The neurophysiology of attention and object recognition in visual scenes

Chapter · October 2014

DOI: 10.13140/2.1.3634.5925

Heida Maria Sigurdardottir
University of Iceland

David L Sheinberg
Brown University

CITATION:

Brooks, D. I., Sigurdardottir, H. M., & Sheinberg, D. L. (2014). The neurophysiology of attention and object recognition in visual scenes. In K. Kveraga & M. Bar (Eds.), *Scene Vision: Making Sense of What We See* (pp. 85-104). Cambridge, Massachusetts: MIT Press.

The Neurophysiology of Attention and Object Recognition in Visual Scenes

Daniel I. Brooks*[1], Heida Maria Sigurdardottir*[1,2], David L. Sheinberg**[1]
1) Deptm. of Neuroscience, Brown University 2) Deptm. of Psychology, University of Iceland

* These authors contributed equally to this work and are listed alphabetically

** To whom correspondence should be addressed: David_Sheinberg@brown.edu

If you are like most academics and scholars, you start your day with a cup of coffee. This task (Land & Hayhoe, 2001), amongst other things, requires you to locate the correct cupboard in the kitchen, open it and search for the can of coffee grounds among the surrounding clutter, and operate the coffee brewer by finding and pushing the correct buttons on the coffee maker. Depending on your view and current goal, the kitchen with its furniture and appliances, the cupboard full of miscellaneous objects, and the coffee brewer with its all too many buttons can be thought of as scenes in which you are searching for and recognizing a particular target object among other irrelevant and distracting ones.

Everyone knows what a scene is, to twist a famous quote by William James, but the concept is nonetheless surprisingly hard to define. When talking about scenes, it is often implied that they depict real-world environments consisting of a background and several objects in a particular spatial arrangement (Henderson & Hollingworth, 1999). While there exists a long list of reports on human scene perception in the aforementioned sense, visual stimuli traditionally used in neurophysiology are rather sparse and artificial in comparison. When reviewing some of the literature on the neurophysiology of searching through and recognizing objects in visual scenes, we have therefore stretched the scene concept to include not only naturalistic environments but also simplified arrays of two or more visual elements.

The evidence presented here will mostly, although not exclusively, be drawn from the visual system of the macaque, one of the most anatomically and functionally comparable with the human visual system (for a detailed overview of the macaque visual system, see e.g. Van Essen, Anderson, & Felleman, 1992). Monkeys rarely search for coffee, but they nonetheless are often given the task of searching for particular things in visual scenes in order to gain other liquid rewards, such as drops of juice. Even though this can be considered rather unusual behavior, macaques in the wild take on many analogous tasks in their everyday environments, such as foraging for food and finding shelter, that require finding objects of interest in a complex scene.

Here, we will review some of the neural processing that a scene image likely undergoes as it passes through the visual system, leading to the identification of an object in the scene. Our attempt will, no doubt, be oversimplified; we focus on only particular processing steps and brain regions, but in reality the steps taken are likely to be both stimulus- and task-dependent, and might involve several iterations where information is sent back and forth between low and high level visual regions (see e.g. Hochstein & Ahissar, 2002; Peters & Payne, 1993; Tsotsos, 2008). Much is yet to be learned about the electrophysiology of searching for and recognizing objects in scenes.

Researchers often go to great lengths to eliminate any temporal and spatial correlations in the stimuli that they use in experiments, because they do not want participants to be able to guess what they are going to be shown based on either what they just saw, or the surrounding visual elements. The natural viewing of scenes, on the other hand, imposes all sorts of contingencies in both space and time (Attneave, 1954; Field, 1987; Kersten, 1987). If you are at the beach now, you are going to be at the beach one second later as well (probably even seeing almost the same view as before), and one grain of sand is most likely going to be next to another grain of sand. Change in the real world is slow. Similarly, the spectrum of naturalistic visual input is not white, i.e. not all frequencies are equally presented in natural scenes; low temporal and spatial frequencies are more abundant than high frequencies (Dong & Atick, 1995; Field, 1987).

What this means is that natural scenes are full of redundancies. It might not be a particularly good strategy to use a large part of your computational resources to essentially code for the same information over and over again (Attneave, 1954; Simoncelli & Olshausen, 2001). Computer scientists know this and have developed methods that make use of the correlational structure of images to compress them into a more easily transferable format. The primate visual system, apparently, also makes use of such contingencies to more efficiently code visual information (but see e.g. Barlow, 2001, for a critical look on the idea that redundancy in natural viewing is used mainly for compressive coding in perceptual systems).

This apparently happens before light even enters the eye. Primates have eyes that can move and thus the ability to influence their own visual input. The eyes are never completely still, even when an attempt is made to keep them fixated at a single spot (Ratliff & Riggs, 1950). Computational modeling work indicates that such fixational instability changes the statistics of neuronal firing in the retina, where the responses of retinal ganglion cells with non-overlapping receptive fields become uncorrelated when images with a spectral density like that of natural images are fixated (Rucci, 2008). Most retinal ganglion cells project to the lateral geniculate nucleus (LGN), a subcortical structure in the thalamus, which then projects mainly to the primary visual cortex (V1; see figure 1). Electrophysiological recordings have shown that the linear filtering properties of neurons in the LGN also lead to a similar temporal decorrelation or whitening of visual information in natural scenes (Dan, Atick, & Reid, 1996).

The whitening of the signal reduces redundancies found in real-world scenes, but it might also have an even more important effect. Since low temporal and spatial frequencies are abundant in natural scenes, the whitening of this input effectively emphasizes the throughput of high frequency information – fast changes – while deemphasizing low frequency information – slow changes. A quick change over time can often signify the appearance of a new object in the scene, and a quick change over space is quite likely to happen at object boundaries. Visual information from a scene that reaches the primary visual cortex is therefore already preprocessed in a manner that accentuates features that are important for finding and segmenting objects in scenes.

It is good to keep in mind that just because a brain region, such as the lateral geniculate nucleus, appears to behave differently under natural conditions than when artificial reduced visual stimuli are used, that does not necessarily indicate that it “knows” anything about the naturalness of its visual inputs. Even in the primary visual cortex (V1), local field potentials and multiunit firing rates are not detectably different for natural movies and noise stimuli with a frequency distribution similar to that of natural scenes (Kayser, Salazar, & Konig, 2003). Neural activity in the primary visual cortex for simple stimuli like bars and gratings, on the other hand, significantly differs from several more complex or naturalistic inputs such as natural movies (Kayser et al., 2003), pink noise (Kayser et al., 2003), random textures (Lehky, Sejnowski, & Desimone, 1992), and three-dimensional surfaces (Lehky et al., 1992). The primary visual cortex might therefore not be adapted to processing natural scenes as such, but to visual input with temporal and spatial statistics or complexities similar to those of real-world environments.

