CHAPTER 9

Finding a face in the crowd: parallel and serial neural mechanisms of visual selection

Narcisse P. Bichot and Robert Desimone*

McGovern Institute for Brain Research, Massachusetts Institute of Technology, Bldg. 46–6121, Cambridge, MA 02139, USA

Abstract: At any given moment, our visual system is confronted with more information than it can process. Thus, attention is needed to select behaviorally relevant information in a visual scene for further processing. Behavioral studies of attention during visual search have led to the distinction between serial and parallel mechanisms of selection. To find a target object in a crowded scene, for example a “face in a crowd”, the visual system might turn on and off the neural representation of each object in a serial fashion, testing each representation against a template of the target object. Alternatively, it might allow the processing of all objects in parallel, but bias activity in favor of those neurons representing critical features of the target, until the target emerges from the background. Recent neurophysiological evidence shows that both serial and parallel selections take place in neurons of the ventral “object-recognition pathway” during visual search tasks in which monkeys freely scan complex displays to find a target object. Furthermore, attentional selection appears to be mediated by changes in the synchrony of responses of neuronal populations in addition to the modulation of the firing rate of individual neurons.

Keywords: attention; selection; saccades; visual search; serial; parallel; area V4; synchrony

Introduction

Vision is of primary importance in gathering information about the surrounding world, and we spend much of our time engaged in visual search to find and process behaviorally relevant information in crowded scenes. When viewers know the location of the relevant object, the brain mechanisms that guide spatial attention to the object are largely overlapping with those for selecting the targets for eye movements (Nobre, 2001; Corbetta and Shulman, 2002), consistent with behavioral studies showing a strong functional link between spatial attention and eye movements (Kowler et al., 1995; Deubel and Schneider, 1996; Liversedge and Findlay, 2000). Selection for attention or eye movements lead to an enhancement of the responses of visual cortex neurons to the relevant object, at the expense of distractors (Luck et al., 1997; Colby and Goldberg, 1999; Seidemann and Newsome, 1999; Treue and Maunsell, 1999; Andersen and Buneo, 2002), leaving object recognition mechanisms in the temporal cortex with only a single relevant stimulus at a time (Desimone and Duncan, 1995). However, in most common visual scenes, viewers rarely know the specific location of the relevant object in advance — instead, they must search for it, based on its distinguishing features such as color or shape, which is commonly termed visual search.

For decades, psychologists have debated on how the brain filters out irrelevant information and focuses attention on information that matters, with
many debates centered around the roles of serial and parallel mechanisms in selection (Shiffrin and Schneider, 1977; Treisman and Gelade, 1980; Nakayama and Silverman, 1986; Wolfe et al., 1989; Townsend, 1990). This distinction can be illustrated by considering a complex visual search such as finding Waldo in a crowded page of a “Where’s Waldo?” book. When searching for Waldo, one possibility is that the brain scans the page spatially (serial processing) like a mental spotlight (Posner et al., 1980) moving across an otherwise dark page. In this model, the attentional spotlight would track across the page, checking each object within its “field of illumination” against a mental image of Waldo. Another possibility is that the brain takes in the entire page at once and gradually zooms in on relevant features such as color and shape (parallel processing). In this model, based for example on a bias towards Waldo’s red-striped shirt, objects with the color red would gradually stand out from the cluttered background. Here, we review recent studies from our lab, as well as others, that have investigated the brain mechanisms underlying top-down, feature-based selection during visual search guided by the knowledge of the target’s visual properties.

Parallel selection during visual search

The key element of parallel search models is a neural bias in favor of stimuli containing features (e.g., color or shape) of the searched-for target that occurs throughout the visual field and throughout the time period of the search, long before a target is identified. We recently investigated the presence of such a bias in the activity of visual cortical neurons in monkeys that freely scanned complex visual search arrays to find a target defined by color (Fig. 1A), shape, or both (Fig. 1B) (Bichot et al., 2005). Recordings were conducted in area V4, a key area of the ventral stream for object recognition (Mishkin et al., 1983), where neurons have smaller receptive fields (RFs) compared with the extremely large RFs in inferior temporal (IT) cortex and are selective for basic stimulus features such as color, orientation, and simple shapes (Desimone et al., 1985; Desimone and Schein, 1986).
1987; Schein and Desimone, 1990). Furthermore, previous studies have shown that attentional modulation in area V4 is larger than in areas V1 and V2 and therefore easier to study (Luck et al., 1997).

