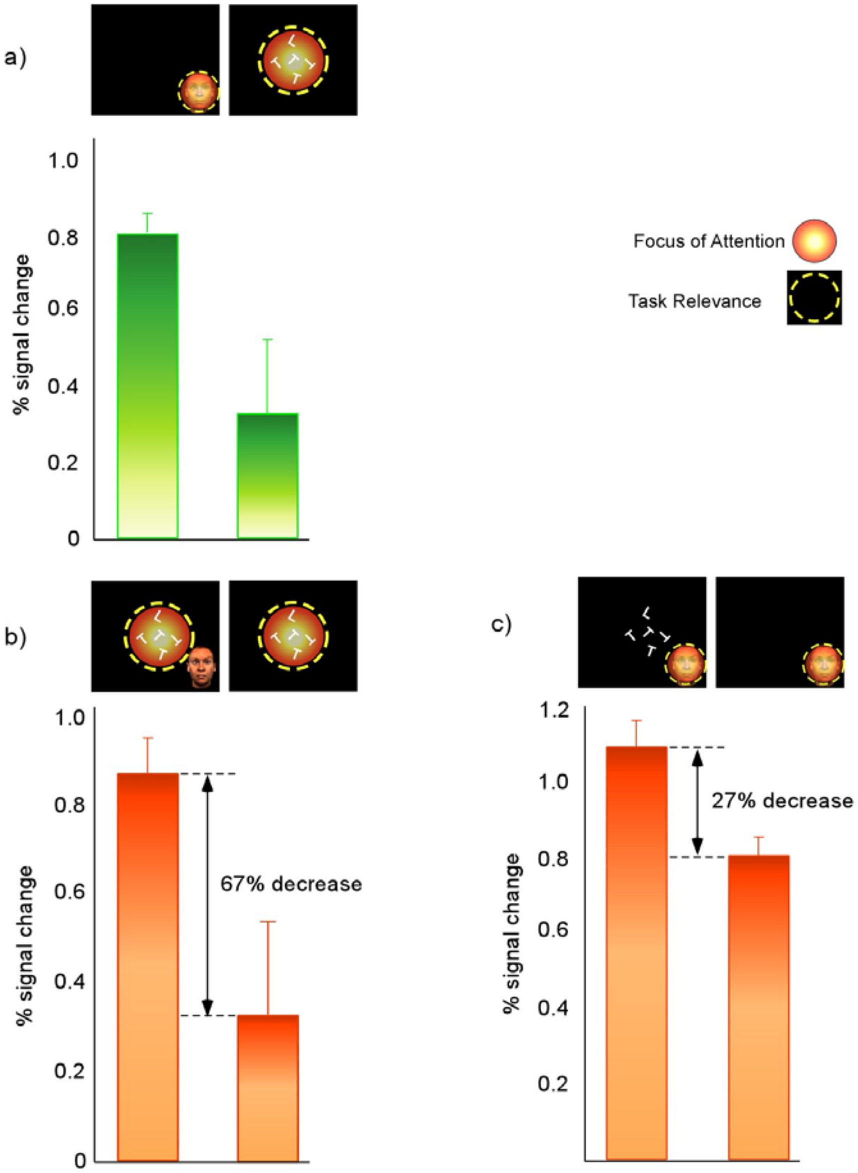


Figure 4.5: Estimating the respective contribution of letters and faces to the observed BOLD signal in the FFA. BOLD activity in the FFA for 3 subjects from the main experiment (Figure 2) and a control experiment in which they were presented with either only letters, or only faces, and performed the corresponding discrimination tasks. a) In the complete absence of the other stimulus, attended (and behaviorally relevant) faces activate the FFA 2.7 times more than attended (and behaviorally relevant) letters. These voxels are thus considerably more responsive to faces. b) The effect of removing the unattended face. The data on the left is obtained from the central task condition for these 3 subjects when the letters were attended but the faces were also presented. In the second bar, data from the control experiment is shown in the condition in which only letters were presented (and attended to) (note this is the same data as on the right in a)). Removing the faces from the display results in a large (67%) decrease in the FFA BOLD signal. c) In contrast, removing unattended letters while faces are presented results in only a 27% decrease. The data on the left is obtained from the peripheral task condition. On the right is data from the control experiment when the faces were attended and the letters were absent (left bar in a).

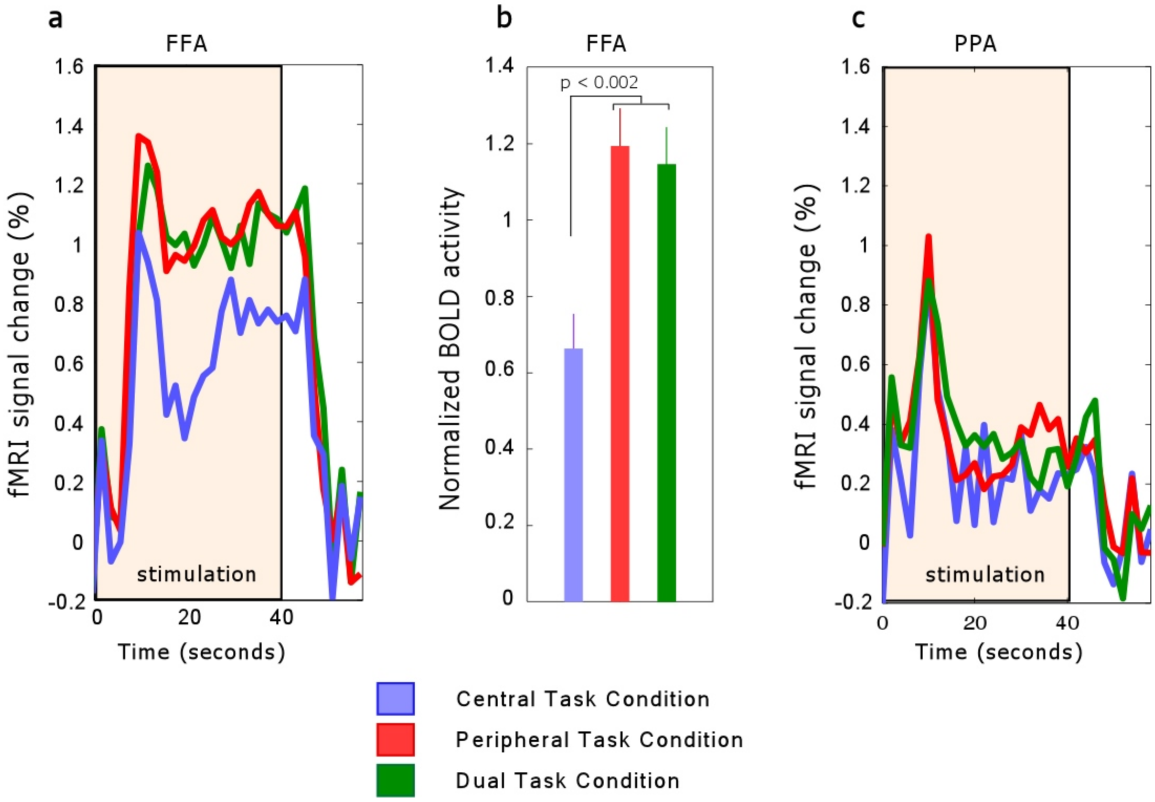


Figure 4.6: BOLD activity for 6 untrained subjects in the dual-task condition. a) Time course of FFA activity. Similar to Figure 3, removing the focus of attention alone does not affect BOLD activity. Instead, making the faces task-irrelevant results in a significant decrease in the signal. b) The same data shown in a different format for the 3 conditions. Note that despite the substantial drop in behavioral performance in the dual-task condition (Fig 4.3c), no corresponding decrease in BOLD activity is observed in this condition. c) BOLD activity in the PPA. The trend observed in the FFA is not observed in another brain area for the untrained subjects, thus ruling out the possibility that the observed effects are due to general arousal.

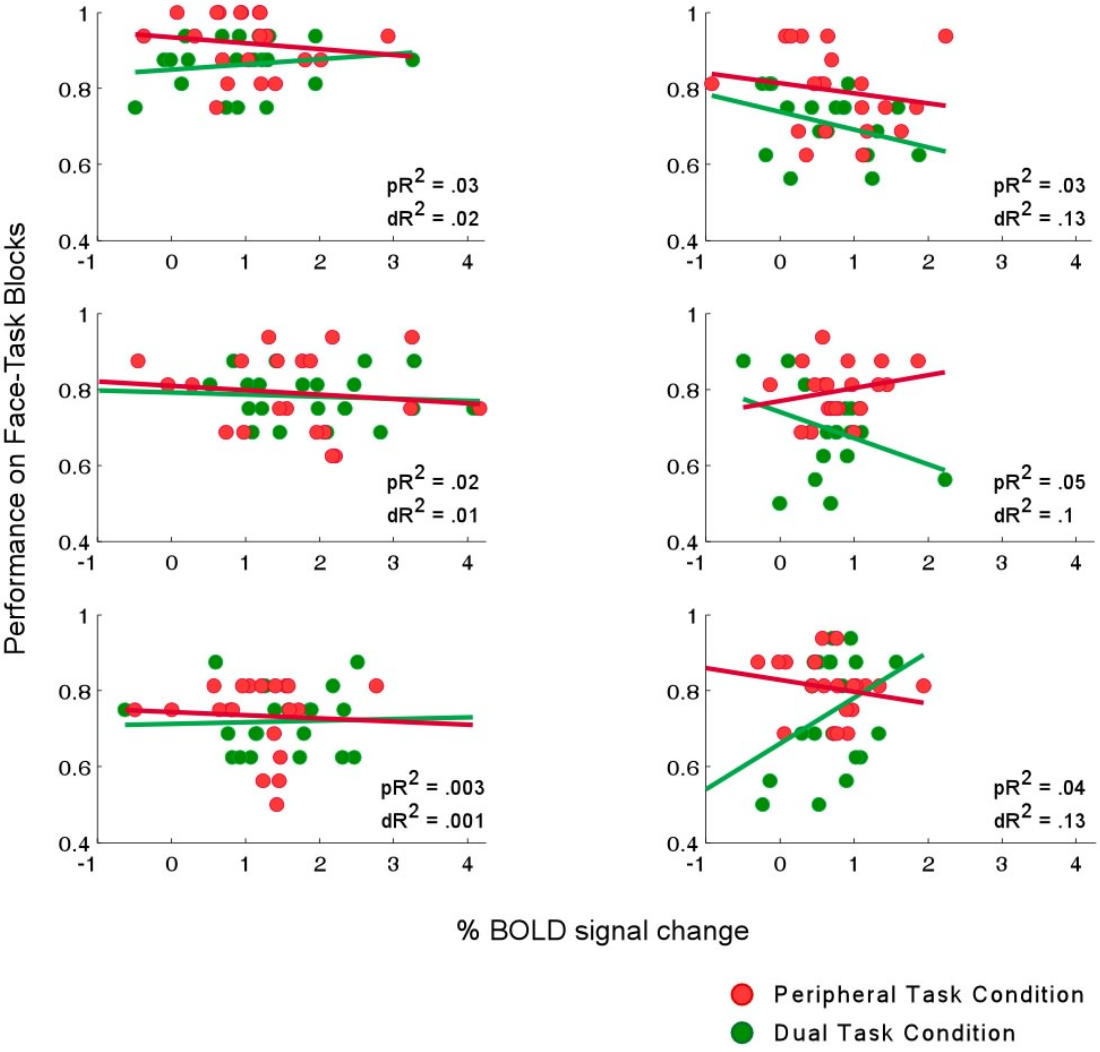


Figure 4.7 Correlating FFA BOLD activity with behavioral performance. The data shown here is for the six 'trained' subjects. For each subject performance on each face task block (i.e., peripheral and dual blocks) is plotted against the % BOLD signal change measured on that block. The red circles and lines represent the peripheral task condition, and the green circles and lines represent the dual task condition. The correlation coefficients for the peripheral (pR^2) and dual-task conditions (dR^2) for the six subjects are indicated on each graph. On average, over the trained and untrained subjects the correlations were weak (r^2<0.13).

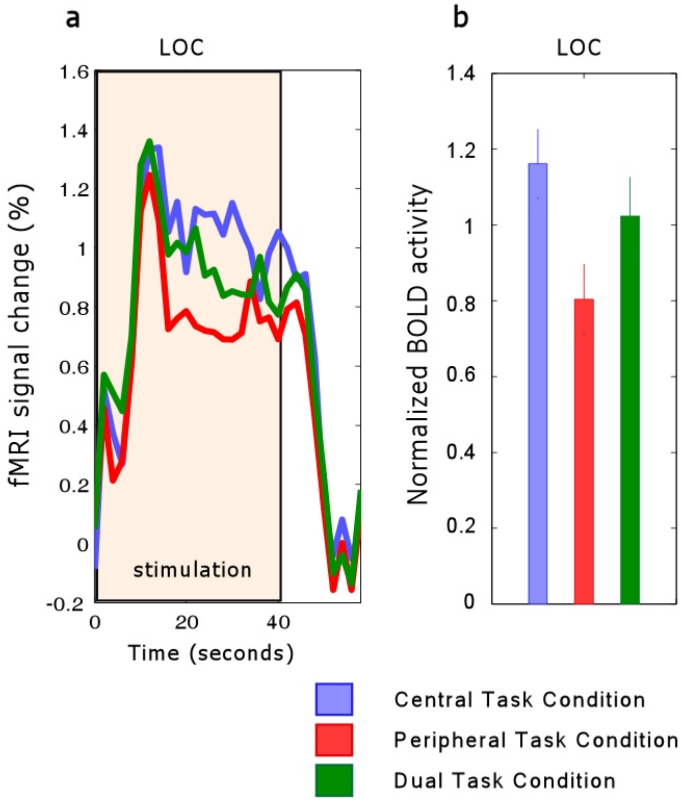


Figure 4.8: BOLD activity for the untrained subjects in the dual-task condition in the LOC. a) Time course of LOC activity. In the LOC, in contrast to the FFA, the opposite trend is observed, namely that BOLD activity in the central-task condition is higher than activity in the peripheral task condition (p=.02), indicating that the attended letters activate these voxels more than attended faces do. The difference in activity between the dual- and peripheral-task conditions is not significant (p=.14). b) The same data shown in a different format for the 3 conditions.

Chapter 5

Visual Representations in the Human MTL

5.1 Introduction

The work presented in the last chapter afforded us a look at how attention modulates brain activity. Although it was a big step forward in directly observing cortical activity, fMRI has its limitations in terms of the temporal and spatial resolution it offers. The fMRI response occurs over several seconds in comparison to the millisecond timescale of the electrical activity of individual neurons, and thus it is always unclear how the dynamics of the BOLD signal relate to the underlying neuronal mechanisms (Logothetis, Pauls et al., 2001; Logothetis & Wandell, 2004). In particular, given this long delay, one cannot assume that the signal being measured is due to a strictly local response to the stimulus. Instead, it is likely that several feedback signals from other processing centers in the brain participate in complex ways to constitute the BOLD signal. In terms of spatial resolution, the smallest voxels we could record from in our fMRI study (3.1 mm^3) would have included approximately 3 x 10^5 neurons[8] and thus the measured signal would reflect the averaged activity over a large neuronal population. While this average signal allows us to sample large parts of cortex, and could potentially represent general properties of a particular brain area, in order to understand how attention affects neuronal responses, it is imperative to study the electrical activity of individual neurons.

---

[8] 1 mm^3 of cortex is estimated to contain approximately 10^5 neurons.

As we have mentioned several times already, most of our understanding of attention has come from electrophysiological studies in awake behaving monkeys as they performed various tasks. By inserting electrodes into the brains of these animals, investigators could measure the underlying neural signal and examine how manipulating the focus of attention changed the response of the neurons. However, the effects of attention on individual neurons in the human brain have not been examined this far. Since the monkey brain is anatomically very similar to the human one, it is generally assumed that findings on monkey and human brains should be comparable. This assumption nevertheless does require one to extrapolate from one system to another, and consequently it is desirable whenever possible to record from the human brain and verify that the assumption is a reasonable one. In addition, certain experiments can be performed with humans that could never be feasible with monkeys (i.e., in relation to conscious perception, memories, etc.), and thus it is important to be able to conduct recordings in the human brain as well.

To complete this dissertation on attention, I will therefore describe work done on human subjects who performed an attentional paradigm while we recorded the activity of single neurons in their brains. However, since an understanding of the effects of attention on visual responses first requires an understanding of the visual responses themselves, in this chapter I will describe experiments that investigated the nature of the visual representations in the neurons we recorded from. The next chapter will describe the effects of attention on these visual responses. Parts of the recordings and analyses discussed in this chapter were performed by Rodrigo Quian Quiroga and are to appear in (Quian Quiroga, Reddy et al., 2005).

5.1.1 Subject population

Our subjects were epileptic patients at the University of California, Los Angeles (UCLA), who suffered from pharmacologically intractable epilepsy. In about 25% of the epileptic population, frequent seizure activity cannot be controlled by existing drugs (Engel, 1996; Ojemann, 1997), and thus the diseased portion of the brain where the focus of the seizure originates is often resected during surgery in a number of these patients. To locate the seizure focus, several non-invasive techniques such as structural MRIs, positron emission topography, and scalp EEG and MEG recordings are first employed. However, in some proportion of patients, these external measures do not yield sufficiently conclusive evidence about the location of a single epileptogenic focus, and patients must therefore be chronically implanted with up to 12 depth electrodes to locate the seizure focus (Fried, MacDonald, & Wilson, 1997; Fried, Wilson et al., 1999). Following electrode implantation, patients are monitored in the neurosurgical ward while their seizure activity is recorded via contacts on the electrodes. Patients typically spend a week to 10 days in the hospital—the length of time depends on clinical constraints, such as whether a sufficient number of seizures have been recorded for the doctors to acquire enough data to identify the seizure focus. During this period, the patients are willing to participate in experiments.

The surgeries at UCLA were performed by neurosurgeon Itzhak Fried. Dr. Fried was also involved at each stage in the planning of these experiments and in their actual implementation. Through the central lumen of each implanted (macro) electrode (1.25 mm in diameter), nine microwires (40 μm in diameter) were inserted from eight of which electrical activity of individual neurons could be recorded (Figure 5.1). The ninth microwire had lower impedance and served as a reference. The macro-electrodes contained 6–7 platinum contacts, which were used to acquire EEG data continuously during the entire length of time the patient spent on the ward. This EEG data was used by doctors to examine seizure activity. All studies conformed to the guidelines of the Medical Institutional Review Board at UCLA.

5.1.2 Electrode locations

The placement of the electrodes was based solely on clinical criteria and was determined by the data obtained from the EEG scalp recordings and the other localization tests performed. The electrode locations were verified by MRI or by computer tomography co-registered to preoperative MRIs. It should be noted that the MRI images are obtained at 1.5 T, which provides sufficient resolution to identify the tip of the macro-electrodes but not the trace of the microwires. Thus the locations we report correspond to the macro-electrodes. Generally, the microwires extended 3–4 mm beyond the tip of the macro-electrode. In most epileptic patients, the medial temporal lobe (MTL) is the primary location in which seizures originate, and consequently most of our electrodes were placed in MTL structures such as the hippocampus, entorhinal cortex, parahippocampal gyrus, and the amygdala. Occasionally, electrodes were also implanted in the orbito-frontal cortex, the supplementary motor area, and the anterior cingulate cortex. See Table 5.1 for location of electrodes.

5.1.3 Anatomy of the MTL

In the monkey brain, the MTL system (also called the limbic system) marks the convergence of inputs from purely visual areas (Felleman & Van Essen, 1991) to the polymodal centers in the brain that receive somatosensory, olfactory, and auditory information as well (Figure 5.2) (Suzuki & Eichenbaum, 2000). In all primates, the MTL consists of several distinct anatomical components that include the hippocampal formation (the dentate gyrus, areas CA3, CA1, and the subiculum); the entorhinal, perirhinal, and parahippocampal cortices; and the amygdala (Van Hoesen, Pandya, & Butters, 1972; Van Hoesen & Pandya, 1975a; Van Hoesen, Pandya, & Butters, 1975; Van Hoesen & Pandya, 1975b; Insausti, Amaral, & Cowan, 1987; Squire & Zola-Morgan, 1991; Suzuki & Amaral, 1994; Lavenex & Amaral, 2000). In the hierarchical organization of these components, the perirhinal and parahippocampal cortices are the structures that first receive input into the MTL. This input consists of information from both higher order areas such as the prefrontal cortex and also from the unimodal sensory areas. In the monkey brain, the perirhinal cortex receives particularly strong projections from visual area TE reflecting the role of this region in visual processing (Saleem & Tanaka, 1996; Suzuki, 1996b; Murray & Richmond, 2001), while the parahippocampal cortex receives prominent projections from dorsal areas and is more involved in spatial processing (Suzuki & Amaral, 1994; Malkova & Mishkin, 2003). The multimodal input from the perirhinal and parahippocampal cortices then feeds into the entorhinal cortex (EC) (Brown & Aggleton, 2001), which also receives some visual input from TE (Suzuki, 1996b). Extensive feedback connections exist between the entorhinal cortex and TE and the perirhinal and parahippocampal cortices.

The EC is the major source of cortical input to the hippocampal structures. Weaker projections from the perirhinal and parahippocampal cortices to the subiculum and CA1 layer of the hippocampus also exist in monkeys, but visual areas such as TE do not project directly to the hippocampus (Suzuki & Amaral, 1990). Instead, the EC appears to gate the flow of information in and out of the hippocampus.

The amygdala receives cortical input through the perirhinal, parahippocampal, and entorhinal cortices (Stefanacci, Suzuki, & Amaral, 1996; Suzuki, 1996a) although it also has other connections with different cortical and sub-cortical structures (LeDoux, 2000). For instance, direct connections to and from TE have been identified (Cheng, Saleem, & Tanaka, 1997).

Given the different amounts and sources of cortical input in each of these areas, it is likely that the different structures process information in distinct ways (e.g. (Baxter & Murray, 2001)). Since we have not recorded from sufficient numbers of neurons in each of these areas, we cannot make qualitative or quantitative distinctions in the nature of the visual responses we observe. However, it is important to keep this fact in mind in the later sections of this chapter where we present the visual properties of our MTL neurons.

5.1.4 Memory and MTL

Ever since the dramatic case of amnesia observed in patient H.M. after bilateral damage to the medial temporal lobe (Scoville & Milner, 1957; Milner, 1972), MTL structures have been linked to memory formation and retrieval. In particular it is now believed that these structures are important for the function of declarative memories (the memory for facts and events as opposed to non-declarative memory of skills and abilities) (Eichenbaum, 1992; Squire, 1992; Squire & Zola, 1996; Eichenbaum, 2000). Evidence for this view comes from patients and monkeys with lesions in MTL structures who reveal severe declarative memory impairments (e.g. a loss of autobiographical memories in humans) (Zola-Morgan, Squire, & Amaral, 1986; Zolamorgan, Squire, & Ramus, 1994; Stefanacci, Buffalo, Schmolck, & Squire, 2000).

Not all aspects of declarative memory function are damaged, however, as a result of these lesions. Immediate memories for items just brought into consciousness are largely intact, even in patients with extensive damage in the MTL (Drachman & Arbit, 1966). Animal studies have also shown that rats with hippocampal lesions can perform normally on tasks that require memory formation during short delays (~4s). However, when the delays become longer (1 or 2 minutes), performance is impaired (Clark, West, Zola, & Squire, 2001). Thus, it appears that MTL structures are involved in the formation of memories over longer intervals of time. Memory of remote events and facts are also usually spared following lesions, although memories acquired just prior to the lesion are generally lost (Squire, Stark, & Clark, 2004). The observation that remote memories are not lost following lesions suggests that the MTL is not the permanent storage site of memory but that long-term memories must be stored elsewhere. Higher visual areas such as TE have been implicated in such visual memory storage, and it is supposed that the medial temporal lobe works together with area TE to establish these visual memories (Mishkin, 1982; Miyashita, 1993; Higuchi & Miyashita, 1996).

Given the multimodal input into the MTL (as discussed in the last section), the hippocampus may also be involved in tasks that depend on pooling information from various sources. Thus the formation of associative memories (i.e. pairing information about different items, such as connecting a face with a name) has been linked to the hippocampus (Stark, Bayley, & Squire, 2002; Stark & Squire, 2003; Wirth, Yanike et al., 2003). Consistent with this role of the MTL in establishing associative memories are findings that report that MTL structures participate in the recall of visually associated information and that lesions to these structures abolish the ability of TE neurons to represent associations between stimuli (Sakai & Miyashita, 1991; Higuchi & Miyashita, 1996).

5.1.5 Visual responses in the MTL

In accordance with the fact that the MTL receives high-level visual input, stimulus specific responses to complex stimuli have been observed in these structures. In monkeys, about 66% of cells in the perirhinal cortex respond selectively to specific visual stimuli while in the entorhinal cortex 12% of cells are stimulus selective (Miller, Li, & Desimone, 1993; Suzuki, Miller, & Desimone, 1997). (This difference in the proportion of selective responses could be related to the fact that the perirhinal cortex receives stronger direct visual input from area TE compared to the entorhinal cortex, as discussed in Section 5.1.3). fMRI studies in the human parahippocampal cortex have revealed that this region responds specifically to visual images of buildings, scenes, and spatial layouts (Aguirre, Detre, Alsop, & Desposito, 1996; Epstein & Kanwisher, 1998; Epstein, Graham et al., 2003).

Relevant to our purposes, previous single unit recordings in humans have also reported visual responses to stimuli such as faces and objects in the MTL, and the nature of these responses tends to be fairly complex (Fried, MacDonald et al., 1997; Kreiman, Koch et al., 2000a). For instance, in the study by Fried et al., some neurons in the hippocampus and entorhinal cortex responded differentially to faces depending on the conjunction of facial expression and gender of the face. In other words, these neurons signaled an association between these different attributes. In another instance of complex representations, Kreiman et al. (2000) reported that visually responsive neurons in the MTL could base their responses on broad category level descriptions of stimuli. Thus, they observed an individual neuron in the entorhinal cortex that responded selectively to all images of animals that were presented, but not to any of the seven other stimulus categories used (i.e. emotional faces, objects, cars, spatial layouts, patterns, famous faces, and drawings of famous faces). Another neuron in the anterior hippocampus responded strongly to all drawings of famous faces, and to a lesser extent to pictures of famous faces, but not to the other categories of stimuli. Overall, 12% of the recorded neurons exhibited similar category selective behavior. Distinguishing images based on their category is clearly a high-level semantic distinction, and these results suggest that rather than base their responses on low-level similarities between different images, MTL cells can encode abstract links between visual stimuli.

In another remarkable study, Kreiman and colleagues (Kreiman, Koch, & Fried, 2000b) also reported that MTL neurons were not only responsive to visual stimulation, but could also signal imagery or recall of the stimuli they represented. Thus when subjects were prompted to imagine a particular stimulus they had previously viewed, the corresponding neuron would show enhanced activity, even though the relevant image was not physically presented. Importantly, in 88% of the neurons that fired selectively during both imagery and recall, these two selectivities were comparable. These results provide further evidence for the observation that MTL neurons encode high-level, conceptual information since their responses were modulated even in the absence of visual input.

In light of these high-level visual responses and the putative role of the MTL in maintaining associations between different stimuli, it is interesting to ponder how abstract representations between various items are organized in the MTL structures. Do individual neurons carry explicit information about several aspects of a particular concept so that they could respond across different expressions of this concept? In other words, are responses invariant to different stimulus manipulations as long as the represented concept is left unmodified? Or is such information present only over distributed networks of neurons? For instance, does the neuron in my brain that codes for my dog do so for several different views of my dog? Or is it tuned for just one particular image? What about the dog’s name? Can this semantic information also be represented in the same neuron? Or will all the different information be spread out over a larger neuronal population? This is the question that we aim to address in this chapter. As is described below, by presenting our patients with several different pictures of individuals, landmarks, and objects, we tried to determine whether MTL neurons encoded information in an invariant manner.

5.2 Methods

5.2.1 Patients and recordings

The data reported here was obtained in 8 patients (8 right-handed, 3 male, 17 to 47 years old) with pharmacologically intractable epilepsy. As described in Section 5.1.1 and 5.1.2, these patients were implanted with chronic depth electrodes and were monitored in the hospital ward for 7–10 days, during which period this study was performed. The electrodes were located in the hippocampus, amygdala, entorhinal cortex, and parahippocampal gyrus (see Table 5.1 for summary of electrode locations).

Our data acquisition system was obtained from Neuralynx, Tucson, Arizona, and allowed us to record data from 64 microwires (in 8 macro-electrodes as discussed earlier) at a time. In 6 of the 8 patients, all 64 microwires were used. The remaining 2 patients were implanted with only 7 macro-electrodes for clinical reasons. The data from the active channels was amplified, usually with an amplification factor of 10,000, although it did change occasionally depending on the quality of the signal. The sampling frequency was always 28 kHz. The data was band-pass filtered by hardware between 1 and 9000 Hz. This continuous broadband signal was then digitally acquired and stored (using the Neuralynx Cheetah software).

5.2.2 Spike sorting

The full details on the spike-sorting procedure are described elsewhere (Quiroga, Nadasdy, & Ben-Shaul, 2004). A briefer description is presented here. Two steps were involved in the processing of the broadband data from the Neuralynx system before we obtained the action potentials. Firstly, from the continuous signal that contained both spikes and background electrical activity and noise, all the spikes that occurred had to be detected. Spike detection was achieved by band-pass filtering the signal between 300 Hz and 1000 Hz and marking all crossings of an amplitude threshold. The threshold was proportional to the median of the signal divided by a constant value. All events in the band-pass filtered signal that crossed this threshold were labeled as candidate spikes, and the times at which the maximum of these events occurred were recorded.

In the second step, the data had to be clustered to separate out spikes that corresponded to different neurons (recorded from the same microwire) or to artifacts. For the purposes of clustering, a different band-pass filter, between 300 Hz and 3000 Hz was necessary. The primary reason for changing this filter setting was to be able to get better resolution for observing the spike shapes. In other words, while the 300–1000Hz filter allowed us to reject all higher frequency artifacts, it also smoothed out the data excessively, which prevented effective clustering (since clustering is based on differences in spike shapes). Thus to get better resolution, the original signal was re-filtered, and datapoints around the spike timestamps (determined in the previous step) were stored for further analysis. For each spike, 64 datapoints (encompassing about 2.3 ms of data; 20 points before and 44 after the maximum value of the spike) were saved. All spikes were aligned to each other such that their maximums occurred at data point 20. Spike maxima were determined from cubic spline interpolated waveforms 256 samples in length. Once the spike shapes were obtained, clustering based on the wavelet transform and super-paramagnetic clustering was performed as described in (Quiroga, Nadasdy et al., 2004). Briefly, for each spike, 64 wavelet coefficients were obtained, and the 10 coefficients that best accounted for the different spike shapes were fed into the clustering program. The clustering algorithm was based on super-paramagnetic clustering, a stochastic clustering procedure, which groups the data into clusters as a function of a parameter called the temperature. The temperature determines the probability of different spikes belonging to a particular cluster. At low temperatures, the probability of spikes being clustered together into one cluster is high whereas at very high temperatures, the clusters can “fragment” and several clusters with a small number of spikes each can form. At some optimal temperature between these two extremes, spikes corresponding to different neurons will form distinct clusters.

As shown in Figure 5.3, the temperature was a parameter that could be changed by the experimenter if the automatic clustering results were not satisfactory. About 20% of the time, we modified the output of the automatic (i.e. unsupervised) clustering program by i) combining clusters from two different temperatures by decreasing this parameter; or ii) splitting existing clusters into two or more smaller clusters by increasing the temperature; and/or iii) assigning membership of unclustered spikes to nearby clusters via a template matching strategy based on the Euclidean distance between the center of each cluster and the unclustered spike.

Once this procedure was complete, we classified each cluster into single or multi-units. Multi-unit clusters were those that reflected the contribution of two or more neurons to the cluster and whose spikes could not be further differentiated due to low signal to noise ratios. This classification into single and multi-units was done visually based on the following criteria: i) the spike shape and its variance; ii) the ratio between the peak of the spike and the noise level; iii) the presence of a refractory period for single units such that less than 1% of the spikes occurred within 3 ms of each other. Additionally, for both single- and multi-units, the ISI distributions were examined for peaks at 16.6 ms that correspond to 60Hz noise.

The results of clustering data from one microwire are shown in Figure 5.3. The blue cluster corresponds to a multi-unit, and the other 3 clusters correspond to single units. Over all 8 patients, a total of 998 units were recorded, of which 346 were classified as single units and 652 as multi-units. On average 47.5 units (16.5 single units, 31 multi-units) were obtained in a single recording session.

5.2.3 Stimulus presentation and behavioral task

Subjects lay in bed while stimuli were presented to them on a laptop Macintosh OS9 G3 powerbook. Each stimulus was presented at the center of the screen for 1 s. The size of the stimuli was about 1.5 degrees of visual angle. Each image was presented 6 times in a pseudo-random order, and the subject had to press either one of two keys (‘Y’ or ‘N’) to report whether the presented image was of a human face or not. This simple task ensured that subjects paid attention to the images.