One aspect of real-world environments to which the primary visual cortex might be adapted is that natural input is not confined to a small spot, as it often is in visual experiments, but encompasses the entire visual field. Natural scenes are therefore bound to stimulate both the classical receptive field of a V1 neuron, a small region within the visual field where stimuli can directly drive the neuron, and the non-classical receptive field, a region that does not evoke a neural response when stimulated alone but can selectively modulate the neuron’s responses to other stimuli within the classical receptive field (Allman, Miezin, & McGuinness, 1985).

Under such more naturalistic viewing conditions, the nonlinear interactions between the classical and non-classical receptive fields of V1 neurons not only produce a sparse code that is energy efficient and minimizes redundancy (Vinje & Gallant, 2000), but they can also be sensitive to scene structure such as whether the classical and non-classical receptive fields are part of a continuous contour (Guo, Robertson, Mahmoodi, & Young, 2005). Such interactions are thought to facilitate contour integration (Guo et al., 2005), which again is important for segmenting objects from the background and from each other. Contours also serve as cues for visual object recognition, as attested by the fact that people can recognize line drawings of objects with relative ease.

In general, neurons in V1 might respond to the greatest extent when the visual features within the classical and non-classical receptive fields are dissimilar, such as when there are orthogonal orientations inside and outside the classical receptive field (Guo et al., 2005; Knierim & van Essen, 1992; Sillito, Grieve, Jones, Cudeiro, & Davis, 1995). Again, this would emphasize the throughput of fast changes in visual scenes more than slow changes. Such contrast between visual features, like a yellow beach ball in a sea of blue, is salient to a human observer and an important cue for finding objects in scenes. We turn next to higher-level brain regions that might use information from V1 about feature contrast, as well as other relevant data such as reward history, or the goal of the current task to prioritize spatial locations in a visual scene. This task is thought to be accomplished mainly by regions within the so-called dorsal visual stream and its interconnectivity with areas of frontal cortex involved in the allocation of attention.

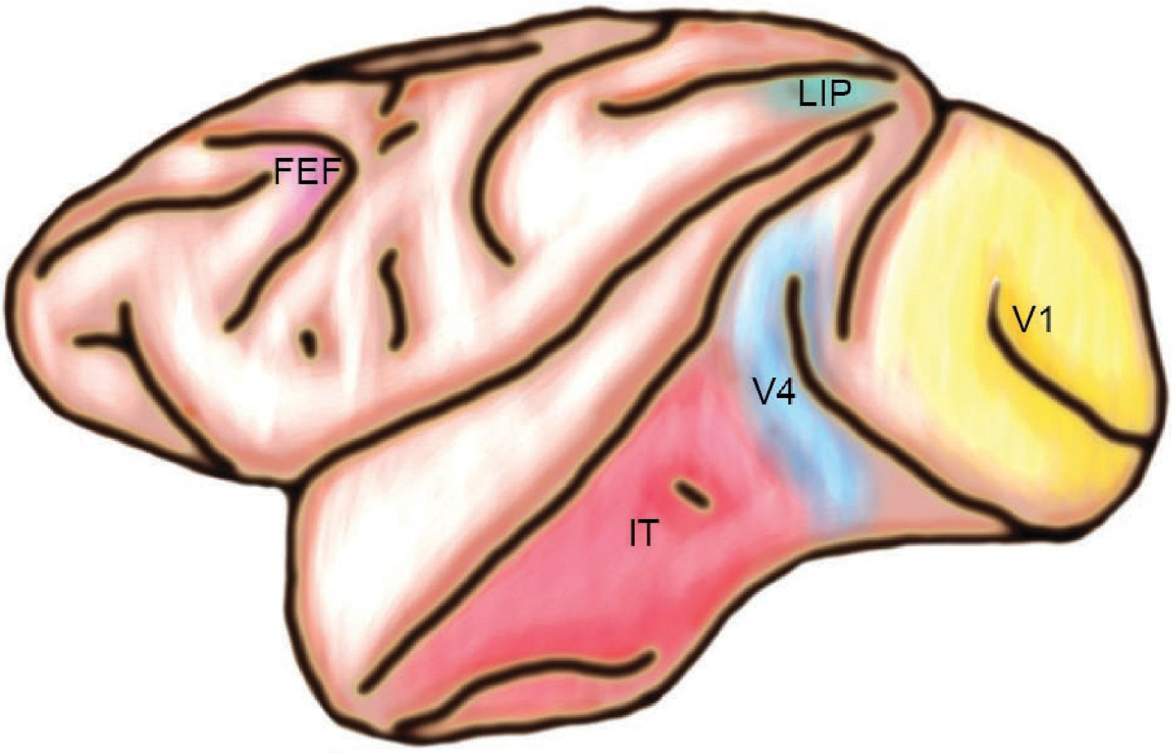

Figure 1. The macaque brain. Neurons in the primary visual cortex (V1) were mainly thought to detect simple features, such as lines of particular orientations, within a very small part of visual space, but are now known to integrate visual information from a much larger area (Allman et al., 1985; Hubel & Wiesel, 1959, 1968). Visual information originating from V1 reaches both a frontoparietal network, including the lateral intraparietal area (LIP) and the frontal eye fields (FEF), and the ventral visual stream, including the fourth visual area (V4) and the inferior temporal cortex (IT). Cells in LIP can respond to visual information, have memory-related responses, respond before, during, or after a saccadic eye movement, or have complex combinations of these types of responses (Barash, Bracewell, Fogassi, Gnadt, & Andersen, 1991a, 1991b). These neurons have a spatial receptive or response field, so in order to evoke a neural response, the thing that is visually presented, attended, looked at, or memorized, needs to have been within a particular spatial location (Barash et al., 1991b; Blatt et al., 1990). Similar to LIP, the FEF also contain neurons with visual responses, oculomotor responses, and a combination of the two (Bruce & Goldberg, 1985). V4 and especially IT likely represent the final output stages of object processing in the primate brain. Unlike the fronto-parietal system, IT neurons often have very large receptive fields (Gross, Bender, & Rocha-Miranda, 1969; but see Rolls et al., 2003). These cells might therefore not carry as much information on the location of an object, but instead they selectively respond to the complex features of objects of interest (Logothetis & Sheinberg, 1996).

The classic characterization of the primate visual system includes a broad segregation into two main processing streams, both of which get input from the primary visual cortex (Goodale & Milner, 1992; for a detailed review on parallel visual processing, see Nassi and Callaway (2009); Ungerleider & Haxby, 1994; Ungerleider & Mishkin, 1982). The ventral ("what") stream is theorized to accomplish the recognition of complex forms and, ultimately, objects, whereas the dorsal ("where" or "how") stream has been implicated in the control of visual attention and spatial guidance of appropriate actions. While these streams are at least partially anatomically and functionally separable, they can be viewed as forming a widely interconnected system that guides the eyes and attention to visually salient or otherwise potentially important features/objects that then can be further scrutinized, recognized, and categorized.

