Experimental Evidence for Top-Down Attentional Selection in the Selective Tuning Model of Visual Attention

Sang-Ah Yoo

A Dissertation submitted to the Faculty of Graduate Studies in Partial Fulfillment
of the Requirements for the Degree of Doctor of Philosophy

Graduate Program in Psychology
York University
Toronto, Ontario

December 2019

ABSTRACT

To overcome limited processing capacity, our visual system facilitates information that relates to the task at hand while inhibiting irrelevant information via selective attention. Among various attention models and theories, the Selective Tuning model of visual attention (ST) is a computation model of visual processing that is based on biological mechanisms. This model emphasizes the role of top-down feedback processing in visual perception and has predicted its unique consequences, such as an attentional surround suppression in which the attentional focus is accompanied by an inhibitory surround. The previous studies have experimentally validated ST's predictions, indicating that the components in ST do reflect actual visual processing in the brain. Nevertheless, many aspects of ST still need to be elaborated and several predictions and assumptions remain untested. The series of works in this dissertation investigate different aspects of top-down feedback processing in visual perception that ST has proposed to corroborate this model and to broaden our understanding of visual attention. The first study examined whether top-down feedback processing is necessary for an attention-demanding, fine-grained visual localization (Chapter 2). The subsequent two studies focused on the properties of different types of the attentional surround suppression, the end-result of top-down feedback processing. The second study suggested the interplay between the location-based and feature-based surround suppression and tested the potential factors that could manipulate the spatial extent of the location-based suppressive surround (Chapter 3). The last study demonstrated feature-based surround suppression in motion processing and its neurophysiological mechanism (Chapter 4). Collectively, this work reinforces functional significance of top-down, attention-mediated feedback for visual processing and supports the validity of ST as well.

ACKNOWLEDGEMENTS

I would like to thank my supervisors Dr. Mazyar Fallah and Dr. John K. Tsotsos who have supported me for the last several years. You have trusted and encouraged me all the time and without your mentoring, I would not have been able to complete my dissertation. Especially, I feel lucky that I had read John's paper while I was searching for a university. That single paper brought me to Canada and provided me invaluable opportunities and experiences of working with awesome people.

I also thank my committee member, Dr. R. Shayna Rosenbaum, for her guidance and reviewing my work. She always gives me detailed and insightful feedback which has improved my research. I extend my gratitude to Dr. Julio Martinez-Trujillo who gave an opportunity to deal with a neurophysiological study. It has made my research much stronger.

I thank my former and current labmates in the Active and Attentive Vision Lab and Visual Perception and Attention Lab. I appreciate research, discussion, conferences, several small and big events, chats, laughs we had together. I hope we can have more good times while I stay in York.

Sincere thanks to everyone who participated in my studies, especially my friends who willingly travelled all the way to York campus. I extend my thanks to the undergraduate research assistants who helped me run experiments.

Special thanks to my parents and the rest of my family in South Korea for their love, prayer, and support. Even though we are miles apart from each other, I can feel your love every day. Also my best friends, thanks for your consistent support that has saved me from all negative feelings.

Finally, and most importantly, all glory to God who gives me strength and leads my life.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1. General introduction

The amount of visual input is overwhelming, and it easily exceeds the limit of processing capacity of the brain. In most cases, however, we can understand the visual world without experiencing difficulties by virtue of selection processes of our visual system. Visual attention plays a crucial role in information selection by facilitating relevant information while inhibiting the irrelevant. Ultimately, attention maximizes our performance in several visual tasks, allowing us to overcome the limited processing capacity. There have been many attempts to elucidate the mechanisms of selective attention (for reviews, see Carrasco, 2011; Desimone & Duncan, 1995; Reynolds & Chelazzi, 2004; Treisman, 1969). Among them, this dissertation focuses on the Selective Tuning model (ST) of visual attention (Tsotsos, 1988, 1990, 1991, 1992, 2011; Tsotsos et al., 1995, 2005) which provides computational and biologically feasible approaches to solve problems in visual perception and proposes concomitant predictions regarding selective attention.

1.1 Overview of the Selective Tuning model

In the ventral pathway of the visual system, neurons in each processing level receive converging input from many other neurons in the previous level and send diverging output to the neurons in the next level (Salin & Bullier, 1995; Tsotsos, 1988). In addition, the sizes of neurons' RFs increase to higher-order areas (Kay, Winawer, Mezer, & Wandell, 2013; Smith, Singh, Williams, & Greenlee, 2001; Zeki, 1978). Due to this structure of the visual system, the representation of a particular object is blurred progressively more in higher-order areas, causing unequal resolution of representation at each processing level in the visual hierarchy (Figure 1.1A, The Blurring Problem (Tsotsos, 2011; Tsotsos et al., 1995)). This issue is aggravated when multiple objects fall within a RF as in our everyday visual experiences. The feed-forward diverging connectivity imposes the Crosstalk Problem that the representations of multiple inputs overlap and interfere with each other (Tsotsos, 2011; Tsotsos et al., 1995), making it difficult to localize the stimulus-of-interest (Figure 1.1B). Therefore, these problems call for a mechanism which filters out irrelevant signals within the RF and binds correct features for intact object perception.

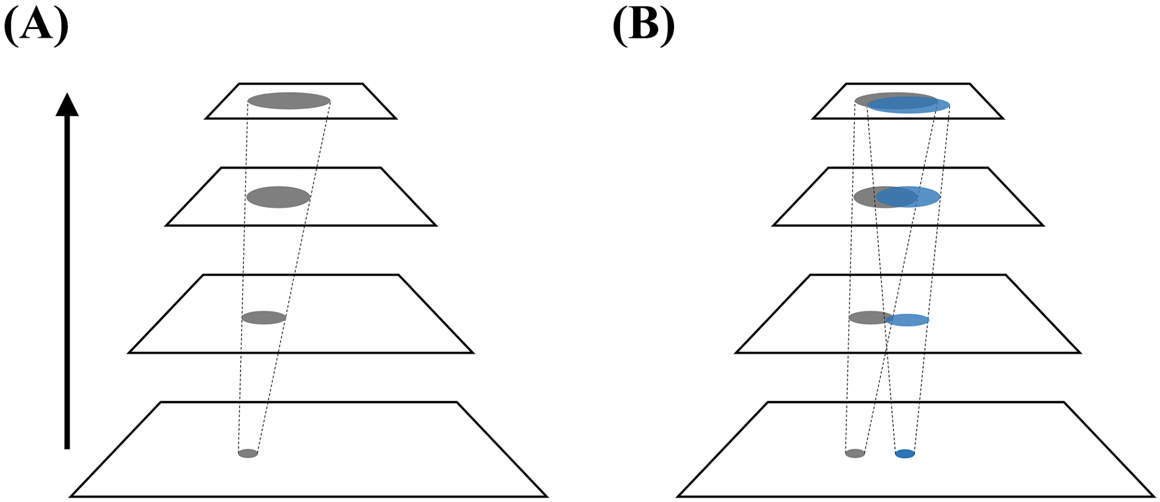


Figure 1.1. Representational problems caused by the structure of the visual system.

(A) Blurring Problem where the representation of a visual input element progressively blurred feed-forwardly due to diverging neural connections. (B) When there are multiple visual inputs their blurred representations would overlap and interfere each other, resulting in the Crosstalk Problem.

A core idea of ST is that top-down feedback (recurrent) processing throughout visual hierarchy plays an important role in attentional selection to resolve such problems in vision. ST explicitly implemented attentional selection processes based on biologically plausible assumptions about neural connectivity in the visual processing hierarchy (Felleman & Van Essen, 1991; Salin & Bullier, 1995). More details about top-down attentional selection in ST will be covered in the following sections of this chapter. Figure 1.2 illustrates how attentional selection processes operate in a hypothetical pyramid processing hierarchy with layers which represent low to high visual areas. In ST, when a stimulus is registered at the input level, it propagates in a feed-forward manner and activates the neurons in the next level that represent the same stimulus. This feed-forward pass results in a diverging cone of activity across the processing hierarchy for each stimulus element. Once the feed-forward stimulus wave reaches the topmost level, top-down winner-take-all (WTA) processes search for the strongest inputs to neurons at each processing level, and then prune away task-irrelevant neural connections around the selected inputs, improving signal-to-noise ratio of the target input. The topmost level is not inhibited by this mechanism. These top-down processes narrow down the attentional focus of relevant input from level to level so that it matches the decreasing size of RF at lower levels of the visual hierarchy. They also operate recursively through the processing hierarchy, hence, the global winning neural connections at the topmost level is traced back to its source at the input level and the selected stimuli at the input level re-propagate. By the end of this top-down competitive selection, the neural connections that are not pruned comprise the pass zone of the attentional beam, while the pruned connections form the attentional suppressive surround (orange annulus at each processing level in Figure 1.2). The other neural connections located far from the attended stimulus that do not interfere with the processing of the selected stimulus remain unaffected. This results in a center-surround attentional profile (hypothesized to abstractly be a Difference of Gaussians) which enhances the attentional resolution and refines location information of the attended stimulus by attenuating interference of nearby distractors.