To test for parallel, feature-based biasing, Bichot et al. (2005) specifically considered responses to stimuli in the RF of neurons at times when the monkey was actively “attending” elsewhere, when the monkey was preparing to make a saccade to a stimulus outside the RF. More specifically, the response to an unselected RF stimulus with the neuron’s preferred or nonpreferred color was compared on trials during which the cue was of the preferred or nonpreferred color for the neuron (Fig. 2A). When a stimulus of the preferred color was in the RF, neurons gave enhanced responses when it matched the cue color. Responses to a RF stimulus of the nonpreferred color, on the other hand, were not modulated by cue color. Similar results were found during shape search taking into account the neurons’ selectivity for stimulus shapes. Thus, neurons responded most strongly when an unselected RF stimulus with the preferred feature was the search target, even though the stimulus was not selected for a saccadic eye movement. For example, if the animal was searching for red, the cells preferred red, and there happened to be a red stimulus inside the RF; this is when firing rate was enhanced, even though the animal was preparing an eye movement to a different stimulus. Furthermore, when the cue was of the preferred color of the neurons, responses to stimuli with colors similar to the cue color also showed some enhancement, explaining why the animals tended to fixate distractors similar to the target.

The feature-based bias observed by Bichot et al. (2005) is consistent with the findings of another study examining neural activity in area V4 during free-viewing visual search (Mazer and Gallant, 2003). In that study, monkeys were trained to search for a target grayscale natural image patch among distractor ones. Across all fixations, they found significant differences in the activity of about one in four V4 neurons across the different search targets. However, it is not clear whether the overall enhancement of neural responses in that study for a particular search target affected all stimuli regardless of features and their similarity to the target, and whether the results were confounded by differences in gaze patterns with different search targets.

As with most studies of the neural substrates of attention (Desimone and Duncan, 1995; Maunsell, 1995), the findings described above show a bias for stimuli that are likely to be the search targets resulting from the increase in the firing rate of
individual neurons in response to those stimuli. However, another potential “amplifier” of the effects of one population of neurons on another is the synchronization of activity in the input population (Gray et al., 1989; Salinas and Sejnowski, 2000; Kara and Reid, 2003). Small changes in high frequency synchronization of spike trains with attention at one stage might lead to pronounced firing rate changes at subsequent stages (Niebur et al., 1993; Salinas and Sejnowski, 2000) because cells generally have short synaptic integration times. Indeed, V4 neurons synchronize their activity when attention is directed to their RFs (Fries et al., 2001), as do neurons in parietal cortex during a memory-saccade task (Pesaran et al., 2002). Also, in monkey somatosensory cortex, it has been reported that cells synchronize their activity when monkeys perform a tactile task compared to a visual task, presumably due to an increase in “tactile attention” in that task (Steinmetz et al., 2000).

To investigate potential changes in neuronal synchronization with feature-based attention, Bichot et al. (2005) measured the coherence between spikes and the local field potential (LFP) (Fries et al., 1997; Jarvis and Mitra, 2001). Spike-field coherence measures phase synchronization between the LFP and spike times as a function of frequency, and is independent of any changes in the firing rate of the spikes and in the power spectrum of the LFP. Coherence for a given frequency ranges from 0 (when the spikes do not have any systematic phase relation to the LFP component at this frequency) to 1 (when all spikes appear at exactly the same phase relation relative to this frequency component). During color search, similar to the effects found on firing rates, neurons increased their synchronization in the gamma-range (30–60 Hz) when an unselected RF stimulus with the preferred color was the target that the animal was searching for (Fig. 2B). There was also a smaller increase in gamma-band synchrony when the preferred color was cued and distractors with similar colors fell in the RF along with a marginal increase for distractors with the nonpreferred color. Similar effects on synchrony were found during shape search. Thus, these results suggest that when the animal is searching for a particular feature, the neurons that prefer that feature begin to synchronize their activity, reaching maximal synchronization when a stimulus with that feature falls within their RF (e.g., when the animal is searching for red, the neurons prefer red, and a red stimulus falls within the RF).