The pictures were obtained from the web and consisted of images of animals, famous and non-famous places and buildings, and faces of famous and non-famous people. These categories of stimuli were chosen based on previous reports of visual responses in the MTL (Kreiman, Koch et al., 2000b, 2000a; Kreiman, 2002) and also the observations made by other researchers about the nature of visual responses in high level areas. In particular, since the hippocampus and parahippocampal areas have been implicated in navigation and the processing of spatial layouts and buildings, we included pictures of this type (O'Keefe & Dostrovsky, 1971; O'Keefe & Conway, 1978; Muller, Kubie, & Ranck, 1987; Wilson & McNaughton, 1993; Aguirre, Detre et al., 1996; Epstein & Kanwisher, 1998; Epstein, Harris, Stanley, & Kanwisher, 1999; Burgess, Maguire, & O'Keefe, 2002; Epstein, Graham et al., 2003). Similarly, since several studies in monkeys have reported that cells in V4 and IT respond to highly familiar objects such as paperclips and other stimuli that the animals had been exposed to over a long period of time (Miyashita & Chang, 1988; Logothetis, Pauls, & Poggio, 1995; Kobatake, Wang, & Tanaka, 1998; Yang & Maunsell, 2004), we included pictures of current celebrities who patients (especially around Hollywood!) would have been familiar with. Based on the reports by Perrett, Rolls, and colleagues that cells in monkey IT and amygdala respond to different faces, we included images of non-famous faces (Rolls, 1984; Leonard, Rolls, Wilson, & Baylis, 1985); (Perrett, Rolls et al., 1982; Perrett, Smith et al., 1984, 1985; Perrett, 1987; Perrett, Hietanen, Oram, & Benson, 1992). Frequently, in discussions with the patients, we were able to find out what their interests were, or what their favorite movies were, etc., and to include relevant pictures based on these conversations. Figure 5.4 represents the overlap in images presented during the various screening sessions for all eight patients.

5.2.4 Screening/testing experiment format

A large proportion of our success in this project was due to a new experimental format that we implemented. In this format, we had two separate sessions, which I will refer to as the screening and testing sessions. In the screening session, we presented subjects with an average of 88.6 (range: 70-110) images from the categories described above. The data obtained from this session was rapidly analyzed, and the stimuli that elicited a response from any of our cells were selected. It should be noted that for this analysis, we did not cluster the data into separate units. Instead, we simply detected the spikes and looked at the multi-unit data obtained from each microwire. Typically, this analysis took about 3–5 hours. Once the selective stimuli for the cells were determined (typically only 3.1% (range: 0.9%–18.0%) of the pictures in a session elicited a response), we prepared the testing session. In this session, we presented between 3 and 8 different views of the stimuli that were selective from the screening session. All the images that elicited a visual response in the screening sessions were included in the testing sessions. Thus the screening session served as a screen for later sessions since, based on their results, we knew the selective stimuli and the responsive cells every testing day. In the testing session, in addition to presenting multiple views of selective stimuli, we also presented a random number of stimuli that did not elicit selective responses from any of the cells. The total number of stimuli presented was determined by the time available with the patient for a single recording session (~30 minutes). We had a total of 21 testing sessions.

5.2.5 Determining visual responsiveness

All trials of the experiment were aligned to stimulus onset (0 ms). The median number of spikes occurring across trials between [300 1000] ms following stimulus onset was determined for each stimulus. Baseline activity was computed as the average number of spikes over all stimuli that occurred in the [-1000 -300] ms interval before stimulus onset. A unit was considered responsive if the activity to at least 1 picture fulfilled two criteria: i) the median number of spikes during stimulus presentation was larger than the average number plus 5 standard deviations of spikes that occurred during the baseline interval; ii) the median number of spikes was ≥ 2. The measure of activity was computed using the median rather than the mean for this experiment to diminish the contribution of outliers.

5.2.6 ROC analysis

The degree of invariance observed for neurons that had selective responses to a given stimulus was analyzed by means of an ROC analysis (Green & Swets, 1966). The idea was to determine how selectively a neuron would respond to *different* pictures of a particular individual (for instance) but not to any other pictures that were presented. This was accomplished by measuring the number of responses to this individual (hits) and the number of responses to other stimuli (false alarms) using different thresholds (T) for the prediction. A response was said to occur when the neuron’s activity crossed the threshold (here the activity was defined as the median number of spikes that occurred over all trials on which a particular picture was presented). The hits were defined as the number of responses to an individual divided by the total number of pictures of this individual. The false alarms were defined as the number of responses to other pictures divided by their total number. By setting different values of the threshold T for counting a response, the number of hits and false alarms could be determined. Thus, at very high thresholds, activity to the individual and to the other pictures would be less than the threshold, and therefore neither hits nor false alarms would occur (lower left hand corner in the ROC plots shown in Figures 5.5–5.9). As the threshold would be decreased, the number of hits and false alarms would change. For a cell that responded exclusively to pictures of a particular individual, hit rates of 1 and 0 false alarms would be observed as the threshold was lowered. This would correspond to steep increases in the ROC curve. On the other hand, for units that responded to a random selection of pictures, similar numbers of hits and false alarms would be observed that would correspond to an ROC curve close to the diagonal. The degree to which a unit could be considered invariant was defined on the basis of the area under the ROC curve. For highly invariant cells, the area under the ROC curve would be close to 1, whereas cells that responded randomly would have areas of about 0.5.

For each cell, we compared the ROC curve obtained as described above with 99 surrogate ROC curves. The surrogate curves were each obtained by randomly choosing a set of *n* different pictures and comparing the hit rates obtained for this set of pictures with the false alarms obtained for the remaining pictures. The value for *n* was identical to the number of pictures of the individual for which invariance was tested for this cell. If a cell was selective for only pictures of a particular individual, then the area of the surrogate curves would be much less than 1. A unit was considered to be invariant to an individual or object if its area was larger than the area of all 99 surrogate curves with $p < .01$.

The ROC analysis could also have been performed by considering responses on a single trial (rather than taking the median across trials). In this case, hits would correspond to the proportion of trials (on which the individual was presented) on which a response occurred. The false alarms would correspond to responses on trials on which other pictures were presented. Similar results were obtained using both ROC measures[9].

A one-way analysis of variance (ANOVA) also yielded similar results. In particular, we tested whether the distribution of median firing rates showed a dependence on the factor identity (i.e. the individual, landmark, or object shown). The different views of each individual or object were the repeated measures[10]. The ANOVA analysis, however, does not explicitly show how invariant the responses were across different views of the individual or object. On the other hand, the ROC analysis explicitly tests the presence of an invariant as well as sparse representation.

The ROC and ANOVA analyses were performed by Rodrigo Quian Quiroga.

---

[9] As is discussed later (Section 5.3), 52 of 137 responsive units showed invariance as defined by means of the ROC analysis computed over the median across all trials. On a single trial level, 56 units were found to be invariant.

[10] According to the ANOVA analysis, 50 units had a significant effect for factor identity with $p<.01$.

5.3 Results

The response of a single unit in the left posterior hippocampus to 30 out of the 87 images presented is shown in Figure 5.5. The cell did not show a statistically significant response to the other pictures. The unit fired selectively to very different images of the actress Jennifer Aniston but not to other famous or non-famous faces, landmarks, objects, or animals. It is interesting to note that this cell did not respond to pictures where Jennifer Aniston appeared with the actor Brad Pitt. The response to Jennifer Aniston was characterized by a sharp burst of 5 to 10 spikes between 300 and 600ms post-stimulus. During the pre-stimulus interval, the cell was almost silent (average of 0.02 spikes in the baseline interval). Figure 5.5b shows the results of the ROC analysis obtained for this cell. The red line corresponds to testing invariance for all seven pictures of Aniston, while the grey lines are the ROC values for the surrogate curves (see Section 5.2.6). As expected from the data shown in Figure 5.5a, the surrogate ROC curves are close to the diagonal while the curve testing for invariance to Aniston has an area of 1.0. Thus this unit fired selectively only to pictures of Aniston and not to a random assortment of stimuli.

Figure 5.6 shows the activity of another selective unit in the right anterior hippocampus. The preferred stimuli of this cell were different images of the actress Halle Berry. There are several interesting points to make about this cell. Firstly, this cell not only responded to various pictures of Halle Berry but also to a caricature of her (although not to other caricatures). Given the large differences between the pictures and caricature, this result is striking. Secondly, the cell responded to images of Halle Berry in character as “Catwoman,” her role in a recent film (but not to other pictures of Catwoman). Since in the images of Catwoman, the actor is masked and unrecognizable in her costume, it is likely that this response arises as a result of the patient’s knowledge of the movie. Finally, the cell also responded to the letter string “Halle Berry”! These results clearly suggest that this invariant pattern of responses is not due to low-level similarities between the images. Rather they represent high-level, semantic knowledge about the relationship between the images. The ROC curves for this cell are shown in Figure 5.6b. The area under the curve was 0.99.

The responses of a cell to famous landmarks are shown in Figures 5.7 and 5.8. The unit in Figure 5.7 was located in the left anterior hippocampus and responds to pictures of the Sydney Opera House and the Bahai Temple in New Delhi. Interestingly the patient identified the images of the Bahai Temple to us as the Sydney Opera House. This unit responded to the letter string “Sydney Opera” as well although not to other words (such as “Eiffel Tower” or “Bahai”). The unit in Figure 5.8 responded to several pictures of the Tower of Pisa.

Interestingly, in contrast to the cell shown in Figure 5.5, which only responded to Jennifer Aniston alone, we observed another cell in the same experimental session that only responded to images of her with Brad Pitt. This cell, shown in Figure 5.9, was located in the right posterior hippocampus (the other side of the brain compared to the Jennifer Aniston cell). The ROC area for this cell was 1.0.

Over all 8 patients, we recorded from a total of 998 units. Of these, 137 neurons (67 single and 70 multi-units) were visually responsive to one or more images according to the criteria defined in Section 5.2.5. The proportion of visually responsive cells we observed (13.7%) is comparable to previous reports (14 %, (Kreiman, 2002)). Of the 137 visually responsive cells, 52 (37.9%, 31 single units and 21 multi-units) showed invariance to a particular individual (38 units), landmark (6 units) animal (6 units) or object (2units). Eight of these invariant units were invariant to two different concepts (e.g. two different people, or a person and an object) in that they responded to all the presented views of these two concepts.

In all 52 cases, the area under the ROC curves was significantly higher than that of the 99 surrogate curves as described above ($p < .01$). The median value for the area under the ROC curves over all cells was 0.94, and these values ranged from 0.76 to 1.0. The distribution of the areas under the ROC curves is shown in Figure 5.10.

The locations of the invariant units as a proportion of the number of cells in each area are shown in Figure 5.11. Relative to the number of cells we recorded from in each area, 45% of cells in the hippocampus, 54% of cells in the parahippocampal gyrus, 25% of amygdala cells, and 22% of entorhinal cortex cells showed invariant responses. As we noted earlier, given the anatomical differences between these areas and observations from monkey electrophysiology about the roles of these areas, it is likely that significant differences exist in the nature of their visual responses. However, we do not have sufficient data to make conclusive claims of this nature.

One of the most striking examples of invariance that we observed was the cell that responded to visual pictures as well as the letter string with the name of the individual. In 18 of the 21 testing sessions, we presented patients with letter strings in addition to the different images. Eight of the 127 responsive units in these sessions showed a selective response to pictures of an individual as well as his/her name. Six of these were in the hippocampus, one in the entorhinal cortex, and one in the amygdala.

5.4 Discussion

We have shown that neurons in the MTL can respond in an invariant manner to several different representations of a particular individual. Given the diversity of images we used (pencil sketches, caricatures, letter strings, photographs taken from different viewpoints, with different backgrounds, etc.) it is unlikely that this degree of invariance is based on common low-level features in the images. Instead, our results here strongly suggest that these neurons encode abstract representations of individuals, landmarks, and so on.

5.4.1 The MTL and high-level visual representations

What is the role of the MTL in storing such high-level visual representations? As we have mentioned previously, the MTL is the site at which information from different sensory areas converges. Given this anatomical basis, these structures could thus be prime candidates for combining information maintained at distinct cortical sites. Thus, these structures could link information that differs significantly in its format but with the same content (e.g. the name and the face of a particular individual). Neurons such as the one that responded to both color photographs of the Sydney Opera House and a letter string with its name, or Halle Berry and her name, clearly establish links between pictorial and semantic sources of information. Indeed, as we discussed in Chapter 4, cortical sites that respond to letter strings and words appear to be distinct from those where faces or buildings are represented, and our results suggest that information from these sites can be pooled together in these particular MTL cells.

On the other hand, the hippocampus and related structures can also form conjunctions between events and stimuli that should ordinarily have no relationship with each other (e.g. different contents and potentially the same format). This is compatible with the role of the MTL proposed by Squire and colleagues (Squire, Shimamura, & Amaral, 1989; Squire & Zola-Morgan, 1991). Examples of these responses are cells that fire to both Jennifer Aniston and Brad Pitt (Figure 5.9), or to Halle Berry and Catwoman (Figure 5.6). These are distinct concepts, and the fact that these cells can link them together based on a high-level knowledge of the current affairs of Hollywood illustrates the role of the MTL in associative memory.

This role, however, is probably only temporary as is evidenced by the fact that very remote memories are unaffected by damage to the medial temporal lobe (Squire, 1992). As we discussed previously, the infero-temporal (IT) cortex is believed to be the site where more permanent storage occurs (Miyashita, 1993; Miyashita & Hayashi, 2000; Miyashita, 2004) and indeed it is now well known that complex stimuli such as ours can be represented in IT. The specific contribution of the MTL could be to maintain online but temporary representations for currently important stimuli. Indeed most of our cells were recorded around the time when Jennifer Aniston was on the cover of most popular magazines, when the movie Catwoman was released, or when the patient had recently visited the Sydney Opera House or the Tower of Pisa. Given that these stimuli had been recently encountered by the patients, it is possible, in accordance with the previously developed theories, that the MTL was engaged in establishing novel associations involving those persons and places to facilitate their later storage in IT. This is not to say that persons or places that have been familiar to the patient for several years will not be represented in the MTL. As long as they are currently relevant to the patient's life, current theories would predict that the MTL would need to encode them. However, they would also predict that items no longer relevant (such as a once-famous actor now forgotten by the media) might be stored in other areas but not in the MTL.[11]

---

[11] Thus the MTL structures might serve as a "temporary buffer" for information before it gets consolidated into long-term memory. Reports of patients with lesions to MTL structures could shed light on how long "temporary" could be. Patients with lesions restricted to the hippocampus suffer from retrograde amnesia that extends up to a couple of years (Kapur & Brooks, 1999). Lesions in other MTL structures such as the perirhinal and parahippocampal cortices are possibly involved in establishing memories over longer time scales since patients with lesions in these areas suffer from more severe retrograde amnesia.

5.4.2 Coding schemes

How neurons represent information is a hotly debated topic in neuroscience. One proposed scheme for neuronal representation, known as population coding, relies on the distributed activity of a large number of neurons. According to this hypothesis, individual neurons do not represent a particular object or concept. Rather, the relevant information is broadly distributed over a population of neurons and becomes available through their concerted activity (Georgopoulos, Kalaska, Caminiti, & Massey, 1982; Georgopoulos, Schwartz, & Kettner, 1986; Georgopoulos, 1987). Evidence for this coding scheme comes from neurons such as those found in the motor cortex. In this area, individual neurons are not finely tuned for a preferred direction of movement. Rather they have broad and overlapping tuning curves in three-dimensional space, which makes it impossible to accurately predict the direction of movement from the activity of any one neuron. However, by combining information from a population of neurons, movement directions can be predicted with much higher precision. Specifically, Georgopoulos and colleagues found that movement directions could be accurately predicted by a population code in which each neuron contributed its preferred direction, and this contribution was weighted by the strength of the neuron’s response. Similar distributed codes have also been proposed for encoding continuous stimulus variables such as orientation.

In contrast to this distributed coding format, another strategy could rely on a local or sparser encoding (Konorski, 1967; Barlow, 1972; Thorpe, 1998; Gross, 2002; Olshausen & Field, 2004). According to this scheme, individual neurons play a more decisive role in the representation of an object or concept. In the extreme case, one individual neuron could signal a particular notion: for instance, the representation of my grandmother could rely on a single neuron in my brain. This neuron would be activated whenever I see my grandmother, independent of the viewpoint, the distance, or her emotional expression, etc. Additionally, this neuron would not fire in response to other concepts even if they were related to her (e.g. my grandfather or another elderly woman). This extreme version of the sparse coding scheme has been sarcastically termed the ‘grandmother scheme.’

Nevertheless, cells in higher-level visual areas such as IT cortex do provide support for the notion of highly specific representations. In the 1970s, Gross and colleagues first reported the existence of cells in IT that fired selectively to hands and faces (Gross, Bender, & Rocha-Miranda, 1969; Gross, Rocha-Miranda et al., 1972). Since then, a number of different studies have confirmed and extended these findings and provided further evidence for the fact that individual neurons can specifically represent particular objects (Perrett, Rolls et al., 1982; Rolls, 1984; Perrett, Smith et al., 1985; Rolls & Baylis, 1986; Perrett, 1987; Yamane, Kaji, & Kawano, 1988). More recently, it has been shown that cells can be trained to show a high degree of specificity for the representation of complex objects, such as paperclips or computer-generated fractal patterns (Logothetis & Sheinberg, 1996; Tanaka, 1996).

The coding strategy of the cells that we have presented here seems to be more in line with the latter scheme. We have observed neurons that maintain an invariant response to several views of an object, thereby explicitly signaling the presence of that object. Further, in contrast to population coding, which predicts that during any percept a large number of cells must be active, we find that most neurons do not respond to the majority of images seen by the patients.

This is not to say, however, that a strict, sparse, explicit and invariant coding strategy is implemented by these neurons such that single neurons must code for one and only one percept. Indeed we do find instances of cells that respond to more than one object. Additionally, given the limited time we have for these experiments, we cannot explore a very large portion of the stimulus space. It could very well be that with slightly longer recording sessions, we would find that many of our cells represented several different concepts. Furthermore, on purely theoretical grounds, a pure grandmother coding strategy would be neither feasible nor robust. If every person, object, place, and animal were to be represented by a dedicated neuron, the brain would soon reach the limits of its capacity to encode information. And if the neuron representing the concept of my grandmother disappeared, would I become unable to recognize my grandmother?

So what coding scheme do we observe in the MTL? In our results, we have observed that individual neurons can carry information in a sparse, invariant, and explicit, although somewhat redundant, manner. However, if only for theoretical reasons, this strategy cannot apply to an infinite number of concepts. Rather, it is more likely that sparse, explicit representations like the ones we have observed here are only formed, through repeated exposure, for highly familiar concepts. The processing of less frequently encountered objects must rely on a more distributed coding scheme.

5.4.3 Explicit representations and the Neuronal Correlates of Consciousness

Zeki first suggested the concept of "essential nodes," the idea that specific regions in the brain were responsible for supporting awareness of particular aspects of vision (Zeki & Bartels, 1999; Zeki, 2001). This suggestion was based on studies with neurological patients with whom it was observed that when particular areas were lesioned, consciousness of certain features was lost. Thus for example, a lesion in area V5/MT would result in a loss of motion perception, or a lesion in the FFA would result in prosopagnosia (the inability to recognize faces). Neurons in these areas would be expected to explicitly encode the relevant feature in their firing rates. That is, it should be possible based on the firing rates of these neurons to decide if that particular feature was present in the stimulus. This, in turn, implies that these neurons “should be invariant to those aspects of the input that do not convey any specific information about the feature symbolized” (p. 27 (Koch, 2004)). It has been proposed that such explicit representations could form the basis for the neuronal correlates of consciousness (NCC) (Crick & Koch, 2003). More specifically, these representations could be a necessary condition for the NCC in that if such representations do not exist, the corresponding percept cannot reach consciousness.

The MTL neurons we have presented in this chapter seem to meet the requirements for explicit representations: they specifically respond to particular concepts (such as “Halle Berry” or the “Sydney Opera House”), and because of their selectivity, the presence of these stimuli can be deduced from the firing rates of these neurons (see for example the ROC curves). Moreover, these responses are invariant to most changes made to the pictures of the individuals, landmarks, etc. In particular, these neurons also respond to the written names of these objects. Given these high-level, abstract, and explicit representations, these cells are candidates of choice for participating in the NCC of these notions.

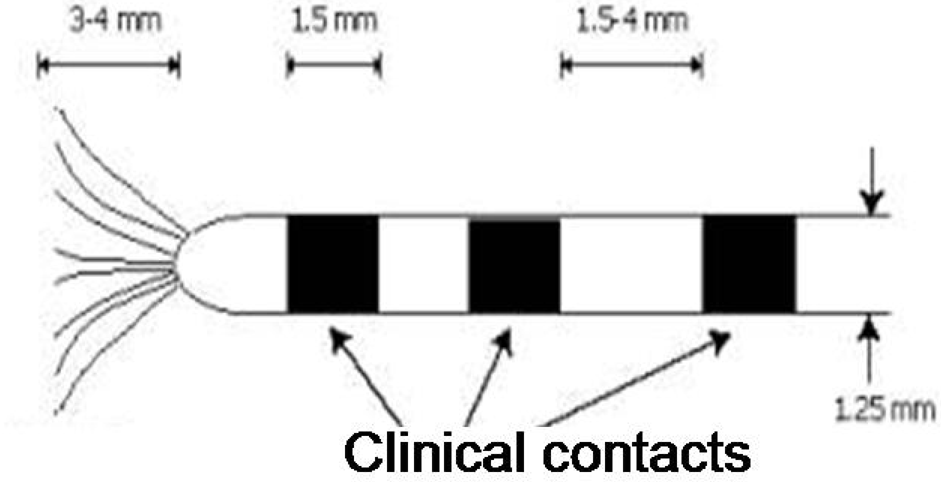


Figure 5.1 An example of the electrodes used. The electrodes were 1.25 mm in diameter. Platinum contacts (1.5 mm in length) along the electrodes were used to collect clinical data. Through each electrode 9 microwires (including a reference) were inserted. The microwires extended 3–4 mm from the tip of the electrode, and lay within a cone subtending 45º. From (Kreiman, 2002).

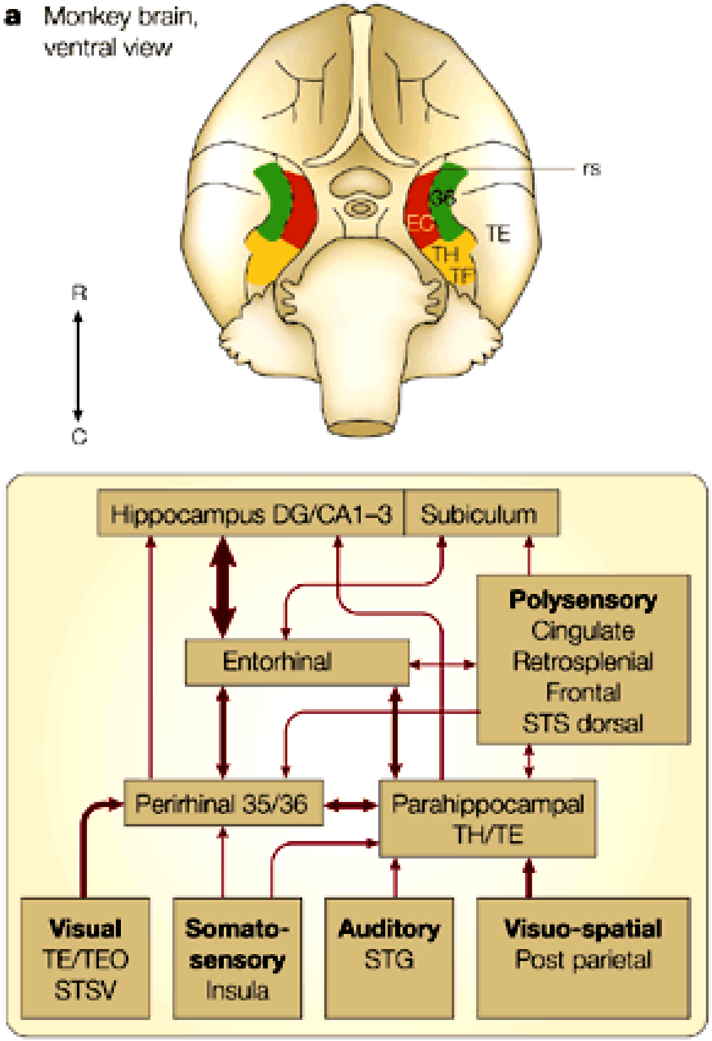


Figure 5.2 Schematic of the anatomy of the MTL in the monkey brain. a) The relative locations of the perirhinal cortex (areas 35 and 36), parahippocampal cortex (TF/TH), entorhinal cortex, and TE in a monkey brain. b) The connection diagram shows the routes by which sensory information projects to the cortical regions and then enters the hippocampus. The thickness of the arrows indicates the size of the projection. Note that several projections are reciprocal. The amygdala is now shown in this figure. (Adapted with permission from Brown & Aggleton, 2001). Copyright, Nature Publishing Group.

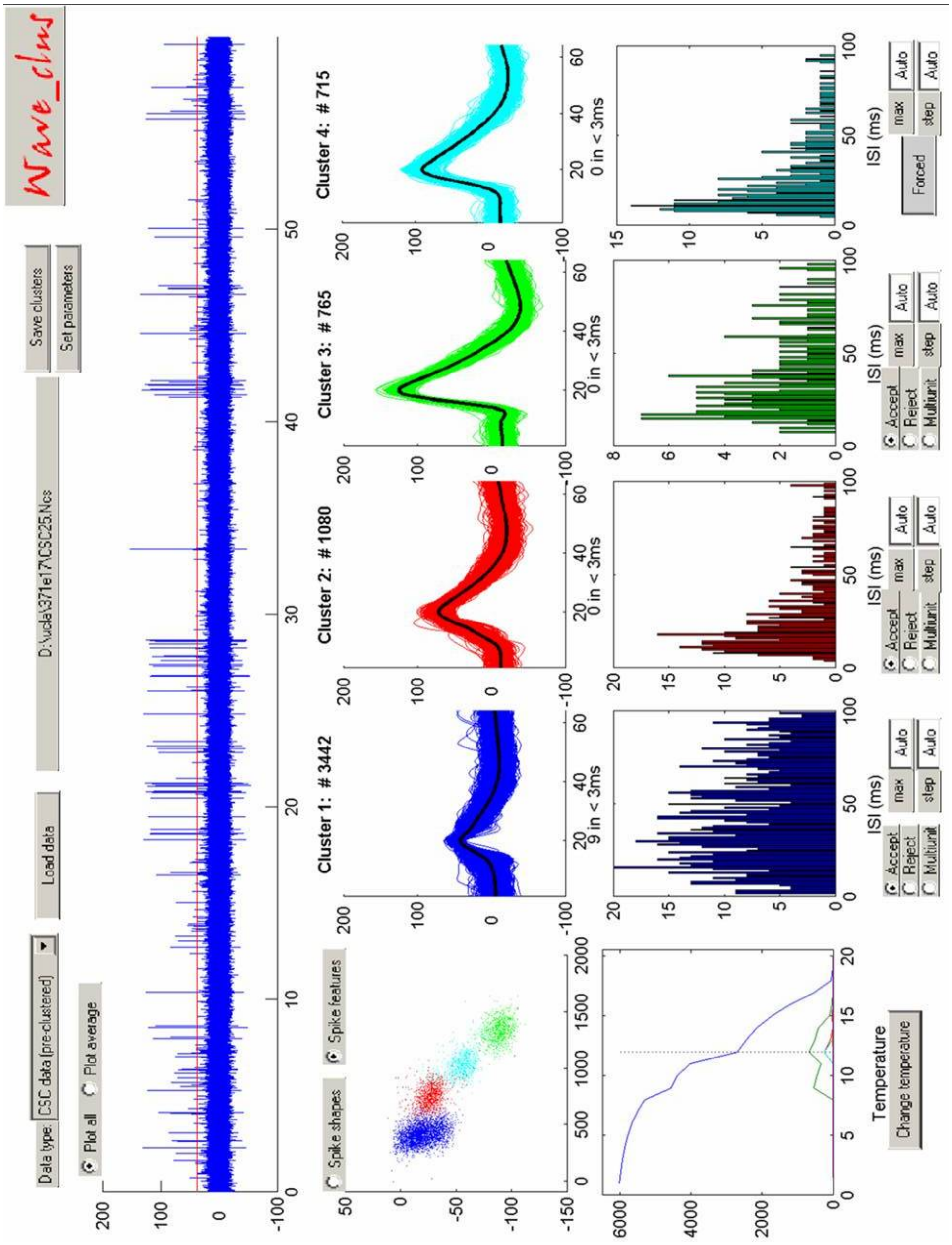


Figure 5.3. The output of the spike sorting algorithm. Upper plot: 60 seconds of continuous data, band-pass filtered between 300 and 3000 Hz. The amplitude of the signal is in μV. The red line represents the amplitude threshold used for spike detection. Middle plots: Distribution of the wavelet coefficients in a 2-D space for four different clusters and the spike shapes corresponding to these clusters. Lower plots: The leftmost plot shows the number of spikes in each cluster as a function of the temperature. The optimal temperature chosen by the algorithm in an unsupervised manner is represented by the dotted line. The other 4 plots correspond to the ISI distribution for the 4 clusters.

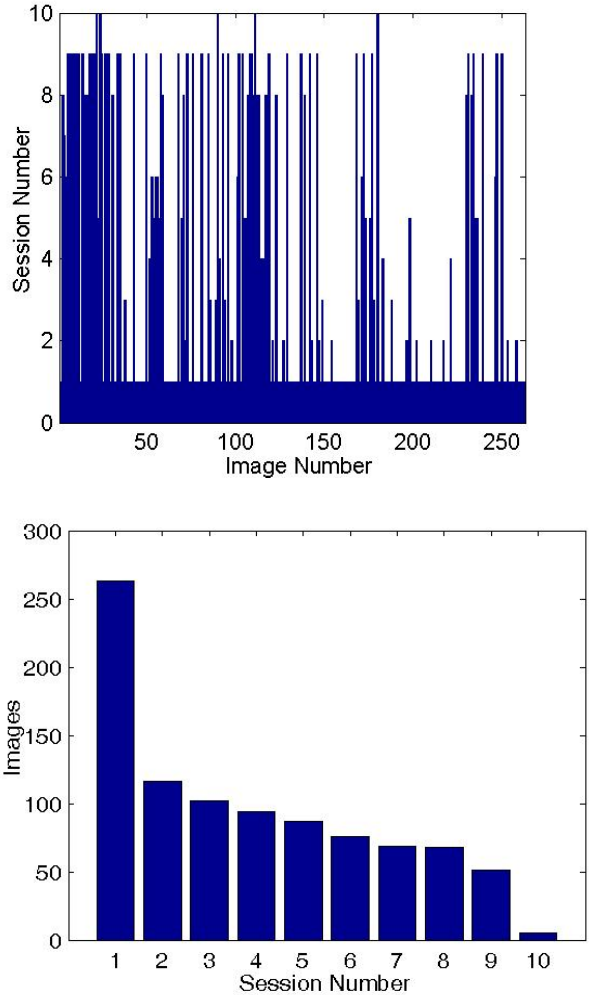


Figure 5.4 Distribution and overlap of images shown in the screening sessions. a) There were 261 unique images (x-axis) presented in 10 Screening Sessions (y-axis). Each bar in the figure represents which session a particular image was shown in. b) The number of images that were shown repeatedly in *at least* X sessions where X = [1 10]. Thus for instance, only 5 images were shown in all 10 sessions, and approximately 55 were shown in at least 9, etc.

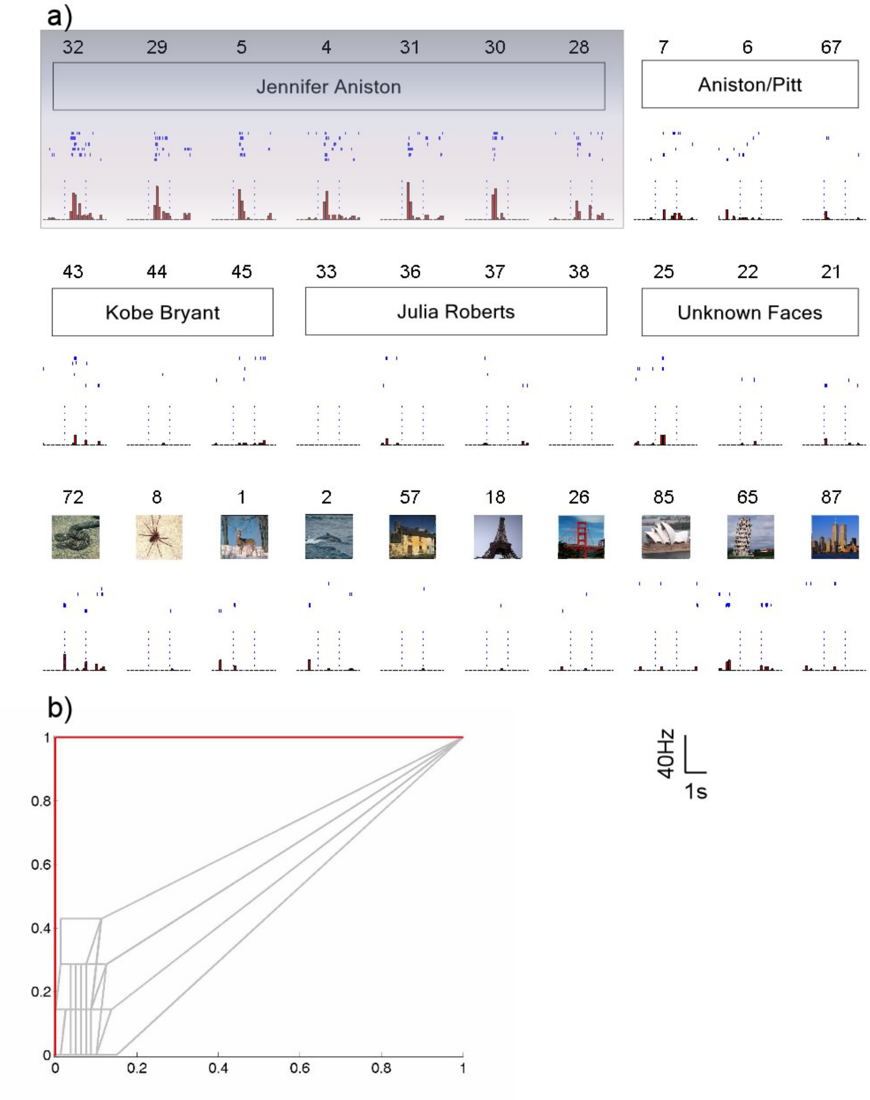


Figure 5.5 Responses of a single unit in the left posterior hippocampus that responds to Jennifer Aniston. a) Responses to 30 out of 87 presented pictures. For each picture presented, the corresponding raster plots (on each of the 6 trials) and post-stimulus time histograms are shown. The vertical dashed lines represent stimulus onset and offset. Stimuli were presented for 1 s. This cell fired exclusively to 7 different pictures of the actress Jennifer Aniston. Note that the cell does not fire to pictures of Aniston with Brad Pitt (images 7, 6, 67). b) The associated ROC curve (red line) and the 99 surrogate curves (grey lines; due to overlap not all 99 curves are visible). The area under the curve is 1.0. Note that in this and other figures in this chapter, we cannot reproduce the original images of celebrities shown to patients due to copyright issues. The white boxes with text in black represent different images of a particular individual that were shown to the patients.