As the dorsal stream projecting from V1 enters the parietal lobe, information is integrated into a fronto-parietal network including the lateral intraparietal area (LIP) in the intraparietal sulcus and the frontal eye fields (FEF) in the prearcuate sulcus of the prefrontal cortex (figure 1). These two regions, as well as a subcortical midbrain structure called the superior colliculus (SC), have been the focus of most research on the neuroanatomical substrates of visual search because each of them is thought to participate in guiding the eyes and attention to important objects or locations within visual scenes (Baluch & Itti, 2011; Bisley & Goldberg, 2010; Fecteau & Munoz, 2006).

One conceptual framework for this network of structures is that together they form a priority map of visual space where each location in a scene is weighed or prioritized according to both its visual salience and probable task relevance (Baluch & Itti, 2011; Bisley & Goldberg, 2010; Fecteau & Munoz, 2006; Goldberg, Bisley, Powell, & Gottlieb, 2006; Itti & Koch, 2001; Thompson & Bichot, 2005). A saliency map was originally proposed as a computational model that uses the distribution of features such as color, brightness, and orientation to make a single representation of visual conspicuity (Itti & Koch, 2000).

As an example of this phenomenon, consider the results of Thomas and Paré (Thomas & Paré, 2007). In that study, monkeys were required to make an eye movement to an oddball stimulus (a red target in an array of green distractors or a green target among red distractors). While cells in the lateral intraparietal area initially responded indiscriminately to a target and a distractor (i.e. whether the object was green or red, or whether it was task relevant or irrelevant), the response of cells shifted to signal whether or not the oddball target was in or out of their receptive fields. Thus, neural activity in the lateral intraparietal area might gradually develop to represent the visual priority of the location occupied by an object in a scene. Similar effects can be observed in the frontal eye fields and the superior colliculus (for reviews on these regions acting as priority maps, see e.g. Baluch & Itti, 2011; Bisley & Goldberg, 2010; Fecteau & Munoz, 2006).

It is a good guess that something that stands out from the background – because it is brighter, darker, of a different color, or differently oriented than the surrounding elements – is a candidate object that might need to be recognized and is thus worth scrutinizing further. However, it is also likely that in a complex environment, dozens of relatively unimportant objects will stand out from the background while a single object (with certain features) is the only behaviorally relevant object. If you are looking for something in particular, such as a red parasol in our hypothetical beach scene, your visual system can prioritize this type of object over objects with different features (green things, square things, etc.) because they are not immediately relevant (see e.g. Wolfe, 2007; Wolfe & Horowitz, 2004). This integration of top-down information with bottom-up saliency likely also happens in the aforementioned brain regions. Since a priority map weights locations based on such information, it can, at least theoretically, be used to select potential objects to which attention and eye gaze should be directed.

For an example of visual selection after attentional prioritization, let us consider the frontal eye fields. The frontal eye fields bear this name because electrical microstimulation within the region readily evokes saccadic eye movements, the endpoints of which depend on which subregion within the frontal eye fields is stimulated (Bruce, Goldberg, Bushnell, & Stanton, 1985). When the frontal eye fields are stimulated at a subthreshold current level – too low to cause eye movements – this modulated activity can nonetheless have a measurable effect on behavior. Moore and Fallah (2004) trained monkeys to detect a change in a small peripheral target in the presence of a continuously flashing distracting stimulus. The researchers then stimulated a subregion of the monkeys' frontal eye fields that at higher current levels would have evoked saccades to the target location. When this was done at the time of the target change, this improved the monkeys' target detection to a degree comparable with removing the distractor altogether. The frontal eye fields therefore causally contribute to the distribution of visuospatial attention.

A single target and a single distractor make a poor scene, if such a display can be considered a scene at all. It is nonetheless likely that activity in the frontal eye fields and the lateral intraparietal area aid target detection in complex and cluttered scenes by giving priority and guiding attention to the location of a potentially important object. Attending to a location means, amongst other things, that finer visual details can be sampled from an object that occupies that location (Montagna, Pestilli, & Carrasco, 2009), details that might be relevant for their successful recognition. Further, it is likely that prioritizing a location directly affects the selection of visual features or candidate objects in the ventral visual pathway. In cluttered scenes, multiple visual objects might compete for detailed representation in the ventral visual pathway. Signals from the frontal eye fields and the lateral intraparietal area could bias this competition (Desimone & Duncan, 1995) in favor of an object that occupies a high priority location.

In order for this to be plausible, brain areas that implement a priority map must somehow communicate with ventral regions that are known to be important for object recognition. In other words, attentional selection should affect neural responses to objects in the ventral stream, and the prioritizing of the locations of candidate objects should precede object recognition.

There is now quite a lot of data that confirms that regions such as the lateral intraparietal area and the frontal eye field have reciprocal structural connections to some regions in the ventral visual stream, such as V4 and parts of the inferior temporal cortex (see figure 1; Blatt, Andersen, & Stoner, 1990; Distler, Boussaoud, Desimone, & Ungerleider, 1993; Lewis & Van Essen, 2000; Schall, Morel, King, & Bullier, 1995; Stanton, Bruce, & Goldberg, 1995; Webster, Bachevalier, & Ungerleider, 1994), that respond selectively to moderately or highly complex visual features or even whole objects and which are thought important for visual object recognition (Logothetis & Sheinberg, 1996). Since these connections exist, the lateral intraparietal area and the frontal eye fields must, at least under some circumstances, exchange information with regions in the ventral visual stream. It is plausible that this happens when the situation requires attentional selection.

Attentional selection strongly modulates object responses in the ventral stream. This is demonstrated by the work of Chelazzi and colleagues (Chelazzi, Duncan, Miller, & Desimone, 1998; Chelazzi, Miller, Duncan, & Desimone, 1993, 2001) who presented monkeys with visual search displays containing two or more objects while they recorded neural activity in lower level (V4) and higher level (inferior temporal cortex) regions of the ventral visual stream (see also e.g. Moran & Desimone, 1985; Reynolds, Chelazzi, & Desimone, 1999). One of the objects was known to elicit a strong response from the neuron being recorded from (an effective stimulus) while another one of the objects tended to elicit a poor or even no response from the same neuron (an ineffective stimulus); note that each neuron can have a unique stimulus preference, so an object that effectively drives one neuron is not necessarily an effective stimulus for another neuron. The monkey was shown one of the two objects, had to keep it in memory, and was then required to find it again in a following stimulus array or scene.

When both objects were present in the scene, the activity of the neuron nonetheless changed to reflect which object the monkey had to attend to, so its activity tended to be greater if the searched-for target was the effective stimulus, and neural activity was lower if the target was the ineffective stimulus. This observation supports the idea that visual attention biases the responses of these object-selective neurons to process mainly one object even when many other objects are present in the neurons' receptive fields (Desimone & Duncan, 1995; Reynolds et al., 1999), a situation that often arises in real-world visual scenes.