The results described so far show that neurons gave enhanced responses and synchronized their activity in the gamma-range whenever a preferred stimulus in their RF was the target the animal was looking for, but had not found as yet. However, it is not clear from these results whether a distractor with a target feature would share in the bias for target features as proposed by parallel models of visual search (Cave and Wolfe, 1990; Desimone and Duncan, 1995). For example, when searching for a red cross among red circles, green crosses, and green circles, we should see evidence for enhancement of responses and/or synchrony when the RF stimulus contains a single feature of the target (i.e., red circle or green cross) but is not, itself, a target as it lacks the other target feature. Indeed, we found that during a conjunction search, neurons increased and synchronized their activity for unselected distractors in the RF with the preferred color when that color was shared with the target (Fig. 3). Interestingly, although sharing in the bias for the target shape also led to neural enhancement, the magnitude of the effect was much smaller, consistent with behavioral evidence that the monkey used the color information more than the shape information in guiding its search to the color–shape conjunction target (i.e., fixated distractors with the target color more often than distractors with the target shape). Altogether, these results suggest that responses are enhanced whenever a RF stimulus contains a preferred feature of the neurons, and that feature is used in guiding the search for the target.

The source of the top-down bias on V4 activity most likely originates, at least in part, in prefrontal cortex, which has been shown to play an important role in working memory and executive control (Miller and Cohen, 2001). Accordingly, during a match-to-sample task, sample-selective delay activity in prefrontal cortex is maintained throughout the trial even when other test stimuli intervene during the delay, whereas delay activity in IT cortex is disrupted by intervening stimuli (Miller.
et al., 1996). Consistent with the idea that feedback inputs from prefrontal cortex to visual cortex bias activity in favor of behaviorally relevant stimuli, recent preliminary evidence shows that prefrontal lesions impair monkeys’ ability to attend to stimuli based on color cues, but not to stimuli that are salient and pop-out (Rossi et al., 2001).

Serial selection during visual search

Although both the behavioral and the neural evidence for parallel processing during the visual search tasks described above is compelling, it is equally clear that these searches have a serial component in that the monkeys make several saccades to find the target (Fig. 1). To test for spatial attention effects on responses, Bichot et al. (2005) compared neural responses to any stimulus in the RF when it was either selected for a saccade or the saccade was made to a stimulus outside the RF during color and shape feature searches. Neurons responded more strongly to a stimulus in the RF when it was the goal of the impending saccade (Fig. 4A), consistent with the findings of Mazer and Gallant (2003), using a similar free-viewing visual search task. Furthermore, both the magnitude (Fig. 4B) and the spectral power of the LFP in the gamma-frequency range were significantly modulated by monkeys’ decision to make a saccade to the RF stimulus. However, despite such

Fig. 3. Color-related enhancement of neuronal activity and spike-field synchronization in area V4 during conjunction search. Conventions are as in Fig. 2.

Fig. 4. Saccadic enhancement in area V4 during feature searches. (a) Normalized firing rates for the population of neurons when a saccade was made to a stimulus inside the RF (solid line) and when a saccade was made to a stimulus outside the RF (dotted line) across all saccades. Data from color and shape searches were combined. (b) and (c) Average normalized LFP and spike-field coherence for the same conditions.
strong modulation of neuronal firing rates and LFPs, spike-field coherence was unaffected by spatial selection (Fig. 4C).

The lack of modulation of neuronal synchrony with spatial selection during visual search stands in contrast to changes in synchrony that were observed with feature selection in the same search. Furthermore, superficially at least, this result seems at odds with a previous finding by Fries et al. (2001) that gamma-frequency synchronization increases and beta-frequency synchronization decreases when a monkey attends to a stimulus inside the RF of neurons (Fig. 5). A critical factor appears to be the length of time for which attention is maintained for a given feature or location. In the Fries et al. study, monkeys monitored and attended to the target location for up to several seconds, whereas during visual search the time that the animal takes to attend to the location of the next stimulus that will be the target of a saccade is only about 250 ms. The effect of feature-based attention on synchrony during visual search is consistent with this explanation as the animal maintains a state of attention to stimulus features lasting several seconds. Thus, it seems plausible that attentional effects on synchrony take longer to develop than attentional effects on firing rate.

The modulation of responses by the locus of spatial attention that we (Bichot et al., 2005) and others (Mazer and Gallant, 2003) have found in V4 likely involve feedback from structures involved in spatial attention and saccade production. One such structure, the frontal eye field (FEF), is reciprocally connected with areas of both the dorsal and the ventral visual processing streams, and these connections are topographically organized (Schall et al., 1995). Consistent with the idea that feedback from FEF to visual cortex plays a role in spatial selection, recent studies by Moore and colleagues have shown that subthreshold stimulation of FEF improves perceptual ability (Moore and Fallah, 2004) and enhances the visual responses of V4 neurons (Moore and Armstrong, 2003), similar to the effects of spatial attention on perception and neuronal responses.