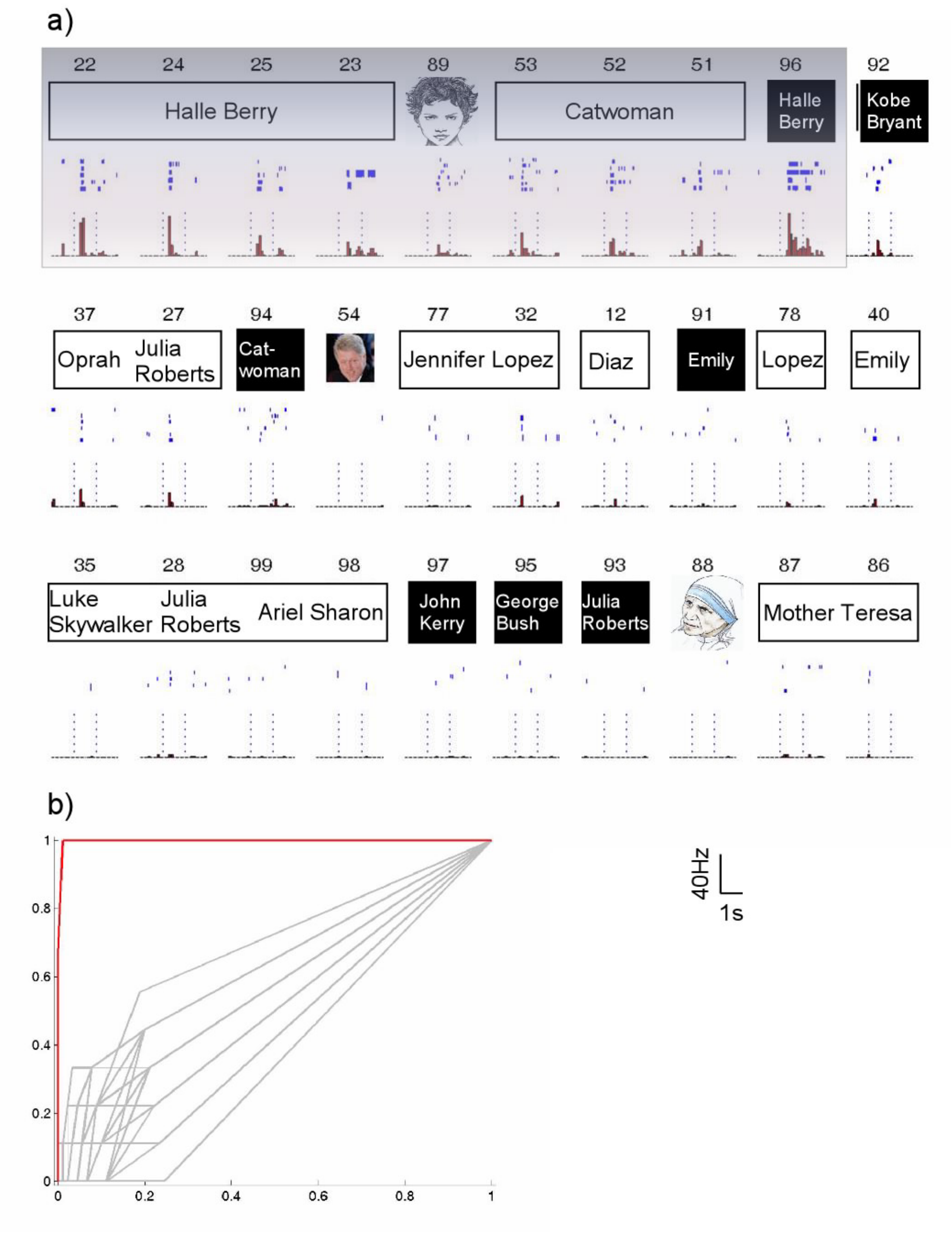


Figure 5.6 Responses of a single unit in the right anterior hippocampus that responds to pictures of Halle Berry. a) This cell responds not only to images of Halle Berry, but also to a caricature of her, images of her dressed as Catwoman, and to a letter string of her name (image 96). All black squares with text in white represent image strings shown to the patient. The format of this figure is the same as above. b) The associated ROC curve (red line) and the 99 surrogate curves (grey lines). The area under the curve is 0.99.

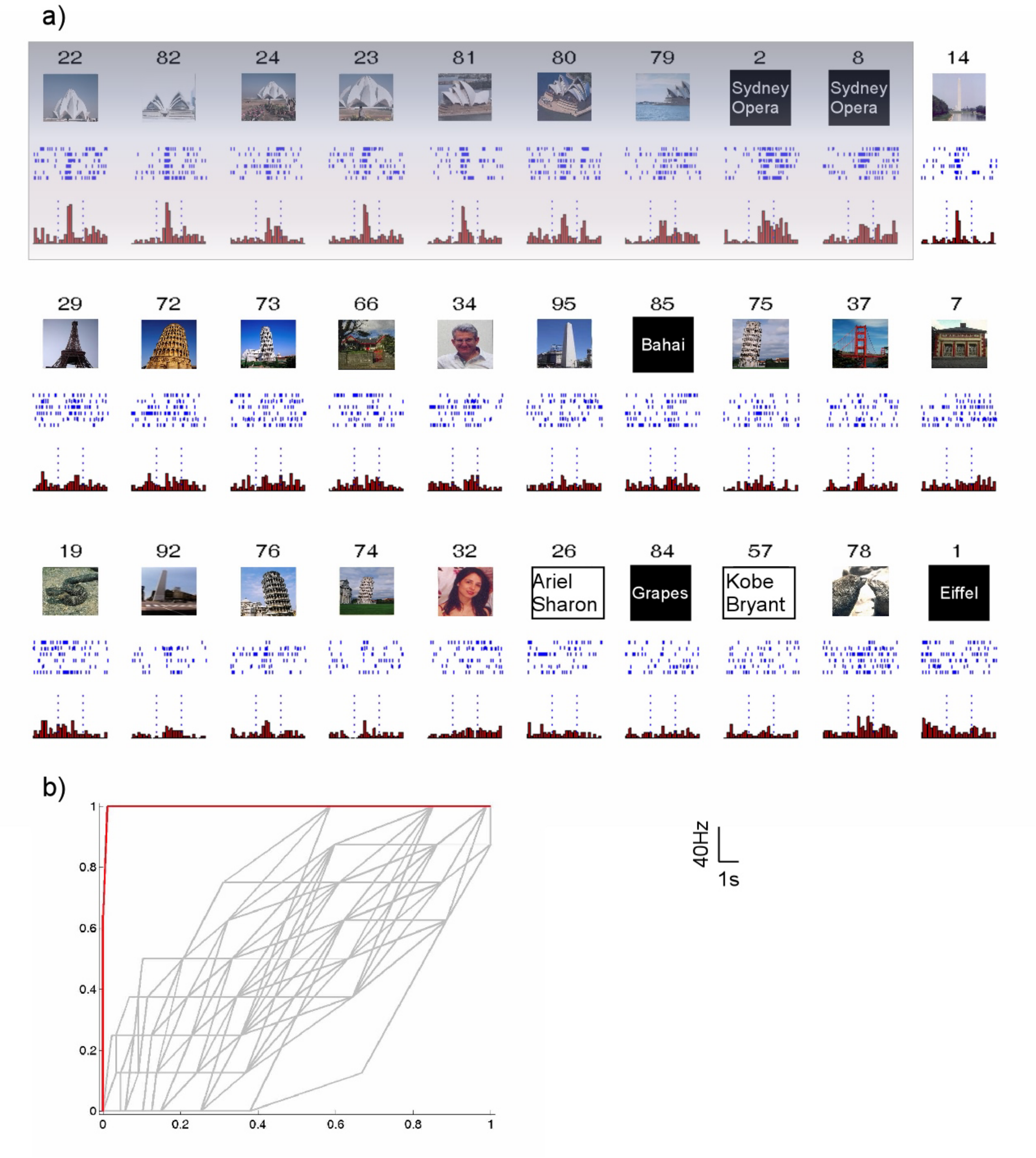


Figure 5.7 Responses of a multi-unit in the left anterior hippocampus that responds to pictures of the Sydney Opera House and the Bahai Temple. a) The format of this figure is the same as above. (The patient identified the Bahai Temple as the Sydney Opera House.) This cell also responded to the text “Sydney Opera” (images 2 and 8). b) The associated ROC curve (red line) and the 99 surrogate curves (grey lines). The area under the curve is 0.97.

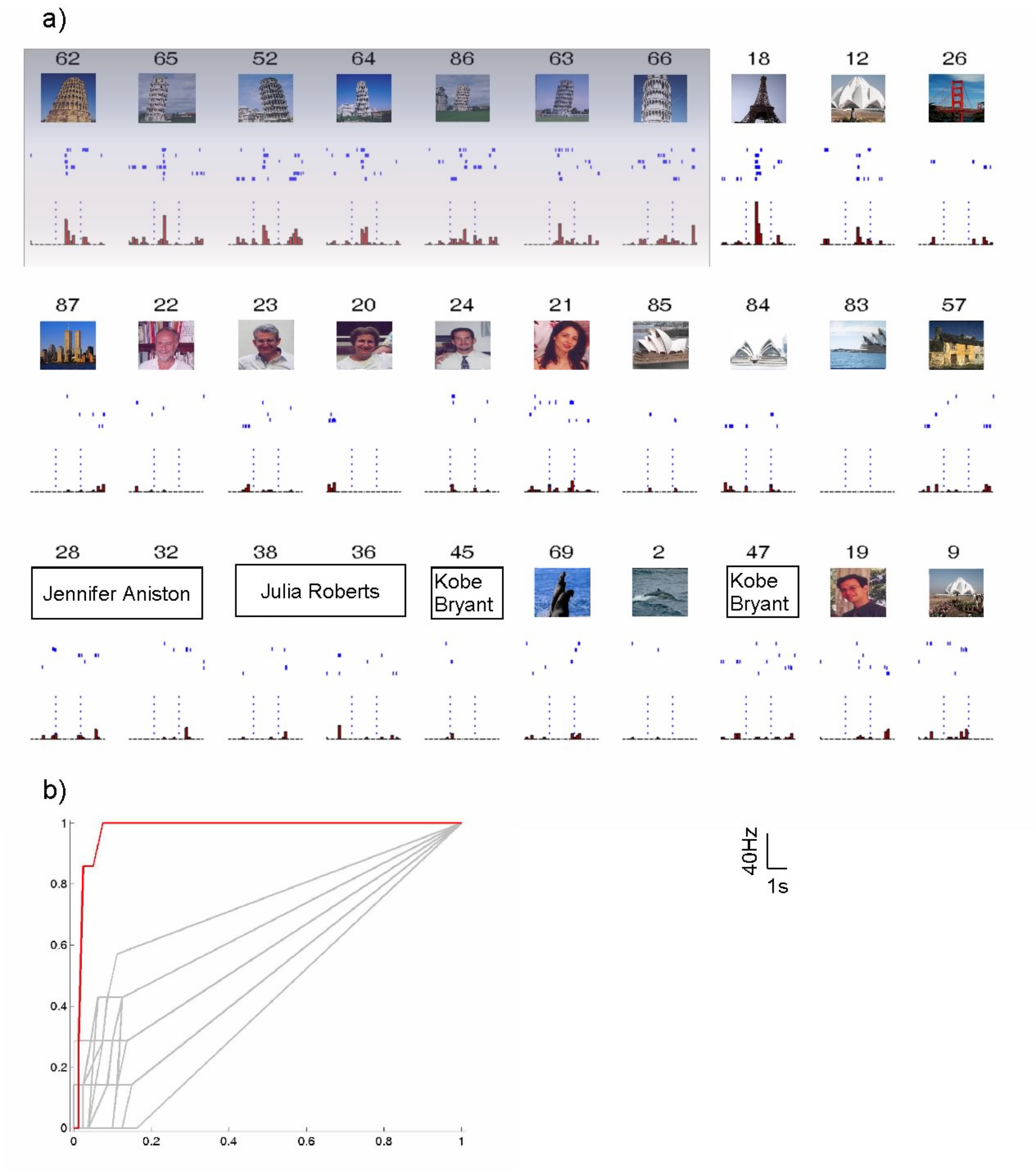


Figure 5.8 Responses of a multi-unit in the right posterior hippocampus that responds to pictures of the tower of Pisa. a) This unit responded to 7 different images of the Tower of Pisa, and also to one picture of the Eiffel Tower. The format of the figure is the same as above. b)The area under the ROC curve is 0.98.

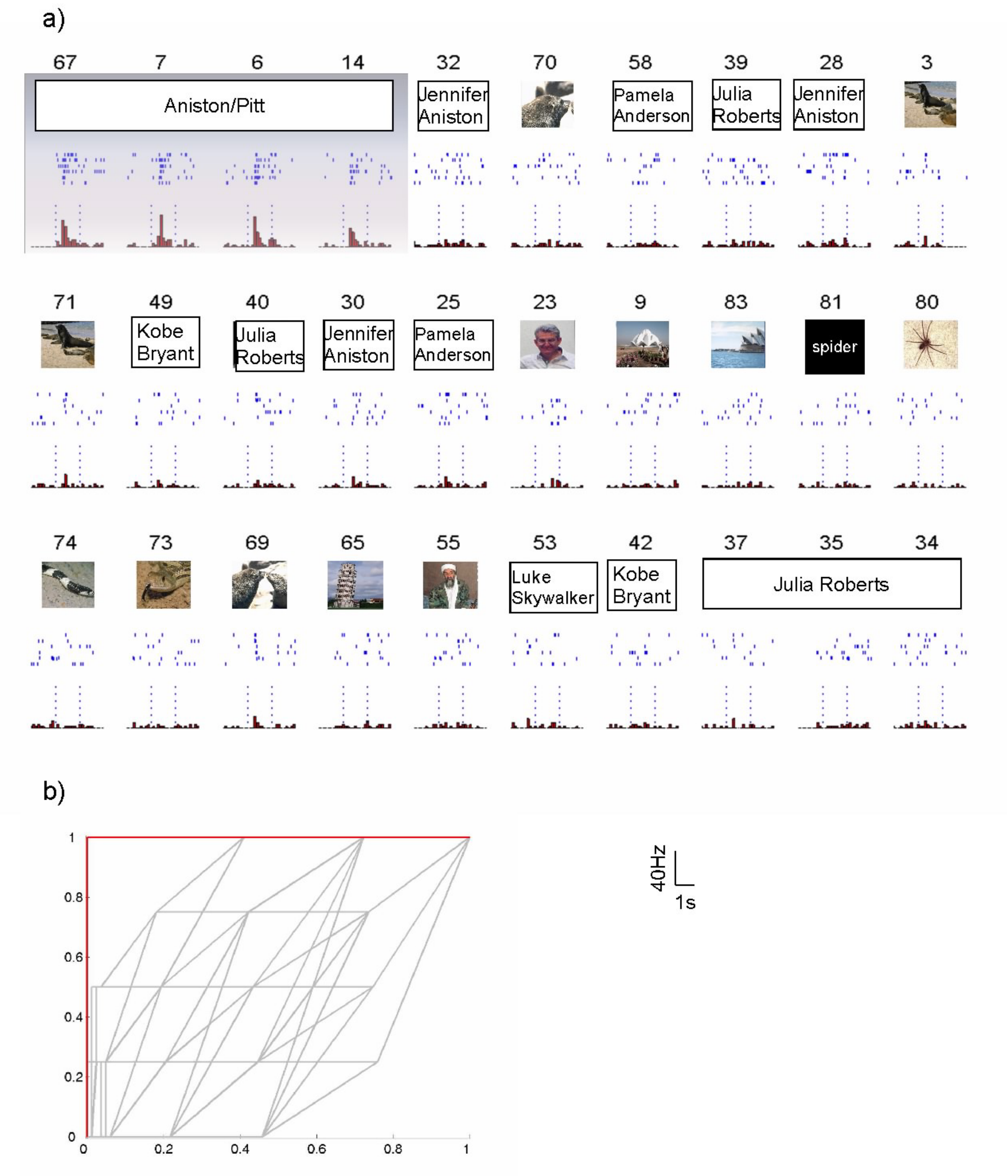


Figure 5.9 Responses of a multi-unit in the right posterior hippocampus that responds to pictures of Brad Pitt and Jennifer Aniston together, but not to images of either of them alone. a) This unit was recorded in the same patient and session as the cell shown in Figure 5.5. The format of this figure is the same as above. b) The area under the ROC curve is 1.0.

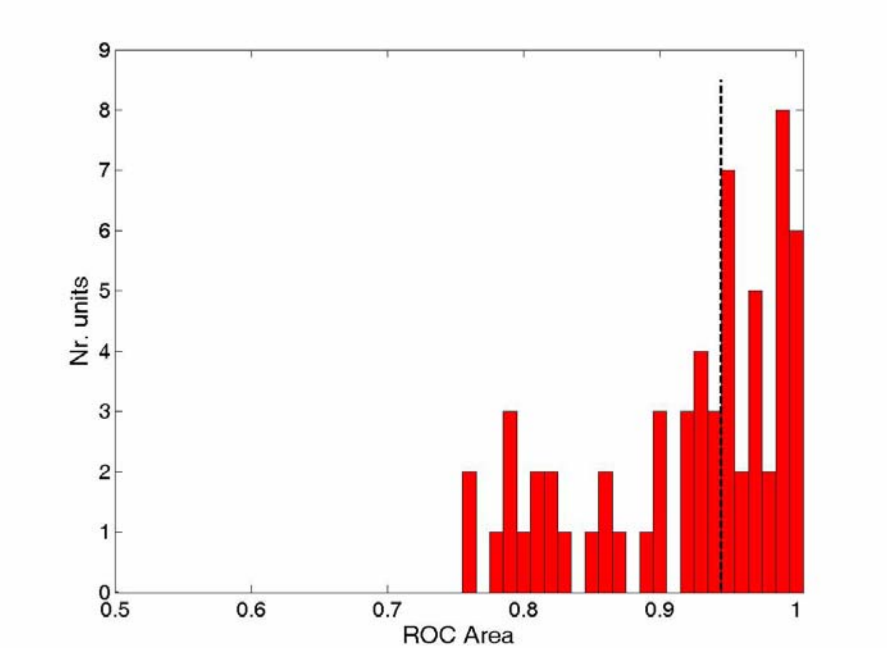


Figure 5.10 Distribution of areas under the ROC curves for the 52 units showing invariant representations. Of these cells, 44 responded to a single individual or object and 8 to 2 individual or objects. The dashed vertical line marks the median of the distribution (0.94).

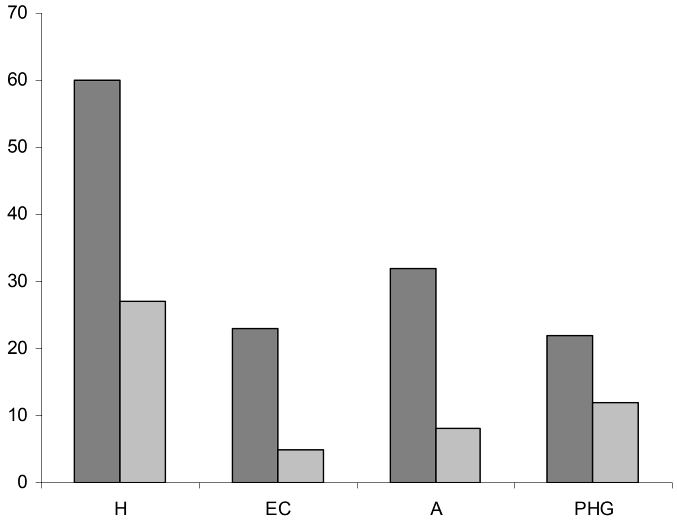


Figure 5.11 Location of the responsive and invariant cells. The dark grey bars correspond to the visually responsive cells found in each brain area. The light gray bars correspond to the number of invariant units. H is the hippocampus, EC entorhinal cortex, A amygdala, and PHG the parahippocampal gyrus. Thus for instance, 60 of our visually responsive cells were located in the hippocampus, and of these 27 were invariant to particular individuals or buildings, etc.

Table 5. 1 Location of electrodes during screening and testing sessions

a) Location of Electrodes during Screening Sessions

Amygdala	17
Entorhinal Cortex	12
Hippocampus	18
Parahippocampal Gyrus	6
Orbito-frontal Cortex	5
Occipital Cortex	4

b) Location of Electrodes during Testing Sessions (Note: there were more Testing sessions, hence the comparatively higher number of electrodes in each location).

Amygdala	38
Entorhinal Cortex	29
Hippocampus	45
Parahippocampal Gyrus	15
Occipital Cortex	6
Transverse Gyrus	8
Anterior Cingulate	2
Supplementary Motor Area	1

Chapter 6

Change Blindness in the Human MTL

6.1 Introduction

In the earlier chapters of this thesis we have addressed the degree of visual processing that can occur outside the focus of attention. However, as has already been pointed out, under some circumstances when attention is not directly involved, we can be unaware of certain aspects of our surroundings. On a regular basis we surprise ourselves by how frequently we fail to notice something that was “right in front of our eyes,” even when we were actively looking for it. This phenomenon, as annoying as it may be, can however serve as an interesting tool to probe visual perception and attention. Consider a familiar situation—you are desperately looking for your car keys on a kitchen table that is long overdue for a good tidying up. All through your frantic search, the keys are before you, but for some interval of time, you simply cannot catch sight of them. Then, all of a sudden your search ends as you spot them, lying in a pile of clutter. Throughout this process, the visual input to your eyes has been constant but your perceptual state changes at some point. That is, somewhere in your brain some group of neurons has changed its response in order to signal the detection of the keys. As we discuss below, visual attention is a key factor that brings about this change. Thus, if one has access to the neurons that encode keys, then the effects of attention on neuronal responses can be directly investigated by comparing responses between the two perceptual states.

In this chapter, we use a paradigm that demonstrates this limitation in perception, outside the focus of attention, to study the effects of attention on neuronal responses in our patient population. We record from single neurons in the human MTL that respond selectively to certain stimuli, and we compare their responses in two conditions: when patients noticed changes made to these stimuli and when they did not.

6.1.1 Change Blindness—another attention paradigm

Subjective experience suggests that we are fully aware of our environment. However, as is often the case with subjective opinions, this is a misleading notion—under some circumstances, we can fail to notice changes that occur in a scene. Although it is difficult to demonstrate this phenomenon in real life where unexpected changes are rare, and almost always attract attention, it can be simply verified in the laboratory under controlled conditions. Movie makers are also well acquainted with this phenomenon, and to some extent it can even be argued that this failure in vision makes their job a little easier. For during moviemaking, different sequences in a scene are not always filmed in order, and consequently a sequence filmed today might be continued tomorrow. Although filmmakers go to extreme lengths to make sure that the details of a scene are as identical as possible across camera cuts, failures do sometimes occur (even in such classics as ‘The Lord of the Rings’ where, amongst other things, in the scene when Aragorn is talking to the dying Boromir, in some shots Boromir’s right hand is gripping Aragorn’s shoulder, while in others it is not; see http://www.movie-mistakes.com). If vision were so finely tuned as to always catch these inconsistencies, we would probably not enjoy our classics as much as we do. However, fortunately for the entertainment industry we can often be blind, even while seeing!

This failure of the visual system to notice changes made to scenes was discussed in the introduction to this thesis, and is known to vision scientists as “change-blindness” (Simons & Levin, 1997; Rensink & Simons, 2000; Simons, 2000; Rensink, 2002) (Figure 1.1). Striking demonstrations of this rather troubling aspect of vision show that the appearance of bizarre figures in a scene (like a gorilla) (Simons & Chabris, 1999) or the disappearance of entire objects (Rensink, O'Regan et al., 1997; O'Regan, Rensink, & Clark, 1999) can go entirely unnoticed. Because of its rather intriguing nature, change-blindness has been extensively studied over the years, under various guises. I will discuss a few of these below.

One version of the experimental paradigms used to study change-blindness is contingent on eye-movements made by observers. For instance, in one experiment, Grimes showed observers various pictures, which they were supposed to study for a later memory test (Grimes, 1996). While observers were looking at the picture, freely moving their eyes from one detail to the next, Grimes made a change to the picture during the saccades, taking advantage of saccadic suppression. He found that under these conditions, observers were surprisingly poor at noticing these changes, even when they were significant (such as two people exchanging heads). Other experimenters have used different types of stimuli to study such saccade-contingent changes (Bridgeman, Hendry, & Stark, 1975; McConkie & Zola, 1979) and have come to the same conclusion: observers fail to detect changes made during saccades, except when the change is made at the targeted location of the eye movement (Currie, McConkie, Carlson-Radvansky, & Irwin, 1995).

Saccades are not necessary for change-blindness to occur. Observers can fail to detect changes made to a scene even when they are fixating. In one set of experiments, O’Regan and colleagues showed that changes go undetected when they are made concurrently with some other global disturbance on the screen (O'Regan, Rensink et al., 1999). In their “mud splash” paradigm, they made changes to a picture and simultaneously presented a few, small, high-contrast patterns in different parts of the picture. These patterns, which resembled mud splashes on a car windshield, never covered the location of the change or obscured it in anyway. Subjects were instructed to fixate on each picture and to press a button as soon as they noticed a change, which could be as large as an object or region moving, changing color, or appearing or disappearing. The results showed that when changes occurred at the same moment as the “mud splashes,” subjects often failed to detect changes.

The situations described in these paradigms differ from everyday life in one crucial aspect—under normal circumstances, when changes occur suddenly in a particular place they cause disturbances or transients at the location of change. These transients are readily detected because they draw attention (Posner, 1980; Klein, Kingstone, & Pontefract, 1992) and thereby signal the location of the change. Thus normally, because the transient motion signals associated with changes can attract our attention, we are fairly efficient at detecting changes. Under the experimental conditions described above, however, transients associated with changes are prevented from grabbing attention.

In the saccade-contingent paradigm, this happens because of a phenomenon that occurs during saccades. Rapid eye movements such as saccades blur the visual image on the retina (Burr, Morrone, & Ross, 1994; Castet & Masson, 2000), thereby also blurring the change-related transient signal. This blurring effectively decreases the strength of the transient signal as a result of which it fails to grab attention. This results in change-blindness. In the “mud splash” paradigm, the transient signal itself is not blurred. Instead, in this case, the “mud splashes” themselves generate a number of other local transients all over the scene. Consequently the transient signal corresponding to the true change is just one in a number of such signals that arise and therefore becomes less salient. Attention is not drawn automatically to the true transient signal, and if it focuses on the location of a “mud splash,” change-blindness occurs.

Another simpler way of diminishing the effective strength of the transient signal to attract attention is demonstrated in the flicker paradigm (Phillips, 1974; Pashler, 1988; Rensink, O'Regan et al., 1997). In this paradigm, the original and changed images are presented to observers, but with a brief blank screen shown between them (Figure 1.1). The images are presented to observers repeatedly until they detect the changing object, which can often take more than 1 minute. In each frame, the flicker introduces transients at every location in the scene, and just as with the previous paradigms, this ensures that the transient motion signals at the actual location of change are not as salient, thus hindering change detection. Additionally, other interesting observations have been made with the flicker paradigm, that further illustrate the role of attention in change detection. Firstly, in their experiment with the flicker paradigm, Rensink and colleagues tried to ascertain the influence of attention on change detection by comparing differences in responses to objects that attracted attention with those that did not. To determine which objects attracted attention, they asked 5 naïve observers to briefly describe a number of scenes verbally. The objects in the scenes that attracted attention were those mentioned by more than 3 observers and were defined as objects of central interest (CI). Objects of marginal interest (MI) were those that were mentioned by none of the participants. 10 different subjects were then tested in the change-blindness experiment using these scenes. As expected, Rensink et al., found that these participants demonstrated change-blindness. However, they also noticed that observers took much longer to detect changes made to the MI objects compared to the CI objects. They concluded that “visual perception of change in an object occurs only when that object is given focused attention …” (Rensink, O'Regan et al., 1997).

To provide further support for this claim, in another experiment, Rensink et al. also studied the influence of cueing attention to different locations in the scene. On some trials, subjects were presented with a valid cue that correctly predicted the location of the change. On other trials, the cue was misleading and directed attention to some other location in the picture. Their results showed that on the valid cue trials reaction times to both the MI and the CI objects were greatly sped up, and in fact there was no difference in these times between the two sets of objects. The invalid cues, on the other hand, had no significant effect on the time it took observers to notice the change. Thus when subjects' attention was directed to an object, they were faster in noticing changes to that object, confirming the conclusion that attention is a critical factor in change detection.

These observations made in the different change-blindness paradigms demonstrate the role of attention in detecting changes. In brief, it appears that on those trials when subjects miss noticing changes, they were less likely to be paying attention to the changing object. Conversely, on change detection trials, attention was more likely to be engaged by the changing object. Thus, along with the visual search and dual-task paradigms, change-blindness can also serve in the study of attention.

6.1.2. Attention in the human MTL

In this chapter, rather than using the dual-task paradigm, we study the effects of attention at the single-cell level using the change-blindness paradigm in our patient population. The main reason behind this switch in paradigms is the difficulty involved in running the dual task paradigm in the hospital setting. In the first place, the dual-task paradigm is very demanding on subjects given the short presentation times of the stimuli in the two tasks. For this reason, we run the risk of tiring patients out in a very short period of time. A logistical difficulty also comes about because of the great amount of time required for training subjects and determining their SOAs in the dual-task. On average, subjects take about 10 hours for training, and since we cannot spend 10 hours before surgery with the patient for the sole purpose of training, running the dual-task paradigm in the hospital setting is not practical. For these reasons it was necessary to engage another simpler attentional paradigm, such as the change-blindness paradigm.

As has been discussed in the previous section, a significant amount of work has gone into the study of change-blindness at the psychophysical level. Its neuronal correlates on the other hand have only recently come under investigation. Much of this work has used fMRI to localize the different brain areas involved in the detection of changes (Ress, Backus, & Heeger, 2000; Beck, Rees, Frith, & Lavie, 2001; Ress & Heeger, 2003; Pessoa & Ungerleider, 2004). However, to the best of my knowledge, the single neuron correlates of this processing have not yet been studied. By using the change-blindness paradigm in the patient population, we thus not only examine the effects of attention on neuronal responses, but also investigate the neuronal correlates of change-blindness at the single cell level.

6.2 Methods

6.2.1 Behavioral task

The experimental timeline for 1 trial in the change-blindness paradigm is shown in Figure 6.1a. Each trial began with a fixation cross that was presented for a random interval between 1000 ms and 1200 ms. Following the fixation cross, 4 images appeared on the screen for 1000 ms as shown in the figure. The pictures were presented at 4 locations on a circle with radius of 6 degrees of visual angle. The midpoint of each picture was located on the circle; each picture subtended approximately 1.5 degrees of visual angle. A blank interval of 1500 ms (black screen) followed the first set of pictures, and then a second display of 4 pictures appeared for 1000 ms. In the second display period, the pictures occupied the same location as the previous set of pictures. Additionally, in roughly half the trials, one of the 4 pictures was changed between the two display periods. Patients were instructed to report at the end of each trial if they noticed a change or not. The presentation times for the two display periods as well as the inter-stimulus interval (ISI) were determined beforehand following extensive psychophysics on a non-clinical population. Six subjects were tested on this paradigm, with ISI intervals of 100 ms and 1500 ms. Additionally, different display sizes (2, 4, or 6 images) were also tested. The performance of the non-clinical population was determined with each of these parameters, and the current set of parameters was chosen because it resulted in a hit rate of about 70% on average, thus ensuring some proportion of incorrect trials (Figure 6.2 a, b) .

The stimuli were presented to patients on a Macintosh G3 laptop computer. The laptop was placed on the patient’s lap or a tray table in front of the patient depending on the patient’s preference. Although we could not precisely constrain the viewing distance for each patient because of constraints in the hospital ward, in all cases, the laptop was roughly 60 cm away from the patient. Patients pressed two keys (‘Y’ and ‘N’) on the keyboard to report their responses on each trial.

During the experiment, patients were always instructed to fixate on the central fixation cross and to only covertly explore the pictures in each display period. Although we did not explicitly control for eye movements because of the difficulty involved with introducing the equipment into the clinical ward, there is evidence that suggests that eye movements do not influence neuronal responses in higher-level visual areas. Recordings in area IT in monkeys have shown that eye movements have virtually no effect on the neuronal representation underlying object recognition (DiCarlo & Maunsell, 2000), particularly in plain backgrounds (as opposed to complex natural scenes) (Rolls, Aggelopoulos, & Zheng, 2003). Ringo et al. (Ringo, Sobotka, Diltz, & Bunce, 1994) have shown some modulation in firing rates due to eye movements in the anterior and medial temporal lobes but found no correlation between cells that were modulated by saccades and cells that were visually responsive. Perhaps most importantly, the strongest evidence for our set-up comes from control experiments carried out by Kreiman for his thesis (Kreiman, 2002). In these experiments, conducted in the same hospital setting as our experiments, Kreiman demonstrated that there was no modulation in firing rates of visually responsive cells as a result of eye movements. Based on his experiments and the literature discussed above, he concluded that “these data strongly support the idea that eye movements do not significantly alter the visual representation in MTL” (p.110 in (Kreiman, 2002)).