The supposition is that biasing signals from spatial priority maps (in the lateral intraparietal area and the frontal eye fields) guide the initial selection of the object that then gets primary representation in brain regions that partake in object recognition (the ventral visual stream). In order for this to be feasible, it is quite important that these priority maps can be computed quickly so that object processing and recognition will be little delayed, even in cluttered scenes with many objects. Electrophysiological work has indeed shown that visual response latency of neurons in the lateral intraparietal area (Tanaka, Nishida, Aso, & Ogawa, 2013) and the frontal eye fields (Schmolesky et al., 1998) can be very short (e.g. frontal eye fields ~50 ms to ~100 ms) while latencies in several areas within the ventral stream tend to be longer (e.g. V4 ~70 ms to ~160 ms, TE1 80 ms to 200 ms; Baylis, Rolls, & Leonard, 1987; Kiani, Esteky, & Tanaka, 2005; Schmolesky et al., 1998; Tamura & Tanaka, 2001).

The response latency of neurons associated with object recognition or spatial attention is nonetheless somewhat variable and might depend on the task, the stimuli being shown (see e.g. Tanaka et al., 2013), and can even vary from trial to trial. In addition, the fact that one event precedes another does not necessitate that the former has a causal effect on the latter. Dual area studies, where neural activity is either simultaneously recorded from two brain regions or manipulated in one region but recorded in another, have the potential of eliminating these concerns. These challenging experiments can answer critical questions related to the timing of visual selection processes and more closely map the interaction between different processing systems.

Armstrong, Fitzgerald, and Moore (2006) electrically stimulated neurons in the frontal eye fields at current levels not strong enough to evoke a saccade but nonetheless known to affect the distribution of covert attention (Moore & Fallah, 2004). At the same time, they recorded neural activity in the ventral stream region V4. When they stimulated in this way, Armstrong and colleagues were able to enhance the V4 representation of an object if it was located at the endpoint of the saccade that the stimulation would have evoked at higher current levels. This was true even when another object was present in the scene and also occupied a nearby location within a V4 neuron’s receptive field, as had previously been demonstrated by manipulating the distribution of attention (Chelazzi et al., 1998; Chelazzi et al., 1993, 2001; Moran & Desimone, 1985; Reynolds et al., 1999). This indicates that neural activity in the frontal eye fields can have a causal effect on visual representations in the ventral stream and might work by prioritizing object processing in a particular location in the visual field.

Despite the apparent causality between the stimulation of the frontal eye fields and ventral stream processing, it was still unclear if activity in the frontal eye fields necessarily preceded that of the ventral stream during active visual search through more complex multi-object scenes. By simultaneously recording from the frontal eye fields and the inferior temporal cortex, Monosov, Sheinberg, and Thompson (2010) attempted to distinguish between an early selection process, in which spatial selection precedes object identification, and a late selection process, in which object identification precedes spatial selection.

In their experiment, monkeys were rewarded for identifying a target object presented with an array of several distractor objects while the monkey maintained central fixation. The target on each trial was known to be either effective or ineffective at driving the responses of an inferior temporal neuron being recorded from. While monkeys were not allowed to look directly to any of the objects, the target location was presumably assigned a high priority, leading to a covert attentional shift. In this task, information on visuospatial selection related to target location could be read out from neurons in the frontal eye fields before information about object identity useful for recognition was available in inferior temporal neurons.

These results are in agreement with those of Zhou and Desimone (Zhou & Desimone, 2011) who simultaneously recorded from both the frontal eye fields and V4. In their task, monkeys were presented with a scene that contained a single searched-for target object and several distractor objects, some of which matched the target in color or shape. Neurons in both the frontal eye fields and V4 tended to respond more to objects that shared features with the target object (and were thus potentially task-relevant), but this feature-based attentional enhancement developed faster in the frontal eye fields than in V4. Feature-based attentional effects in V4 could therefore be caused by top-down selection from the frontal eye fields.

What about the lateral intraparietal area? Buschman and Miller (2007) recorded the activity of neurons in the lateral intraparietal area and the frontal eye fields while monkeys searched for a target within a visual scene. In some cases, the target was not particularly conspicuous because it shared some features with the distractors (similar to the experiment by Zhou & Desimone, 2011); the monkeys therefore needed to find and identify the target based on its remembered appearance. In other cases, the target was of a color different from that of the distractors and was therefore almost immediately noticeable based on bottom-up feature contrast. Buschman and Miller (2007) found that in the former case, neurons of the frontal eye fields signaled the location of the target at an earlier time point than neurons in the lateral intraparietal area. In the latter case, when the target was visually salient, neurons in the lateral intraparietal area “found” the target faster than neurons in the frontal eye fields.

The results of the experiments described above are consistent with the idea that visual recognition mechanisms in the ventral stream are biased toward certain objects in visual scenes because they occupy a location that has already been assigned a high priority by regions such as the lateral intraparietal area and the frontal eye fields. The two regions might both implement a priority map, and these priority maps might very well update each other through the strong connections that exist between these cortical areas (Ferraina, Pare, & Wurtz, 2002; Lewis & Van Essen, 2000; Stanton et al., 1995). They might nonetheless also divide some of their responsibilities, so that priority signals based on bottom-up saliency could possibly arrive first to the ventral stream from the lateral intraparietal area, while priority based on task-relevant yet inconspicuous features might initially be sent from the frontal eye fields. If objects compete for representation in the ventral visual stream when complex multi-object scenes are viewed, then weighing them by their assigned priority can bias this competition in favor of those likely to be behaviorally important.

So far we have talked about how locations of objects might be prioritized so that neural machinery in the ventral stream is mostly used to process objects that are deemed most important at any given time. But how does the ventral stream actually behave when objects need to be found and recognized in realistic scenes?

When objects are presented alone, selectivity of inferior temporal neurons is remarkably insensitive to retinal location (Rolls, Aggelopoulos, & Zheng, 2003). Individual neurons, however, do tend to respond to several different objects, so if these objects were all presented to such a neuron at once, the resulting neural activity might be expected to end in a total cacophony.

This, however, does not actually seem to occur. Rolls and colleagues (Rolls et al., 2003) visually presented an effective object (known to drive the neuron) at various eccentricities away from the center of the visual field. Inferior temporal neurons tended to robustly respond to the object when it was shown at the center of gaze, regardless of whether it was presented alone on a blank background or as part of a scene. The neural firing rate, however, sharply fell off with increased object eccentricity in the complex scene, but stayed relatively high even when the effective object was shown in the periphery of the visual field on a blank background.

The interpretation of Rolls and colleagues (Rolls et al., 2003) is that receptive fields of inferior temporal neurons shrink in complex natural scenes and that neural output is weighed toward objects closest to the fovea. However, the center of gaze often, but not always, coincides with the center of visuospatial attention. While a central location might always be given a relatively high attentional priority, the priority given to a single object on a blank background, even a peripheral one, is also likely to be high since there are no other potential objects that compete with it. It is quite likely that a single object on a blank background will immediately attract the monkey’s attention, and under most normal circumstances, the object would also become the target of the monkey’s next saccade. In a complex scene, the priority given to the object in question is likely to be comparatively lower since multiple other stimuli are present and are also given some weight.