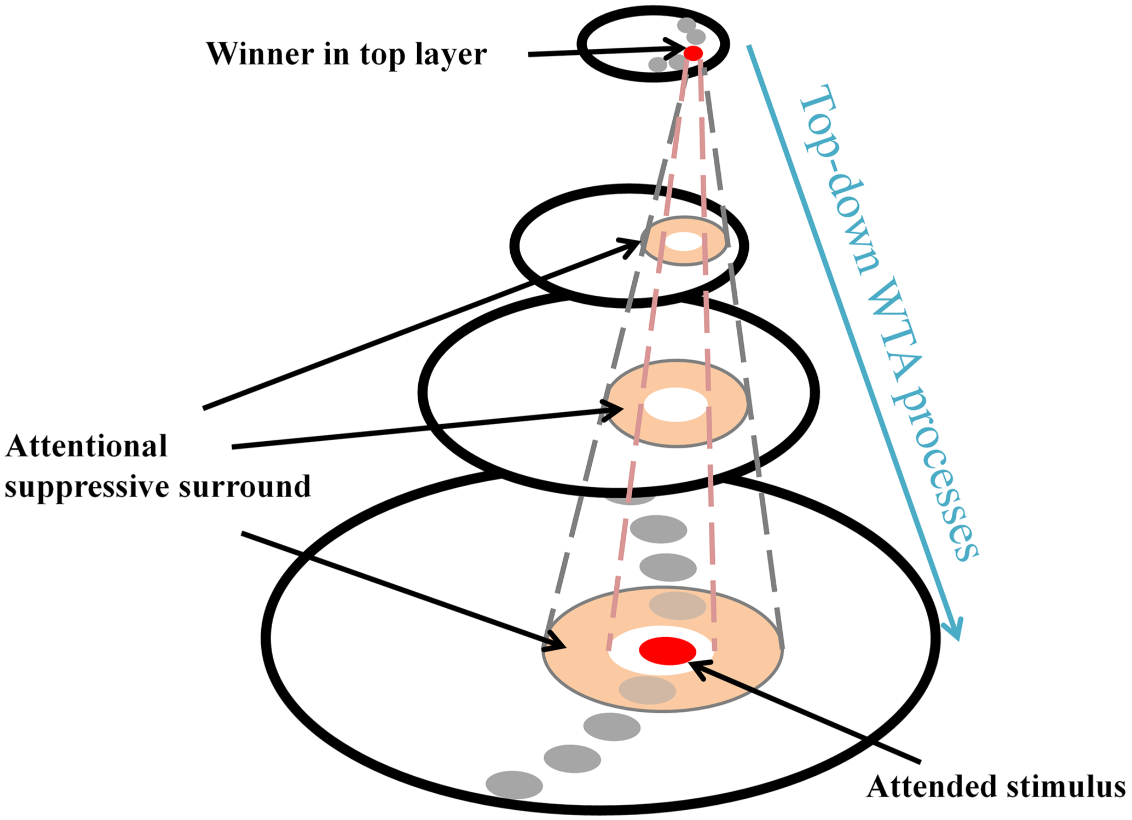


Figure 1.2. Schematic of attentional selection as suggested by the Selective Tuning model. Once the input signal from an attended stimulus reaches the topmost level in a feed-forward manner, a top-down winner-take-all (WTA) mechanism operates to localize this stimulus. During this process, the mechanism inhibits irrelevant connections that surround the attended stimulus at each processing level and it consequently gives rise to the attentional surround suppression (adapted from Hopf et al. (2006) with permission).

Unlike many other attention models (see Rothenstein & Tsotsos (2014) for detailed comparisons between ST and the other models), ST is not a data-fitting model which relies on a set of equations that provides good fit to the existing experimental data. Because of this factor, ST does not give quantitative predictions such as actual response times, neural response curves, or accuracies that can be directly compared to experimental results. Importantly, however, this also means ST can give counter-intuitive predictions that are not of the same character as existing experiments (Braun, Koch, & Davis, 2001). Several predictions of this kind have already been shown to lead to new knowledge about attentional processing, such as attentional surround suppression something which data-based models may not provide.

1.2 Attentional binding strategies in the Selective Tuning model

Our visual system can extract meaningful elements from a visual scene and associate them to represent a unified object. This binding process seems to be achieved effortlessly in most cases. Nevertheless, understanding its underlying mechanisms has been a great challenge (binding problem; see *Neuron*, 24(1), 1999 for reviews) and different solutions to the binding problem have been proposed. For example, *convergent hierarchical coding* suggests that object recognition occurs when a specialized neuron or a group of neurons for each possible combination of the features is activated by convergent input from the neurons of earlier processing level (Epstein & Kanwisher, 1998; Ghose & Maunsell, 1999; Kanwisher, McDermott, & Chun, 1997; Riesenhuber & Poggio, 1999, 2002). The critical problem of this coding scheme is that it requires a myriad of neurons to represent all possible features with details, leading to a combinatorial explosion. In addition, exhaustive learning of features should be needed to form high object specificity. Alternatively, *population coding* (distributed coding) proposes that an object is represented by a particular firing pattern of a population of distributed neurons, making the binding processes more economical (Carlson, Hogendoorn, Fonteijn, & Verstraten, 2011; Haxby et al., 2001; Pasupathy & Connor, 2002; Quian Quiroga, Reddy, Kreiman, Koch, & Fried, 2005; Young & Yamane, 1992). A problem with population coding arises when more than one object appears simultaneously. Due to coactivation of neurons for multiple features, information of their membership in the original objects becomes ambiguous, causing the superposition catastrophe (von der Malsburg, 1999). In the scheme of *temporal coding*, neurons representing features that belong to an object fire synchronously, thus, no superposition catastrophe happens (Gray, 1999; Gray, König, Engel, & Singer, 1989; Singer, 1999). However, it remains unclear whether neuronal synchronous firing indeed accomplishes feature binding, or it is merely an artifact. Also, temporal coding cannot maintain location information of combined features. *Feature integration theory* suggests that focused attention plays a crucial role in feature binding (Treisman, 1998; Treisman & Gelade, 1980; Wolfe & Cave, 1999). Individual features of an objects are bound together by means of focused attention to this object's location. Failure of allocating attention to object location leads to binding errors (Treisman & Schmidt, 1982). While location information is emphasized in the feature integration theory, this theory does not deal with spatial overlap of multiple objects and changes in precision of location information across the visual hierarchy caused by Blurring and Crosstalk Problems (Tsotsos, Rodríguez-Sánchez, Rothenstein, & Simine, 2008).