6.2.2 Stimulus presentations and trial types

In most paradigms that study attention in monkeys (e.g. (Moran & Desimone, 1985; Reynolds, Chelazzi et al., 1999)), for each recorded cell, stimuli that drive the cell well and weakly (preferred and non-preferred stimuli, respectively) are first selected, and then used to investigate the effects of attention on the cell’s response. Similarly, in this experiment, it was necessary to select a set of preferred (usually, only one) and non-preferred (4 to 8 on average) stimuli for each targeted cell. This was greatly facilitated by the screening/testing experiment format described in Chapter 5. Given the data from the screening session, we usually had a good idea of the visually responsive neurons we were recording from and the stimuli that drove these neurons strongly or weakly. Based on this information, we could select the stimuli to use in the change-blindness experiment. As a result of our time constraints, we were limited in the number of preferred pictures we could select for each session, since we wanted to ensure that we had enough repetitions of the change and no-change trials for each preferred picture that we presented. Thus, given the 30-minute recording session on average, and taking the length of each trial into account, this limited us to 4 preferred pictures in each session.

The limitation on the number of preferred pictures also constrained the number of neurons we could target for our study during a single experimental session. In some sessions, the 4 stimuli we chose drove 4 independent neurons that would count towards the total number of neurons in the change-blindness experiment. In most cases though, in the screening session only two or three neurons would have the strongest responses and would respond significantly to several pictures. Accordingly, in these cases, a smaller number of neurons would contribute to the total count. All in all, we recorded 17 change detection sessions in 9 patients. In these sessions, although we had recorded approximately 110 visually responsive neurons (for a definition of responsiveness, see Section 6.3.4) from a total of approximately 740 cells in the corresponding screening sessions, because of the limitations explained previously, only 49 of these neurons were targeted (i.e. had been pre-selected on the basis of the screening session, and preferred and non-preferred stimuli had been determined for these cells). Only these 49 targeted cells were included in our analysis.

The general design of the experiment also mimicked monkey experiments in that, on each trial, only one of the stimuli was selected from one cell’s set of preferred pictures. The other stimuli in each display period were chosen from the group of non-preferred pictures (common to all targeted cells). Based on task demands, from the subject’s point of view, there were only two different trial types—change and no-change trials. However, as Figure 6.1b shows, from the point of view of a particular neuron, 4 different trial types could be determined with respect to its preferred stimulus. For example, consider a neuron that responds selectively to a particular picture of Bill Clinton. In the first type of trial, the picture of Clinton could have been one of the four images in the first display period, but could have been replaced by another picture in the second display period. We will call these types of trials “disappear trials.” In the “appear trials”, the preferred picture of the cell, could have been absent in the first set of four pictures but then have appeared in the second display period. On trials when no change occurred, the preferred picture of the cell would have been present in both display periods—the “both trials”. And finally, the preferred picture could have been absent all together on the “none trials”. In order to study the effects of attention on the cell encoding this picture of Clinton, we will have to consider its activity during all the relevant display periods (where the preferred stimulus was present), that is, the first display period of all the “disappear” trials, the second period of all the “appear” trials, and both display periods of the “both” trials. Over these intervals, to get at the effects of attention on this cell, we will then look at the difference in firing rates between correct and incorrect trials since, as discussed before, the probability of attention being focused on this picture is higher on the correct trials than the incorrect ones. For every preferred stimulus, there were 20 “disappear”, 20 “appear,” and 14 “both” trials over the recording session. Overall for all preferred stimuli, there were at least 50 “none” trials.

6.2.3 Recordings

The procedures for electrode implantation, and data acquisition and pre-processing have been discussed previously in Chapter 5. The current chapter describes data obtained in 9 patients (3 female, all right-handed, 17-47 years old) who participated in a total of 17 sessions of the change-blindness experiment.

6.2.4 Data analysis

Visual responsiveness

As was discussed in Section 6.3.2, for each cell, preferred and non-preferred stimuli were determined based on the results of the screening experiment. In the testing experiment, we verified if the preferred stimulus from the screening session still elicited significant increases in firing rates by computing a paired t-test. This test was uncorrected for multiple comparisons because only one hypothesis was being tested. The t-test compared the distribution of firing activity during the [-1000, -300] ms baseline interval and the [300, 1000] ms interval during which the stimulus was presented over all trials. (0 ms represents the time of stimulus onset.) A cell was considered to have remained visually responsive if the p value of the t-test was < 0.05.

Population responses

Population responses were computed using the normalized spike density function (sdf) (Richmond, Optican, & Spitzer, 1990; Sheinberg & Logothetis, 1997). For each neuron, the sdf was obtained by convolving the spike train on each trial with a 200 ms fixed width Gaussian and then averaging over all trials. The spike trains were binned in 5 ms bins before convolution. For each unit, the sdf was normalized by dividing by its peak activity. The normalized sdf was then averaged over the population of cells. Although, for each neuron the maximum of this normalized sdf was equal to 1, over the population of cells, the maximum value is less than 1 because the activity in different cells peaked at different times. The average responses were computed separately for correct and incorrect trials in all trial types. Cells were included in this analysis in each condition only if there was a minimum of 4 trials in that condition. This method is similar to the one used by Kreiman (Kreiman, 2002).

6.3 Attention in the MTL

6.3.1. Visual responses in a crowded environment

During the change-blindness experiments, we targeted 49 neurons that had been visually responsive in the screening sessions in 9 patients. Of the 49 targeted neurons, 29 were still responsive during the testing sessions. In this section, we will first describe the responsiveness of these 29 cells, and then speculate on why the other 20 neurons appear to have “lost” their visual selectivity.

6.3.1.1 Reduction in the strength of responses

Of the 49 visually responsive neurons from the screening sessions, 29 were still responsive when presented with 4 other images and were further analyzed (see Table 6.1 for the location of these neurons). An example of such a neuron, located in the right amygdala is shown in Figure 6.3. The response of this neuron in the screening session is shown in Figure 6.3a while the response in the change-blindness experiment is shown in Figure 6.3b. In comparison to the response in the screening session, in the testing session the response is much weaker when the same stimulus is placed together with 3 other stimuli known to drive this cell poorly. While in the screening session, the neuron had an average firing rate of about 14 Hz; the response in the testing session was about 7 Hz on average. The response of another unit located in the right hippocampus is shown in Figure 6.4. Over all 29 cells that were visually responsive in the testing sessions, the relationship between firing activity in the screening and testing sessions is shown in Figure 6.5. As the graph shows, there is a substantial drop in the firing activity over the population of neurons in the change-blindness sessions compared to their response in the screening experiments. The slope of the graph is 0.2 signifying that the responses in the testing session were 5 times lower on average.

This significant drop in firing rates observed in the testing session can be accounted for by the competitive interactions that would arise amongst our stimuli. As is well known from monkey electrophysiology, the response of a neuron to a preferred stimulus in a crowded environment is much lower compared to its response to the same stimulus presented in isolation. In these experiments (see Fig 1.5b), a drop in firing rate is observed if even one additional stimulus is placed in competition with the preferred stimulus of the cell (Moran & Desimone, 1985; Sato, 1989; Treue & Maunsell, 1996; Luck, Chelazzi et al., 1997; Missal, Vogels, Li, & Orban, 1999; Reynolds, Chelazzi et al., 1999; Treue & Maunsell, 1999; Chelazzi, Miller, Duncan, & Desimone, 2001; Rolls, Aggelopoulos et al., 2003). In comparison, in our experiment, 3 additional stimuli known to generate a poor response in the targeted neuron were presented along with the preferred stimulus of the cell. Thus, in effect, the environment in the change-blindness experiment was more cluttered compared to these classic electrophysiology experiments, increasing the competitive interactions among stimuli. These interactions between stimuli would result in reduced firing activities.

6.3.1.2 Reduction in the number of responsive cells

Of the 49 cells we targeted from the screening session, only 29 maintained statistically significant responses (i.e. $p < .05$) in the testing session, while the other 20 lost their responsiveness. This loss in the number of responsive cells could simply be a consequence of the decrease in response amplitudes observed in the cluttered environment (as discussed above), which would make the detection of any signal more difficult. Furthermore, as we have discussed in Chapter 1, unattended stimuli in cluttered scenes typically get filtered out. The absence of focal attention on certain stimuli in our cluttered environment could similarly reduce the selectivity of the relevant neurons.[12] Another reason could be the time delay between the screening and testing sessions. In general, we were fairly successful in running the testing session shortly after the screening session. Typically, the delay was on the order of 3-4 hours, the time it took us to analyze the data acquired in the screening session. However, on a few occasions, we had to postpone the testing experiment till much later in the day, or even the next day, generally because the patient did not want to test right away. Of the 17 change-blindness sessions we ran, 7 were tested within 4 hours of the screening experiments, 6 were tested 6 hours later, while the other 4 were tested 12 or more hours later. These delays between screening and testing sessions could also account for the differences in neuronal responses we observed. It is possible that during these delays, an electrode could have moved relative to the skull, and that in the testing session we were no longer recording from the same neuron (for an in-depth discussion on the stability of recordings in our clinical setting, see (Kreiman, 2002)). Additionally, it is not necessarily true that a neuron would always maintain the same selective responses. One would assume *a priori* that a response would be stable over days, but even in our screening sessions we observed neurons that had more plastic responses. For example, there were instances when after a few days of recording, cells developed responses to the experimenters whom the patient had only ever seen in the previous few days.

---

[12] As we have already discussed, on incorrect trials, attention is less likely to be focused on the relevant stimuli.

6.3.2. Behavioral responses

Over the population of subjects during the change-blindness experiments, the average hit rate (i.e. the probability of reporting a change when it occurred) was 77.8 ± 2.0% (mean ± s.e.m.). The average false alarm rate (falsely reporting a change when there was none) was 18.8 ± 2.0% (Figure 6.2c). The average hit rate was close to that obtained for our non-clinical population (see Section 6.2.1), suggesting that despite being in the hospital and under medication, the clinical population performed fairly well on this task. The patients' percent correct was 78.0 ± 1.7% on average.[13]

6.3.3 The neural correlates of change blindness and change detection.

As we discussed in the introduction to this chapter, the question we are interested in here is what happens at the level of individual neurons in the MTL when observers detect a change, compared to when they do not. The response of one neuron in the anterior hippocampus is shown in Figure 6.6a, and another unit in the right amygdala is shown in Figure 6.6b. The responses in each of the 4 conditions are shown in these figures where the histograms in red correspond to the correct trials in each condition and the black histograms reflect the incorrect trials. The dashed vertical lines represent the time intervals of the first display period, the ISI interval, and the second display period, respectively. A similar analysis was conducted for the entire population of responsive cells. The average response over the population in the 4 separate trial types is shown in Figure 6.7. As discussed in Section 6.3.3, this figure shows the mean

---

[13] Note that chance performance was not always at 50% because the design of the experiment included a slightly higher number of change trials compared to no-change trials. On average, chance performance was 56.6 ± 0.9 (mean ± sem).

normalized spike density function averaged over all cells and all trial types, and computed separately for the correct and incorrect trials. The red trace shows the average activity in the correct trials, and the black trace corresponds to incorrect trials. The shaded regions reflect the standard error of the mean. It is evident from this figure that for all trial types, neuronal responses over our population of 29 cells were higher in correct trials compared to the incorrect trials, during the intervals that contained the preferred stimulus. For example, during the “disappear” trials, cell firing rates were significantly higher ($p < 0.001$)[14] during the first, 1s long display period (when the preferred stimulus was presented to subjects) for the correct trials compared to the incorrect ones. Similarly during the “appear” trials, firing rates were significantly higher ($p < 0.01$) in the second display period for the correct trials, and in the “both” trials, firing rates for correct trials were higher during both display periods ($p < 0.05$). These results thus indicate that on correct trials, the preferred stimuli were represented at the neuronal level more strongly than on the incorrect trials, regardless of the display interval they appeared in. On the other hand, during those display periods when the preferred stimuli were not present, no significant difference was observed between correct and incorrect trials ($p > .05$), indicating that the observed results are specific to the selectivity of the neurons and are not due to an increased level of arousal on correct trials. Interestingly, increased activity on correct trials was also observed during the 1.5s long blank interval between the two display periods on the “disappear” and “both” trials that approached significance ($p = 0.06$).[15] In both these types of trials, a preferred stimulus was presented during the first display period, and this increased activity after the stimulus was removed might reflect a memory of the stimulus encoded in the neuronal trace.

---

[14] Significance values were computed using a two-tailed t-test to compare activity on correct and incorrect trials. Neuronal activity was integrated over a 1s interval for the periods when the preferred stimuli were present (starting at 300ms, to take into account the latency in responses).

[15] Significance values were computed as before using a two-tailed t-test to compare activity during the [1000 2500ms] time interval.

Figure 6.8a presents (in a summarized form) the average activity over our population of cells in all intervals when the preferred stimuli were present. In other words, the data shown here is averaged over the first display period of the “disappear” trials, the second display period of the “appear” trials, and both periods of the “both” trials. The red and black traces again correspond to correct and incorrect trials. As mentioned earlier, the results show that firing rates were significantly higher on correct trials compared to incorrect trials when the preferred stimulus was present on the screen ($p < 0.0002$). These data thus demonstrate the correlates of change-blindness and change detection in the human MTL. When subjects correctly detected changes, firing rates in the population of cells encoding the relevant stimuli were significantly higher during the time intervals when these stimuli were presented to the subjects. On the other hand, neuronal activity during these time intervals was reduced when subjects were blind to changes. In contrast to Figure 6.8a, the results in Figure 6.8b summarize the average neuronal activity in those time intervals when the preferred stimuli of the cell were absent. In this condition, no significant difference in firing rates are observed between correct and incorrect trials ($p > 0.4$), which is expected since only those stimuli that drove these cells weakly were presented to subjects.

6.3.4 Predicting change

The data presented thus far is averaged over many trials and many cells, and shows that on average the activity of individual neurons differs between correct and incorrect trials. However, on each trial, an individual cell encoding a particular stimulus does not have the liberty to compute averages. Instead, if it participates in change detection, it must contribute to the decision based on information it receives on a trial-by-trial basis. How could this neuron tell us on each trial, whether a change occurred or not? A very simple model of change detection could assume that on each trial, a decision-making system would monitor the activity of neurons encoding the presented stimuli over the length of the trial. It would then compare the firing rates between the two display periods, and based on this comparison it could signal the presence or absence of a change. A change in activity between the first and second display intervals would signal that a change had occurred in one of the stimuli, while similar firing rates in the two intervals would indicate no change.

How reliably can the activity of a neuron on individual trials tell us whether or not a change occurred? We addressed this issue quantitatively using a receiver operating characteristic (ROC) analysis that is used classically in psychophysics (Green & Swets, 1966). For each neuron that we recorded from, we determined how well an ideal observer could use the firing rate on individual trials to predict whether or not a change had occurred. On each trial, we computed the difference between the firing activity in the two display periods[16] and used a sliding threshold to determine the probability of false alarms and the probability of correct detections. More precisely, on each trial, if we denoted the firing activity in the two display periods as r_1 and r_2 respectively, and the event of a change as X, we were interested in determining the quantity $P(X \mid \Delta r)$ for different values of a threshold T, (where Δr is the difference between r_1 and r_2). For each integer value of the threshold T, a difference in spike number between the two display periods greater than T, (i.e. $\Delta r > T$), would mean that the neuron signaled that a change had occurred on that trial. Depending on if a change had actually occurred or not, this signal would then count either as a “correct detection” or a “false alarm”. We computed these values for all trials together, as well as separately for correct and incorrect trials.

---

[16] We computed the ROC over different periods of time in both display periods (Figure 6.13). To be precise, the number of spikes in the interval [300 X] ms (aligned to the onset of each display period) was computed separately for X = [700, 1000, 1500, 2000, 2500]. For all the data presented in this chapter, ROCs were evaluated using spike counts in the [300 1500] ms interval.

The ROC curves we thus obtained for each of our cells are shown in Figure 6.9a and 6.9c. In these plots, the probability of correct detection is plotted against the probability of false alarms for different values of the threshold T. The area under each curve is a measure of the ability of each neuron to estimate whether a change occurred. A value of 0.5 would correspond to chance performance, while a value of 1 would reflect perfect accuracy in signaling a change. Over the population of cells, when computed over correct trials, the average value was 0.67 ± 0.02. In contrast, the average ROC area computed over incorrect trials was significantly lower (0.49 ± 0.04; $p<.005$) and not significantly different from chance ($p=0.95$).

The individual values for each cell in our population over correct and incorrect trials are plotted in Figure 6.9b and 6.9d, respectively. Over the correct trials, the distribution of the area under the ROC curves (Figure 6.9b) is centered at 0.67 and is significantly shifted to the right of 0.5 ($p < 0.0001$). On incorrect trials, however, the mean of this distribution is no different from chance (Figure 6.9d). These figures thus indicate that the ability to predict whether or not a change occurred on individual trials depends on whether the patient made correct or incorrect behavioral reports on the corresponding trials. In other words, on trials when the subject correctly reported the presence or absence of a change, the MTL cells also appeared to be able to correctly make this judgment. On the other hand, when subjects were in error, the MTL units did not appear to carry the relevant information either. These results are therefore compatible with the notion that MTL neurons could contribute to subjects' perception during change-blindness and change detection: on trials when subjects were "blind" to changes, the neurons appear to be "blind" as well, whereas on correct trials, both subjects and MTL neurons were able to detect changes.

Additionally, this data suggests that at the neuronal level, attention plays a key role in a neuron's ability to correctly signal a change. Under the assumption that on correct trials attention was more likely to be focused on the relevant stimuli, the ROC values indicate that with the influence of attention, we can generally guess above chance whether or not a change occurred simply by observing the firing rates of individual neurons on individual trials. On the other hand, when attention is presumably not focused on the relevant stimuli, we are as likely to correctly predict a change as not. The significant, beneficial effect of attention in predicting a change is thus demonstrated by the difference of the two distributions shown in Figure 6.9b and d ($p<.005$).

This ability of individual cells to signal the occurrence of a change can be compared to the behavioral performance of our subjects. On average over subjects, the percentage correct in change detection was 0.78 ± 0.01. This value is higher than the performance of individual neurons (0.65 ± 0.02%),[17] but this is not unreasonable, given that the behavioral performance of subjects presumably reflects information contributed by more than one neuron. Although it has been observed that single neurons can sometimes be as sensitive as subjects in some tasks (Newsome, Britten, & Movshon, 1989; Britten, Shadlen, Newsome, & Movshon, 1992), in most cases it is reasonable to suppose that performance of single neurons is bounded by overall behavioral performance.

The performance achieved by the ROC analysis in predicting the occurrence of a change can be further understood by comparing the firing activity of each unit in the two display intervals. Theoretically, on change trials, the larger the differences in firing activity between the intervals when the preferred stimulus was present versus absent, the higher the probability of predicting a change. Conversely, on no-change trials when the preferred stimulus was present in both intervals, for high prediction values, firing activity should be almost identical in the two display periods.

---

[17] The ROC value presented here is computed over both correct and incorrect trials together. This value over all trials is closer to the ROC value for correct trials as a result of the fact that there were many more correct than incorrect trials.

For our set of cells, the relationship between firing activity in the two display periods of the change trials is shown in Figure 6.10a. It is evident from the figure that activity in the interval when the preferred stimulus was absent was significantly lower ($p < .0001$) than the corresponding activity in preferred stimulus "present" intervals. On average, firing activity during the former period was 60% of the activity in the latter interval. The relationship between firing activity in the two display intervals of the no-change trials when the preferred stimuli were present (i.e. "both" trials) is shown in Figure 6.10b. This figure demonstrates that firing activity during the second interval was significantly ($p < .00001$) lower in the second display period compared to the first. This difference in firing activity might appear to be surprising since the physical stimulus was identical in the two display periods. However, previous reports of decreases in firing activity with prolonged exposure to a particular stimulus (e.g. due to adaptation or habituation) are consistent with this observation (Miller, Li, & Desimone, 1991; Lueschow, Miller, & Desimone, 1994; Miller & Desimone, 1994; Desimone, 1996).

Another way of examining the data that is summarized in the ROC plots is to compare the distributions of the difference of firing activity between the two display intervals in the change and no-change trials. In principle, the further apart these distributions are, the easier it is to correctly predict the occurrence of a change. The distribution of firing activity in the change and no-change trials is shown in Figure 6.10c, d for the correct and incorrect trials, respectively. The black distributions in these plots, centered at 0, correspond to the "null" distribution—the distribution of differences in firing activity between the two display periods of the no-change trials. Because of our earlier observation that in the ("both") no-change trials, firing activities are not identical in the two display periods, to obtain the null distributions, we have averaged the two distributions obtained by subtracting r_1 from $r_{2,}$ and vice versa. By thus averaging, we remove the bias introduced by choosing either one of the two subtractions. Indeed, in the change trials, with which we wish to compare these null distributions, the target picture occurs with equal probability in the first or second periods.

To estimate the ability of cells to be able to predict changes, the distributions of firing activity during change trials can be compared to these “null” distributions. The distributions for the change trials are shown in red. These distributions correspond to differences in firing activity between the stimulus present and stimulus absent intervals of the change trials. The mean of this distribution over correct trials is centered at 2.6, indicating that on average, for each cell, the difference in firing activity between the two display intervals is 2.6 spikes on change trials. Although this value appears to be fairly low at first, it is interesting to recall reports that suggest that subjects are able to distinguish between stimuli on the basis of 1-2 spikes in the somatosensory system (Vallbo & Johansson, 1976; Johansson & Vallbo, 1979; Parker & Newsome, 1998). On incorrect trials, however, the mean of the distribution is much smaller (0.8 spikes), again demonstrating that on trials when subjects made errors, the difference in firing activity between the two display periods on change trials was not readily distinguishable from the corresponding difference in activity on no-change trials.

6.3.5 Predicting behavior

The firing activities of individual neurons in the MTL are thus able to give us information about the visual stimulation presented to subjects—they are able to signal above chance whether or not a change occurred. But to what extent do these cells participate in the subjects’ decisions? Does their activity on individual trials allow us to guess above chance the behavioral report of the subject? In other words, using the terminology mentioned in the previous section, and denoting the behavioral report of subjects by Y, we are interested in estimating $P(Y \mid \Delta r)$.

To understand this issue, we conducted an ROC analysis to determine the correlation between neuronal activity and the behavioral report of subjects on a trial-by-trial basis. For this analysis we again estimated the probability of whether or not a change had occurred by comparing spike counts in the first and second display periods, as described above. We then compared this estimate not with the actual stimulus, but with the behavioral report of the subjects, for each value of the sliding threshold T. A “correct detection” or “false alarm” would be counted depending on whether the prediction from the spike counts matched the subjects’ behavioral reports or not.

The ROC curves for this analysis are shown in Figure 6.11a. As before, the probability of correct detection is plotted against the probability of false alarms for each value of the threshold. The average area under the curve in this case is 0.58 ± 0.01. The distribution of these values for all cells is shown in Figure 6.9b and as before is significantly shifted to the right of 0.5 ($p < 0.001$). This indicates that on a trial-by-trial basis, the behavior of subjects, (their perception of the change, or the corresponding change-blindness) can be predicted above chance by individual neurons in this population.

The relationship between predicting changes in the stimulus and behavioral decisions reflecting the percept of the subject is shown in Figure 6.12 for each cell. On average, the population of cells appears to be better at predicting stimulus related information than the behavioral percept of the subjects (slope = 0.4). Thus, in a continuum of processing from the retina (which would be perfect at predicting changes in the stimulus) to the output motor neurons (which would presumably perfectly reflect the behavioral decision of the subject), MTL cells appear to be closer to the sites of stimulus rather than perceptual/decision-making processing.

This conclusion appears to be at odds with a previous report that demonstrated that the activity of MTL neurons correlates strongly with the behavioral percept of subjects during flash suppression (Kreiman, Fried, & Koch, 2002). However, the high correlation between neuronal activity and subjective percept observed in that study could be related to the relatively unnatural manner in which stimuli are presented during flash suppression. Flashing an image to one eye could completely suppress neural activities in response to the other stimulus, thereby giving rise to the observation that suppressed (i.e. unperceived) stimuli elicit no neuronal response in the MTL. Under more natural viewing conditions, however, MTL neurons could maintain weak (and non-zero) responses to stimuli that are not perceived, thus suggesting that these neurons are not strongly correlated with perception.

6.4 Discussion

In this experiment, we set out to investigate the effects of attention on neuronal responses in the human MTL. Our results demonstrate that in a change-blindness paradigm, when subjects correctly detect the presence or absence of a change made to particular stimuli, firing activity of neurons encoding these stimuli are high during stimulus presentation. On the other hand, neuronal activity during presentation of the relevant stimuli is significantly reduced when change-blindness occurs. Thus, we demonstrate for the first time at the single-cell level the neuronal correlates of change-blindness and change detection. Furthermore, under the reasonable assumption that on correct trials attention was more likely to be focused on the relevant stimuli than on incorrect trials, these results also demonstrate the effect of attention on the population of neurons. From our data it appears that attention enhances firing rates of neurons encoding the attended stimuli. This attentional effect, which is consistent with observations made in monkeys (Moran & Desimone, 1985; Treue & Maunsell, 1996;

Reynolds, Chelazzi et al., 1999) is demonstrated here for the first time in single neurons of the human brain.

Our results also showed a tendency towards higher firing rate on correct trials during the blank interval (ISI) between the two stimulus periods. This activity was observed on those trials when the preferred stimulus of the cell was presented in the display period preceding the ISI (i.e. the “disappear” and “both” trials). This enhanced activity which persists after the stimulus is removed might reflect a memory for the stimulus, that is more strongly encoded on correct rather than incorrect trials. It is not unreasonable to suppose that a stronger memory of stimuli from the preceding interval could contribute to better performance on a trial. Indeed some theories of change detection assume a stable representation of a limited number of objects in a scene, which is used for online comparison with incoming information (Rensink, 2000, 2002). The memory trace observed after the offset of the stimulus could serve as a neuronal correlate of this representation. This interpretation is also in line with the role of the hippocampus and other medial temporal lobe structures in the formation of short-term memories (Squire & Zola-Morgan, 1991; Kandel & Hawkins, 1992; Zola-Morgan & Squire, 1993; Kandel, Schwartz, & Jessell, 2000; Buckner & Wheeler, 2001). Damage to these structures results in learning deficits and impairs formation of recent memories, as is strikingly demonstrated in the case of patient H.M. (Penfield & Milner, 1958; Milner, 1972). Thus, MTL structures are involved in the temporary storage of information during visual processing, and the trace we observe during the ISI interval could represent this storage.

Based on the neurons they record from, monkey electrophysiologists have often tried to determine how the firing of individual neurons relates to the animal’s behavior (Celebrini & Newsome, 1994; Britten, Newsome, Shadlen, Celebrini, & Movshon, 1996; Shadlen, Britten, Newsome, & Movshon, 1996). Similarly, given the neural correlates of change detection in the MTL cells, it is interesting to ask how close these MTL neurons are to the site where the decision is made regarding change perception. On the one hand, if the decisions were made in the MTL itself, one could reasonably expect that the firing rate of MTL cells should be highly correlated with the behavioral report of the subjects on a trial-by-trial basis. However, as our ROC analysis demonstrates, while individual neurons can predict the subjects' behaviors significantly above chance, this performance is correct only 58% of the time.[18] Furthermore, if MTL cells perfectly reflected the percept of subjects, on correct trials one would expect high firing rates during the presentation of the relevant stimulus, but on those trials when subjects made a mistake and presumably did not perceive the relevant stimuli, the neurons representing these stimuli should be silent. However, our data does not indicate that neurons are silent on incorrect trials. Instead, the data shows that even on these trials, neuronal activity in response to the preferred stimuli was significantly higher when compared to baseline activity. It is as though in these situations, although the subject does not see the particular stimulus (resulting in change-blindness), the recorded neuron does to a certain extent. Thus, at the neuronal level, the subjective percept and neuronal activity do not match up exactly, resulting in lower levels of performance in predicting behavioral reports.

Another factor to explain the moderate performance value from the ROC analysis could be that this analysis reflects a behavioral prediction that is based on the activity of a single neuron. In contrast, the decision systems of the brain probably monitor large populations of neurons simultaneously. The limited information provided by individual

---

[18] Note however, that this value is comparable with several earlier reports of the relationship between neural responses and perceptual judgments in monkey area MT (Britten, Newsome et al., 1996; Cook & Maunsell, 2002). In these studies, the authors report that MT neurons could predict the animal's behavioral choice roughly 56% to 60% of the time.

cells might not be sufficient to always reflect behavior (Britten, Newsome et al., 1996; Cook & Maunsell, 2002).

In addition, if one assumes that a neuron carries visual information that contributes to the behavioral decision, then the information it can carry about the change is likely to be bounded by what the neuron “sees” of the change. In other words, its ability to accurately signal the presence or absence of a change might be constrained by its ability to reliably encode the presence or absence of its preferred stimulus. As we saw in the ROC analysis, based on the firing rate of individual neurons, changes in the presence or absence of preferred stimuli can be guessed 65% of the time on average. Given our assumption that the MTL neurons contribute to the decision of change perception, their performance at predicting behavior would then have an upper bound of 65%. Under this constraint, it turns out that the MTL neurons would actually be fairly efficient in using the stimulus-related information available to them—their performance in estimating behavior being 77% of their ability to signal a change.[19] (Note again that this upper bound is only valid under the assumption that we are dealing with neurons that are involved in the perceptual and cognitive operations underlying change detection. For example, a motor neuron encoding subjects’ button presses would be accurate 100% of the time in predicting the behavior of subjects, without carrying any independent stimulus-related information.)

Alternatively, the seat of the behavioral decision could be located in other brain regions. Evidence for this hypothesis comes from recent fMRI work investigating the neuronal correlates of change-blindness and change detection. These studies have reported that visual areas in the ventral stream are activated strongly during change

---

[19] This calculation scales performance in behavioral prediction relative to performance in signaling a change based on stimulus-related information, while keeping chance at 50%. The exact formula used is: 0.5+0.5*(B-0.5)/(S-0.5) where B is performance on behavioral prediction (for us 0.58), and S is performance on signaling change (0.65).

detection and to a lesser extent (but still significantly) during change-blindness (as we show here for the medial temporal lobe cells). However, activity in dorsal areas was more clearly differentiated depending on whether subjects perceived the change or not. In fronto-parietal sites for instance, enhanced activity was observed during change detection, but no activity was detected during change-blindness (Beck, Rees et al., 2001; Pessoa & Ungerleider, 2004). Given this pattern of activation, it is thus possible that individual neurons in dorsal areas would be better correlated with subjects' behavioral reports than the MTL neurons we record from. Additionally, these dorsal networks are less likely to carry stimulus-related information since they are mostly silent during change-blindness trials, even though the visual stimulation is identical to change detection trials.

Taken together these results suggest that the mere activation of ventral stream and MTL neurons might not directly predict change detection (since correlation with the stimulus is higher than with the behavioral outcome). Rather, the ventral and MTL neurons could carry information related to the visual input (such as what we observe during both change detection and change-blindness trials), but the contents of awareness would arise through interactions with areas in the dorsal pathway, as suggested by Rees and colleagues (Rees, Kreiman, & Koch, 2002).

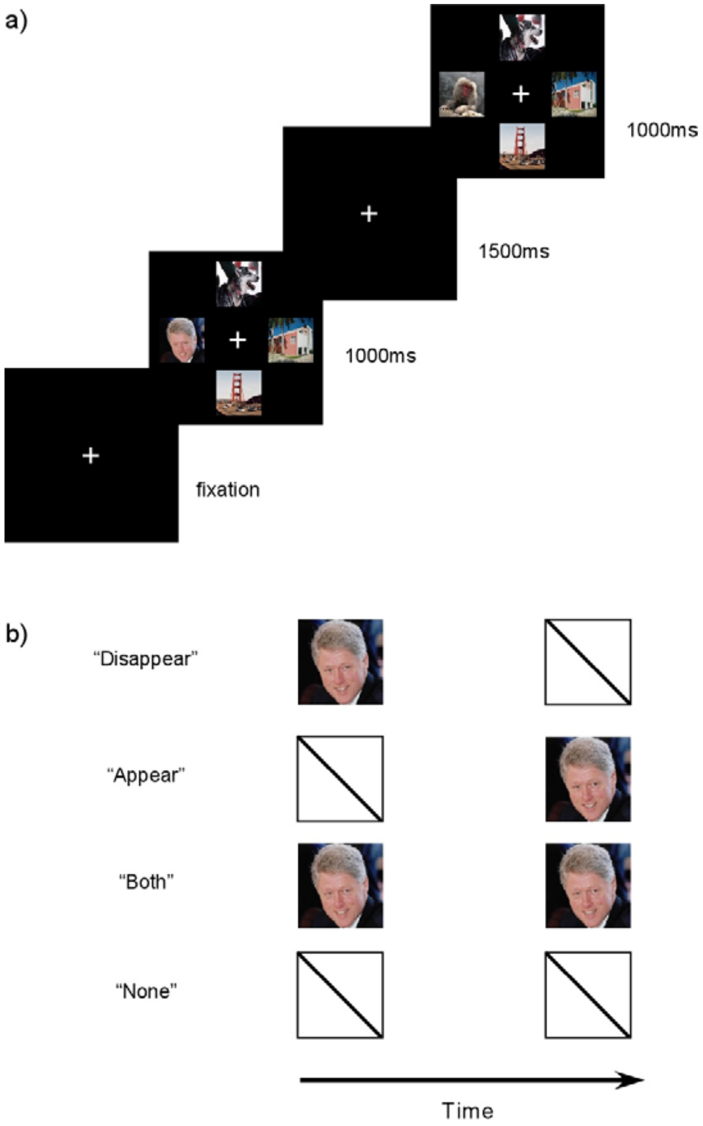


Figure 6.1 Experimental timeline for one trial in the change-blindness experiment. a) Each trial begins with a fixation cross, followed by four pictures that are presented for 1000ms. A blank interval for 1500ms follows and is replaced by another set of 4 pictures, again presented for 1000ms. On roughly half of the trials, one of the four pictures changes from the first display to the second (in this case the picture of Bill Clinton changes). At the end of the trial, subjects have to report whether or not they noticed a change. b) From the point of view of a cell selective to Bill Clinton, 4 different trial types could occur: In the change trials, the picture of Clinton could be present in the first display period and then be absent in the next (“disappear trials”). Or it could be absent in the first and appear in the second (“appear trials”). In the no-change trials, the picture could be present in both display periods (“both periods”) or could be absent all together (“none trials”). By comparing activity between correct and incorrect trials, during the intervals when the picture was present, the effects of attention on the responses of this neuron can be examined.