Another possible way of thinking about this phenomenon, therefore, is that inferior temporal neurons process mainly the object that is at the center of attention. This is in alignment with the results of Sheinberg and Logothetis (2001) who found that inferior temporal neurons represent an object in a complex scene before it is actually fixated, as long as it is the target of the next fixation and has thus presumably become the center of attention.

Sheinberg and Logothetis (2001) recorded from inferior temporal neurons while monkeys searched for and identified objects in a complex visual scene (figure 2A). One of dozens of possible target objects was either presented alone on a blank background, or blended into a complex realistic scene. In the latter case, the monkey had to make several saccades within the scene before he would find and recognize the target, just as one would have to do when searching for an object in a cluttered real-world environment.

Under these conditions, they found that responses of inferior temporal neurons appeared to be linked to the monkeys’ recognition performance, even in the complex scenes. They also saw that object selectivity of inferior temporal neurons was similar, regardless of whether a target object was presented alone or in a complex scene. In both cases, the neurons often responded vigorously to a few objects and almost not at all to many others, and the objects that effectively drove the neurons when presented on the blank background also tended to be the preferred objects in the complex scene. This object selectivity did not unfold until just before the monkey made a saccadic eye movement to the target object in a scene, presumably because this was the time when the monkey noticed that the object was there and could recognize it. Once the monkeys seemingly attended to the target's location, these neurons behaved almost as if the rest of the complex environment had been filtered out. In the complex scene condition, it was not unusual for the eyes to be within a few degrees of an effective target object without eliciting a response from a neuron if the monkey's search behavior indicated that it was not actually noticed.

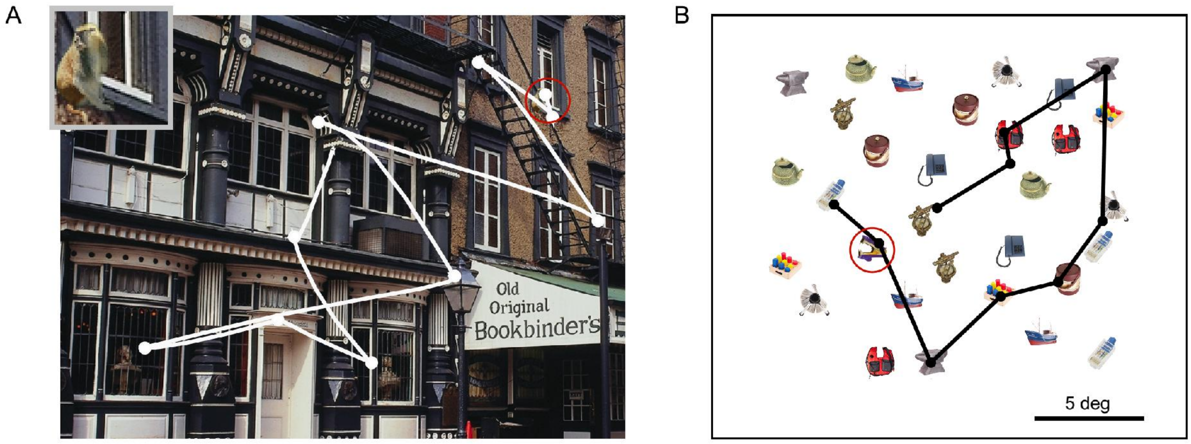

Figure 2. Paradigms for studying visual search in neurophysiological studies. The range of tasks used to understand the physiological contributions of particular neural areas involved in visual search constitutes a large class of related, but distinct, experimental paradigms. Many of these focus on the ability to select a region of space in the context of a simple environment: the subject is presented with arrays of simple stimuli, such as colored patches or lines in various orientations that form the basis for either pop-out or conjunctive search displays (Treisman & Gelade, 1980). Here, the recognition of an object is not a critical component of the task. This broad class of experimental studies has primarily investigated the contributions of the fronto-parietal network with the objective of understanding the physiological process of attentional selection. Some other tasks are designed to investigate object recognition in a scene with other distracting objects or complex backgrounds. These tasks can be used to investigate either the dorsal or ventral visual streams and, sometimes, directly look at the interaction of these two pathways. The visual search for a realistic object embedded within a natural scene is very little used because of the complexities involved in teaching such as task to non-human primates, and the difficulty of analyzing neural responses from an animal engaged in an active search through a complex scene. Two examples of this kind are depicted. A) Sheinberg and Logothetis (2001) trained monkeys to search for and then categorize objects embedded in real-world scenes. Each category of possible objects was arbitrarily assigned to either a left or a right lever press (e.g. pull the left lever whenever you find a monkey in the scene). A saccadic scan path from a single trial is shown; the search target in this case was an image of a monkey. B) In the experiment by (Mruczek and Sheinberg (2007)), monkeys had to search for any one of a number of possible target objects among either familiar or unfamiliar task-irrelevant objects. Once the target object was found, the monkey pushed one of two buttons, only one of which had been arbitrarily deemed to be the correct choice for that particular object.

Effectively filtering out several salient objects that nonetheless happen to be irrelevant for one's current goal might require some learning. Visual object representations in the inferior temporal cortex do appear to be plastic and develop as a function of experience with particular objects. In this region, neural representations for familiar objects appear to be more selective than neural representations for unfamiliar objects (Freedman, Riesenhuber, Poggio, & Miller, 2006; Kobatake, Wang, & Tanaka, 1998; Woloszyn & Sheinberg, 2012), which is likely related to the increased ability to recognize objects that one has encountered many times before.

However, these studies were conducted by presenting isolated objects at fixation. In real-world scenes, a task-relevant object often needs to be found and recognized even in the presence of several unimportant objects that are best left ignored, such as when you need to find the key to your house among your car key, your office key, and the key to your garage. Familiarity with such irrelevant and distracting objects appears to promote more efficient visual search through an environment. For example, Mruczek and Sheinberg (2005) found that people search more effectively through scenes with familiar distracting objects than through scenes with distracting objects that are unfamiliar to them.

Mruczek and Sheinberg (Mruczek & Sheinberg, 2007) then tested whether such familiarity with the contents of a scene affected object processing in the inferior temporal cortex. They trained monkeys to search for a particular object, effective at driving neural responses, hidden among either familiar objects (which had been seen hundreds of times) or novel objects (which were new for each particular testing session; see figure 2B). Mruczek and Sheinberg found that effective target objects could be detected by individual inferior temporal cortical neurons at greater eccentricities when the distracting objects were familiar than when they were unfamiliar. Furthermore, this enlarged functional visual receptive field was predictive of the monkeys' performance, which showed greater efficiency at finding objects among familiar distractors compared to novel ones, just as had been shown for human observers (Mruczek &

Sheinberg, 2005).