ST proposed different types of binding strategies integrated with top-down attentional selection mechanisms described in the previous section (Rothenstein, Rodríguez-Sánchez, Simine, & Tsotsos, 2008; Tsotsos, 2011; Tsotsos et al., 2008). Importantly, ST suggested that the nature of a given visual task (e.g., the level of processing that the task requires) determines what binding strategy should be employed and that processing time required to complete each task should also vary depending on the type of the binding strategy. For example, people can rapidly determine whether a particular object category is present in a briefly presented image and it is true even if the visibility of the image is reduced due to backward-masking or rapid serial visual presentation (RSVP) (Fabre-Thorpe, Delorme, Marlot, & Thorpe, 2001; Grill-Spector & Kanwisher, 2005; Keysers, Xiao, Földiák, & Perrett, 2001; Potter, 1976; Potter, Wyble, Hagmann, & McCourt, 2014; Rousselet, Macé, & Fabre-Thorpe, 2003; Thorpe, Fize, & Marlot, 1996; VanRullen & Koch, 2003; VanRullen & Thorpe, 2001). Searching for a pop-out element in visual scene is also very fast (Treisman, 1985; Treisman & Gormican, 1988; Treisman & Souther, 1985). These sorts of tasks do not require precise location information of features as it is sufficient to check the presence/absence of the task-related features of the object. Therefore, they are likely to be performed within the feed-forward temporal range, about 150 ms after the stimulus onset or even substantially earlier (Cichy, Pantazis, & Oliva, 2014; Fabre-Thorpe, Richard, & Thorpe, 1998; Hung, Kreiman, Poggio, & DiCarlo, 2005; Isik, Meyers, Leibo, & Poggio, 2014; Linkenkaer-Hansen et al., 1998; Liu, Agam, Madsen, & Kreiman, 2009; Liu, Harris, & Kanwisher, 2002; Mouchetant-Rostaing, Giard, Delpuech, Echallier, & Pernier, 2000; Riesenhuber & Poggio, 1999; Schendan, Ganis, & Kutas, 1998; Serre, Oliva, & Poggio, 2007). In ST's framework, *Convergence Binding* enables those tasks by determining the strongest neural representations at the highest level in the processing hierarchy in a feed-forward fashion (Tsotsos et al., 2008). Convergence Binding suffices for a task only when visual inputs represented within an RF are not too similar and they do not interfere each other. Since this type of binding is based on the pure feed-forward pass, it does not reduce interference from the task-irrelevant distractors via the attentional surround suppression (i.e., the end-result of a top-down selection).

Other types of visual tasks could ask for more details about stimuli. For example, if one needs to identify what kind of dog s/he currently sees, information such as color, shape, or size is necessary to analyze the dog. These features are lower-level and would be found in earlier areas of the visual hierarchy, hence, the visual system needs downward signals to access these earlier areas where task-relevant features are represented. The extent of the feedback traversal is determined by the task requirements – i.e., how much detailed information is required to achieve a task, and the time taken for feedback processing can indicate the stage of recognition because each step in the visual hierarchy requires additional processing time (c.f., Johnson & Olshausen, 2003). If additional information that is needed to perform a certain visual task is represented in intermediate cortical areas, *Partial Recurrence Binding* allows the access to this information. Partial Recurrence Binding may recover coarse location information. Neuroimaging studies demonstrated that coarse location information is available in higher object-selective areas (Carlson et al., 2011; Cichy, Chen, & Haynes, 2011; DiCarlo & Maunsell, 2003; Grill-Spector et al., 1999; Hemond, Kanwisher, & Op de Beeck, 2007; MacEvoy & Epstein, 2007; Niemeier, Goltz, Kuchinad, Tweed, & Vilis, 2005; Sayres & Grill-Spector, 2008; Schwarzlose, Swisher, Dang, & Kanwisher, 2008). Partial downward signals to these areas might be able to complete a visual task which asks for approximate object location (Crouzet, Kirchner, & Thorpe, 2010; Evans & Treisman, 2005; Kirchner & Thorpe, 2006). Partial Recurrence Binding can occur not only because of task requirements but also because a full traversal to the lowest area is interrupted by the following new visual inputs or tasks so the traversal is not allowed to complete. If this binding process is interrupted, it may result illusory conjunctions or even absence of awareness for the visual features.

When precise location information is required for a visual task, *Full Recurrence Binding* which involves the full top-down traversal to the earliest visual areas is needed. Full Recurrence Binding retraces the neural connections that elicits the strongest response at the top of the processing hierarchy, and processing time required for this is about 250 ms after stimulus onset (Boehler, Schoenfeld, Heinze, & Hopf, 2008; Boehler, Tsotsos, Schoenfeld, Heinze, & Hopf, 2009; Di Russo, Martinez, & Hillyard, 2003; Martínez et al., 1999, 2001; Noesselt et al., 2002; Tsotsos et al., 2008). This process is what Figure 1.2 illustrates. As previously emphasized, a critical consequence of this top-down traversal is the attentional surround suppression which actively clears noise and improves localization precision. Experimentally, it has been demonstrated that the attentional surround suppression occurs only if a visual task is attention-demanding. Furthermore, the attentional surround suppression is associated with a clear feedback sweep across the visual hierarchy, down to area V1 (Boehler et al., 2009) and it is manifested at about 250 ms after stimulus onset. This delay is beyond the temporal range of the initial feed-forward sweep in human visual system (Foxe & Simpson, 2002) and it is in line with the temporal range of attention-driven recurrent activity modulations in early visual cortex (Boehler et al., 2008; Di Russo et al., 2003; Martínez et al., 1999, 2001; Mehta, Ulbert, & Schroeder, 2000; Noesselt et al., 2002).

These three binding strategies described above would ideally work when stimuli are simple, well separated, and can be easily segregated from the background. However, visual scenes we are confronted in real world consist of several, more complex elements so multiple saccades or attentional shifts should be engaged to achieve visual tasks in these environments.

The complexity of visual scenes and recognition tasks may call for multiple iterations of the feed-forward and feedback processes until the task goal is satisfied – i.e., *Iterative Recurrence Binding*. There are several tasks that require Iterative Recurrence Binding. One intuitive example would be visual search, specifically one that involves visits to multiple spatial locations (e.g., conjunction search with numbers of distractors). Another example that shows the need of Iterative Recurrence Binding is the recovery of a shape from motion (form-from-motion) (Rothenstein et al., 2008; Tsotsos et al., 2008). Imagine a square which consists of random dots placed on a similar random dot (static) background. The square is initially indistinguishable from the background but once the dots in the square show translational motion, this square is naturally perceived even though its edge information is not explicitly presented. The initial feed-forward pass (Convergence Binding) detects the presence of the motion and the subsequent feedback processing (Full Recurrence Binding) localizes the motion in the input level and suppresses its surround. This surround suppression enables detection of the contour information necessary for recognizing the shape. The next feed-forward/feedback cycle refines the edges of the shape and eventually binds it with the motion.

1.3 Kinds of feedback processes

Several different terms are often used interchangeably with “feedback” processing, such as “top-down”, “recurrent”, “reentry”, and “backpropagation”. Although they conceptually indicate the signals from higher cortical areas that influence lower areas, the nuances and the roles of feedback processing could vary depending on contexts. Feedback signals could be triggered solely by endogenous high-level cognition, not necessarily requiring the precedent feed-forward pass of the input stimulus.

Table 1.1 lists some potential factors that cause this type of feedback signals, and the related empirical studies and theories/models.

Category	**Empirical studies**	**Theories and models**
Task-dependency	Betz, Kietzmann, Wilming, & König (2010), Morishima et al. (2009), Watanabe et al. (1998)	Gilbert & Li (2013)
Memory-guided	**Working memory** Awh, Jonides, & Reuter-Lorenz (1998), Chelazzi, Duncan, Miller, & Desimone (1998), Downing (2000) **Implicit memory** Chun (2000), Chun & Jiang (1998, 2003), Wolfe, Butcher, Lee, & Hyle (2003) **Long-term memory** Moores, Laiti, & Chelazzi (2003), Summerfield, Lepsien, Gitelman, Mesulam, & Nobre (2006), Tomita, Ohbayashi, Nakahara, Hasegawa, & Miyashita (1999)	Hamker (2005), Wolfe (1994)
Expectation	**Attentional cueing and preparation** Esterman & Yantis (2010), Kastner, Pinsk, De Weerd, Desimone, & Ungerleider (1999), Kim & Cave (1999), Peelen & Kastner (2011), Puri, Wojciulik, & Ranganath (2009),	LaBerge, Auclair, & Sieroff (2000), Posner (1980), Wolfe (1994), Wolfe, Horowitz, & Michod (2007)

	Stokes, Thompson, Cusack, & Duncan (2009)	
	Predictive coding (hypothesis testing) Alink, Schwiedrzik, Kohler, Singer, & Muckli (2010), Kok, Rahnev, Jehee, Lau, & de Lange (2012), Summerfield, Trittschuh, Monti, Mesulam, & Egner (2008)	Friston & Kiebel (2009), Lee & Mumford (2003), Mumford (1991), Rao & Ballard (1999), Spratling (2010)
	Context Auckland, Cave, & Donnelly (2007), Biederman, Mezzanotte, & Rabinowitz (1982), Boyce, Pollatsek, & Rayner (1989), Chun (2000), Chun & Jiang (1998, 2003), Davenport & Potter (2004), Palmer (1975)	Bar (2004), Friedman (1979), Torralba, Oliva, Castelhano, & Henderson (2006)

Table 1.1. A list of potential factors that can induce feedback signals without perceptual inputs and their supporting evidence.