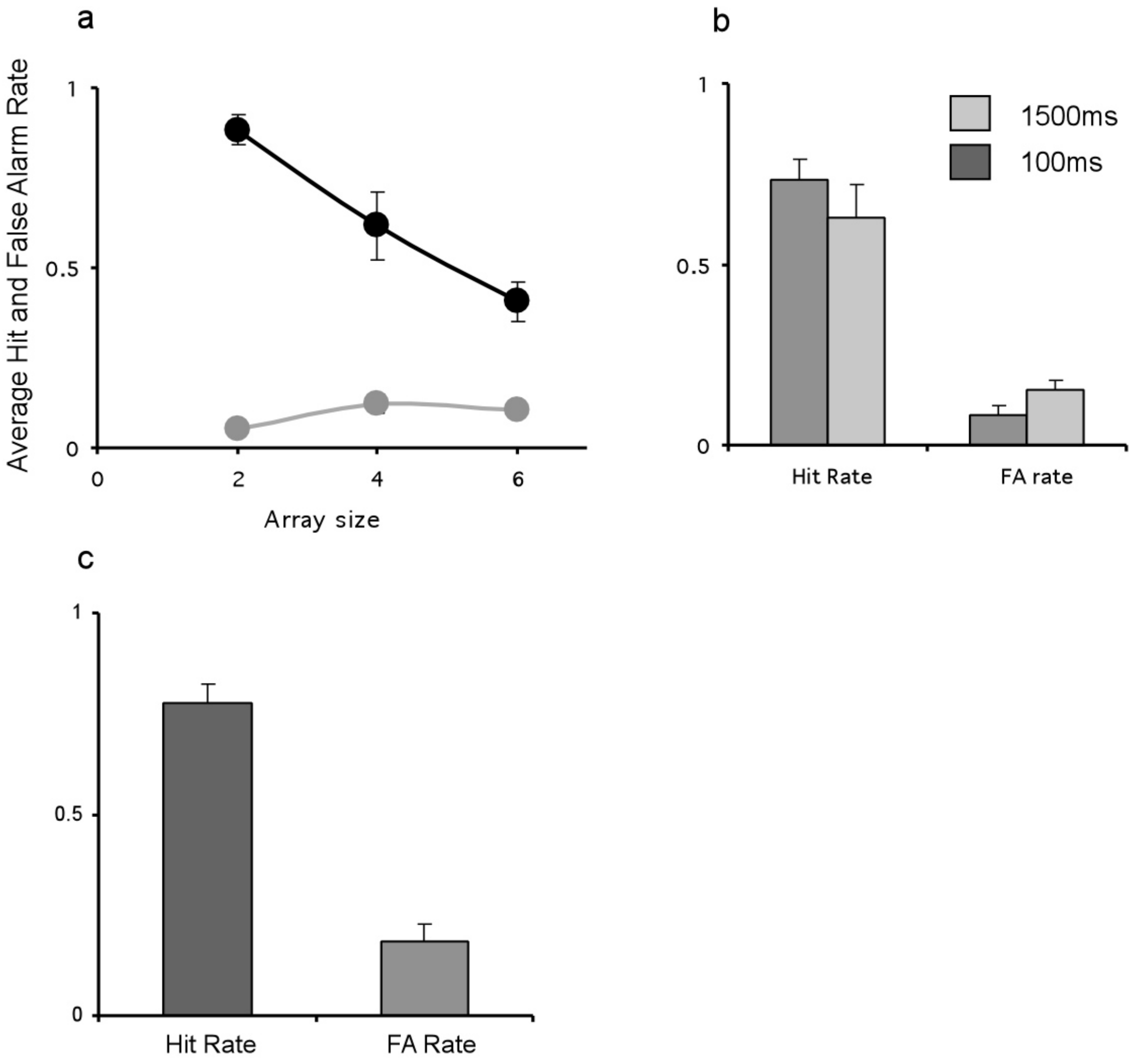


Figure 6.2 Behavioral responses of the clinical and non-clinical populations. a) Average hit rate (black) and false alarm rate (gray) for the non-clinical population (n=6 subjects) for different sizes of the stimulus display (2, 4, or 6 images in each display period; data for an ISI interval =1.5s is shown here). b) The hit rates and false alarm rates for the 6 subjects, when tested with two different ISI intervals (100ms and 1500ms; data with a 4-image display in each display period is shown here). c) Average hit rate and false alarm rate observed for the clinical population (n=9 subjects).

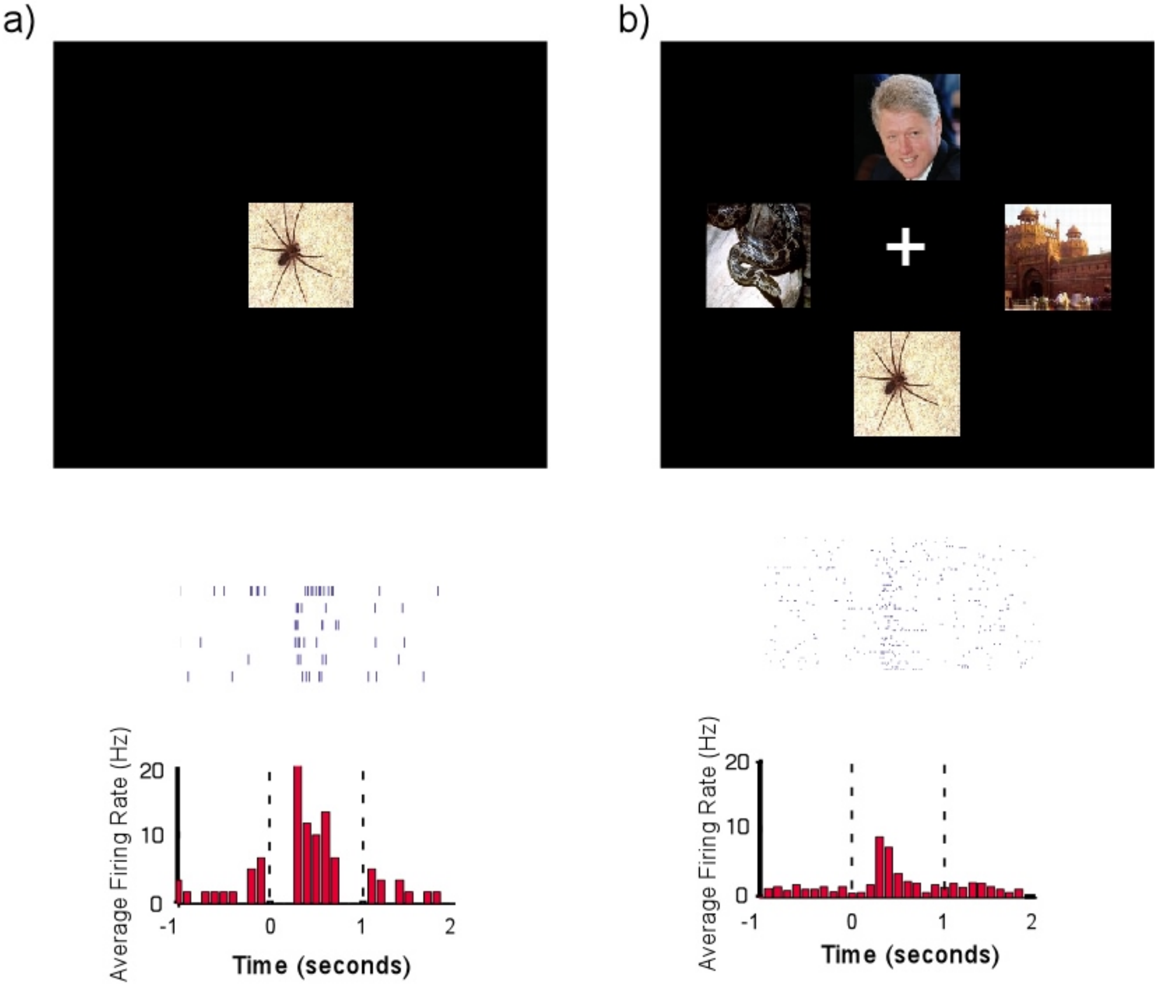


Figure 6.3 Comparing visual responses between screening and testing sessions. This unit is located in the right amygdala. a) In the screening session, this unit showed a significant increase in firing activity to an image of a spider that was presented foveally and in isolation (p<.001). The image appeared at t=0 and was presented for 1 second. b) The response of the same unit to the spider during the change-blindness set up is still significantly above baseline (p<.0001) but is about half as strong as the response in a). During the change-blindness experiment, the preferred image was presented peripherally along with 3 others known to drive the cell weakly. Visual responsiveness was determined by comparing firing activity during the pre-stimulus baseline interval with firing activity during the intervals when the stimulus was presented (see Section 6.3.4).

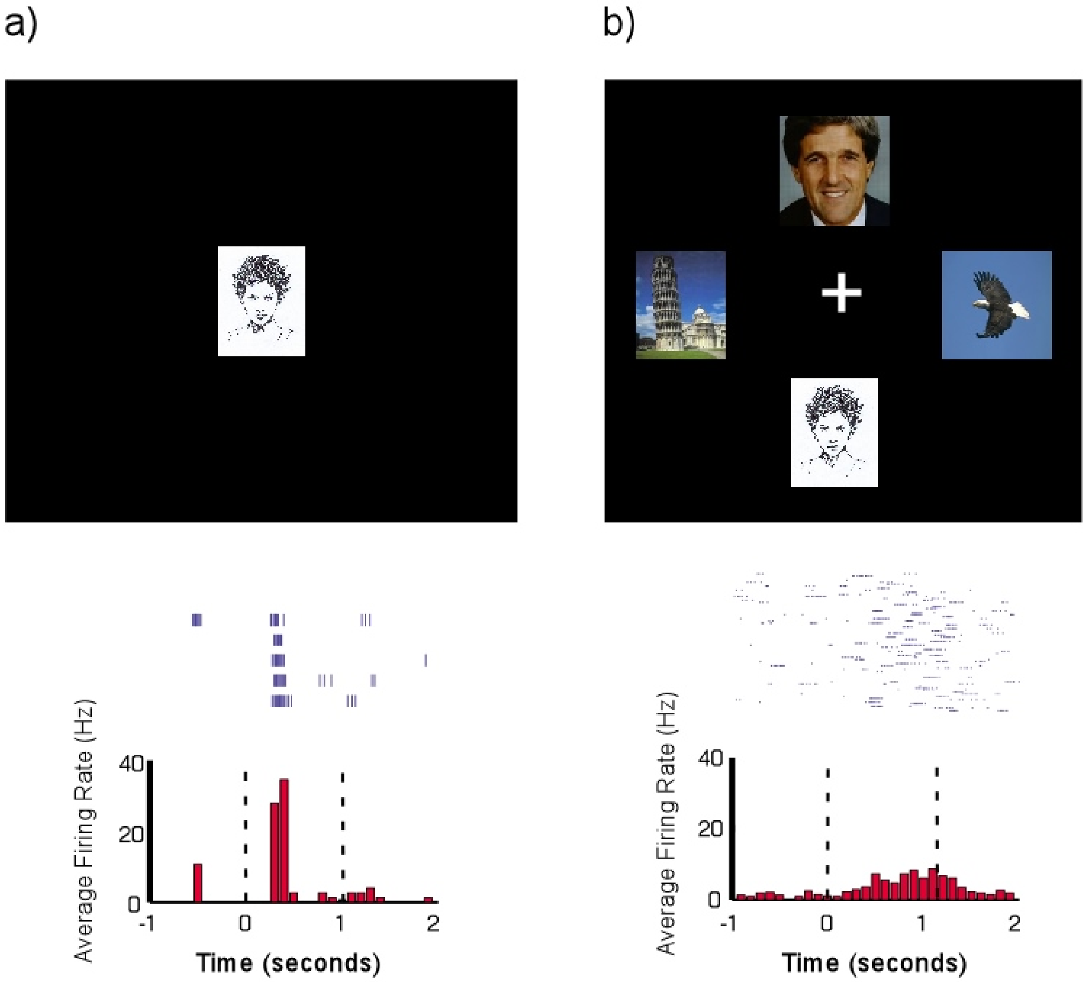


Figure 6.4 Comparing visual responses between screening and testing sessions. This unit is located in the right hippocampus. a) In the screening session, this unit showed a significant increase in firing activity to an image of actress Halle Berry that was presented foveally and in isolation (p<.005). The image appeared at t=0 and was presented for 1 second. b) The response of the same unit to Halle Berry during the change-blindness set up is still significantly above baseline (p<.001) but is much weaker than the response in a). Note that the actual image of the actress that was presented to the patients is different from the one shown here; we cannot show the original image because of image-copyright issues.

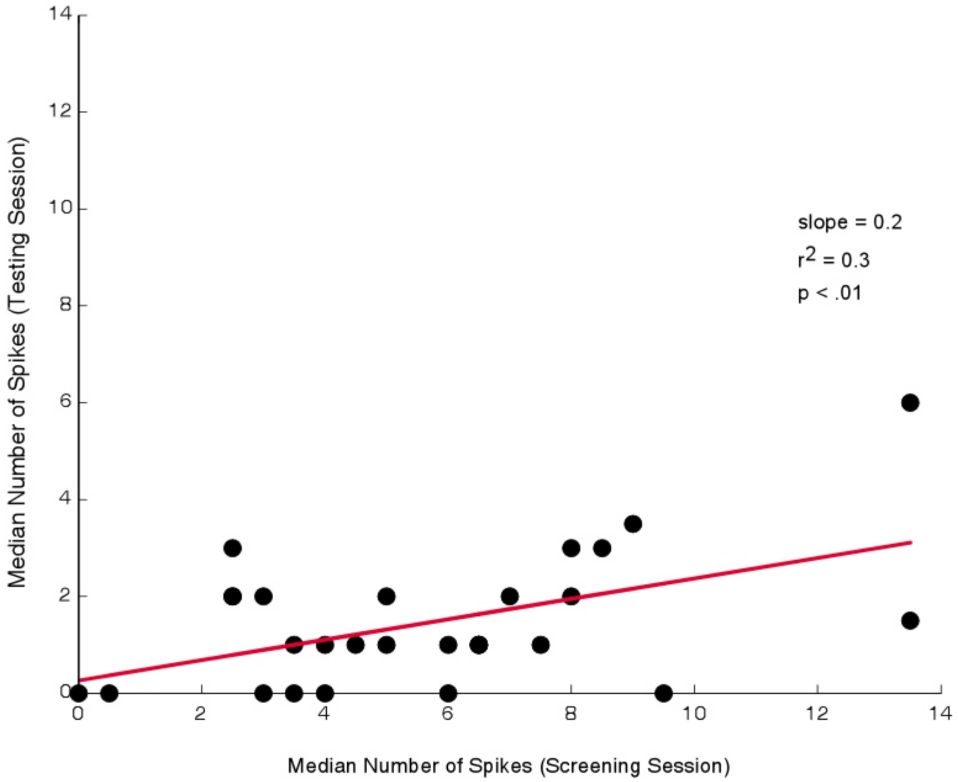


Figure 6.5 A comparison of the firing activity of all 29 neurons during the testing and screening sessions. The neuronal activity in the testing session is plotted against activity during the screening session. Over the population of cells, there is a significant decrease in activity in the testing session. This can be explained by the fact that in the testing session, each preferred stimulus was presented peripherally along with 3 other stimuli known to drive the cell weakly, while in the screening session, a single image was foveally presented (see Figures 6.3 and 6.4). On average the response to the preferred stimulus during the testing session was 20% of the response during the screening session.

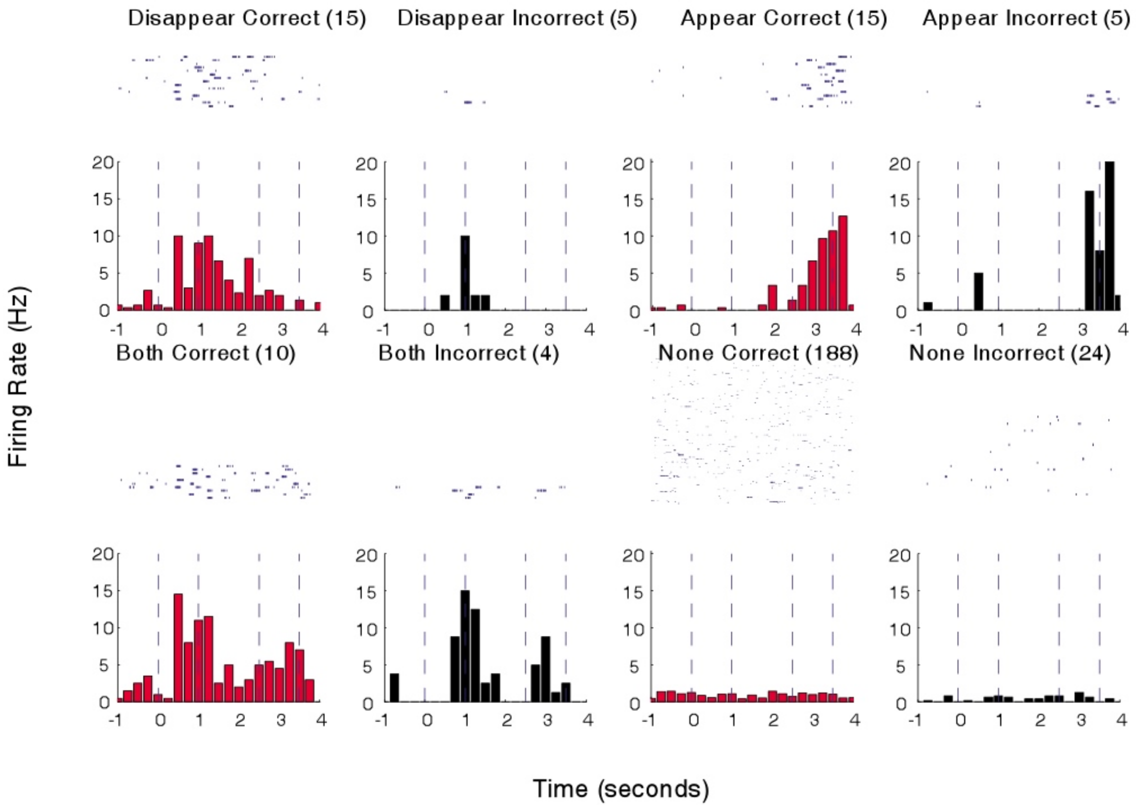


Figure 6.6 Responses during the 4 different trial types of the change-blindness experiment. a) Response of 1 unit located in the right amygdala. The red histograms correspond to correct trials in each trial type. The black histograms correspond to incorrect trials. Firing rate in Hz is plotted against time (ms). The first dashed vertical line represents the onset of the first set of 4 pictures, which stay on the screen for 1s (second vertical line). Following an ISI interval of 1.5s the second set of 4 pictures appears for 1s (last two vertical lines). A pre-stimulus baseline of 1s is also shown. The numbers in parenthesis represent the number of different trials in each condition. Although the effect is not significant at the single-cell level, the average neuronal response during the interval when the preferred stimulus is present tends to be higher on correct compared to incorrect trials. The average response time in reporting whether a change had occurred was 4.3 ±0.1s over the trials presented in this figure, aligned relative to the onset of the first set of 4 stimuli.

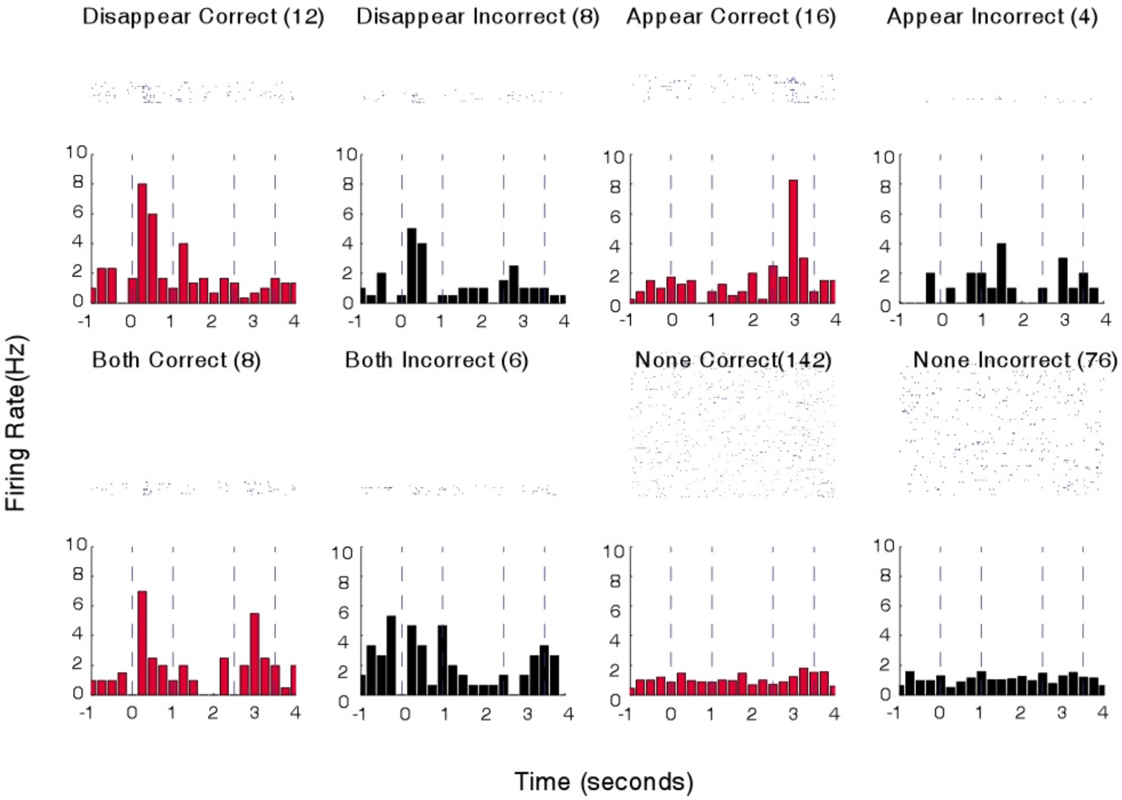


Figure 6.6b) Responses of another unit during the 4 different trial types of the change-blindness experiment. The data is presented in the same format as the previous figure. The average response time in reporting whether a change had occurred was 5.1 ± 0.1s over the trials presented in this figure, aligned relative to the onset of the first set of four stimuli. Again, activity on correct trials tends to be higher than on incorrect trials during those display periods when the preferred stimulus was present.

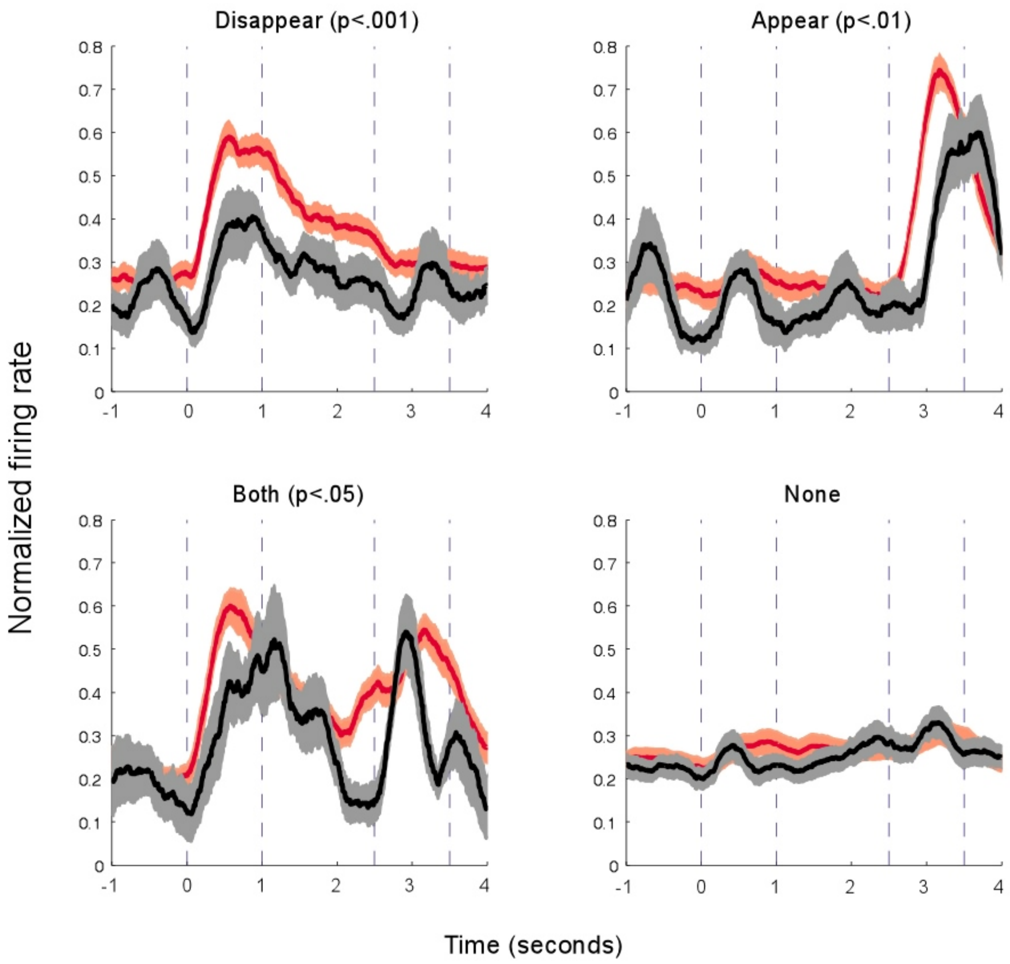


Figure 6.7 Responses during change detection and blindness for all cells. The mean normalized spike density function was calculated over the 29 cells in the 4 trial conditions for the correct (red curve) and incorrect (black curve) trials. Population responses were computed using the normalized spike density function (sdf). For each neuron, the sdf was obtained by convolving the spike train on each trial with a 200 ms fixed width Gaussian and then averaging over all trials. The spike trains were binned in 5 ms bins before convolution. For each unit, the sdf was normalized by dividing by its peak activity. The normalized sdf was then averaged over the population of cells. The average responses were computed separately for correct and incorrect trials in all trial types. The shaded areas represent the s.e.m. Note the significantly higher levels of activity on correct trials during all intervals when the preferred stimuli were present (corresponding significance values are presented at the top of each plot). During intervals when the preferred stimuli were not presented, there was no significant difference between correct and incorrect trials (p>.05). The average response time in reporting whether a change had occurred was 4.9 ± 0.1s over all patients. The difference in reaction times between correct and incorrect trials is not significant (p>.05).

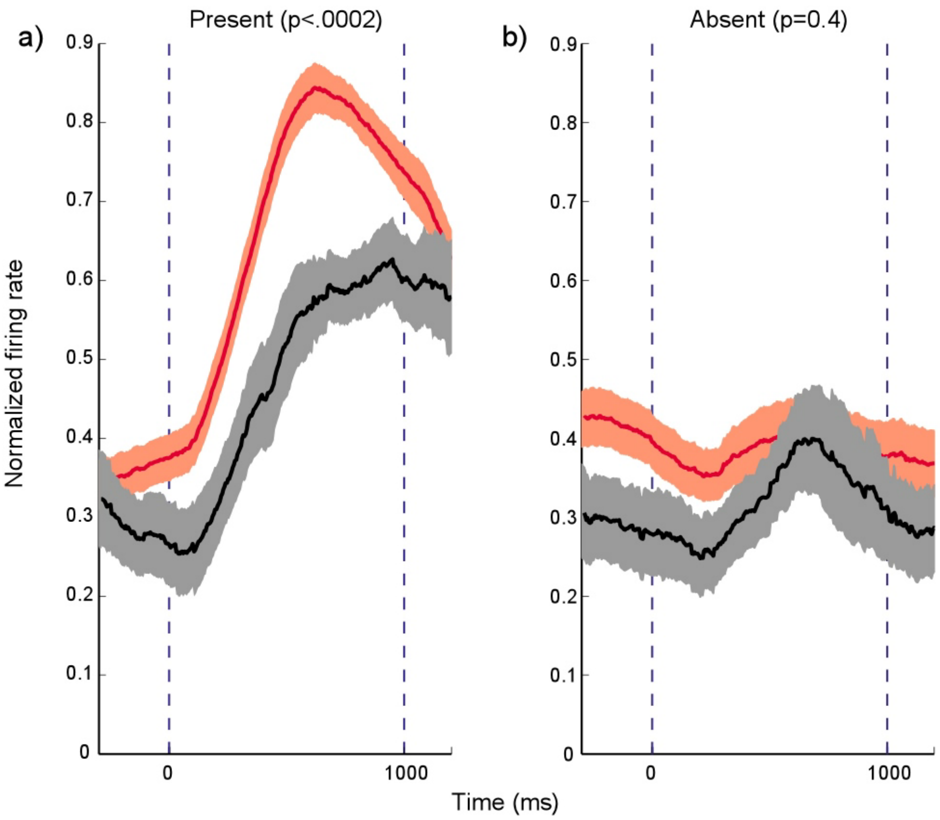


Figure 6.8 Population responses averaged over a) all stimulus intervals when the preferred stimuli were present and b) all stimulus intervals when the preferred stimuli were absent. In a), the data is averaged over the first display period of the “disappear” trials, the second period of the “appear” trials, and both periods of the “both” trials. In b), the data is averaged over the second interval of the disappear trials, the first interval of the appear trials, and both intervals of the none trials. The x-axis represents time in milliseconds. The stimulus appeared on the screen at 0ms and stayed on for 1000ms (the length of each display period). The y-axis represents normalized firing rates. The firing rates during the correct trials are significantly higher than firing rates during incorrect trials. In b) there is no significant difference in firing rates between correct and incorrect trials, which is expected since the preferred stimuli were absent during these intervals.

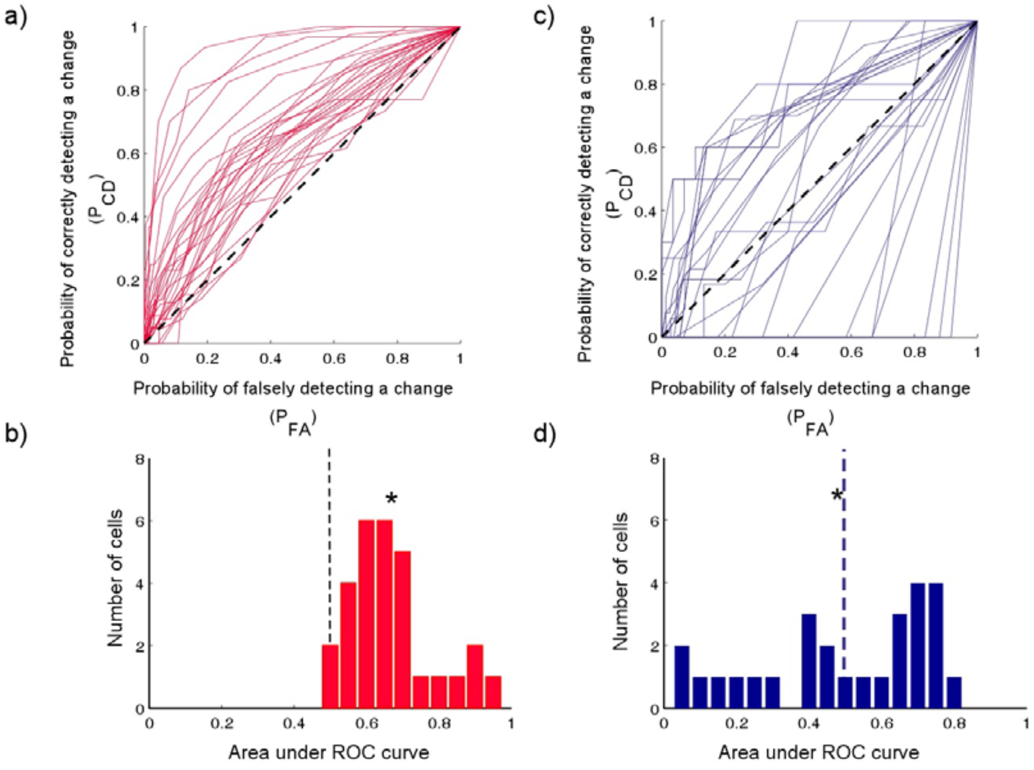


Figure 6.9 Predicting a change–ROC analysis over correct and incorrect trials. a) On each correct trial, the difference in the number of spikes occurring in the two display intervals was computed. If the absolute value of this difference exceeded a certain threshold T (for successive integer values of T), a change was predicted. This prediction was then compared to the actual visual stimulation—if a change had actually occurred, the signal would count as a "correct detection"; otherwise it would count as a "false alarm." The probability of classifying a trial as a change trial, P_{CD} ("correct detection") is plotted against the probability of "false alarms," P_{FA} (falsely detecting a change). The dashed line indicates chance performance ($P_{CD} = P_{FA}$). The different lines show the result of this calculation for each cell. b) The distribution of the area under the curves for each cell over correct trials is significantly shifted to the right of 0.5 indicating that the population of cells can signal a change above chance on a trial-by-trial basis ($p<.0001$). The mean area under the ROC curves is marked by a * and equals 0.67 ± 0.02. c) ROC curves for all cells calculated over incorrect trials. d) The mean area under the ROC curves (*) is 0.5 ± 0.04. In contrast to data in (b), on incorrect trials, the population of cells were at chance at signaling a change ($p=.95$). Under the assumption that attention was more likely to be focused on the relevant stimuli on correct trials, the significant difference ($p<.005$) between ROC areas on correct versus incorrect trials reflects the advantageous effect of attention at predicting a change. The data presented demonstrates that only on those trials when patients detected changes made to the relevant stimuli (i.e. the correct trials) could cells signal a change significantly above chance on a trial-by-trial basis.