Interestingly, the largest differences in firing rate for the target object in the presence of novel and familiar distracting objects were found when the target was shown in the periphery, not in the center of the visual field. This is reminiscent of the results by Rolls et al. (2003) who found that objects in the visual periphery were less effective for driving the responses of inferior temporal neurons when presented in a scene than when presented alone, while the difference between object responses with and without a scenic background was not as apparent when the object was at the center of gaze. It is possible that top-down priority signals are used to down-weight familiar distractors so that a peripheral target object becomes more noticeable, attracts more attention, and is thus more likely to be processed by inferior temporal neurons. A relevant object in a familiar scene might be treated by inferior temporal neurons almost as if the object was the only thing present.

While the inferior temporal cortex likely plays a role in object recognition in visual scenes, it might not always be crucial for this task. Meyers et al. (2010) recorded neural activity in the anterior part of the inferior temporal cortex while monkeys viewed natural scenes, some of which contained animals and some that did not. Meyers and colleagues then looked to see if the neural data contained enough information to judge whether the monkey had been looking at a scene with or without an animal. While decoding accuracy was above what would have been expected by mere chance, the information useful for this classification appeared quite late in the responses of inferior temporal neurons (around 125 ms after the visual onset of a scene). When given the choice of two scenes, only one of which contains an animal, both people (Kirchner & Thorpe, 2006) and monkeys (Girard, Jouffrais, & Kirchner, 2008) can reliably make a saccade to the animal scene before any information about object category is present in inferior temporal neural responses to similar scenes. Such rapid object recognition in natural scenes seems to rely on regions other than the inferior temporal cortex.

This kind of object recognition in visual scenes might not even depend on selective attention. Current theories of visual search (Wolfe, Vo, Evans, & Greene, 2011) have now begun to acknowledge that searching through scenes might involve both a selective and a non-selective path. While the former requires the selection of individual objects, the latter makes use of summary statistics from the whole scene. This global information might in some cases suffice for simple categorization, such as distinguishing between an animal and a non-animal, and could provide the basis for a scene-based guidance of visual search.

Lastly, we must consider the identification and processing of spatial contexts themselves. Kornblith et al. (2013) recently reported electrophysiological data from two newly identified ventral stream areas termed the lateral place patch (LPP) and medial place patch (MPP), a possible macaque homologue to the human parahippocampal place area (PPA). These scene-processing regions contained cells that were strongly responsive not just to individual objects (like neurons in the inferior temporal cortex) or the spatial distribution of object salience or priority (like neurons in the fronto-parietal attention systems) but instead to the environmental structure and texture of both familiar and unfamiliar scenes. Stimulation studies done in concert with the neural recordings suggest that these regions lie within a neural network that feeds back to other regions that process visual information, such as V4. Interestingly, these scene-based responses were observed even while the monkey engaged in a passive fixation task, suggesting that active search or exploration through the environment was not necessary to activate brain regions that process scene stimuli. Performing recordings in these scene selective areas while animals engage in active behavioral tasks may further elucidate the contribution of this processing to search and recognition.

This chapter has attempted to follow a scene as it passes through the visual hierarchy until an object in the scene is found and successfully recognized. We have focused on two divergent yet highly interactive neural systems, a frontoparietal network and the ventral visual stream, one of which allocates and directs visual attention to important features of the environment, while the other processes and identifies the objects in that environment. This division of labor by the two systems is supported by the work reviewed here, as well as behavioral work that indicates that spatial selection and target identification are separable parts of finding objects in visual scenes (Ghorashi, Enns, Klein, & Di Lollo, 2010).

While some progress has recently been made in understanding how the visual system finds and identifies objects in complex environments, several aspects of this work make it difficult to generalize from these experimental results to the more human undertaking of "scene processing".

First, due to the difficulty of analyzing data from experiments with sufficiently complex stimuli, a large majority of electrophysiological work makes use of simplified stimuli that are easier to control but are quite impoverished in comparison to the richness of visual input we encounter in our everyday lives. Since neural responses can be highly nonlinear, it is almost impossible to know how well experimental results derived from highly artificial conditions translate to the types of environments that the visual system actually evolved, developed, and learned to process.

Second, while most human observers no doubt recognize a high resolution photograph of a "beach scene" as a real-world place with three-dimensional structure and individuated objects, and as a setting for particular memories consistent with individual experiences, the animals studied in the research lab arguably lack the former of these traits and definitely lack the latter. How this translation of a two-dimensional photographic representation into a three-dimensional environment with shared experiences changes the outcome or process of neural activity is difficult to say.

Last, even pictures of real-world scenes are in many cases a poor approximation of the genuine thing, because the real world is ever changing. When searching for a friendly face while walking through a crowd of people, we are not only moving within the environment but the environment is also moving, in chaotic ways, around us. How the visual system keeps track of contexts with many moving objects and still manages to search in an efficient and effective manner – and how what has been discovered about searching for and recognizing objects in static scenes translates into dynamic scenes – is for the most part still an open question. It is also the case that we often search in service of some larger behavioral goal (e.g., “make a cup of coffee”), but it is rarer to search per se with no larger goal in mind (e.g., “locate the sugar on the shelf”). How these priority maps dynamically change and update to reflect the nature of our shifting behavioral goals is still challenging to explain.

These open issues deal mainly with the interpretative difficulty in studying the very human experience of “scene perception” in nonhuman animals. In some ways, a comparative psychology and neurobiology of scene perception will always suffer from the difficulty of translating naturalistic vision into the laboratory. But despite this issue, the last decade has seen enormous progress in understanding how the brain makes sense of complex environments and the objects within them and has seen the discovery of multiple homologous human and animal pathways.

Thus, by designing experiments that require animals to search through complex environments, we can better understand how visual processing occurs through the dynamic interaction between multiple brain regions within the more naturalistic context of a visual scene.

References

Allman, J., Miezin, F., & McGuinness, E. (1985). Stimulus specific responses from beyond the classical receptive field: neurophysiological mechanisms for local-global comparisons in visual neurons. *Annu Rev Neurosci, 8*, 407-430. doi: 10.1146/annurev.ne.08.030185.002203

Armstrong, K. M., Fitzgerald, J. K., & Moore, T. (2006). Changes in visual receptive fields with microstimulation of frontal cortex. *Neuron, 50*(5), 791-798. doi: 10.1016/j.neuron.2006.05.010

Attneave, F. (1954). Some informational aspects of visual perception. *Psychol Rev, 61*(3), 183-193.

Baluch, F., & Itti, L. (2011). Mechanisms of top-down attention. *Trends Neurosci, 34*(4), 210-224. doi: 10.1016/j.tins.2011.02.003

Barash, S., Bracewell, R. M., Fogassi, L., Gnadt, J. W., & Andersen, R. A. (1991a). Saccade-related activity in the lateral intraparietal area. I. Temporal properties; comparison with area 7a. *J Neurophysiol, 66*(3), 1095-1108.

Barash, S., Bracewell, R. M., Fogassi, L., Gnadt, J. W., & Andersen, R. A. (1991b). Saccade-related activity in the lateral intraparietal area. II. Spatial properties. *J Neurophysiol, 66*(3), 1109-1124.