Studies of neural selection in FEF during visual search have led to the proposal that this area represents the behavioral significance of stimuli...
regardless of their visual features (Thompson and Bichot, 2005), much like the concept of a “salience map” found in many models of visual search (Koch and Ullman, 1985; Treisman, 1988; Cave and Wolfe, 1990; Olshausen et al., 1993; Itti and Koch, 2001). During conjunction search for example, visually responsive FEF neurons initially respond indiscriminately to the target and distractors of the search array (Bichot and Schall, 1999), consistent with the fact that FEF neurons are not selective for visual features such as color (Mohler et al., 1973). However, over time, these neurons not only discriminate the target from distractors, but also discriminate among distractors on the basis of their similarity to the target, even though a saccade is only made to the target location (Fig. 6). In other words, while the highest activation was observed when the target was in the neurons’ RF, the activity in response to RF distractors that shared the target color or shape was also relatively enhanced. Such a spatial map of potential targets would result from topographically organized, convergent input from visually selective neurons with activity biased for target features, and in turn, spatial selection within this map would modulate the activity of visual neurons through feedback connections enhancing the representation of a stimulus selected for an eye movement (Hamker, 2005). Other structures that likely encode a map of behavioral relevance for spatial selection include the lateral intraparietal area (Gottlieb et al., 1998) and the superior colliculus (Findlay and Walker, 1999; McPeek and Keller, 2002).

Conclusion

We have reviewed recent neurophysiological studies showing that both serial and parallel processing of visual information takes places in the brain during visual search, consistent with hybrid models of visual selection (Cave and Wolfe, 1990; Hamker, 2005). Furthermore, these processes are observed in the same brain area and in the same neurons (e.g., area V4) (Mazer and Gallant, 2003; Bichot et al., 2005).

The search for a target based on its features appears to enhance and synchronize the activity of populations of V4 neurons that are selective for and respond preferentially to those features. As a result, stimuli that are similar to the target or that share target features are better represented in the cortex, leading to serial selection among candidate stimuli (Fig. 7). These results offer an explanation for why some visual search tasks are difficult, including some, where targets are defined by the conjunction of different features. Of course, if the results of parallel processing were as clean as depicted in Fig. 7, and the target selection mechanism picked the stimulus location with the highest activation (i.e., winner-take-all), finding the search target would require only one attentional shift or saccadic eye movement. However, evidence from the reviewed studies clearly shows a serial component to the visual search in that several stimuli are fixated before the target is found. Several sources of variability in the activation map from which a target is selected
would account for the inability of the visuomotor system to detect the target immediately. For example, the guided-search model of visual selection has attributed this variability to simple “noisiness” in the firing rates of neurons (Cave and Wolfe, 1990), consistent with the known variability of neuronal responses (Henry et al., 1973; Tolhurst et al., 1983; Shadlen et al., 1996; Bichot et al., 2001). The feature-gate model of selection, on the other hand, suggests the possibility that bottom-up interactions between stimuli are a source of variability in their neural representation (Cave et al., 2005). Furthermore, decreased acuity for peripheral stimuli and cortical magnification also affect the neural representation of stimuli, and it has been shown that high probability detection of targets occurs only within a restricted area surrounding the fixation point (Motter and Belky, 1998).

Finally, several studies have shown that neuronal synchrony (especially in the gamma-frequency range) plays an important role in spatial selection (Fries et al., 2001; Pesaran et al., 2002) as well as featural selection (Bichot et al., 2005), suggesting that synchronizing signals could be a general way the brain focuses on important information. Gamma-frequency synchronization may also contribute to the increased “activation” found in functional magnetic resonance imaging (fMRI) studies of attention, based on findings from a recent study suggesting that the BOLD signal measured in fMRI is closely correlated with synchronous activity in the gamma-range (Kayser et al., 2004). Furthermore, the contribution of neural synchrony to feature-based selection during visual search (Bichot et al., 2005) lends additional support to the idea that synchronized activity has an amplifying role in relaying certain types of behaviorally relevant information from one neural population to the next. Although it has been argued that gamma-frequency synchronization solves the binding problem in visual perception (Engel et al., 1997), the results described here suggest that synchrony may play a much broader role in neural processing.

References