On the other hand, feedback processing could be also subjected to prior feed-forward processing and this is what the current dissertation focuses on. Table 1.2 summarizes major cognitive entities where the roles of feedback processing caused by the initial feed-forward sweep have been studied. Among these cognitive entities, ST specifically deals with visual recognition (feature binding) and attentional selection via recurrent feedback.

Category	Empirical studies	Theories and models
Visual recognition	**Figure-ground segmentation** Hupé et al. (1998), Poort et al. (2012), Self, van Kerkoerle, Supèr, & Roelfsema (2013)	Jehee, Lamme, & Roelfsema (2007), Layton, Mingolla, & Yazdanbakhsh (2014), Roelfsema, Lamme, Spekreijse, & Bosch (2002), Zipser, Lamme, & Schiller (1996)
	Perceptual grouping (contour integration) Altmann, Bülthoff, & Kourtzi (2003), Chen et al. (2014), Li, Piëch, & Gilbert (2006), Seghier & Vuilleumier (2006), Shpaner, Molholm, Forde, & Foxe (2013), Volberg & Greenlee (2014)	Roelfsema, (2006), Roelfsema & Houtkamp (2011)
	Feature binding Bouvier & Treisman (2010), Juan & Walsh (2003), Koivisto & Silvanto (2011, 2012)	Rothenstein et al. (2008), Treisman (1996), Tsotsos et al. (2008), Wyatte, Herd, Mingus, & O'Reilly (2012b)
	Object perception Bar et al. (2006), Barceló, Suwazono, & Knight (2000), Kveraga, Boshyan, & Bar (2007)	Bar et al. (2006), O'Reilly, Wyatte, Herd, Mingus, & Jilk (2013), Ullman (1995)
	Scene perception Koivisto, Railo, Revonsuo, Vanni, & Salminen-Vaparanta (2011), Peyrin et al. (2010), Schyns & Oliva (1994)	Bullier (2001)
Perceptual learning	Ahissar & Hochstein (1997, 2000), Ito, Westheimer, & Gilbert (1998), Sigman & Gilbert (2000)	Hochstein & Ahissar (2002), Layher, Giese, & Neumann (2014a), Layher, Schrodt, Butz, & Neumann (2014b)

Visual attention	**Backpropagation of attentional modulation to lower cortical areas (delayed feedback)** Boehler et al. (2009), Buffalo, Fries, Landman, Liang, & Desimone (2010), Di Russo et al. (2003), Martínez et al. (2001), Mehta et al. (2000), Noesselt et al. (2002), Roelfsema, Lamme, & Spekreijse (1998), Saalmann, Pigarev, & Vidyasagar (2007), Vidyasagar (1998) **Top-down attentional selections** Boehler et al. (2009), Hopf et al. (2006), Hopf, Boehler, Schoenfeld, Heinze, & Tsotsos (2010)	Deco & Zihl (2001), Fukushima (1986), Navalpakkam & Itti (2006), Spratling & Johnson (2004), Tsotsos (1988, 1990), Tsotsos et al. (1995)
Visual awareness	Boehler et al. (2008); Fahrenfort, Scholte, & Lamme (2007); Koivisto, Mäntylä, & Silvanto (2010), Lamme, Zipser, & Spekreijse (2002), Pascual-Leone & Walsh (2001), Ro, Breitmeyer, Burton, Singhal, & Lane (2003), Silvanto, Cowey, Lavie, & Walsh (2005)	Di Lollo, Enns, & Rensink (2000), Hochstein & Ahissar (2002), Silverstein (2015)

Table 1.2. Cognitive entities where the role of feedback processing following the feed-forward pass of visual inputs is emphasized.

Different types of feedback processes can be also defined based on their time courses. While attention-mediated feedback processing to early visual areas is slow (occurs ~250 ms after stimulus onset), rapid local recurrent processing happens much earlier, as quick as 10 ms after the initial feed-forward sweep (for a review, see Wyatte, Jilk, & O'Reilly (2014)). This local recurrent processing is known to be involuntary and independent of attentional effects (Boehler et al., 2008; Roland, 2010; Roland et al., 2006). Studies demonstrated that local recurrent activity plays roles in amplifying visual inputs under challenging conditions, such as filling-in missing information of degraded visual inputs (Mohsenzadeh, Qin, Cichy, & Pantazis, 2018; Wyatte et al., 2014), and gating visual awareness (Boehler et al., 2008).

Early on, theories and computational models of the visual system mainly focused on bottom-up, feed-forward processes of visual information processing (Itti & Koch, 2000; Itti, Koch, & Niebur, 1998; Koch & Ullman, 1987). On the other hand, later theories and models tried to implement the different roles of feed-forward and feedback connections based on anatomical evidence (Felleman & Van Essen, 1991; Lamme, Supèr, & Spekreijse, 1998), and ST also explicitly incorporates feedback processing. Nevertheless, the ways of implementing feedback processing are different depending on theories/models and they even predict different outcomes. The table below (Table 1.3) shows the list of theories/models that incorporate feedback processing (or top-down modulation). It also summarizes the roles of feedback in these models and how they are different from ST.

Theories / Models	**Role of feedback processing**	**Comparisons with ST**
Feature integration theory (Treisman, 1996)	Feedback (reentrant) processing retraces features to the early visual areas to	The original FIT did not include feedback processing as its component (Treisman &

	confirm whether feature conjunction is accurate.	Gelade, 1980), and it was added later. FIT does not provide specific explanations/mechanisms of feedback processing.
Reverse hierarchy theory (Hochstein & Ahissar, 2002)	Feedback processing (*vision with scrutiny*) accesses lower cortical areas for detailed information and conscious perception.	RHT provides only general idea of abstract feed-forward vision and detailed feedback vision.
Computational model of object substitution (Di Lollo et al., 2000)	Feedback pass verifies a hypothesis on input by comparing high-level representation with the initial pattern of activity at lower level.	CMOS does not fully implement selective attention, except for set size effect (more focuses on visual consciousness/awareness).
Neural network model of selective attention (Fukushima, 1986)	Feedback (efferent) signals reach the input layer and facilitate only the afferent paths relevant to the target pattern.	Unlike Fukushima's model, ST does not have a facilitatory mechanism (facilitation is the consequence of inhibition).
Biased competition model (Reynolds, Chelazzi, & Desimone, 1999) and Normalization model of attention (Reynolds & Heeger, 2009)	In BC, top-down attentional modulation increases input strength for the attended stimulus by multiplicatively modulating the synaptic strengths of inter-regional connections.	BC's mechanism for the multiplicative modulation of synaptic weights is unknown (data-fitting model). Also, Normalization model is presented as a mathematical abstraction, without specifying the underlying biophysical mechanisms or neural circuitry.

	Divisive normalization could be implemented via feed-forward or via recursive lateral and/or feedback connections.	ST does not have a facilitatory mechanism and attentional suppression happens during feedback processing.
Feedback model of visual attention (Spratling & Johnson, 2004)	Top-down signal modulates the activations of nodes, rather than the efficiencies of synapses, and then the attended information is routed to higher cortical areas (based on biased competition).	Feedback model is used to simulate neurophysiological data, so values for the synaptic weights are arbitrarily chosen to provide the best fit (similar to BC).
Reentry hypothesis (Hamker, 2005)	Reentry signal represents the expectation of the input (information in VWM) and transfer this information in higher areas to the areas in the ventral stream, controlling gain of the feed-forward signal.	This model specifically simulated neuronal responses during the memory-guided search (Chelazzi et al., 1998).