Barlow, H. (2001). Redundancy reduction revisited. *Network, 12*(3), 241-253.

Baylis, G. C., Rolls, E. T., & Leonard, C. M. (1987). Functional subdivisions of the temporal lobe neocortex. *J Neurosci, 7*(2), 330-342.

Bisley, J. W., & Goldberg, M. E. (2010). Attention, intention, and priority in the parietal lobe. *Annu Rev Neurosci, 33*, 1-21. doi: 10.1146/annurev-neuro-060909-152823

Blatt, G. J., Andersen, R. A., & Stoner, G. R. (1990). Visual receptive field organization and cortico-cortical connections of the lateral intraparietal area (area LIP) in the macaque. *J Comp Neurol, 299*(4), 421-445. doi: 10.1002/cne.902990404

Bruce, C. J., & Goldberg, M. E. (1985). Primate frontal eye fields. I. Single neurons discharging before saccades. *J Neurophysiol, 53*(3), 603-635.

Bruce, C. J., Goldberg, M. E., Bushnell, M. C., & Stanton, G. B. (1985). Primate frontal eye fields. II. Physiological and anatomical correlates of electrically evoked eye movements. *J Neurophysiol, 54*(3), 714-734.

Buschman, T. J., & Miller, E. K. (2007). Top-down versus bottom-up control of attention in the prefrontal and posterior parietal cortices. *Science, 315*(5820), 1860-1862. doi: 10.1126/science.1138071

Chelazzi, L., Duncan, J., Miller, E. K., & Desimone, R. (1998). Responses of neurons in inferior temporal cortex during memory-guided visual search. *J Neurophysiol, 80*(6), 2918-2940.

Chelazzi, L., Miller, E. K., Duncan, J., & Desimone, R. (1993). A neural basis for visual search in inferior temporal cortex. *Nature, 363*(6427), 345-347. doi: 10.1038/363345a0

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.

Dan, Y., Atick, J. J., & Reid, R. C. (1996). Efficient coding of natural scenes in the lateral geniculate nucleus: experimental test of a computational theory. *The Journal of Neuroscience, 16*(10), 3351-3362.

Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annu Rev Neurosci, 18*, 193-222. doi: 10.1146/annurev.ne.18.030195.001205

Distler, C., Boussaoud, D., Desimone, R., & Ungerleider, L. G. (1993). Cortical connections of inferior temporal area TEO in macaque monkeys. *J Comp Neurol, 334*(1), 125-150. doi: 10.1002/cne.903340111

Dong, D. W., & Atick, J. J. (1995). Statistics of natural time-varying images. *Network: Computation in Neural Systems, 6*(3), 345-358.

Fecteau, J. H., & Munoz, D. P. (2006). Salience, relevance, and firing: a priority map for target selection. *Trends Cogn Sci, 10*(8), 382-390. doi: 10.1016/j.tics.2006.06.011

Ferraina, S., Pare, M., & Wurtz, R. H. (2002). Comparison of cortico-cortical and cortico-collicular signals for the generation of saccadic eye movements. *J Neurophysiol, 87*(2), 845-858.

Field, D. J. (1987). Relations between the statistics of natural images and the response properties of cortical cells. *J. Opt. Soc. Am. A, 4*(12), 2379-2394.

Freedman, D. J., Riesenhuber, M., Poggio, T., & Miller, E. K. (2006). Experience-dependent sharpening of visual shape selectivity in inferior temporal cortex. *Cereb Cortex, 16*(11), 1631-1644. doi: 10.1093/cercor/bhj100

Ghorashi, S., Enns, J. T., Klein, R. M., & Di Lollo, V. (2010). Spatial selection and target identification are separable processes in visual search. *J Vis, 10*(3), 7 1-12. doi: 10.1167/10.3.7

Girard, P., Jouffrais, C., & Kirchner, C. (2008). Ultra-rapid categorisation in non-human primates. *Animal cognition, 11*(3), 485-493.

Goldberg, M. E., Bisley, J. W., Powell, K. D., & Gottlieb, J. (2006). Saccades, salience and attention: the role of the lateral intraparietal area in visual behavior. *Prog Brain Res, 155*, 157-175. doi: 10.1016/S0079-6123(06)55010-1

Goodale, M. A., & Milner, A. D. (1992). Separate visual pathways for perception and action. *Trends Neurosci, 15*(1), 20-25.

Gross, C. G., Bender, D. B., & Rocha-Miranda, C. E. (1969). Visual receptive fields of neurons in inferotemporal cortex of the monkey. *Science, 166*(3910), 1303-1306.

Guo, K., Robertson, R. G., Mahmoodi, S., & Young, M. P. (2005). Centre-surround interactions in response to natural scene stimulation in the primary visual cortex. *Eur J Neurosci, 21*(2), 536-548. doi: 10.1111/j.1460-9568.2005.03858.x

Henderson, J. M., & Hollingworth, A. (1999). High-level scene perception. *Annu Rev Psychol, 50*, 243-271. doi: 10.1146/annurev.psych.50.1.243

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

Hubel, D. H., & Wiesel, T. N. (1959). Receptive fields of single neurones in the cat's striate cortex. *The Journal of physiology, 148*(3), 574-591.

Hubel, D. H., & Wiesel, T. N. (1968). Receptive fields and functional architecture of monkey striate cortex. *The Journal of physiology, 195*(1), 215-243.

Itti, L., & Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Res, 40*(10-12), 1489-1506.

Itti, L., & Koch, C. (2001). Computational modelling of visual attention. *Nat Rev Neurosci, 2*(3), 194-203. doi: 10.1038/35058500

Kayser, C., Salazar, R. F., & Konig, P. (2003). Responses to natural scenes in cat V1. *J Neurophysiol, 90*(3), 1910-1920. doi: 10.1152/jn.00195.2003

Kersten, D. (1987). Predictability and redundancy of natural images. *J Opt Soc Am A, 4*(12), 2395-2400.

Kiani, R., Esteky, H., & Tanaka, K. (2005). Differences in onset latency of macaque inferotemporal neural responses to primate and non-primate faces. *J Neurophysiol, 94*(2), 1587-1596. doi: 10.1152/jn.00540.2004

Kirchner, H., & Thorpe, S. J. (2006). Ultra-rapid object detection with saccadic eye movements: Visual processing speed revisited. *Vision Res, 46*(11), 1762-1776.

Knierim, J. J., & van Essen, D. C. (1992). Neuronal responses to static texture patterns in area V1 of the alert macaque monkey. *J Neurophysiol, 67*(4), 961-980.

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

Land, M. F., & Hayhoe, M. (2001). In what ways do eye movements contribute to everyday activities? *Vision Res, 41*(25-26), 3559-3565.

Lehky, S. R., Sejnowski, T. J., & Desimone, R. (1992). Predicting responses of nonlinear neurons in monkey striate cortex to complex patterns. *J Neurosci, 12*(9), 3568-3581.