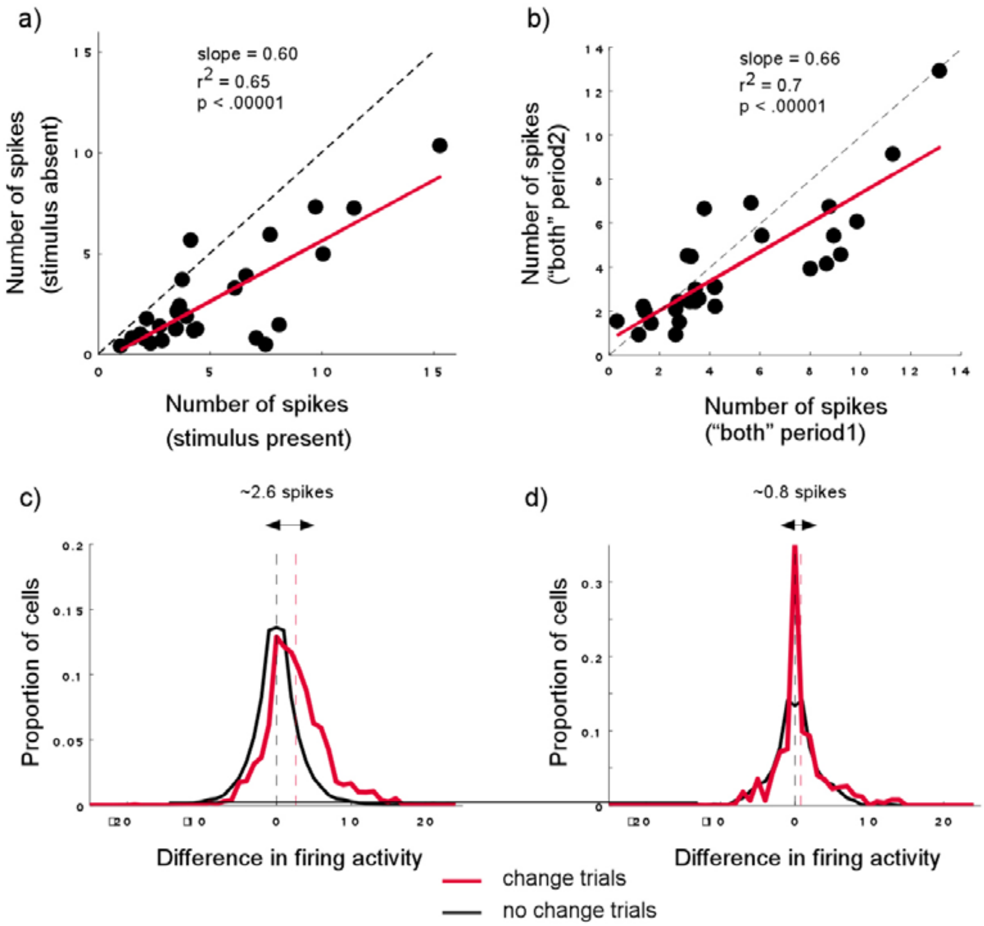


Figure 6.10 Firing activity during change and no-change trials. a) Activity for each cell during change trials. In this plot, the number of spikes fired in the absence of the preferred stimuli is plotted against activity in the presence of the preferred stimuli. The slope of 0.6 reflects that on average, firing activity in the absence of the preferred stimuli was six-tenths of activity when the stimuli were present. b) Activity in the two display periods of the no-change trials when the preferred stimuli were present (i.e. “both” trials). On average, the activity of each cell is lower during the second display period compared to the first, even though the visual stimulation was identical in both periods. c, d) The distribution of the difference in firing activity between the two display intervals for the no-change (black) and change (red) trials. These distributions are shown in c) for correct trials and in d) for incorrect trials. The distribution on correct change trials (i.e. activity in “present” —activity in “absent” intervals) is shifted to the right of distribution for no-change trials by about 2.6 spikes. Thus, on average, on correct change trials, the difference in activity between the two display periods is ~ 2.6 spikes. On incorrect change trials, however, the mean difference in activity between the two display intervals is only ~0.8 spikes indicating again that on incorrect trials, firing activity between change and no-change trials is not easily distinguishable.

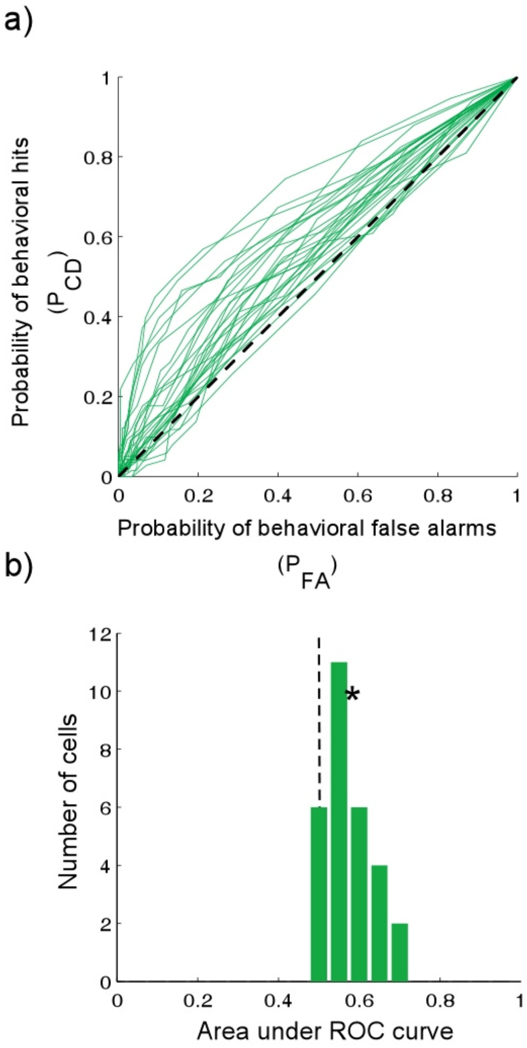


Figure 6.11 Predicting Behavior–ROC analysis. a) As before (Figure 6.8) on each trial, the difference in the number of spikes occurring in the two display intervals were computed, and the probability of a change occurring was predicted. This prediction was then compared with the behavior of the patient—if the neuron and patient reported that a change had occurred, the signal would count as a "correct detection"; otherwise if the neuron predicted a change but the patient reported no-change, the signal would count as a "false alarm." The probability of predicting behavior correctly, P_{CD} ("correct detection") is plotted against the probability of incorrect predictions ("false alarms," P_{FA}). The dashed line indicates chance performance ($P_{CD} = P_{FA}$). The different lines show the result of this calculation for each cell. b) The distribution of the area under the curve for each cell. The histogram is significantly shifted to the right of 0.5 indicating that the population of cells can predict behavior above chance on a trial-by-trial basis ($p<.001$). The mean area is marked by a * and equals 0.58 ± 0.01.

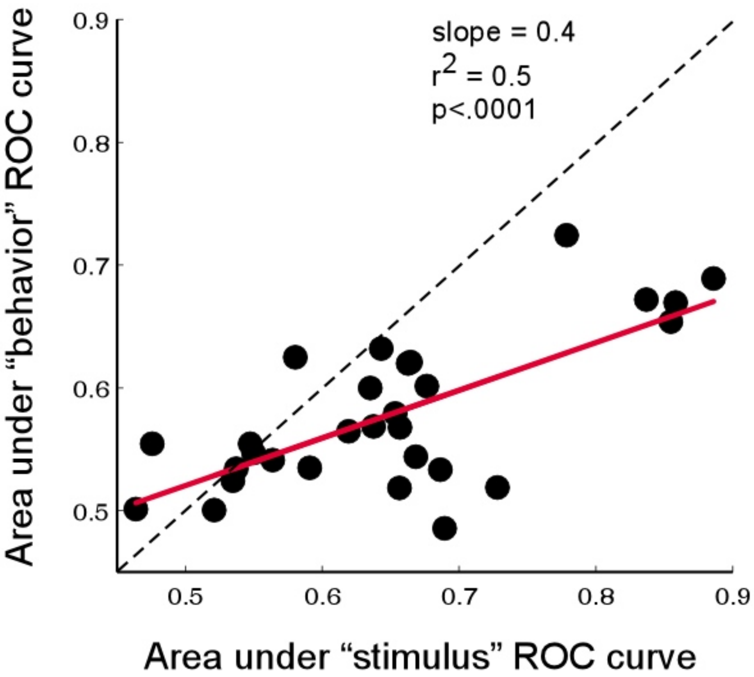


Figure 6.12 Comparison of the ability to predict change and the ability to predict behavior on a trial-by-trial basis, for each cell. On average, each cell is better able to predict changes in the stimuli rather than the behavioral choice of the subject on a trial-by-trial basis.

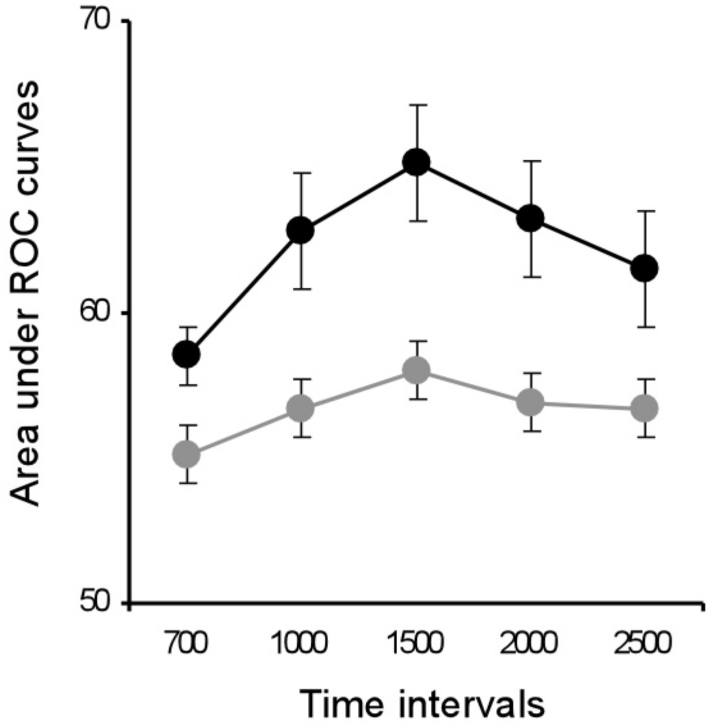


Figure 6.13 Comparison of the area under the stimulus and behavior ROC curves. ROC curves for predicting the stimulus (black line, corresponding to data in Figure 6.9a, b), and behavior (gray line, corresponding to data in Figure 6.11a) were computed over different time intervals. Each point on the graph corresponds to the ROC area when computed between [300 X] ms (aligned to the onset of the relevant display period) for 5 different values of X (shown on the x-axis). The ROC data reported in this chapter was computed over [300 1500] ms, which is optimal for both stimulus and behavior predictions.

Table 6. 1 Distribution of responsive cells during the change-blindness paradigm

Brain Area	Number of cells
RA	5
REC	2
LPG	4
LEC	1
RAH	11
RMH	1
RPG	1
LMH	2

The brain areas are as follows: RA: Right Amygdala, REC: Right Entorhinal Cortex, LPG: Left Parahippocampal Gyrus, LEC: Left Entorhinal Cortex, RAH: Right Anterior Hippocampus, RMH: Right Medial Hippocampus, RPG: Right Parahippocampal Gyrus, LMH: Left Medial Hippocampus.

Chapter7.

Closing Thoughts.

7.1 The processing of natural stimuli is favored by the brain

The goal of this thesis was to understand the role of attention in the processing of natural stimuli at both the behavioral and neuronal levels. As we mentioned at the beginning, much of our current understanding of attention has been shaped by how the visual system processes simple geometric shapes. However, given that most of these stimuli are rarely encountered on a daily basis, it is important to expand our investigations to the realm of natural stimuli. The experiments discussed in this thesis have shown that in contrast to the artificial geometric shapes on which many theories of attention are founded, natural stimuli are treated specially by the visual system.

We have observed that natural stimuli can be processed efficiently even when focal attention is not fully available to them. Following the original demonstration by Li et al. that animals or vehicles could be detected in natural scenes in the near-absence of attention (Li, VanRullen et al., 2002), we have shown that this processing ability holds for even finer discriminations of natural stimuli (in particular face stimuli), right down to the level of the identification of individuals. These results are surprising from a computational point of view since identifying a face constitutes an enormous challenge for machine vision systems. However, given that we engage in face identification countless times a day, it shouldn’t come as a surprise that the human visual system is well tuned for such tasks. Furthermore, it is advantageous for the visual system to be able to perform these routine tasks (such as face identification, gender discrimination, or animal or vehicle categorizations) with the least effort possible, in order to free up resources for the processing of unexpected and possibly threatening events and objects.

7.2 Underlying neuronal substrates for the efficient processing of natural stimuli

In contrast to the processing of natural stimuli, computationally simpler tasks (such as telling a bisected red-green disk apart from its mirror image) are nearly impossible for observers unless attention is engaged. This fact, that under identical conditions (i.e. dual-task, where the peripheral stimuli are presented in isolation) some tasks suffer while others are effortless, could reflect important aspects of the architecture underlying human vision. What about the structure of the visual system could support the processing of natural stimuli that we observe?

In Chapter 5, we presented evidence that demonstrates that natural and familiar stimuli are represented explicitly within single neurons in high level areas. The representations we observed were of an abstract nature, in that they corresponded to specific high-level, semantic concepts rather than a particular visual property. It is possible that the hardwired selectivity of neurons such as these, established through familiarity with the stimuli, could underlie the efficient processing we have observed in the dual-task experiments. A similar hypothesis was proposed by VanRullen and colleagues (VanRullen, Reddy et al., 2004). In other words, the activity of a neuron that specifically encodes all elements from the category “animal” could be sufficiently strong, even in the near-absence of attention, to perform an animal/non-animal discrimination task. Although we did not control for attention in the experiments described in Chapter 5, this premise is supported by several early monkey experiments, which have observed category specific responses in IT (e.g. to faces) (Gross, Rocha-Miranda et al., 1972). These studies were performed in anesthetized animals where attention was unlikely to have played a significant role in generating the responses. Thus, engaging the attentional resource could be superfluous to achieving good performance, in the case of single isolated stimuli.

According to this idea that natural scenes (presented in isolation) are exempt from attentional requirements because they have explicit neuronal representations, human vision should also be efficient for other stimulus features that are represented explicitly (VanRullen, Reddy et al., 2004). And indeed, there is evidence that provides support for this premise. For instance, simple orientation, color, or motion detection tasks are known to occur pre-attentively, and adequate neuronal representations for these features are found in early visual areas such as V1 and V5/MT (Baker, Petersen, Newsome, & Allman, 1981; Maunsell & VanEssen, 1983; Zeki, 1983). On the other hand, for stimuli that do not correspond to explicit neuronal representations, performance suffers tremendously unless attention is fully available. In these cases it is likely that attention could play the critical role of binding different features together (as proposed by the Feature Integration Theory, (FIT)) at the neuronal level to create the relevant representations.

Compatible with these theories, the data presented in Chapter 4 lend support to the idea that the near-absence of attention does not necessarily weaken the representation of isolated stimuli in their “essential nodes” (the specific regions in the brain that support awareness of particular facets of vision; see Chapter 5). In conditions when the stimuli were relevant to the observer, high levels of BOLD activity were observed in the relevant brain area (FFA), even when attention was not focused on the stimuli. Under the influence of top-down factors such as behavioral relevance, these high levels of activity could form a basis for the efficient processing we observe at the behavioral level.

These results have important implications because they challenge some aspects of the influential FIT. One of the strong premises of the FIT is that while the pre-attentive processing of low-level features can be directly supported by the visual system, complex object recognition tasks and their corresponding neuronal representations must rely on attention. Our behavioral data (along with previous results (Li, VanRullen et al., 2002; Rousselet, Fabre-Thorpe et al., 2002)) clearly contradicts this assertion. Additionally, at the cortical level, our fMRI study shows that strong cortical representations can be maintained in areas known to support high-level tasks even in the near-absence of spatial attention. Taken together, in contrast to the claims of the FIT, our behavioral and fMRI findings suggest that high-level complex stimuli, presented in isolation, do not pose a major problem for the visual system.

For everyday life, these data thus suggest that certain stimulus types, when relevant or meaningful to the observer, can automatically trigger their corresponding neuronal populations without engaging the relatively laborious focal attention mechanism. Given the limitations of the attentional resource, this is indeed a desirable property.

7.3 Why then, does processing fail on occasion when attention is unavailable?

The discussion so far has focused on the idea that natural stimuli can be processed outside the focus of attention without suffering behavioral deficits and without decreases in the associated neural activities. However, as we mentioned in the introduction to this thesis, in several situations, observers can be oblivious to objects (including natural ones) unless attention is directly involved. This phenomenon was also demonstrated in our change-blindness (CB) experiments (Chapter 6), where we observed that both behavioral performance and the neuronal signal were lower on incorrect trials when attention was less likely to be engaged. Thus, we are faced with an apparent contradiction: the data from the first portion of this thesis suggests that natural stimuli can escape the attentional requirement, but the last chapter seems to demonstrate just the opposite. What factors could explain this conflict?

7.3.1 Top-down behavioral relevance (or the lack thereof) during CB

This discrepancy is reminiscent of the apparent inconsistency between the findings of our fMRI study and many previous reports: we demonstrate that BOLD activity does not necessarily decrease significantly in the near-absence of attention, but other reports have observed substantial drops. As we discussed in Chapter 4, this contradiction can be explained by taking into account the top-down influence of behavioral relevance. Perhaps, the same factor can help address the differences between our dual-task and change-blindness results.

During the dual-task condition of the experiments described in Chapters 2 and 3, subjects always expected a stimulus in the periphery and knew which aspects of it were relevant to their task. For instance, in the face-gender discrimination experiment, subjects' expectations included the knowledge that the stimulus would be a face in the periphery, and that the gender of the face was the key factor. During CB, however, subjects' expectations were not similarly constrained—neither by prior knowledge of the location of the relevant picture, nor by the type of stimuli that would be presented, nor by which aspects could change (e.g. subjects could not know beforehand that an apple would be presented, or whether its color (or some other feature) or the entire apple itself could change). Consequently the top-down signal was likely to have been a much weaker influence in the change-blindness experiments. In the absence of this signal, we have shown in Chapter 4 that the lack of focal attention results in a significant decrease in cortical activity. Thus, similarly during change-blindness, neuronal activity could have been lower on the incorrect trials (when attention was less likely to be engaged) due to this lack of top-down expectation.

7.3.2. The effects of cluttered environments

Another factor that could explain the discrepancy in the effects of attention between the dual-task and change-blindness paradigms could be related to the manner in which the stimuli are presented in these experiments. In the DT experiment, the face stimuli were presented in isolation in the periphery. Monkey electrophysiology studies have shown that moving the focus of attention onto or away from (well-contrasted) isolated stimuli has a small effect on neuronal responses (Reynolds, Pasternak et al., 2000). Thus, during the DT experiment, the fact that focal attention had no significant effect on the representation of the face stimuli is compatible with these observations.

However, in the CB experiment, all stimuli were presented in cluttered environments. In these conditions, competitive interactions arise between stimuli, and it is known that focal attention significantly affects neuronal responses by biasing competition. As we discussed in the introduction to this thesis, presenting a “good” and a “bad” stimulus together introduces competition between the stimuli that weakens the influence of the “good” stimulus on the cell’s response. In the CB experiment, we presented each “preferred” stimulus along with 3 stimuli that we knew drove the cell weakly. Thus competition would be especially strong in this situation. Therefore, the differences in neuronal activity we observe between attended and unattended conditions during the CB paradigm could be a result of the stronger influence of attention in cluttered scenes.

7.4 Focal attention and behavioral relevance revisited

The differential effects of focal attention and behavioral relevance were encountered in the fMRI dual-task experiment in Chapter 4. It was remarked there that when stimuli are processed in the near-absence of focal attention (i.e. the central and dual-task conditions) top-down influences such as behavioral relevance can significantly impact the BOLD signal. In particular, we observed that even in the near-absence of focal attention, the top-down influence of behavioral relevance was sufficient for sustaining strong cortical representations of the stimuli.

However, the dual-task paradigm did not allow us to discern whether focal attention and behavioral relevance can have equivalent effects on neuronal activity. In other words, we could not investigate the reverse interaction between these two factors, namely if in the near-absence of *behavioral relevance*, focal attention could similarly enhance the activity of underlying neuronal populations. The change-blindness paradigm however affords us a look at this question.

As we have recently argued, during change-blindness situations, subjects' expectations of potential stimuli and events are relatively unconstrained and consequently it could be maintained that these stimuli are processed in the near-absence of behavioral relevance. Under these conditions therefore, the effects of manipulating focal attention can be investigated relatively independently of top-down influences. Accordingly, as we have demonstrated in Chapter 6, during change-blindness situations shifting the focus of attention to particular stimuli does indeed increase their underlying neuronal response.

Thus, resembling the effects of behavioral relevance in the near-absence of focal attention, focal attention appears to be able to strengthen neuronal representations in the near-absence of behavioral relevance. In other words, these two factors could work independently of each other, although they appear to have equivalent effects on neuronal representations. Engaging either one could be sufficient to strengthen and maintain representations of objects at the neuronal level, but in the absence of both factors, neuronal representations would tend to suffer.

7.5 A proposed framework for attention and the processing of natural scenes

Information from scenes rich in complex objects enters our eyes constantly to be processed by the visual system. In some cases, attention plays a critical role in this processing. In others, it is not fully necessary. In all cases, however, the relevant objects and events must be represented at the neuronal level so that accurate percepts can be generated. How does the visual system accomplish this task?

Based on the experiments described in this thesis, I propose the following framework for natural stimulus processing. This framework, which is summarized in Figure 7.1, is of course a simplified interpretation of our data, but I hope it will serve as a useful tool for summing up our conclusions. The assertions of the framework are not likely to apply to unnatural and unfamiliar stimuli that are not explicitly represented at the neuronal level (Chapter 5).

In the following framework, we consider two extreme situations in which an observer could encounter natural stimuli: in isolation, or as a part of a cluttered scene. (Since objects are normally encountered in a variety of conditions, these two extremes are used here only as a first approximation.) In these situations, depending on the various attentional conditions, different outcomes are possible as has been discussed throughout this thesis. These outcomes are illustrated in Figure 7.1 and detailed further below. Each of the following points corresponds to a particular situation (stimulus/attentional condition) as indicated by the corresponding letter in the figure.

A. When an object is present in isolation, and it is both attended to and task relevant or meaningful to the observer, strong cortical representations are created (peripheral-task condition of the dual-task paradigm; Chapter 4). These representations could provide the necessary support for efficient processing of these stimuli as was manifested at the behavioral level in the dual-task paradigm (Chapters 2 and 3).

B. Even if the focus of attention is not on the object, as long as it is behaviorally relevant to the observer, the integrity of the underlying neuronal representations is not necessarily compromised (dual-task condition in Chapter 4). Instead, the neuronal activities triggered by the presentation of the object appear to be strong enough to generate percepts of the object and support its processing (Chapters 2 and 3).

C. However, if the object is not meaningful to the observer, the underlying cortical representation is substantially weaker (central task condition in Chapters 2, 3, and 4). This is not to suggest that cortical representations are not formed for this object. Instead, the object still appears to activate the relevant neurons significantly above baseline. This activity could be maintained online, and in the event that the corresponding information is required by the observer, top-down signals or focal attention could strengthen the underlying neuronal signals (Chapter 4).

D. For objects that are presented in cluttered scenes in the near-absence of both top-down expectations and focal attention (chapter 6; see also point H), the strong competitive interactions between objects significantly weaken the corresponding neuronal signals, and these objects are consequently in danger of falling into oblivion.

E. In these cluttered environments and outside the influence of top-down signals, focal attention can play a critical role in object representation. By selecting out a certain subset of stimuli, focal attention effectively biases the neuronal competition in favor of these objects. Based on our findings in Chapter 6, and also several monkey electrophysiology studies, attention appears to have the effect of filtering out the influence of the competing objects. Thus to all intents and purposes, the influence of focal attention brings us back to the condition of the "isolated stimulus."

F. Since attention is now focused on this stimulus, even though top-down expectations are minimal (which makes this point different from point A), the activity of the relevant neuronal population is strengthened. As was shown in Chapter 6, these higher levels of activity could support the generation of accurate percepts for this stimulus.

G. Presumably, adding top-down signals, such as behavioral relevance to the existing influence of focal attention in cluttered scenes, would continue to bias the competition in favor of the attended and relevant object, again taking us back to the case of the "isolated stimulus" (as shown in A). This possibility is just speculation, however, since our data (Chapter 6) does not provide us with direct evidence to support this claim.

H. The independent effect of behavioral relevance in crowded environments is an open question. Under our experimental conditions, we could not directly examine the influence of this factor on neuronal activity (this factor was either absent altogether (Chapter 6), or present in conjunction with focal attention (Chapters 2, 3, and 4). However, it is plausible that these signals would affect crowded scenes and isolated stimuli similarly and favor the relevant objects thereby strengthening the underlying neuronal activity.

This framework is an attempt to succinctly summarize our findings about how the brain could process naturalistic stimuli. Some aspects of the framework might appear to be grossly simplified for, as we have said earlier, the proposed framework uses two extreme situations as an approximation for the myriad possibilities in real life. However, it is the larger picture that I would like to leave my readers with. The results of this thesis, along with a few other previous studies, seem to strongly suggest that the processes underlying the perception of natural objects are not mere extensions of the mechanisms we are familiar with from experiments on simple geometric shapes. Instead, the visual system seems to be well adapted for these stimuli and can often process them without calling upon additional resources such as focal attention. Hopefully, future studies will elucidate these processes further, and thereby bring us to a deeper understanding of how the brain tackles its constantly changing, rich visual environment on a daily basis.

7.6 Honestly, closing thoughts

It's another glorious day in Southern California. As we walk out of the basement, into an afternoon flooded with light, we are once again taken aback by the nature that surrounds us. Struck by the complexity of all that we (almost) effortlessly see, our thoughts turn to the attentional mechanisms and cortical areas that generate our experience of the world (at least they should, given that we’ve spent our morning mulling over the visual system). We realize with gratitude that we are fortunate these complex systems function normally in our lives — systems so complex that we still don't understand them. And we marvel at the design of it all.

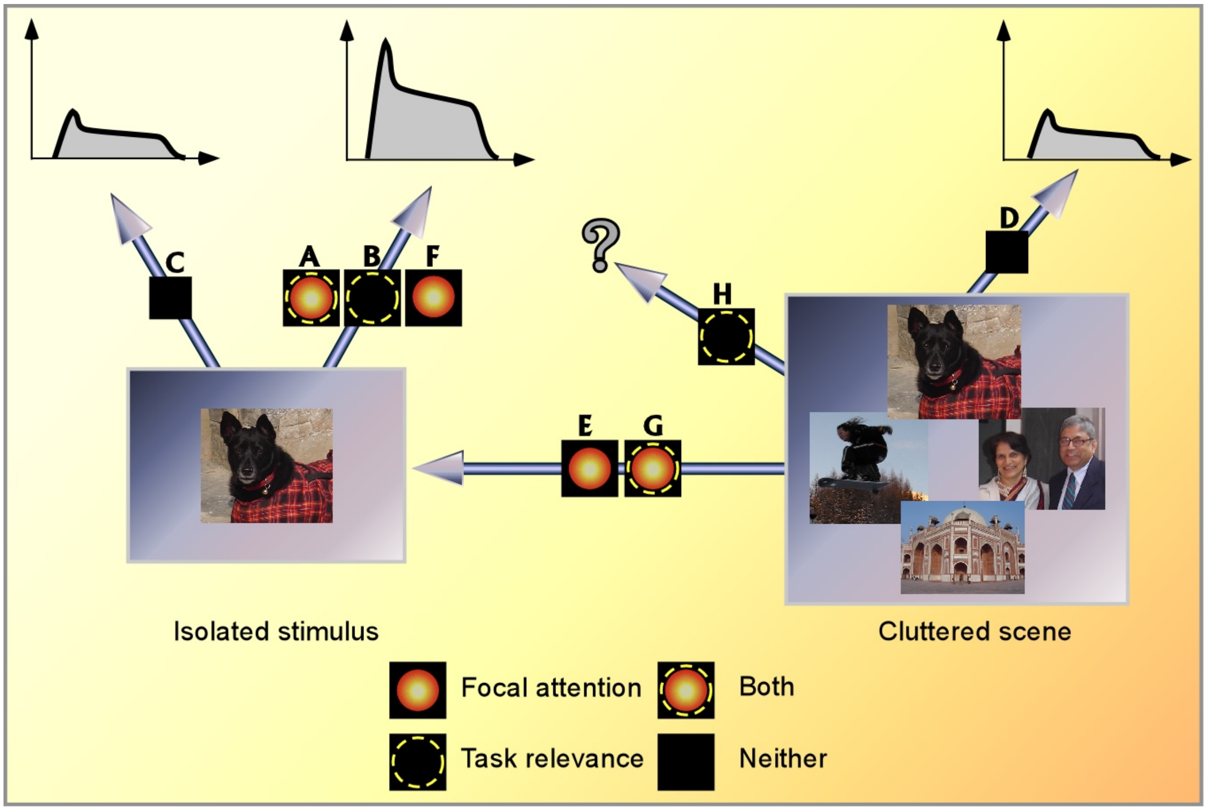


Figure 7.1 A proposed framework for the processing of natural stimuli when presented in isolation or in cluttered environments. Processing of stimuli in isolation is depicted on the left, and processing of stimuli in cluttered environments is shown on the right. Focal attention and top-down task relevance can have different effects on the neuronal representations of natural stimuli. For an isolated stimulus, the influence of focal attention and/or behavioral relevance is sufficient for eliciting high levels of cortical activity (A, B, F). If neither influence is available, neuronal signals are weaker (C). In cluttered scenes, the neuronal signal is weakened as a result of competition between stimuli (D). However, shifting attention to an object filters out the influence of unattended objects. This has virtually the same effect as presenting an object in isolation (E, G). The effect of top-down behavioral relevance to cluttered scenes in the absence of focal attention is an open question, since our experiments do not allow us to speak to this issue (H).

Appendix I: Looking a little closer at discriminating faces and bisected color disks

1.1 Introduction

Throughout the early chapters of this thesis where the dual-task paradigm was used to study attention, we stressed the importance of verifying that the central letter discrimination task was effective at engaging the attentional resource. To make certain that attention was made unavailable to the periphery, we used a bisected red-green discrimination task and demonstrated that in the dual-task condition, performance on this task was dramatically impaired. In contrast, subjects achieved high levels of performance on face-processing tasks in the near-absence of attention. From a computational perspective, this result is very surprising because a red-green bisected disk discrimination task (telling a red-green disk from a green-red one) can be easily accomplished with a few lines of Matlab code. On the other hand, for current machine vision algorithms, face or object recognition is still a very difficult problem.

Given the efficiency of natural stimulus processing and the dismal failure of the visual system at discriminating the red-green colored disks, our question still remains: For natural stimuli, where does the ability to efficiently process them break down? In the earlier chapters of this thesis, we had tried to identify the limits of pre-attentive processing by making the face tasks more difficult (an approach that admittedly is based on rather subjective criteria for difficulty). In this section, we instead try to make the face discrimination task more similar to disk discrimination task—the rationale being that, at some point in the continuum between face discrimination and disk discrimination, face task performance will have to break down in the absence of attention. In what follows, I will present some very preliminary data from two experiments that were conducted in an attempt to address this question.

In the first experiment, we asked whether the symmetry of the disk-discrimination task was in part responsible for the low levels of behavioral performance. In other words, did the task's difficulty lie in the fact that the visual system is not adept at distinguishing between two objects that are mirror images of each other? To test this hypothesis, three subjects performed a dual-task experiment with the same setup as described in Chapter 2. The only difference in this experiment was that subjects were presented with the 45 degree viewpoints of faces (relative to an axis in the plane of the image; in other words, faces were seen somewhat in profile), and instead of reporting the gender of the faces, they had to report if the face was facing leftwards or rightwards. Thus this task involved discriminating between symmetrical views of the face stimuli.

The normalized performance from this experiment is presented in Figure A1 below. The figure shows that all three subjects achieved high performance on discriminating face viewpoints (mean normalized performance = 99.8 ± 7.8%). The performance of each subject was also not significantly different between single and dual-task conditions ($p=.07$, $p=.09$, $p=1$ for the three subjects). The results of this experiment demonstrate that in contrast to bisected color disks, making symmetry judgments of face stimuli is relatively effortless in the near-absence of attention.

What then is the key factor that differentiates disk processing from face processing? Is it that the red-green disks are defined in the red-green color domain, while the faces are characterized by natural colors and strong luminance variations? To address this question, I will present data from a second experiment in which we carried out two orthogonal manipulations: black-white bisected disk discrimination and a red-green face orientation judgment.

In our second experiment, we attempted to reduce the information of the face stimuli to the domain of red and green colors, of luminance comparable to the red-green disks that have been used in several experiments described in this thesis (Chapters 2, 3, and 4). This was accomplished by converting all color information from the original faces to grayscale values and then converting gray levels such that black was mapped to red and white to green. Examples of the face stimuli and mask we used are shown in Figure A2a. In separate sessions, subjects also discriminated between vertically bisected black-white and white-black disks.