Table 1.3. A list of models that implemented feedback processing and the comparisons between these models and ST

1.4 Location-based surround suppression

There have been several attempts to characterize the spatial distribution of visual attention. Major models of attention have proposed different metaphorical terms to describe the distribution of spatial attention. For example, the "spotlight" model (Posner, Snyder, &

Davidson, 1980; Shulman, Remington, & McLean, 1979) claims that attentional enhancement of visual information occurs within a circumscribed region of visual space or within the confines of an attended object. While information processing inside the attentional spotlight is facilitated, information outside the spotlight is neglected. The “zoom lens” model is an extension of the spotlight model which suggests that the spatial extent of the attentional focus can vary and that its size is inversely related to attentional resource within the focus (Eriksen & St. James, 1986; Eriksen & Yeh, 1985). In addition, the gradient model proposes that attentional resources are concentrated in the center of the attentional focus, whereas they are gradually reduced with distance from the center (Downing, 1988; Laberge & Brown, 1986; LaBerge & Brown, 1989). Although these models have different views on how attentional resources are allocated within the attentional focus, they all suggest that only information within the attentional focus is processed while information outside the attentional focus disappears. However, mounting evidence has suggested that these versions of the attentional spotlight may not always be correct by demonstrating that the form of the attentional distribution could be more complex (Awh & Pashler, 2000; Egly & Homa, 1984; Eimer & Grubert, 2014; McMains & Somers, 2004; Müller & Hübner, 2002; Müller, Malinowski, Gruber, & Hillyard, 2003). Furthermore, there is accumulating behavioral and neurophysiological evidence for the center-surround attentional distribution where the focus of attention is accompanied by an annulus suppressive surround (Bahcall & Kowler, 1999; Boehler, Tsotsos, Schoenfeld, Heinze, & Hopf, 2011; Boehler et al., 2009; Caputo & Guerra, 1998; Cave & Zimmerman, 1997; Cutzu & Tsotsos, 2003; Hopf et al., 2006, 2010; Mounts, 2000a, 2000b; Müller & Kleinschmidt, 2004; Müller, Mollenhauer, Rösler, & Kleinschmidt, 2005; Puckett & DeYoe, 2015; Schall, Sato, Thompson, Vaughn, & Juan, 2004; Vanduffel, Tootell, & Orban, 2000). Unlike the spotlight metaphor of the attentional distribution, the center-surround distribution proposes that the suppressive mechanism only attenuates visual processing within the suppressive surround (i.e., the region adjacent to attentional focus) but visual processing beyond this inhibitory zone is not affected. Note that the attentional surround suppression in ST, which is due to top-down influences rather than sensory, horizontal, or lateral influences (Boehler et al., 2009; Cutzu & Tsotsos, 2003; Hopf et al., 2010), is different from the surround suppression in which visual stimuli located outside of the classical RF modulate the neuron's response to stimuli within the RF (Adesnik, Bruns, Taniguchi, Huang, & Scanziani, 2012; Haider et al., 2010; Ozeki, Finn, Schaffer, Miller, & Ferster, 2009; Self et al., 2014). Indeed, the attentional surround suppression manipulates the center-surround structure of a neuron's classical RF where an attended stimulus lies (Tsotsos, 1990, 2011; Tsotsos et al., 1995) and such major changes in the RF structure also have been observed elsewhere (Anton-Erxleben, Stephan, & Treue, 2009; Womelsdorf, Anton-Erxleben, Pieper, & Treue, 2006; Womelsdorf, Anton-Erxleben, & Treue, 2008). ST also predicts that the spatial extent of the suppressive surround is determined by the size of a neuron's RF that best represents the attended stimulus (Tsotsos, 2011).

One of the pioneering studies that demonstrated location-based surround suppression in ST's framework was conducted by Cutzu and Tsotsos (2003). In this study, the participants' task was to discriminate whether a cued stimulus and a probe presented on an invisible annulus were the same or different while the distance between them systematically varied. The hypothesis was that there would be a suppressive surround around the cued item where attention was initially directed, and that target discrimination would suffer if a *nearby* probe falls within this suppressive surround. The result demonstrated that target discrimination accuracy was lowest when the cued item and the probe were closest to each other, but accuracy gradually improved as the inter-target distance increased, supporting the existence of the location-based attentional surround suppression. Follow-up magnetoencephalographic (MEG) studies showed that the electromagnetic response to the probe increased when it was presented at the location of the attended stimulus, decreased when it was presented next to the attended stimulus, and then recovered when the separation between the attended stimulus and the probe became greater (Boehler et al., 2009; Hopf et al., 2006, 2010). Location-based surround suppression occurs if and only if a visual task requires top-down spatial localization of the attended stimulus to regain finer spatial resolution. For example, a pop-out color detection task that can be performed without precise feature binding/localization (Treisman, 1998; Treisman & Gelade, 1980) does not produce an attentional suppressive surround, whereas orientation discrimination task forms a stable suppressive surround. In addition, the surround suppression is manifested after 250 ms relative to stimulus onset (Boehler et al., 2009) and it corresponds to the time course that top-down, attention-mediated feedback processing reaches earlier visual areas (Boehler et al., 2008; Martínez et al., 1999, 2001; Mehta et al., 2000; Noesselt et al., 2002; Wyatte et al., 2014).

The idea of attentional surround suppression due to top-down attentive localization first appeared in (Tsotsos, 1988, 1990), earlier than most other models of attention. However, one might ask if other attentional models could also account for the attentional surround suppression. Specifically, the biased competition model (Desimone & Duncan, 1995; Reynolds et al., 1999) proposes that directing attention to one stimulus attenuates the processing of the other unattended stimulus within the same neuron's RF and this notion appears to be similar to location-based surround suppression in ST. These two models, however, have fundamentally different mechanisms. Attentional enhancement of the target stimulus is an explicit element of the biased competition model, whereas ST only includes a top-down suppressive mechanism without implementing attentional facilitation. In ST, attentional facilitation is a “side-effect” of suppression (Boehler et al., 2009; Tsotsos, 1990). Another difference is that the two models imply the opposite direction of attentional modulation along the visual processing hierarchy. Attentional modulation in the biased competition model selectively impacts *inputs* of the target neurons, suggesting that attention first targets early cortical areas and then its effect is passed on successively to later areas (Reynolds et al., 1999; Reynolds & Heeger, 2009). In contrast, attentional effects in ST progress backward from higher-order to lower-order areas as supported by several studies (Boehler et al., 2009; Buffalo et al., 2010; Lauritzen, D’Esposito, Heeger, & Silver, 2009; Maunsell & Cook, 2002; Mehta et al., 2000). Furthermore, the attentional surround suppression happens even without distractors around the target stimulus (Boehler et al., 2011), whereas suppression is not required in the absence of competing stimuli in the biased competition model. Besides the biased competition model, the other attention models that explicitly implement top-down attentional modulation are not likely to predict location-based surround suppression (Beuth & Hamker, 2015; Hamker, 2005; Miconi & VanRullen, 2016; Spratling & Johnson, 2004). These models do not incorporate critical components pertaining to the attentional surround suppression, such as connection pruning via top-down winner-take all processes.

1.5 Feature-based surround suppression

Attention can be allocated to certain visual features and facilitates their processing throughout the visual field while inhibiting the processing of other unattended features, which is called feature-based attention (Anllo-Vento & Hillyard, 1996; Egeth, 1966; Hopf, Boelmans, Schoenfeld, Luck, & Heinze, 2004; Maunsell & Treue, 2006; Sàenz, Buracas, & Boynton, 2002;

Sàenz, Buraĉas, & Boynton, 2003). A potential mechanism of feature-based attentional modulation would be to enhance the responses of neurons that preferentially respond to the attended feature so that the representation of stimuli that contain the attended feature can be strengthened globally. Single cell recording studies in the motion-selective area (middle temporal, MT) demonstrated that feature-based attention changes the gain of a neuron depending on the similarity between the attended feature and the feature that the neuron selectively responds (Martinez-Trujillo & Treue, 2004; Treue & Martinez-Trujillo, 1999). For example, if a MT neuron prefers upward motion, directing attention to upward motion facilitates its response, whereas the neuronal response is suppressed when downward motion (the neuron's anti-preferred direction) is attended. Hence, this finding suggests that feature-based attentional modulation is a monotonic function of the similarity between the attended feature and the neuron's preferred one, proposing the feature-similarity gain hypothesis (Boynton, 2005; Martinez-Trujillo & Treue, 2004; Treue & Martinez-Trujillo, 1999).