Lewis, J. W., & Van Essen, D. C. (2000). Corticocortical connections of visual, sensorimotor, and multimodal processing areas in the parietal lobe of the macaque monkey. *J Comp Neurol, 428*(1), 112-137.

Logothetis, N. K., & Sheinberg, D. L. (1996). Visual object recognition. *Annu Rev Neurosci, 19*, 577-621. doi: 10.1146/annurev.ne.19.030196.003045

Meyers, E., Embark, H., Freiwald, W., Serre, T., Kreiman, G., & Poggio, T. (2010). Examining high level neural representations of cluttered scenes.

Monosov, I. E., Sheinberg, D. L., & Thompson, K. G. (2010). Paired neuron recordings in the prefrontal and inferotemporal cortices reveal that spatial selection precedes object identification during visual search. *Proc Natl Acad Sci U S A, 107*(29), 13105-13110. doi: 10.1073/pnas.1002870107

Montagna, B., Pestilli, F., & Carrasco, M. (2009). Attention trades off spatial acuity. *Vision Res, 49*(7), 735-745.

Moore, T., & Fallah, M. (2004). Microstimulation of the frontal eye field and its effects on covert spatial attention. *J Neurophysiol, 91*(1), 152-162. doi: 10.1152/jn.00741.2002

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

Mruczek, R. E., & Sheinberg, D. L. (2005). Distractor familiarity leads to more efficient visual search for complex stimuli. *Percept Psychophys, 67*(6), 1016-1031.

Mruczek, R. E., & Sheinberg, D. L. (2007). Context familiarity enhances target processing by inferior temporal cortex neurons. *J Neurosci, 27*(32), 8533-8545. doi: 10.1523/JNEUROSCI.2106-07.2007

Nassi, J. J., & Callaway, E. M. (2009). Parallel processing strategies of the primate visual system. *Nat Rev Neurosci, 10*(5), 360-372. doi: 10.1038/nrn2619

Peters, A., & Payne, B. R. (1993). Numerical relationships between geniculocortical afferents and pyramidal cell modules in cat primary visual cortex. *Cerebral Cortex, 3*(1), 69-78.

Ratliff, F., & Riggs, L. A. (1950). Involuntary motions of the eye during monocular fixation. *Journal of experimental psychology, 40*(6), 687.

Reynolds, J. H., Chelazzi, L., & Desimone, R. (1999). Competitive mechanisms subserve attention in macaque areas V2 and V4. *J Neurosci, 19*(5), 1736-1753.

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.

Rucci, M. (2008). Fixational eye movements, natural image statistics, and fine spatial vision. *Network: Computation in Neural Systems, 19*(4), 253-285.

Schall, J. D., Morel, A., King, D. J., & Bullier, J. (1995). Topography of visual cortex connections with frontal eye field in macaque: convergence and segregation of processing streams. *J Neurosci, 15*(6), 4464-4487.

Schmolesky, M. T., Wang, Y., Hanes, D. P., Thompson, K. G., Leutgeb, S., Schall, J. D., & Leventhal, A. G. (1998). Signal timing across the macaque visual system. *J Neurophysiol, 79*(6), 3272-3278.

Sheinberg, D. L., & Logothetis, N. K. (2001). Noticing familiar objects in real world scenes: the role of temporal cortical neurons in natural vision. *J Neurosci, 21*(4), 1340-1350.

Sillito, A. M., Grieve, K. L., Jones, H. E., Cudeiro, J., & Davis, J. (1995). Visual cortical mechanisms detecting focal orientation discontinuities. *Nature, 378*(6556), 492-496. doi: 10.1038/378492a0

Simoncelli, E. P., & Olshausen, B. A. (2001). Natural image statistics and neural representation. *Annu Rev Neurosci, 24*, 1193-1216. doi: 10.1146/annurev.neuro.24.1.1193

Stanton, G. B., Bruce, C. J., & Goldberg, M. E. (1995). Topography of projections to posterior cortical areas from the macaque frontal eye fields. *J Comp Neurol, 353*(2), 291-305. doi: 10.1002/cne.903530210

Tamura, H., & Tanaka, K. (2001). Visual response properties of cells in the ventral and dorsal parts of the macaque inferotemporal cortex. *Cereb Cortex, 11*(5), 384-399.

Tanaka, T., Nishida, S., Aso, T., & Ogawa, T. (2013). Visual response of neurons in the lateral intraparietal area and saccadic reaction time during a visual detection task. *Eur J Neurosci, 37*(6), 942-956. doi: 10.1111/ejn.12100

Thomas, N. W., & Paré, M. (2007). Temporal processing of saccade targets in parietal cortex area LIP during visual search. *J Neurophysiol, 97*(1), 942-947. doi: 10.1152/jn.00413.2006

Thompson, K. G., & Bichot, N. P. (2005). A visual salience map in the primate frontal eye field. *Prog Brain Res, 147*, 251-262. doi: 10.1016/S0079-6123(04)47019-8

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

Tsotsos, J. K. (2008). *What roles can attention play in recognition?* Paper presented at the Development and Learning, 2008. ICDL 2008. 7th IEEE International Conference on.

Ungerleider, L. G., & Haxby, J. V. (1994). 'What' and 'where' in the human brain. *Curr Opin Neurobiol, 4*(2), 157-165.

Ungerleider, L. G., & Mishkin, M. (1982). Two cortical visual systems. *Analysis of visual behavior*, 549-586.

Van Essen, D. C., Anderson, C. H., & Felleman, D. J. (1992). Information processing in the primate visual system: an integrated systems perspective. *Science, 255*(5043), 419-423.

Vinje, W. E., & Gallant, J. L. (2000). Sparse coding and decorrelation in primary visual cortex during natural vision. *Science, 287*(5456), 1273-1276.

Webster, M. J., Bachevalier, J., & Ungerleider, L. G. (1994). Connections of inferior temporal areas TEO and TE with parietal and frontal cortex in macaque monkeys. *Cereb Cortex, 4*(5), 470-483.

Wolfe, J. M. (2007). Guided Search 4.0: Current Progress with a model of visual search. In W. Gray (Ed.), *Integrated Models of Cognitive Systems* (pp. 99-119): Oxford.

Wolfe, J. M., & Horowitz, T. S. (2004). What attributes guide the deployment of visual attention and how do they do it? *Nature Reviews Neuroscience, 5*(6), 495-501.

Wolfe, J. M., Vo, M. L., Evans, K. K., & Greene, M. R. (2011). Visual search in scenes involves selective and nonselective pathways. *Trends Cogn Sci, 15*(2), 77-84. doi: 10.1016/j.tics.2010.12.001

Woloszyn, L., & Sheinberg, D. L. (2012). Effects of long-term visual experience on responses of distinct classes of single units in inferior temporal cortex. *Neuron, 74*(1), 193-205. doi: 10.1016/j.neuron.2012.01.032

Zhou, H., & Desimone, R. (2011). Feature-based attention in the frontal eye field and area V4 during visual search. *Neuron, 70*(6), 1205-1217. doi: 10.1016/j.neuron.2011.04.032