Two subjects performed left-right view-discrimination task with these face stimuli in the dual-task paradigm. They were trained for 2 hours on this task before data was collected. Their normalized behavioral performance on the red-green face and white-black disk discrimination tasks is shown in Figure A2b. In comparison to the data presented in Chapters 2 and 3, subjects did show significant drops in performance on red-green face discrimination in the dual-task condition ($p<.007$). However, it must be noted, that the performance with these faces (mean normalized performance = 85.0 ± 7.0%) was still substantially higher than performance on the black-white disk discrimination task (mean normalized performance = 51.7 ± 1.0%).

It is important to emphasize again that the data presented in this appendix is preliminary, and I do not intend to make any decisive conclusions about it. However, these preliminary results do indicate that low-level differences between the disk and face tasks are not responsible for the amazing discrepancy observed in behavioral performance. Red-green disks are not more difficult for the visual system because they involve vertical orientation judgments between symmetric alternatives (Figure A.1), nor because they are defined by colors targeting primarily the parvocellular pathways, rather than luminance contrasts (optimized for the magnocellular pathways) (Figure A.2). Instead, the data presented here lends support to the notion that the critical difference between the two tasks lies in how “meaningful” these stimuli are to the visual system. Once in the red-green domain, the main difference between faces and disks is that for the faces, the distribution of red and green pixels defines something meaningful for the visual system, but the same is not true for the disk stimuli. At the neuronal level, this might imply that stimuli, which are explicitly represented at the neuronal level, escape stringent attentional requirements.

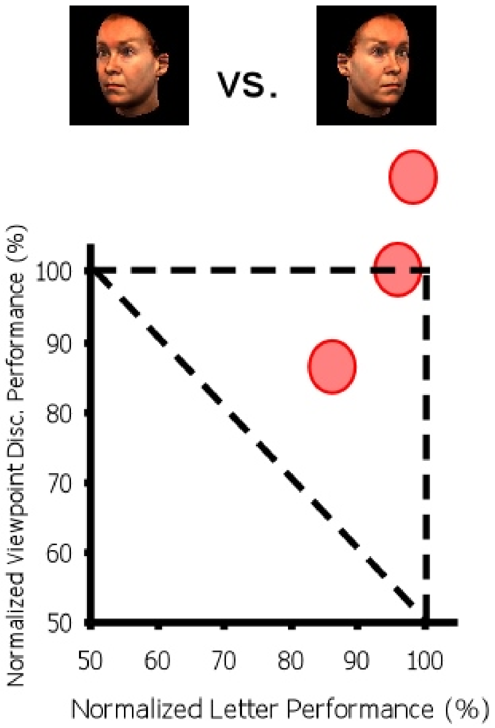


Figure A.1 Performance on view-point discrimination with faces. The data for 3 subjects on this task is shown here. Subjects achieve a high level of performance on this task in the dual-task condition (mean normalized performance = 99.8 ± 7.8%), demonstrating that discriminating between symmetric views of faces in the near-absence of attention is not a difficult problem for the visual system. These subjects had been previously trained in the face-gender discrimination task and performed this task without any prior training.

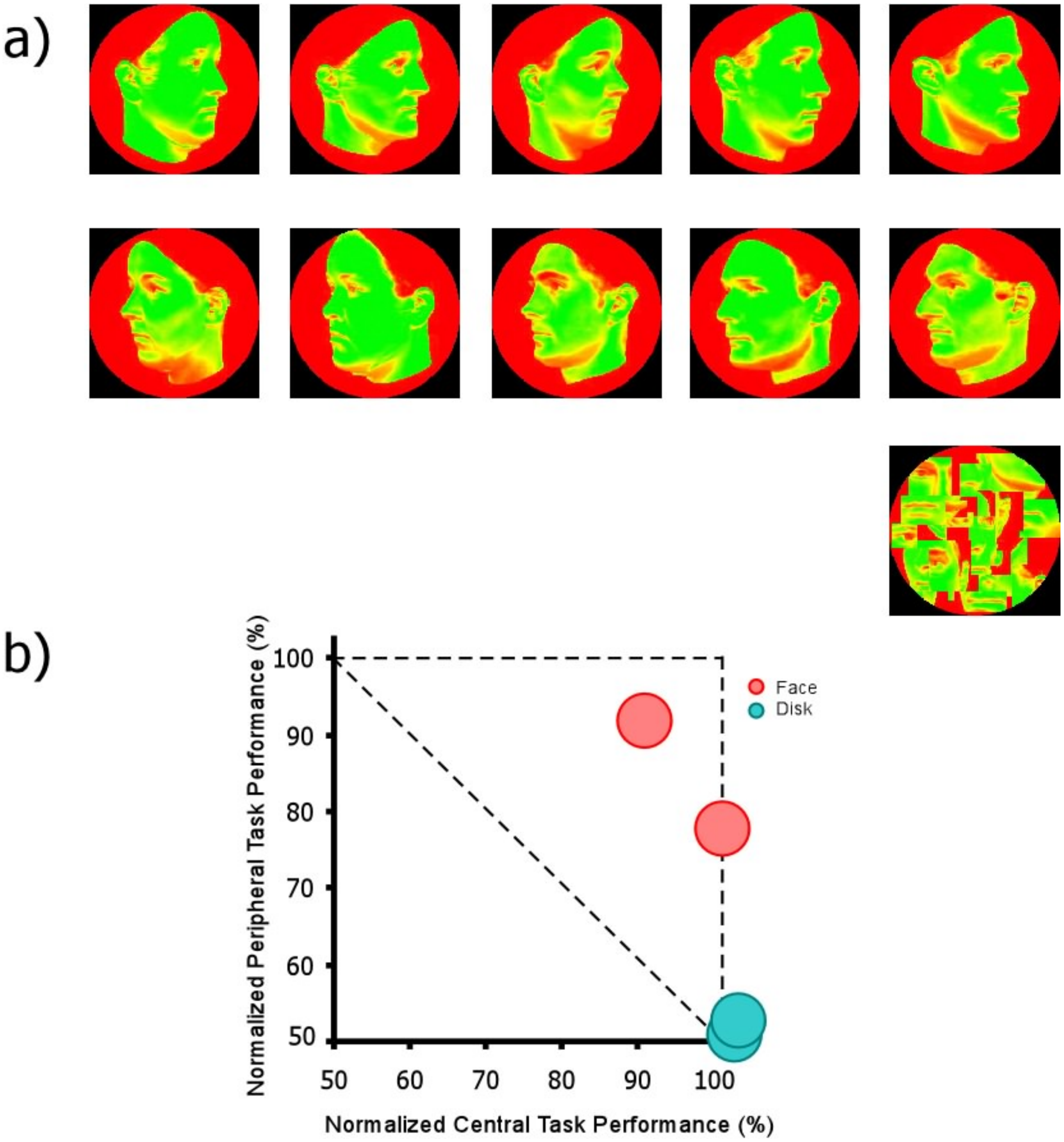


Figure A.2 Performance on red-green face and black-white disk discrimination. a) Examples of the face stimuli and the mask used for this experiment. The original faces (e.g. Figure 2.2) were converted to grayscale values and then gray levels were converted such that black mapped to red, and white to green. b) The data for 2 subjects on this task is shown here. Subjects achieve a high level of performance on the red-green face viewpoint discrimination task in the dual-task condition (mean normalized performance = 85.0 ± 7.0%). However, performance on a vertically bisected black-white disk discrimination task is at chance (mean normalized performance = 51.7 ± 1.0%). These two subjects, who had been previously trained on the face-gender discrimination task, were trained on the two tasks shown here for 2 hours. The data shown here was collected over 5 hours (15 blocks of each condition for each task).

References

1. Abbott, L. F., & Nelson, S. B. (2000). Synaptic plasticity: Taming the beast. *Nat Neurosci, 3 Suppl*, 1178-1183.

2. Aguirre, G. K., Detre, J. A., Alsop, D. C., & Desposito, M. (1996). The parahippocampus subserves topographical learning in man. *Cerebral Cortex, 6*(6), 823-829.

3. Allport, D. A. (1980). Attention and performance. In G. Claxton (Ed.), *Cognitive psychology: New directions* (pp. 43-67). London: Routledge & Kegan Paul.

4. Anllo-Vento, L., Luck, S. J., & Hillyard, S. A. (1998). Spatio-temporal dynamics of attention to color: Evidence from human electrophysiology. *Hum Brain Mapp, 6*(4), 216-238.

5. Awh, E., & Pashler, H. (2000). Evidence for split attentional foci. *J Exp Psychol Hum Percept Perform, 26*(2), 834-846.

6. Baker, J. F., Petersen, S. E., Newsome, W. T., & Allman, J. M. (1981). Visual response properties of neurons in 4 extrastriate visual areas of the owl monkey (aotus-trivirgatus)—a quantitative comparison of medial, dorsomedial, dorsolateral, and middle temporal areas. *Journal Of Neurophysiology, 45*(3), 397-416.

7. Bar, M., Tootell, R. B., Schacter, D. L., Greve, D. N., Fischl, B., Mendola, J. D., et al. (2001). Cortical mechanisms specific to explicit visual object recognition. *Neuron, 29*(2), 529-535.

8. Barlow, H. B. (1972). Single units and sensation: A neuron doctrine for perceptual psychology? *Perception, 1*(4), 371-394.

9. Baudouin, J. Y., & Tiberghien, G. (2002). Gender is a dimension of face recognition. *J Exp Psychol Learn Mem Cogn, 28*(2), 362-365.

10. Baxter, M. G., & Murray, E. A. (2001). Opposite relationship of hippocampal and rhinal cortex damage to delayed nonmatching-to-sample deficits in monkeys. *Hippocampus, 11*(1), 61-71.

11. Baylis, G. C., & Driver, J. (1993). Visual-attention and objects—evidence for hierarchical coding of location. *Journal Of Experimental Psychology-Human Perception And Performance, 19*(3), 451-470.

12. Beck, D. M., Rees, G., Frith, C. D., & Lavie, N. (2001). Neural correlates of change detection and change blindness. *Nat Neurosci, 4*(6), 645-650.

13. Bergen, J. R., & Julesz, B. (1983). Parallel versus serial processing in rapid pattern discrimination. *Nature, 303*(5919), 696-698.

14. Bi, G., & Poo, M. (2001). Synaptic modification by correlated activity: Hebb's postulate revisited. *Annu Rev Neurosci, 24*, 139-166.

15. Bogen, J. E., & Berker, E. (2002). Face modules, face network: The cognitive architecture of the brain revealed through studies of face processing. *Neurology, 59*(4), 652-653; discussion 653.

16. Braun, J. (1994). Visual search among items of different salience: Removal of visual attention mimics a lesion in extrastriate area v4. *J Neurosci, 14*(2), 554-567.

17. Braun, J. (2003). Natural scenes upset the visual applecart. *Trends Cogn Sci, 7*(1), 7-9.

18. Braun, J., & Julesz, B. (1998). Withdrawing attention at little or no cost: Detection and discrimination tasks. *Percept Psychophys, 60*(1), 1-23.

19. Braun, J., & Sagi, D. (1990). Vision outside the focus of attention. *Percept Psychophys, 48*(1), 45-58.

20. Braun, J., & Sagi, D. (1991). Texture-based tasks are little affected by second tasks requiring peripheral or central attentive fixation. *Perception, 20*(4), 483-500.

21. Brefczynski, J. A., & DeYoe, E. A. (1999). A physiological correlate of the 'spotlight' of visual attention. *Nat Neurosci, 2*(4), 370-374.

22. Bridgeman, B., Hendry, D., & Stark, L. (1975). Failure to detect displacement of the visual world during saccadic eye movements. *Vision Res, 15*(6), 719-722.

23. Britten, K. H., Newsome, W. T., Shadlen, M. N., Celebrini, S., & Movshon, J. A. (1996). A relationship between behavioral choice and the visual responses of neurons in macaque mt. *Vis Neurosci, 13*(1), 87-100.

24. Britten, K. H., Shadlen, M. N., Newsome, W. T., & Movshon, J. A. (1992). The analysis of visual motion: A comparison of neuronal and psychophysical performance. *J Neurosci, 12*(12), 4745-4765.

25. Brown, E., & Perrett, D. I. (1993). What gives a face its gender? *Perception, 22*(7), 829-840.

26. Brown, M. W., & Aggleton, J. P. (2001). Recognition memory: What are the roles of the perirhinal cortex and hippocampus? *Nat Rev Neurosci, 2*(1), 51-61.

27. Brown, V., Huey, D., & Findlay, J. M. (1997). Face detection in peripheral vision: Do faces pop out? *Perception, 26*(12), 1555-1570.

28. Bruce, C. (1982). Face recognition by monkeys: Absence of an inversion effect. *Neuropsychologia, 20*(5), 515-521.

29. Bruce, V., Burton, A. M., Hanna, E., Healey, P., Mason, O., Coombes, A., et al. (1993). Sex discrimination: How do we tell the difference between male and female faces? *Perception, 22*(2), 131-152.

30. Bruce, V., & Young, A. (1986). Understanding face recognition. *Br J Psychol, 77 (Pt 3)*, 305-327.

31. Bruyer, R., Galvez, C., & Prairial, C. (1993). Effect of disorientation on visual analysis, familiarity decision and semantic decision on faces. *Br J Psychol, 84 (Pt 4)*, 433-441.

32. Buckner, R. L., & Wheeler, M. E. (2001). The cognitive neuroscience of remembering. *Nat Rev Neurosci, 2*(9), 624-634.

33. Burgess, N., Maguire, E. A., & O'Keefe, J. (2002). The human hippocampus and spatial and episodic memory. *Neuron, 35*(4), 625-641.

34. Burr, D. C., Morrone, M. C., & Ross, J. (1994). Selective suppression of the magnocellular visual pathway during saccadic eye movements. *Nature, 371*(6497), 511-513.

35. Cairney, P. T. (1975). Bisensory order judgement and the prior entry hypothesis. *Acta Psychol (Amst), 39*(5), 329-340.

36. Campbell, R., Benson, P. J., Wallace, S. B., Doesbergh, S., & Coleman, M. (1999). More about brows: How poses that change brow position affect perceptions of gender. *Perception, 28*(4), 489-504.

37. Castet, E., & Masson, G. S. (2000). Motion perception during saccadic eye movements. *Nat Neurosci, 3*(2), 177-183.

38. Celebrini, S., & Newsome, W. T. (1994). Neuronal and psychophysical sensitivity to motion signals in extrastriate area mst of the macaque monkey. *J Neurosci, 14*(7), 4109-4124.

39. Chao, L. L., Martin, A., & Haxby, J. V. (1999). Are face-responsive regions selective only for faces? *Neuroreport, 10*(14), 2945-2950.

40. Chelazzi, L. (1995). Neural mechanisms for stimulus selection in cortical areas of the macaque subserving object vision. *Behav Brain Res, 71*(1-2), 125-134.

41. Chelazzi, L., Miller, E. K., Duncan, J., & Desimone, R. (2001). Responses of neurons in macaque area v4 during memory-guided visual search. *Cereb Cortex, 11*(8), 761-772.

42. Cheng, K., Saleem, K. S., & Tanaka, K. (1997). Organization of corticostriatal and corticoamygdalar projections arising from the anterior inferotemporal

area te of the macaque monkey: A phaseolus vulgaris leucoagglutinin study. *J Neurosci, 17*(20), 7902-7925.

43. Clark, R. E., West, A. N., Zola, S. M., & Squire, L. R. (2001). Rats with lesions of the hippocampus are impaired on the delayed nonmatching-to-sample task. *Hippocampus, 11*(2), 176-186.

44. Connor, C. E., Preddie, D. C., Gallant, J. L., & Van Essen, D. C. (1997). Spatial attention effects in macaque area v4. *J Neurosci, 17*(9), 3201-3214.

45. Cook, E. P., & Maunsell, J. H. (2002). Dynamics of neuronal responses in macaque mt and vip during motion detection. *Nat Neurosci, 5*(10), 985-994.

46. Crick, F., & Koch, C. (2003). A framework for consciousness. *Nat Neurosci, 6*(2), 119-126.

47. Currie, C., McConkie, G. W., Carlson-Radvansky, L. A., & Irwin, D. E. (1995). *Maintaining visual stability across saccades: Role of the saccade target object.*: Beckman Institute, University of Illinois.

48. Cutzu, F., & Tsotsos, J. K. (2003). The selective tuning model of attention: Psychophysical evidence for a suppressive annulus around an attended item. *Vision Res, 43*(2), 205-219.

49. Desimone, R. (1996). Neural mechanisms for visual memory and their role in attention. *Proc Natl Acad Sci U S A, 93*(24), 13494-13499.

50. Desimone, R., Albright, T. D., Gross, C. G., & Bruce, C. (1984). Stimulus-selective properties of inferior temporal neurons in the macaque. *J Neurosci, 4*(8), 2051-2062.

51. Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annu Rev Neurosci, 18*, 193-222.

52. DiCarlo, J. J., & Maunsell, J. H. (2000). Form representation in monkey inferotemporal cortex is virtually unaltered by free viewing. *Nat Neurosci, 3*(8), 814-821.

53. Drachman, D. A., & Arbit, J. (1966). Memory and the hippocampal complex. Ii. Is memory a multiple process? *Arch Neurol, 15*(1), 52-61.

54. Dubois, S., Rossion, B., Schiltz, C., Bodart, J. M., Michel, C., Bruyer, R., et al. (1999). Effect of familiarity on the processing of human faces. *Neuroimage, 9*(3), 278-289.

55. Duncan, J. (1980a). Demonstration of capacity limitation. *Cognit Psychol, 12*(1), 75-96.

56. Duncan, J. (1980b). The locus of interference in the perception of simultaneous stimuli. *Psychol Rev, 87*(3), 272-300.

57. Duncan, J. (1984). Selective attention and the organization of visual information. *Journal Of Experimental Psychology-General, 113*(4), 501-517.

58. Eichenbaum, H. (1992). The hippocampal system and declarative memory in animals. *Journal Of Cognitive Neuroscience, 4*(3), 217-231.

59. Eichenbaum, H. (2000). A cortical-hippocampal system for declarative memory. *Nature Reviews Neuroscience, 1*(1), 41-50.

60. Ellis, H. D. (1986). Processes underlying face recognition. In R. Bruyer (Ed.), *The neuropsychology of face perception and facial expression.* (pp. 1-27). Hillsdale (NJ): Lawrence Erlbaum.

61. Engel, J. (1996). Current concepts—surgery for seizures. *New England Journal Of Medicine, 334*(10), 647-652.

62. Epstein, R., Graham, K. S., & Downing, P. E. (2003). Viewpoint-specific scene representations in human parahippocampal cortex. *Neuron, 37*(5), 865-876.

63. Epstein, R., Harris, A., Stanley, D., & Kanwisher, N. (1999). The parahippocampal place area: Recognition, navigation, or encoding? *Neuron, 23*(1), 115-125.

64. Epstein, R., & Kanwisher, N. (1998). A cortical representation of the local visual environment. *Nature, 392*(6676), 598-601.

65. Eriksen, C. W., & St James, J. D. (1986). Visual attention within and around the field of focal attention: A zoom lens model. *Percept Psychophys, 40*(4), 225-240.

66. Eriksen, C. W., & Yeh, Y. Y. (1985). Allocation of attention in the visual field. *J Exp Psychol Hum Percept Perform, 11*(5), 583-597.

67. Farah, M. J. (1995). Dissociable systems for visual recognition: A cognitive neuropsychology approach. In S. M. Kosslyn, & Osherson, D.N. (Ed.), *Visual cognition: An invitation to cognitive science* (2 ed., Vol. 2, pp. 101-119). Cambridge, MA: MIT Press.

68. Farah, M. J., Wilson, K. D., Drain, M., & Tanaka, J. N. (1998). What is "special" about face perception? *Psychol Rev, 105*(3), 482-498.

69. Felleman, D. J., & Van Essen, D. C. (1991). Distributed hierarchical processing in the primate cerebral cortex. *Cereb Cortex, 1*(1), 1-47.

70. Fried, I., MacDonald, K. A., & Wilson, C. L. (1997). Single neuron activity in human hippocampus and amygdala during recognition of faces and objects. *Neuron, 18*(5), 753-765.

71. Fried, I., Wilson, C. L., Maidment, N. T., Engel, J., Behnke, E., Fields, T. A., et al. (1999). Cerebral microdialysis combined with single-neuron and electroencephalographic recording in neurosurgical patients—technical note. *Journal Of Neurosurgery, 91*(4), 697-705.

72. Gauthier, I., Curran, T., Curby, K. M., & Collins, D. (2003). Perceptual interference supports a non-modular account of face processing. *Nat Neurosci, 6*(4), 428-432.

73. Gauthier, I., & Logothetis, N. K. (2000). Is face recognition not so unique after all? *Cognitive Neuropsychology, 17*(1-3), 125-142.

74. Gauthier, I., & Nelson, C. A. (2001). The development of face expertise. *Curr Opin Neurobiol, 11*(2), 219-224.

75. Gauthier, I., Skudlarski, P., Gore, J. C., & Anderson, A. W. (2000). Expertise for cars and birds recruits brain areas involved in face recognition. *Nat Neurosci, 3*(2), 191-197.

76. Gauthier, I., & Tarr, M. J. (1997). Becoming a "greeble" expert: Exploring mechanisms for face recognition. *Vision Res, 37*(12), 1673-1682.

77. Gauthier, I., Tarr, M. J., Anderson, A. W., Skudlarski, P., & Gore, J. C. (1999). Activation of the middle fusiform 'face area' increases with expertise in recognizing novel objects. *Nat Neurosci, 2*(6), 568-573.

78. George, N., Dolan, R. J., Fink, G. R., Baylis, G. C., Russell, C., & Driver, J. (1999). Contrast polarity and face recognition in the human fusiform gyrus. *Nat Neurosci, 2*(6), 574-580.

79. Georgopoulos, A. P. (1987). Cortical mechanisms subserving reaching. *Ciba Found Symp, 132*, 125-141.

80. Georgopoulos, A. P., Kalaska, J. F., Caminiti, R., & Massey, J. T. (1982). On the relations between the direction of two-dimensional arm movements and cell discharge in primate motor cortex. *J Neurosci, 2*(11), 1527-1537.

81. Georgopoulos, A. P., Schwartz, A. B., & Kettner, R. E. (1986). Neuronal population coding of movement direction. *Science, 233*(4771), 1416-1419.

82. Gorno-Tempini, M. L., & Price, C. J. (2001). Identification of famous faces and buildings: A functional neuroimaging study of semantically unique items. *Brain, 124*(Pt 10), 2087-2097.

83. Green, D., & Swets, J. (1966). *Signal detection theory and psychophysics.* New York: Wiley.

84. Grill-Spector, K., Knouf, N., & Kanwisher, N. (2004). The fusiform face area subserves face perception, not generic within-category identification. *Nat Neurosci, 7*(5), 555-562.

85. Grill-Spector, K., Kushnir, T., Edelman, S., Itzchak, Y., & Malach, R. (1998). Cue-invariant activation in object-related areas of the human occipital lobe. *Neuron, 21*(1), 191-202.

86. Grill-Spector, K., Kushnir, T., Hendler, T., & Malach, R. (2000). The dynamics of object-selective activation correlate with recognition performance in humans. *Nat Neurosci, 3*(8), 837-843.

87. Grimes, J. (1996). On the failure to detect changes in scenes across saccades. In K. Akins (Ed.), *Perception* (Vol. 5). New York: Oxford University Press.

88. Gross, C. G. (2002). Genealogy of the "grandmother cell". *Neuroscientist, 8*(5), 512-518.

89. Gross, C. G., Bender, D. B., & Rocha-Miranda, C. E. (1969). Visual receptive fields of neurons in inferotemporal cortex of the monkey. *Science, 166*(910), 1303-1306.

90. Gross, C. G., Rocha-Miranda, C. E., & Bender, D. B. (1972). Visual properties of neurons in inferotemporal cortex of the macaque. *J Neurophysiol, 35*(1), 96-111.

91. Haenny, P. E., Maunsell, J. H., & Schiller, P. H. (1988). State dependent activity in monkey visual cortex. Ii. Retinal and extraretinal factors in v4. *Exp Brain Res, 69*(2), 245-259.

92. Halgren, E., Dale, A. M., Sereno, M. I., Tootell, R. B., Marinkovic, K., & Rosen, B. R. (1999). Location of human face-selective cortex with respect to retinotopic areas. *Hum Brain Mapp, 7*(1), 29-37.

93. Hansen, C. H., & Hansen, R. D. (1988). Finding the face in the crowd: An anger superiority effect. *J Pers Soc Psychol, 54*(6), 917-924.

94. Hasselmo, M. E., Rolls, E. T., & Baylis, G. C. (1989). The role of expression and identity in the face-selective responses of neurons in the temporal visual cortex of the monkey. *Behav Brain Res, 32*(3), 203-218.

95. Hasson, U., Harel, M., Levy, I., & Malach, R. (2003). Large-scale mirror-symmetry organization of human occipito-temporal object areas. *Neuron, 37*(6), 1027-1041.

96. Haxby, J. V., Gobbini, M. I., Furey, M. L., Ishai, A., Schouten, J. L., & Pietrini, P. (2001). Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science, 293*(5539), 2425-2430.

97. Haxby, J. V., Hoffman, E. A., & Gobbini, M. I. (2000). The distributed human neural system for face perception. *Trends Cogn Sci, 4*(6), 223-233.

98. Haxby, J. V., Hoffman, E. A., & Gobbini, M. I. (2002). Human neural systems for face recognition and social communication. *Biol Psychiatry, 51*(1), 59-67.

99. Haxby, J. V., Horwitz, B., Ungerleider, L. G., Maisog, J. M., Pietrini, P., & Grady, C. L. (1994). The functional organization of human extrastriate cortex: A pet-rcbf study of selective attention to faces and locations. *J Neurosci, 14*(11 Pt 1), 6336-6353.

100. Haxby, J. V., Ungerleider, L. G., Clark, V. P., Schouten, J. L., Hoffman, E. A., & Martin, A. (1999). The effect of face inversion on activity in human neural systems for face and object perception. *Neuron, 22*(1), 189-199.

101. Higuchi, S., & Miyashita, Y. (1996). Formation of mnemonic neuronal responses to visual paired associates in inferotemporal cortex is impaired by perirhinal and entorhinal lesions. *Proceedings Of The National Academy Of Sciences Of The United States Of America, 93*(2), 739-743.

102. Hikosaka, O., Miyauchi, S., & Shimojo, S. (1991). Focal visual-attention produces motion sensation in lines. *Investigative Ophthalmology & Visual Science, 32*(4), 716-716.

103. Hikosaka, O., Miyauchi, S., & Shimojo, S. (1993). Focal visual attention produces illusory temporal order and motion sensation. *Vision Res, 33*(9), 1219-1240.

104. Hillyard, S. A., & Anllo-Vento, L. (1998). Event-related brain potentials in the study of visual selective attention. *Proc Natl Acad Sci U S A, 95*(3), 781-787.

105. Hochstein, S., & Ahissar, M. (2002). View from the top: Hierarchies and reverse hierarchies in the visual system. *Neuron, 36*(5), 791-804.

106. Hoffman, E. A., & Haxby, J. V. (2000). Distinct representations of eye gaze and identity in the distributed human neural system for face perception. *Nat Neurosci, 3*(1), 80-84.

107. Insausti, R., Amaral, D. G., & Cowan, W. M. (1987). The entorhinal cortex of the monkey: Ii. Cortical afferents. *J Comp Neurol, 264*(3), 356-395.

108. Ishai, A., Haxby, J. V., & Ungerleider, L. G. (2002). Visual imagery of famous faces: Effects of memory and attention revealed by fmri. *Neuroimage, 17*(4), 1729-1741.

109. Ishai, A., Ungerleider, L. G., Martin, A., & Haxby, J. V. (2000). The representation of objects in the human occipital and temporal cortex. *J Cogn Neurosci, 12 Suppl 2*, 35-51.

110. Ishai, A., Ungerleider, L. G., Martin, A., Schouten, J. L., & Haxby, J. V. (1999). Distributed representation of objects in the human ventral visual pathway. *Proc Natl Acad Sci U S A, 96*(16), 9379-9384.

111. Itti, L., & Koch, C. (2001). Computational modelling of visual attention. *Nat Rev Neurosci, 2*(3), 194-203.

112. Jenkinson, M., Bannister, P., Brady, M., & Smith, S. (2002). Improved optimization for the robust and accurate linear registration and motion correction of brain images. *Neuroimage, 17*(2), 825-841.

113. Johansson, R. S., & Vallbo, A. B. (1979). Detection of tactile stimuli. Thresholds of afferent units related to psychophysical thresholds in the human hand. *J Physiol, 297*(0), 405-422.

114. Kandel, E. R., & Hawkins, R. D. (1992). The biological basis of learning and individuality. *Sci Am, 267*(3), 78-86.

115. Kandel, E. R., Schwartz, J., & Jessell, T. (2000). *Principles of neural science.* New York: McGraw-Hill.

116. Kanwisher, N. (2000). Domain specificity in face perception. *Nat Neurosci, 3*(8), 759-763.

117. Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: A module in human extrastriate cortex specialized for face perception. *J Neurosci, 17*(11), 4302-4311.

118. Kanwisher, N., Stanley, D., & Harris, A. (1999). The fusiform face area is selective for faces not animals. *Neuroreport, 10*(1), 183-187.

119. Kanwisher, N., Woods, R. P., Iacoboni, M., & Mazziotta, J. C. (1997). A locus in human extrastriate cortex for visual shape analysis. *Journal Of Cognitive Neuroscience, 9*(1), 133-142.

120. Kapur, N., & Brooks, D. J. (1999). Temporally-specific retrograde amnesia in two cases of discrete bilateral hippocampal pathology. *Hippocampus, 9*(3), 247-254.

121. Kastner, S., De Weerd, P., Desimone, R., & Ungerleider, L. G. (1998). Mechanisms of directed attention in the human extrastriate cortex as revealed by functional mri. *Science, 282*(5386), 108-111.

122. Kastner, S., Pinsk, M. A., De Weerd, P., Desimone, R., & Ungerleider, L. G. (1999). Increased activity in human visual cortex during directed attention in the absence of visual stimulation. *Neuron, 22*(4), 751-761.

123. Kastner, S., & Ungerleider, L. G. (2000). Mechanisms of visual attention in the human cortex. *Annu Rev Neurosci, 23*, 315-341.

124. Kayser, C., Kording, K. P., & Konig, P. (2004). Processing of complex stimuli and natural scenes in the visual cortex. *Curr Opin Neurobiol, 14*(4), 468-473.

125. Kelley, T. A., Chun, M. M., & Chua, K. P. (2003). Effects of scene inversion on change detection of targets matched for visual salience. *J Vis, 3*(1), 1-5.

126. Klein, R., Kingstone, A., & Pontefract, A. (1992). Orienting of visual attention. In K. Rayner (Ed.), *Eye movements and visual cognition: Scene perception and reading.* New York.: Springer.

127. Klein, R. M. (2000). Inhibition of return. *Trends Cogn Sci, 4*(4), 138-147.

128. Kobatake, E., Wang, G., & Tanaka, K. (1998). Effects of shape-discrimination training on the selectivity of inferotemporal cells in adult monkeys. *Journal Of Neurophysiology, 80*(1), 324-330.

129. Koch, C. (2004). *The quest for consciousness: A neurobiological approach.* Englewood, CO: Roberts & Company Publishers.

130. Koch, C., & Ullman, S. (1985). Shifts in selective visual attention: Towards the underlying neural circuitry. *Hum Neurobiol, 4*(4), 219-227.

131. Konorski, J. (1967). *Integrative activity of the brain: An interdisciplinary approach*. Chicago: University of Chicago Press.

132. Kourtzi, Z., & Kanwisher, N. (2000). Cortical regions involved in perceiving object shape. *J Neurosci, 20*(9), 3310-3318.

133. Kourtzi, Z., & Kanwisher, N. (2001). Representation of perceived object shape by the human lateral occipital complex. *Science, 293*(5534), 1506-1509.

134. Kreiman, G. (2002). *On the neuronal activity in the human brain during visual recognition, imagery and binocular rivalry.*, California Institute of Technology, Pasadena,CA.

135. Kreiman, G., Fried, I., & Koch, C. (2002). Single-neuron correlates of subjective vision in the human medial temporal lobe. *Proc Natl Acad Sci U S A, 99*(12), 8378-8383.

136. Kreiman, G., Koch, C., & Fried, I. (2000a). Category-specific visual responses of single neurons in the human medial temporal lobe. *Nat Neurosci, 3*(9), 946-953.

137. Kreiman, G., Koch, C., & Fried, I. (2000b). Imagery neurons in the human brain. *Nature, 408*(6810), 357-361.

138. Kuehn, S. M., & Jolicoeur, P. (1994). Impact of quality of the image, orientation, and similarity of the stimuli on visual search for faces. *Perception, 23*(1), 95-122.

139. Lavenex, P., & Amaral, D. G. (2000). Hippocampal-neocortical interaction: A hierarchy of associativity. *Hippocampus, 10*(4), 420-430.

140. LeDoux, J. E. (2000). Emotion circuits in the brain. *Annu Rev Neurosci, 23*, 155-184.

141. Lee, D. K., Koch, C., & Braun, J. (1997). Spatial vision thresholds in the near absence of attention. *Vision Res, 37*(17), 2409-2418.

142. Lee, D. K., Koch, C., & Braun, J. (1999). Attentional capacity is undifferentiated: Concurrent discrimination of form, color, and motion. *Percept Psychophys, 61*(7), 1241-1255.

143. Leonard, C. M., Rolls, E. T., Wilson, F. A., & Baylis, G. C. (1985). Neurons in the amygdala of the monkey with responses selective for faces. *Behav Brain Res, 15*(2), 159-176.

144. Leveroni, C. L., Seidenberg, M., Mayer, A. R., Mead, L. A., Binder, J. R., & Rao, S. M. (2000). Neural systems underlying the recognition of familiar and newly learned faces. *J Neurosci, 20*(2), 878-886.