ST predicts a non-monotonic, center-surround profile of feature-based attention (Tsotsos, 1990; Tsotsos et al., 1995), similar to location-based surround suppression. This idea of the "feature-based" surround suppression is that attention to a particular visual feature causes greater suppression of other features that are represented closely to the attended one in the same feature map, while features that are represented further away from the attended one are less affected. For instance, orientation-selective neurons in area V1 are organized in a columnar structure (Hubel & Wiesel, 1968, 1974), and these neurons respond best to their preferred orientations but the responses gradually decrease as orientation becomes dissimilar from the preferred one, showing Gaussian-like tuning curves (Figure 1.3A). Once attention is directed to a given orientation (e.g., vertical), feature-based surround suppression inhibits the processing of neighboring orientations without affecting the processing of orientations that are represented far from the attended orientation (Figure 1.3B). This attentional modulation changes the overall population tuning curve in the orientation map into a center-surround profile and sharpens the tuning to the attended orientation.

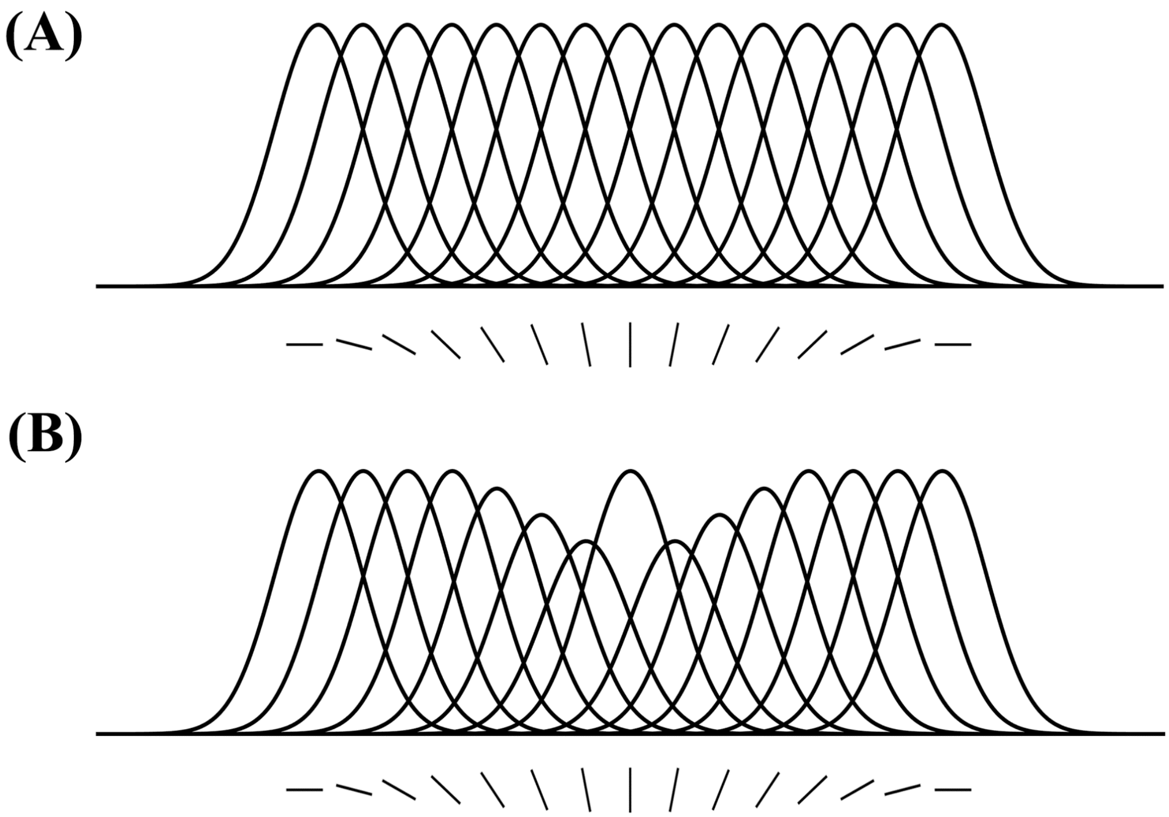


Figure 1.3. Feature-based attentional surround suppression in the orientation map.

(A) Hypothetical feature tuning curves for different orientations. (B) Attention to the vertical orientation inhibits neighboring orientations and changes the overall tuning profile in the orientation map (adapted from Tsotsos (2011) with permission).

Evidence for feature-based surround suppression has been found in several feature domains. The first empirical study on feature-based surround suppression targeted the orientation domain (Tombu & Tsotsos, 2008). This study demonstrated that judgments about the oriented stripes (i.e., whether it is straight or jagged) suffered when the orientation of the stripes was similar to the initially attended orientation. The same attentional modulation was observed in the color (Bartsch et al., 2017; Störmer & Alvarez, 2014) and motion domains (Ho, Brown, Abuyo, Ku, & Serences, 2012; Wang, Miller, & Liu, 2015) by both psychophysical and neuroimaging methods. One of these studies also demonstrated that feature-based surround suppression is due to backward-propagating attentional modulation in the visual processing hierarchy (Bartsch et al., 2017). This is consistent with the reverse hierarchical direction of spatial attentional modulation and the underlying mechanism of location-based surround suppression (Boehler et al., 2009; Buffalo et al., 2010; Lauritzen et al., 2009; Maunsell & Cook, 2002; Mehta et al., 2000). Interestingly, features stored in visual working memory influences perceptual judgments through the center-surround mechanism (Kiyonaga & Egner, 2016) and feature-based surround suppression in the sensory domain may be transferred to the motor system (Loach, Frischen, Bruce, & Tsotsos, 2008). These findings suggest that feature-based surround suppression is not confined to the perceptual domain, but it could be extended to other cognitive domains.

1.6 Purpose of the current work

The current dissertation tests the role of top-down feedback processing and the properties of attentional surround suppression in ST's framework that have not been clarified to promote understanding of the visual system and selective attention. The first study (Chapter 2) investigated whether top-down feedback processing to earlier cortical areas is necessary for fine-grained localization (i.e., Full Recurrence Binding) as ST predicted (Rothenstein et al., 2008; Tsotsos et al., 2008). Performance of the object localization task and even that of the object categorization task in cluttered, attention-demanding scenes gradually improved when additional time was given to allow top-down feedback processing to be involved. The second study (Chapter 3) examined the factors that influence the profile of the location-based suppressive surround using the experimental paradigm introduced by Cutzu and Tsotsos (2003). Two experiments suggested that the profile of the surround suppression is the result of the interaction between the location- and feature-based surround suppression and that the size of the location-based suppressive surround is modulated by stimulus eccentricity but not stimulus size. The third study (Chapter 4) looked for feature-based surround suppression in the motion processing using both psychophysical and neurophysiological methods. Feature-based surround suppression reduced motion repulsion (Marshak & Sekuler, 1979), an illusory motion perception, when the directions of the two superimposed motion surfaces were close to each other. This attentional modulation was also affected by bottom-up factors such as motion speed and the color of the motion surfaces. Furthermore, when MT/MST neurons' preferred directions were attended, their tuning curve exhibited a center-surround profile, providing a neuronal mechanism of feature-based surround suppression.

CHAPTER 2. Feed-forward visual processing suffices for coarse localization but fine-grained localization in an attention-demanding context needs feedback processing

(Adapted from Yoo, Tsotsos, & Fallah (2019))

2.1 Abstract

It is well known that simple visual tasks, such as object detection or categorization can be performed within a short period of time, suggesting the sufficiency of feed-forward visual processing. However, more complex visual tasks, such as fine-grained localization may require high-resolution information available at the early processing levels in the visual hierarchy. To access this information using a top-down approach, feedback processing would need to traverse several stages in the visual hierarchy and each step in this traversal takes processing time. In the present study, we compared the processing time required to complete object categorization and localization by varying presentation duration and complexity of natural scene stimuli. We hypothesized that performance would be asymptotic at shorter presentation durations when feed-forward processing suffices for visual tasks, whereas performance would gradually improve as images are presented longer if the tasks rely on feedback processing. In Experiment 1, where simple images were presented, both object categorization and localization performance sharply improved until 100 ms of presentation then it leveled off. These results are a replication of previously reported rapid categorization effects, but they do not support the role of feedback processing in localization tasks, indicating that feed-forward processing enables coarse localization in relatively simple visual scenes. In Experiment 2, the same tasks were performed

but more attention-demanding and ecologically valid images were used as stimuli. Unlike in Experiment 1, both object categorization performance and localization precision gradually improved as stimulus presentation duration became longer. This finding suggests that complex visual tasks that require visual scrutiny call for top-down feedback processing.