145. Levy, I., Hasson, U., Avidan, G., Hendler, T., & Malach, R. (2001). Center-periphery organization of human object areas. *Nat Neurosci, 4*(5), 533-539.

146. Li, F. F., VanRullen, R., Koch, C., & Perona, P. (2002). Rapid natural scene categorization in the near absence of attention. *Proc Natl Acad Sci U S A, 99*(14), 9596-9601.

147. Liu, J., Harris, A., & Kanwisher, N. (2002). Stages of processing in face perception: An meg study. *Nat Neurosci, 5*(9), 910-916.

148. Logothetis, N. K., Pauls, J., Augath, M., Trinath, T., & Oeltermann, A. (2001). Neurophysiological investigation of the basis of the fmri signal. *Nature, 412*(6843), 150-157.

149. Logothetis, N. K., Pauls, J., & Poggio, T. (1995). Shape representation in the inferior temporal cortex of monkeys. *Current Biology, 5*(5), 552-563.

150. Logothetis, N. K., & Sheinberg, D. L. (1996). Visual object recognition. *Annual Review Of Neuroscience, 19*, 577-621.

151. Logothetis, N. K., & Wandell, B. A. (2004). Interpreting the bold signal. *Annu Rev Physiol, 66*, 735-769.

152. Lu, Z. L., & Dosher, B. A. (1998). External noise distinguishes attention mechanisms. *Vision Res, 38*(9), 1183-1198.

153. Luck, S. J., Chelazzi, L., Hillyard, S. A., & Desimone, R. (1997). Neural mechanisms of spatial selective attention in areas v1, v2, and v4 of macaque visual cortex. *J Neurophysiol, 77*(1), 24-42.

154. Lueschow, A., Miller, E. K., & Desimone, R. (1994). Inferior temporal mechanisms for invariant object recognition. *Cereb Cortex, 4*(5), 523-531.

155. Mack, A., & Rock, I. (1998). *Inattentional blindness*. Cambridge MA: MIT Press.

156. Malach, R., Levy, I., & Hasson, U. (2002). The topography of high-order human object areas. *Trends Cogn Sci, 6*(4), 176-184.

157. Malach, R., Reppas, J. B., Benson, R. R., Kwong, K. K., Jiang, H., Kennedy, W. A., et al. (1995). Object-related activity revealed by functional magnetic resonance imaging in human occipital cortex. *Proc Natl Acad Sci U S A, 92*(18), 8135-8139.

158. Malkova, L., & Mishkin, M. (2003). One-trial memory for object-place associations after separate lesions of hippocampus and posterior parahippocampal region in the monkey. *J Neurosci, 23*(5), 1956-1965.

159. Marois, R., Yi, D. J., & Chun, M. M. (2004). The neural fate of consciously perceived and missed events in the attentional blink. *Neuron, 41*(3), 465-472.

160. Martinez-Trujillo, J., & Treue, S. (2002). Attentional modulation strength in cortical area mt depends on stimulus contrast. *Neuron, 35*(2), 365-370.

161. Martinez, A., Anllo-Vento, L., Sereno, M. I., Frank, L. R., Buxton, R. B., Dubowitz, D. J., et al. (1999). Involvement of striate and extrastriate visual cortical areas in spatial attention. *Nat Neurosci, 2*(4), 364-369.

162. Maunsell, J. H., Sclar, G., Nealey, T. A., & DePriest, D. D. (1991). Extraretinal representations in area v4 in the macaque monkey. *Vis Neurosci, 7*(6), 561-573.

163. Maunsell, J. H. R., & VanEssen, D. C. (1983). Functional-properties of neurons in middle temporal visual area of the macaque monkey.1. Selectivity for stimulus direction, speed, and orientation. *Journal Of Neurophysiology, 49*(5), 1127-1147.

164. McAdams, C. J., & Maunsell, J. H. (1999a). Effects of attention on orientation-tuning functions of single neurons in macaque cortical area v4. *J Neurosci, 19*(1), 431-441.

165. McAdams, C. J., & Maunsell, J. H. (1999b). Effects of attention on the reliability of individual neurons in monkey visual cortex. *Neuron, 23*(4), 765-773.

166. McConkie, G. W., & Zola, D. (1979). Is visual information integrated across successive fixations in reading. *Perception & Psychophysics, 25*(3), 221-224.

167. McMains, S. A., & Somers, D. C. (2004). Multiple spotlights of attentional selection in human visual cortex. *Neuron, 42*(4), 677-686.

168. Mervis, C. B., & Rosch, E. (1981). Categorization of natural objects. *Annual Review Of Psychology, 32*, 89-115.

169. Miller, E. K., & Desimone, R. (1994). Parallel neuronal mechanisms for short-term memory. *Science, 263*(5146), 520-522.

170. Miller, E. K., Li, L., & Desimone, R. (1991). A neural mechanism for working and recognition memory in inferior temporal cortex. *Science, 254*(5036), 1377-1379.

171. Miller, E. K., Li, L., & Desimone, R. (1993). Activity of neurons in anterior inferior temporal cortex during a short- term memory task. *J. Neurosci., 13*(4), 1460-1478.

172. Milner, B. (1972). Disorders of learning and memory after temporal lobe lesions in man. *Clin Neurosurg, 19*, 421-446.

173. Mishkin, M. (1982). A memory system in the monkey. *Philos Trans R Soc Lond B Biol Sci, 298*(1089), 83-95.

174. Missal, M., Vogels, R., Li, C. Y., & Orban, G. A. (1999). Shape interactions in macaque inferior temporal neurons. *J Neurophysiol, 82*(1), 131-142.

175. Miyashita, Y. (1993). Inferior temporal cortex—where visual-perception meets memory. *Annual Review Of Neuroscience, 16*, 245-263.

176. Miyashita, Y. (2004). Cognitive memory: Cellular and network machineries and their top-down control. *Science, 306*(5695), 435-440.

177. Miyashita, Y., & Chang, H. S. (1988). Neuronal correlate of pictorial short-term memory in the primate temporal cortex. *Nature, 331*(6151), 68-70.

178. Miyashita, Y., & Hayashi, T. (2000). Neural representation of visual objects: Encoding and top-down activation. *Current Opinion In Neurobiology, 10*(2), 187-194.

179. Moran, J., & Desimone, R. (1985). Selective attention gates visual processing in the extrastriate cortex. *Science, 229*(4715), 782-784.

180. Motter, B. C. (1993). Focal attention produces spatially selective processing in visual cortical areas v1, v2, and v4 in the presence of competing stimuli. *J Neurophysiol, 70*(3), 909-919.

181. Motter, B. C. (1994a). Neural correlates of attentive selection for color or luminance in extrastriate area v4. *J Neurosci, 14*(4), 2178-2189.

182. Motter, B. C. (1994b). Neural correlates of feature selective memory and pop-out in extrastriate area v4. *J Neurosci, 14*(4), 2190-2199.

183. Mouchetant-Rostaing, Y., Giard, M. H., Bentin, S., Aguera, P. E., & Pernier, J. (2000). Neurophysiological correlates of face gender processing in humans. *Eur J Neurosci, 12*(1), 303-310.

184. Muller, M. M., Malinowski, P., Gruber, T., & Hillyard, S. A. (2003). Sustained division of the attentional spotlight. *Nature, 424*(6946), 309-312.

185. Muller, N. G., Bartelt, O. A., Donner, T. H., Villringer, A., & Brandt, S. A. (2003). A physiological correlate of the "zoom lens" of visual attention. *J Neurosci, 23*(9), 3561-3565.

186. Muller, R. U., Kubie, J. L., & Ranck, J. B. (1987). Spatial firing patterns of hippocampal complex-spike cells in a fixed environment. *Journal Of Neuroscience, 7*(7), 1935-1950.

187. Murray, E. A., & Richmond, B. J. (2001). Role of perirhinal cortex in object perception, memory, and associations. *Curr Opin Neurobiol, 11*(2), 188-193.

188. Murray, S. O., & Wojciulik, E. (2004). Attention increases neural selectivity in the human lateral occipital complex. *Nat Neurosci, 7*(1), 70-74.

189. Newsome, W. T., Britten, K. H., & Movshon, J. A. (1989). Neuronal correlates of a perceptual decision. *Nature, 341*(6237), 52-54.

190. Nothdurft, H. C. (1993). Faces and facial expressions do not pop out. *Perception, 22*(11), 1287-1298.

191. O'Connor, D. H., Fukui, M. M., Pinsk, M. A., & Kastner, S. (2002). Attention modulates responses in the human lateral geniculate nucleus. *Nat Neurosci, 5*(11), 1203-1209.

192. O'Craven, K. M., Downing, P. E., & Kanwisher, N. (1999). Fmri evidence for objects as the units of attentional selection. *Nature, 401*(6753), 584-587.

193. O'Keefe, J., & Conway, D. H. (1978). Hippocampal place units in freely moving rat—why they fire where they fire. *Experimental Brain Research, 31*(4), 573-590.

194. O'Keefe, J., & Dostrovsky, J. (1971). Hippocampus as a spatial map—preliminary evidence from unit activity in freely-moving rat. *Brain Research, 34*(1), 171-&.

195. O'Regan, J. K., Rensink, R. A., & Clark, J. J. (1999). Change-blindness as a result of 'mudsplashes'. *Nature, 398*(6722), 34.

196. Ojemann, G. A. (1997). Treatment of temporal lobe epilepsy. *Annual Review Of Medicine, 48*, 317-328.

197. Olshausen, B. A., & Field, D. J. (2004). Sparse coding of sensory inputs. *Curr Opin Neurobiol, 14*(4), 481-487.

198. Parker, A. J., & Newsome, W. T. (1998). Sense and the single neuron: Probing the physiology of perception. *Annu Rev Neurosci, 21*, 227-277.

199. Pashler, H. (1984). Processing stages in overlapping tasks—evidence for a central bottleneck. *Journal of Experimental Psychology—Human Perception and Performance, 10*(3), 358-377.

200. Pashler, H. (1988). Familiarity and visual change detection. *Percept Psychophys, 44*(4), 369-378.

201. Pashler, H. (1994). Dual-task interference in simple tasks—data and theory. *Psychological Bulletin, 116*(2), 220-244.

202. Penfield, W., & Milner, B. (1958). Memory deficit produced by bilateral lesions in the hippocampal zone. *AMA Arch Neurol Psychiatry, 79*(5), 475-497.

203. Perrett, D. I., Hietanen, J. K., Oram, M. W., & Benson, P. J. (1992). Organization and functions of cells responsive to faces in the temporal cortex. *Philos Trans R Soc Lond B Biol Sci, 335*(1273), 23-30.

204. Perrett, D. I., Mistlin, A. J., Chitty, A. J., Smith, P. A., Potter, D. D., Broennimann, R., et al. (1988). Specialized face processing and hemispheric asymmetry in man and monkey: Evidence from single unit and reaction time studies. *Behav Brain Res, 29*(3), 245-258.

205. Perrett, D. I., Mistlin, A.J., Chitty, A.J. (1987). Visual neurons responsive to faces. *Trends in Neurosciences, 10*(9), 358-364.

206. Perrett, D. I., Rolls, E. T., & Caan, W. (1982). Visual neurones responsive to faces in the monkey temporal cortex. *Exp Brain Res, 47*(3), 329-342.

207. Perrett, D. I., Smith, P. A., Potter, D. D., Mistlin, A. J., Head, A. S., Milner, A. D., et al. (1984). Neurones responsive to faces in the temporal cortex: Studies of functional organization, sensitivity to identity and relation to perception. *Hum Neurobiol, 3*(4), 197-208.

208. Perrett, D. I., Smith, P. A., Potter, D. D., Mistlin, A. J., Head, A. S., Milner, A. D., et al. (1985). Visual cells in the temporal cortex sensitive to face view and gaze direction. *Proc R Soc Lond B Biol Sci, 223*(1232), 293-317.

209. Pessoa, L., McKenna, M., Gutierrez, E., & Ungerleider, L. G. (2002). Neural processing of emotional faces requires attention. *Proc Natl Acad Sci U S A, 99*(17), 11458-11463.

210. Pessoa, L., & Ungerleider, L. G. (2004). Neural correlates of change detection and change blindness in a working memory task. *Cereb Cortex, 14*(5), 511-520.

211. Phillips, W. A. (1974). Distinction between sensory storage and short-term visual memory. *Perception & Psychophysics, 16*(2), 283-290.

212. Pinsk, M. A., Doniger, G. M., & Kastner, S. (2004). Push-pull mechanism of selective attention in human extrastriate cortex. *J Neurophysiol, 92*(1), 622-629.

213. Posner, M. I. (1980). Orienting of attention. *Q J Exp Psychol, 32*(1), 3-25.

214. Posner, M. I., & Cohen, Y. (1984). Components of visual orienting. In H. Bouma & D. Bouwhuis (Eds.), *Attention and performance.* (Vol. X, pp. 531-556): Erlbaum.

215. Posner, M. I., & Petersen, S. E. (1990). The attention system of the human brain. *Annu Rev Neurosci, 13*, 25-42.

216. Posner, M. I., Snyder, C. R., & Davidson, B. J. (1980). Attention and the detection of signals. *J Exp Psychol, 109*(2), 160-174.

217. Puce, A., Allison, T., Asgari, M., Gore, J. C., & McCarthy, G. (1996). Differential sensitivity of human visual cortex to faces, letterstrings, and textures: A functional magnetic resonance imaging study. *J Neurosci, 16*(16), 5205-5215.

218. Puce, A., Allison, T., Gore, J. C., & McCarthy, G. (1995). Face-sensitive regions in human extrastriate cortex studied by functional mri. *J Neurophysiol, 74*(3), 1192-1199.

219. Purcell, D. G., Stewart, A. L., & Skov, R. B. (1996). It takes a confounded face to pop out of a crowd. *Perception, 25*(9), 1091-1108.

220. Pylyshyn, Z. W., & Storm, R. W. (1988). Tracking multiple independent targets: Evidence for a parallel tracking mechanism. *Spat Vis, 3*(3), 179-197.

221. Quian Quiroga, R., Reddy, L., Kreiman, G., Koch, C., & Fried, I. (2005). Invariant visual representations by single neurons in the human brain. *Nature, in press*.

222. Quiroga, R. Q., Nadasdy, Z., & Ben-Shaul, Y. (2004). Unsupervised spike detection and sorting with wavelets and superparamagnetic clustering. *Neural Computation, 16*(8), 1661-1687.

223. Reddy, L., Wilken, P., & Koch, C. (2004). Face-gender discrimination is possible in the near-absence of attention. *J Vis, 4*(2), 106-117.

224. Rees, G., Frith, C. D., & Lavie, N. (1997). Modulating irrelevant motion perception by varying attentional load in an unrelated task. *Science, 278*(5343), 1616-1619.

225. Rees, G., Kreiman, G., & Koch, C. (2002). Neural correlates of consciousness in humans. *Nat Rev Neurosci, 3*(4), 261-270.

226. Rensink, R., & Simons, D. (2000). Change detection, attention, and the contents of awareness. *Consciousness And Cognition, 9*(2), S18-S19.

227. Rensink, R. A. (2000). The dynamic representation of scenes. *Visual Cognition, 7*(1-3), 17-42.

228. Rensink, R. A. (2002). Change detection. *Annu Rev Psychol, 53*, 245-277.

229. Rensink, R. A., O'Regan, J. K., & Clark, J. J. (1997). To see or not to see: The need for attention to perceive changes in scenes. *Psychological Science, 8*(5), 368-373.

230. Ress, D., Backus, B. T., & Heeger, D. J. (2000). Activity in primary visual cortex predicts performance in a visual detection task. *Nat Neurosci, 3*(9), 940-945.

231. Ress, D., & Heeger, D. J. (2003). Neuronal correlates of perception in early visual cortex. *Nat Neurosci, 6*(4), 414-420.

232. Reynolds, J. H., Chelazzi, L., & Desimone, R. (1999). Competitive mechanisms subserve attention in macaque areas v2 and v4. *J Neurosci, 19*(5), 1736-1753.

233. Reynolds, J. H., & Desimone, R. (2003). Interacting roles of attention and visual salience in v4. *Neuron, 37*(5), 853-863.

234. Reynolds, J. H., Pasternak, T., & Desimone, R. (2000). Attention increases sensitivity of v4 neurons. *Neuron, 26*(3), 703-714.

235. Richmond, B. J., Optican, L. M., & Spitzer, H. (1990). Temporal encoding of two-dimensional patterns by single units in primate primary visual cortex. I. Stimulus-response relations. *J Neurophysiol, 64*(2), 351-369.

236. Ringo, J. L., Sobotka, S., Diltz, M. D., & Bunce, C. M. (1994). Eye movements modulate activity in hippocampal, parahippocampal, and inferotemporal neurons. *J Neurophysiol, 71*(3), 1285-1288.

237. Ro, T., Russell, C., & Lavie, N. (2001). Changing faces: A detection advantage in the flicker paradigm. *Psychol Sci, 12*(1), 94-99.

238. Rock, I., & Gutman, D. (1981). The effect of inattention on form perception. *Journal Of Experimental Psychology-Human Perception And Performance, 7*(2), 275-285.

239. Rock, I., Linnett, C. M., Grant, P., & Mack, A. (1992). Perception without attention: Results of a new method. *Cognit Psychol, 24*(4), 502-534.

240. Rolls, E. T. (1984). Neurons in the cortex of the temporal lobe and in the amygdala of the monkey with responses selective for faces. *Hum Neurobiol, 3*(4), 209-222.

241. Rolls, E. T. (2004). Consciousness absent and present: A neurophysiological exploration. *Prog Brain Res, 144*, 95-106.

242. Rolls, E. T., Aggelopoulos, N. C., & Zheng, F. (2003). The receptive fields of inferior temporal cortex neurons in natural scenes. *J Neurosci, 23*(1), 339-348.

243. Rolls, E. T., & Baylis, G. C. (1986). Size and contrast have only small effects on the responses to faces of neurons in the cortex of the superior temporal sulcus of the monkey. *Exp Brain Res, 65*(1), 38-48.

244. Rolls, E. T., & Tovee, M. J. (1994). Processing speed in the cerebral cortex and the neurophysiology of visual masking. *Proc R Soc Lond B Biol Sci, 257*(1348), 9-15.

245. Rolls, E. T., Tovee, M. J., & Panzeri, S. (1999). The neurophysiology of backward visual masking: Information analysis. *J Cogn Neurosci, 11*(3), 300-311.

246. Rossion, B. (2002). Is sex categorization from faces really parallel to face recognition? *Visual Cognition, 9*(8), 1003-1020.

247. Rossion, B., Schiltz, C., Robaye, L., Pirenne, D., & Crommelinck, M. (2001). How does the brain discriminate familiar and unfamiliar faces? A pet study of face categorical perception. *J Cogn Neurosci, 13*(7), 1019-1034.

248. Rousselet, G. A., Fabre-Thorpe, M., & Thorpe, S. J. (2002). Parallel processing in high-level categorization of natural images. *Nat Neurosci, 5*(7), 629-630.

249. Saenz, M., Buracas, G. T., & Boynton, G. M. (2002). Global effects of feature-based attention in human visual cortex. *Nat Neurosci, 5*(7), 631-632.

250. Sakai, K., & Miyashita, Y. (1991). Neural organization for the long-term-memory of paired associates. *Nature, 354*(6349), 152-155.

251. Saleem, K. S., & Tanaka, K. (1996). Divergent projections from the anterior inferotemporal area te to the perirhinal and entorhinal cortices in the macaque monkey. *J Neurosci, 16*(15), 4757-4775.

252. Sato, T. (1989). Interactions of visual stimuli in the receptive fields of inferior temporal neurons in awake macaques. *Exp Brain Res, 77*(1), 23-30.

253. Schendan, H. E., Ganis, G., & Kutas, M. (1998). Neurophysiological evidence for visual perceptual categorization of words and faces within 150 ms. *Psychophysiology, 35*(3), 240-251.

254. Scoville, W. B., & Milner, B. (1957). Loss of recent memory after bilateral hippocampal lesions. *J Neurochem, 20*(1), 11-21.

255. Sengpiel, F., & Hubener, M. (1999). Visual attention: Spotlight on the primary visual cortex. *Curr Biol, 9*(9), R318-321.

256. Sergent, J. (1986). Microgenesis of face perception. In H. D. Ellis, M. A. Jeeves, F. Newcombe & A. Young (Eds.), *Aspects of face processing*. Nijhoff,Dordrecht,The Netherlands.

257. Sergent, J., Ohta, S., & MacDonald, B. (1992). Functional neuroanatomy of face and object processing. A positron emission tomography study. *Brain, 115 Pt 1*, 15-36.

258. Shadlen, M. N., Britten, K. H., Newsome, W. T., & Movshon, J. A. (1996). A computational analysis of the relationship between neuronal and behavioral responses to visual motion. *J Neurosci, 16*(4), 1486-1510.

259. Sheinberg, D. L., & Logothetis, N. K. (1997). The role of temporal cortical areas in perceptual organization. *Proc Natl Acad Sci U S A, 94*(7), 3408-3413.

260. Simons, D. J. (2000). Current approaches to change blindness. *Visual Cognition, 7*(1-3), 1-15.

261. Simons, D. J., & Chabris, C. F. (1999). Gorillas in our midst: Sustained inattentional blindness for dynamic events. *Perception, 28*(9), 1059-1074.

262. Simons, D. J., & Levin, D. T. (1997). Change blindness. *Trends Cogn Sci, 1*(7), 261-267.

263. Simons, D. J., & Levin, D. T. (1998). Failure to detect changes to people during a real-world interaction. *Psychonomic Bulletin & Review, 5*(4), 644-649.

264. Somers, D. C., Dale, A. M., Seiffert, A. E., & Tootell, R. B. (1999). Functional mri reveals spatially specific attentional modulation in human primary visual cortex. *Proc Natl Acad Sci U S A, 96*(4), 1663-1668.

265. Sperling, G., & Melchner, M. J. (1978). The attention operating characteristic: Examples from visual search. *Science, 202*(4365), 315-318.

266. Sperling, G., & Weichselgartner, E. (1990). Episodic theory of visual-attention. *Bulletin Of The Psychonomic Society, 28*(6), 482-482.

267. Spitzer, H., Desimone, R., & Moran, J. (1988). Increased attention enhances both behavioral and neuronal performance. *Science, 240*(4850), 338-340.

268. Squire, L. R. (1992). Memory and the hippocampus: A synthesis from findings with rats, monkeys, and humans. *Psychol Rev, 99*(2), 195-231.

269. Squire, L. R., Shimamura, D. G., & Amaral, D. G. (1989). In J. Byrne & W. Berry (Eds.), *Neural models of plasticity* (pp. 208-239). New York: Academic Press.

270. Squire, L. R., Stark, C. E., & Clark, R. E. (2004). The medial temporal lobe. *Annu Rev Neurosci, 27*, 279-306.

271. Squire, L. R., & Zola-Morgan, S. (1991). The medial temporal lobe memory system. *Science, 253*(5026), 1380-1386.

272. Squire, L. R., & Zola, S. M. (1996). Structure and function of declarative and nondeclarative memory systems. *Proc Natl Acad Sci U S A, 93*(24), 13515-13522.

273. Stark, C. E., Bayley, P. J., & Squire, L. R. (2002). Recognition memory for single items and for associations is similarly impaired following damage to the hippocampal region. *Learn Mem, 9*(5), 238-242.

274. Stark, C. E., & Squire, L. R. (2003). Hippocampal damage equally impairs memory for single items and memory for conjunctions. *Hippocampus, 13*(2), 281-292.

275. Stefanacci, L., Buffalo, E. A., Schmolck, H., & Squire, L. R. (2000). Profound amnesia after damage to the medial temporal lobe: A neuroanatomical and neuropsychological profile of patient e.P. *Journal Of Neuroscience, 20*(18), 7024-7036.

276. Stefanacci, L., Suzuki, W. A., & Amaral, D. G. (1996). Organization of connections between the amygdaloid complex and the perirhinal and parahippocampal cortices in macaque monkeys. *J Comp Neurol, 375*(4), 552-582.

277. Suzuki, S., & Cavanagh, P. (1995). Facial organization blocks access to low-level features—an object inferiority effect. *Journal of Experimental Psychology—Human Perception and Performance, 21*(4), 901-913.

278. Suzuki, W. A. (1996a). The anatomy, physiology and functions of the perirhinal cortex. *Current Opinion In Neurobiology, 6*(2), 179-186.

279. Suzuki, W. A. (1996b). Neuroanatomy of the monkey entorhinal, perirhinal and parahippocampal cortices: Organization of cortical inputs and interconnections with amygdala and striatum. *Seminars In The Neurosciences, 8*(1), 3-12.

280. Suzuki, W. A., & Amaral, D. G. (1990). Cortical inputs to the ca1 field of the monkey hippocampus originate from the perirhinal and parahippocampal cortex but not from area te. *Neurosci Lett, 115*(1), 43-48.

281. Suzuki, W. A., & Amaral, D. G. (1994). Perirhinal and parahippocampal cortices of the macaque monkey: Cortical afferents. *J Comp Neurol, 350*(4), 497-533.

282. Suzuki, W. A., & Eichenbaum, H. (2000). The neurophysiology of memory. *Ann N Y Acad Sci, 911*, 175-191.

283. Suzuki, W. A., Miller, E. K., & Desimone, R. (1997). Object and place memory in the macaque entorhinal cortex. *Journal Of Neurophysiology, 78*(2), 1062-1081.

284. Tanaka, K. (1996). Inferotemporal cortex and object vision. *Annual Review Of Neuroscience, 19*, 109-139.

285. Tarr, M. J., & Gauthier, I. (2000). Ffa: A flexible fusiform area for subordinate-level visual processing automatized by expertise. *Nat Neurosci, 3*(8), 764-769.

286. Thesen, S., Heid, O., Mueller, E., & Schad, L. R. (2000). Prospective acquisition correction for head motion with image-based tracking for real-time fmri. *Magn Reson Med, 44*(3), 457-465.

287. Thorpe, S. (1998). Localized versus distributed representatons. In *The handbook of brain theory and neural networks* (pp. 549-552): MIT Press.

288. Thorpe, S., Fize, D., & Marlot, C. (1996). Speed of processing in the human visual system. *Nature, 381*(6582), 520-522.

289. Tootell, R. B., Hadjikhani, N., Hall, E. K., Marrett, S., Vanduffel, W., Vaughan, J. T., et al. (1998). The retinotopy of visual spatial attention. *Neuron, 21*(6), 1409-1422.

290. Tovee, M. J. (1998). Is face processing special? *Neuron, 21*(6), 1239-1242.

291. Treisman, A. (1996). The binding problem. *Curr Opin Neurobiol, 6*(2), 171-178.

292. Treisman, A. (1998). Feature binding, attention and object perception. *Philos Trans R Soc Lond B Biol Sci, 353*(1373), 1295-1306.

293. Treisman, A. (1999). Solutions to the binding problem: Progress through controversy and convergence. *Neuron, 24*(1), 105-110, 111-125.

294. Treisman, A., & Schmidt, H. (1982). Illusory conjunctions in the perception of objects. *Cognit Psychol, 14*(1), 107-141.

295. Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognit Psychol, 12*(1), 97-136.

296. Treue, S., & Martinez Trujillo, J. C. (1999). Feature-based attention influences motion processing gain in macaque visual cortex. *Nature, 399*(6736), 575-579.

297. Treue, S., & Maunsell, J. H. (1996). Attentional modulation of visual motion processing in cortical areas mt and mst. *Nature, 382*(6591), 539-541.

298. Treue, S., & Maunsell, J. H. (1999). Effects of attention on the processing of motion in macaque middle temporal and medial superior temporal visual cortical areas. *J Neurosci, 19*(17), 7591-7602.

299. Valentine, T. (1988). Upside-down faces: A review of the effect of inversion upon face recognition. *Br J Psychol, 79 (Pt 4)*, 471-491.

300. Valentine, T., & Bruce, V. (1988). Mental rotation of faces. *Mem Cognit, 16*(6), 556-566.

301. Vallbo, A. B., & Johansson, R. S. (1976). Skin mechanoreceptors in the human hand: Neural and psychophysical thresholds. In Y. Zotterman (Ed.), *Sensory functions of the skin in primates.* (pp. 185-199). Oxford: Pergamon.

302. Van Hoesen, G., & Pandya, D. N. (1975a). Some connections of the entorhinal (area 28) and perirhinal (area 35) cortices of the rhesus monkey. I. Temporal lobe afferents. *Brain Res, 95*(1), 1-24.

303. Van Hoesen, G., Pandya, D. N., & Butters, N. (1975). Some connections of the entorhinal (area 28) and perirhinal (area 35) cortices of the rhesus monkey. Ii. Frontal lobe afferents. *Brain Res, 95*(1), 25-38.

304. Van Hoesen, G. W., & Pandya, D. N. (1975b). Some connections of the entorhinal (area 28) and perirhinal (area 35) cortices of the rhesus monkey. Iii. Efferent connections. *Brain Res, 95*(1), 39-59.

305. Van Hoesen, G. W., Pandya, D. N., & Butters, N. (1972). Cortical afferents to the entorhinal cortex of the rhesus monkey. *Science, 175*(29), 1471-1473.

306. VanRullen, R. (2003). Visual saliency and spike timing in the ventral visual pathway. *J Physiol Paris, 97*(2-3), 365-377.

307. VanRullen, R., & Koch, C. (2003). Competition and selection during visual processing of natural scenes and objects. *J Vis, 3*(1), 75-85.

308. VanRullen, R., Reddy, L., & Koch, C. (2004). Visual search and dual tasks reveal two distinct attentional resources. *J Cogn Neurosci, 16*(1), 4-14.

309. VanRullen, R., & Thorpe, S. J. (2001). The time course of visual processing: From early perception to decision-making. *J Cogn Neurosci, 13*(4), 454-461.

310. Vetter, R. (1998). Synthesis of novel views from a single face image. *International Journal of Computer Vision, 28*(2), 103-116.

311. von der Malsburg, C. (1981). *The correlation theory of brain function.* (Internal Report No. 81-2): MPI Biophysical Chemistry.

312. Vuilleumier, P. (2000). Faces call for attention: Evidence from patients with visual extinction. *Neuropsychologia, 38*(5), 693-700.

313. Vuilleumier, P., Armony, J. L., Driver, J., & Dolan, R. J. (2001). Effects of attention and emotion on face processing in the human brain: An event-related fmri study. *Neuron, 30*(3), 829-841.

314. Vuilleumier, P., Sagiv, N., Hazeltine, E., Poldrack, R. A., Swick, D., Rafal, R. D., et al. (2001). Neural fate of seen and unseen faces in visuospatial neglect: A combined event-related functional mri and event-related potential study. *Proc Natl Acad Sci U S A, 98*(6), 3495-3500.

315. Watanabe, T., Harner, A. M., Miyauchi, S., Sasaki, Y., Nielsen, M., Palomo, D., et al. (1998). Task-dependent influences of attention on the activation of human primary visual cortex. *Proc Natl Acad Sci U S A, 95*(19), 11489-11492.

316. Wilson, M. A., & McNaughton, B. L. (1993). Dynamics of the hippocampal ensemble code for space. *Science, 261*(5124), 1055-1058.

317. Wirth, S., Yanike, M., Frank, L. M., Smith, A. C., Brown, E. N., & Suzuki, W. A. (2003). Single neurons in the monkey hippocampus and learning of new associations. *Science, 300*(5625), 1578-1581.

318. Wojciulik, E., Kanwisher, N., & Driver, J. (1998). Covert visual attention modulates face-specific activity in the human fusiform gyrus: Fmri study. *J Neurophysiol, 79*(3), 1574-1578.

319. Wolfe, J. M. (1994). Guided search 2.0—a revised model of visual-search. *Psychonomic Bulletin & Review, 1*(2), 202-238.

320. Yamaguchi, M. K., Hirukawa, T., & Kanazawa, S. (1995). Judgment of gender through facial parts. *Perception, 24*(5), 563-575.

321. Yamamoto, S., & Kashikura, K. (1999). Speed of face recognition in humans: An event-related potentials study. *Neuroreport, 10*(17), 3531-3534.

322. Yamane, S., Kaji, S., & Kawano, K. (1988). What facial features activate face neurons in the inferotemporal cortex of the monkey? *Exp Brain Res, 73*(1), 209-214.

323. Yang, T. M., & Maunsell, J. H. R. (2004). The effect of perceptual learning on neuronal responses in monkey visual area v4. *Journal Of Neuroscience, 24*(7), 1617-1626.

324. Yi, D. J., Woodman, G. F., Widders, D., Marois, R., & Chun, M. M. (2004). Neural fate of ignored stimuli: Dissociable effects of perceptual and working memory load. *Nat Neurosci, 7*(9), 992-996.

325. Yin, R. K. (1969). Looking at upside down faces. *Journal of Experimental Psychology, 81*(1), 141-145.

326. Yovel, G., & Kanwisher, N. (2004). Face perception: Domain specific, not process specific. *Neuron, 44*(5), 889-898.

327. Zeki, S. (1983). The distribution of wavelength and orientation selective cells in different areas of monkey visual cortex. *Proc R Soc Lond B Biol Sci, 217*(1209), 449-470.

328. Zeki, S. (2001). Localization and globalization in conscious vision. *Annu Rev Neurosci, 24*, 57-86.

329. Zeki, S., & Bartels, A. (1999). Toward a theory of visual consciousness. *Conscious Cogn, 8*(2), 225-259.

330. Zola-Morgan, S., & Squire, L. R. (1993). Neuroanatomy of memory. *Annu Rev Neurosci, 16*, 547-563.

331. Zola-Morgan, S., Squire, L. R., & Amaral, D. G. (1986). Human amnesia and the medial temporal region: Enduring memory impairment following a bilateral lesion limited to field ca1 of the hippocampus. *J Neurosci, 6*(10), 2950-2967.

332. Zolamorgan, S., Squire, L. R., & Ramus, S. J. (1994). Severity of memory impairment in monkeys as a function of locus and extent of damage within the medial temporal-lobe memory system. *Hippocampus, 4*(4), 483-495.