2.2 Introduction

The human visual system is known to be very rapid and efficient at analyzing some types of visual information. People can determine whether a briefly flashed image contains a depiction of a certain object category and categorization performance still holds even if another visual pattern immediately follows the target image by backward-masking or rapid serial visual presentation (RSVP) (Delorme, Richard, & Fabre-Thorpe, 2000; Fabre-Thorpe et al., 2001; Grill-Spector & Kanwisher, 2005; Keysers et al., 2001; Potter, 1976; Potter et al., 2014; Rousselet et al., 2003; Thorpe et al., 1996; VanRullen & Koch, 2003; VanRullen & Thorpe, 2001). Since object categorization is performed within a very short period of time, this process is thought to rely on feed-forward visual processing (Riesenhuber & Poggio, 1999; Serre et al., 2007). Besides behavioral evidence, electroencephalographic activity demonstrates the same point. ERP analysis revealed a divergence in voltage between category-present and category-absent trials that developed after 150 ms of stimulus onset (Fabre-Thorpe et al., 2001; Rousselet, Fabre-Thorpe, & Thorpe, 2002; Thorpe et al., 1996). Studies using classifier-based readout techniques also demonstrated that information about object category and identity can be decoded from human temporal cortex and macaque inferior temporal area (IT) as early as 100 ms after stimulus onset, suggesting that hierarchical feed-forward processing is sufficient for rapid object categorization (Hung et al., 2005; Isik et al., 2014; Liu et al., 2009).

The architecture of the visual system also impacts how object location information is represented. It is well-established that the ventral pathway of the visual system is structured as a layered pipeline where each area, from the retina to the temporal cortex, features increasing receptive field (RF) sizes (nicely documented in Kay et al. (2013)) and fewer representational columns, as a visual signal traverses the pipeline in its feed-forward journey. All neurons within each area receive converging input from the previous area organized in a spatially-limited RF and provide diverging input to many neurons in a reciprocal spatially-limited manner in the next area (see Salin & Bullier (1995) for neuroanatomical and Tsotsos (1988) for computational discussions). This structure blurs location precision, e.g. the Blurring Problem (Tsotsos et al., 1995). Thus, the precision of location representation is necessarily different for each area of this pipeline, with areas least affected by blurring (earliest) having the most precise location representation and those most affected by blurring (highest) the worst. Recent neuroimaging research supports that high-order object areas represent coarse object locations (Carlson et al., 2011; Cichy et al., 2011; DiCarlo & Maunsell, 2003; Grill-Spector et al., 1999; Hemond et al., 2007; MacEvoy & Epstein, 2007; Niemeier et al., 2005; Sayres & Grill-Spector, 2008; Schwarzlose et al., 2008). Coarse location information in higher visual areas would be sufficient if a visual task asks for approximate object location, and it could be accessed rapidly. For instance, saccades toward the visual hemifield where the animal target was presented could be initiated in as little as 120 ms (Kirchner & Thorpe, 2006) and saccadic latency was even shorter when human faces were the target (Crouzet et al., 2010), meaning that the coarse target location was necessarily processed before the minimum saccadic latency.

On the other hand, more precise object localization would require access to early areas of the visual processing hierarchy since that is where this information is best represented. These early areas cannot represent a complex object as a whole, while higher visual areas see an entire object, but they encode only coarse location information. Thus, higher visual areas need to access fine-grained location information in early visual areas to determine an object's precise location (e.g., Selective Tuning model (Tsotsos, 1990, 2011; Tsotsos et al., 2008)). The Selective Tuning model suggested that different types of visual recognition tasks would go through different processing steps across the visual hierarchy, and that the amount of time it takes to achieve the task indicates the stage of recognition. For example, a single feed-forward pass would suffice for simple discrimination or categorization tasks (in ST, Convergence Binding (Tsotsos et al., 2008)) and its time course would be consistent with the time courses that the previous studies on ultra-rapid object categorization have demonstrated (Fabre-Thorpe et al., 2001; Rousselet et al., 2002; Thorpe et al., 1996). However, if a task requires detailed feature binding or localization, subsequent top-down feedback signals should reach earlier areas to recover ambiguous location information thereby increasing processing time (in ST, Full Recurrence Binding). Even though the visual hierarchy pipeline does indeed have both short-range and long-range feedback connections between several areas, it necessarily takes additional time to use those connections for access to early areas. This is so because the nature of connectivity just described imposes crosstalk (the Crosstalk Problem (Tsotsos et al., 1995)); multiple converging signals interfere and corrupt each other. If there is no mechanism to select one signal over another, the blurring of location and ambiguity of category information would persist. The choice is whether the visual system simply decodes a corrupted signal or actively attempts to clean the signal before its interpretation. The Selective Tuning model takes the latter position and provides such a mechanism that combines competitive attentional selection with suppression of interference that progresses in a top-down manner from higher to lower cortical areas (Boehler et al., 2009; Hopf et al., 2010; Tsotsos, 1990; Tsotsos et al., 1995). This progression is responsible for the additional time observed when more precise location information is required by a given visual task. Figure 2.1 illustrates the hierarchical structure of neuronal RFs and a schematic feedback processing suggested by ST. Here, feedback processing means the top-down, attention-mediated signal that is temporally and functionally dissociated from rapid local recurrent processing within the ventral visual stream (Boehler et al., 2008, 2009; Hopf et al., 2006, 2010; Mohsenzadeh et al., 2018; Wyatte et al., 2014). Evans and Treisman (Evans & Treisman, 2005) asked participants to detect a target object in a rapid serial visual presentation (RSVP, each stimulus was presented for 75 ms) sequence and then report its identity and spatial location (left, center or right side of the image). Similar to ST's claim, their participants could not report even the rough position of a given object in the image, although its detection was successful. It implies that localization requires top-down feedback processing, thus, additional processing time.

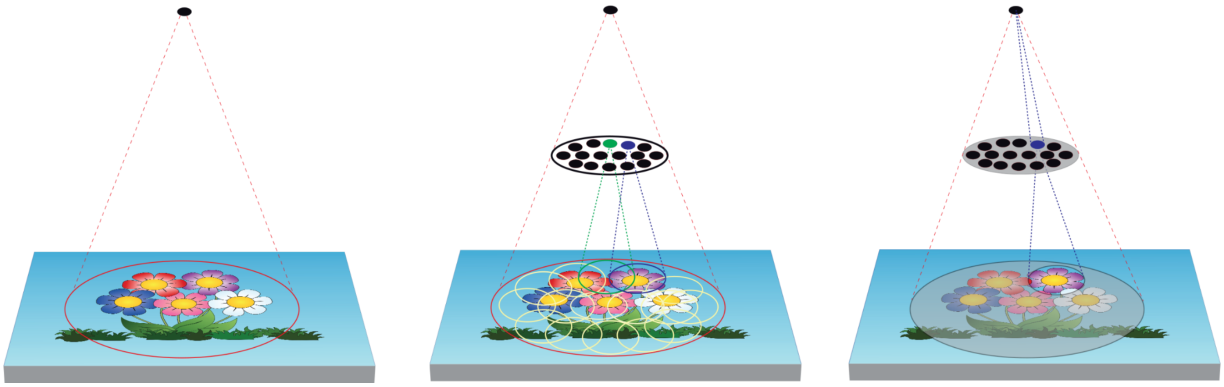


Figure 2.1. A schematic illustration of top-down feedback processing in visual hierarchy. (Leftmost) A hypothetical neuron whose RF sees the whole bunch of flowers. This neuron does not selectively respond to individual flowers. For example, if that neuron were trained to prefer a single flower with yellow center and purple petals, it would respond only partially because of all

the other flowers in the same RF (Moran & Desimone, 1985). (Middle) Each neuron in an intermediate layer has a smaller RF. If the neuron in green were tuned to the flower with yellow center and purple petals, it would respond only partially. On the other hand, the neuron in blue with the same tuning property would respond maximally. All the other neurons regardless of what kind of flower they might be tuned to, would also respond partially or fully, each to their own tuning. (Rightmost) The neuron in the top layer receives all inputs from the intermediate neurons. If precise localization of the flower with yellow center and purple petals is needed, the neuron in the top layer could selectively receive input from the intermediate neuron which maximally respond to the target (neuron in blue). In ST, it is achieved by suppressing the inputs coming from all the other neurons in a top-down manner (i.e., attentional surround suppression). This top-down inhibition process operates layer by layer until it reaches the earliest one where the finest location information required for the task is represented.

In the present study, we examined whether feed-forward processing is sufficient for object localization or whether the subsequent top-down, attention-mediated feedback processing should be involved in localization as ST predicts. We conducted an animal detection task as a replication of the previously reported ultra-rapid visual categorization task (Fabre-Thorpe et al., 2001; Rousselet et al., 2002; Thorpe et al., 1996; VanRullen & Thorpe, 2001) and a novel localization task that asks participants to report the spatial location of a given animal feature (e.g., "Where is the tail?). Critically, we constrained the presentation durations of images so that we can determine how much processing time is needed to perform each type of task. We hypothesized that if feed-forward signals enable both visual tasks, performance would level off from relatively shorter presentation durations, whereas performance would continuously improve as presentation duration becomes longer if feedback signals enhance recognition processes. Stimulus complexity varied between Experiments 1 and 2.

2.3 Experiment 1

In Experiment 1, each participant performed animal detection tasks to replicate the previous findings and animal feature localization tasks to examine the role of feedback processing in object localization. In the feature localization tasks, participants reported the location of a certain feature by clicking on a screen location using a mouse. Presentation duration of the stimuli varied to manipulate the visual processing time.

2.3.1 Materials and Methods

2.3.1.1 Participants

Forty-eight naïve participants (12 men, 36 women), between the ages of 17 and 39 years completed the experiment. 24 participants performed the *before* condition and the other 24 participants performed the *after* condition where a task question was shown before or after stimulus presentation, respectively. They had normal or corrected-to-normal vision and their color vision was also intact. Informed written consent was obtained from all participants. Participants, who were recruited from the Undergraduate Research Participants Pool of York University, received course credit for their study participation and the other participants were paid $15 CAD. The research was approved by York University's Human Participants Review Committee. According to the human research guidelines of York University, older minor age participants (e.g. 17 years old) do not require parental consent to participate in this research, as it fell under the category of minimal risk research.

2.3.1.2 Apparatus and stimuli

Experiments were conducted in a dark room. Participants sat 57 cm from a CRT monitor (21” View Sonic G225f, 1280 x 1024, 85 Hz) and their heads were stabilized on a head and chin rest (Headspot, UHCOtech, Houston, TX). Participants wore an infrared eye tracker (Eyelink II, SR Research, 500 Hz, Mississauga, ON, Canada) monitoring the left eye position. Experimental control was maintained by Presentation (Neurobehavioral Systems, Berkeley, CA).

We used 400 images (200 animal-present, 200 animal-absent) that are the same images used in Thorpe et al. (1996) or similar images collected from the Internet. The images subtended roughly 16° visual angle in width and 22° visual angle in height. Animal targets included mammals, birds, insects, fish, amphibians, and reptiles. Only real animals were counted as targets and humans were not categorized as animals in this experiment. Each animal-present image contained one animal (humans were not presented together), located in the central area of the image. The distractor images included natural landscapes (e.g., mountains, forests, lakes, and oceans), cityscapes, plants, buildings, and other man-made objects. Each image had a corresponding masking stimulus created by randomly scrambling pixels from the original image.

2.3.1.3 Procedure

Figure 2.2 depicts the procedure of Experiment 1. In the *before* condition, participants viewed a task question first. If the task of a given trial was animal detection, the question was “is there an animal?” or if the task was feature localization, participants were asked to localize a certain feature of an animal (e.g., “where is the beak?”). Animal detection and feature localization trials were randomly interleaved. Participants had to look at the fixation cross appearing in the center of the screen before stimulus presentation. When their eyes were fixated, an image could be presented for 20, 100, 170, or 250 ms and then immediately masked for 500 ms. Each image was presented only once throughout the experiment. Participants who were assigned to the *after* condition saw the task question after stimulus presentation. If the task was animal detection, participants clicked the left or right mouse button to report whether the animal target was present or absent, respectively. If the task was feature localization, a white reference frame that was equally-sized with the images was presented in the center of the screen. Participants were instructed to click on the location within that frame that corresponded to the location of the target feature in the test image. For both tasks, there was no time constraint for responses. Each participant completed 80 animal detection trials (40 animal-present and 40 animal-absent trials) and 80 feature localization trials (all animal-present) in a single session.

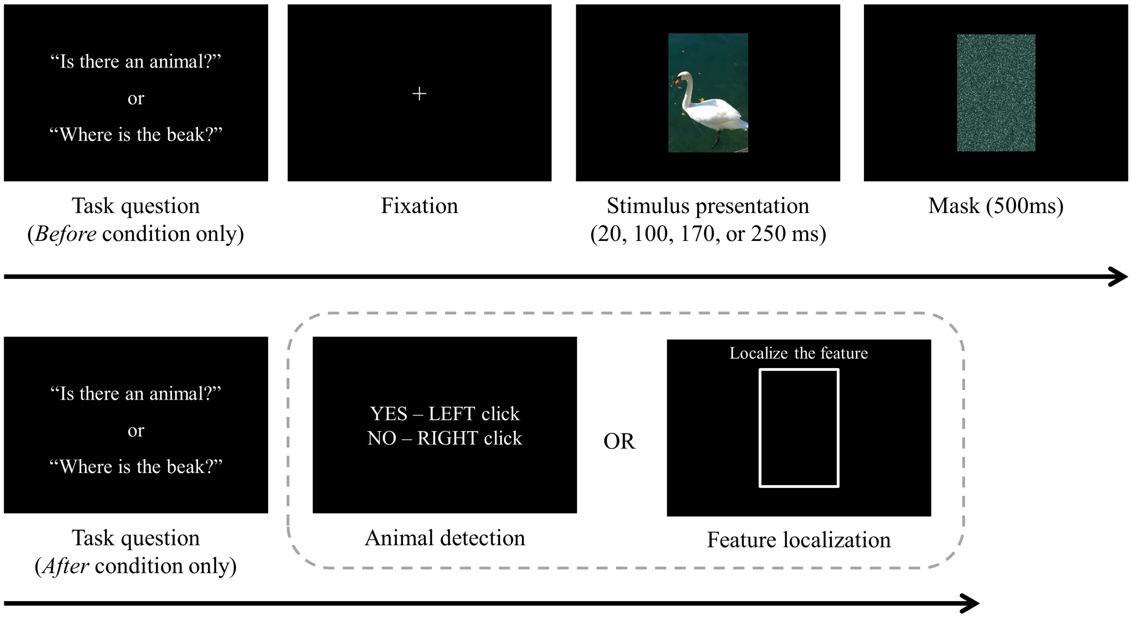


Figure 2.2. Procedure of Experiment 1.

Participants performed randomly interleaved animal detection and feature localization trials. The stimulus was presented for 20, 100, 170, or 250 ms and it was immediately masked. A task question was shown either before or after stimulus presentation depending on which condition a participant was assigned to. Participants reported the presence of the animal target or localized a given animal feature using a mouse.

2.3.2 Results

2.3.2.1 Animal detection

Mean animal detection accuracy was significantly affected by stimulus presentation duration (Figure 2.3A; Greenhouse-Geisser correction ($\varepsilon = .668$); $F(2, 92.18) = 56.372$, $p < .001$), but the main effect of the *before-after* conditions ($F(1, 46) = .274$, $p = .603$) and the interaction between presentation duration and the *before-after* conditions were not significant ($F(2, 92.18) = .376$, $p = .688$). Viewing the task question prior to stimulus presentation is not necessary for accurate animal detection because even participants assigned to the *after* condition could judge the presence/absence of an animal easily once they saw an image. In both *before* and *after* conditions, animal detection accuracy sharply improved between 20 ms to 100 ms (all ps $< .001$) and then it leveled off (all ps $> .05$). We do not provide the reaction time (RT) data here because there is no comparable starting point for RTs between the *before* and *after* conditions. While RTs in the *before* condition can be measured after mask offset, RTs in the *after* condition unavoidably vary depending on how quickly participants read and understand task questions after viewing the stimulus. In addition, there was no time constraint for responses and the participants were not required to make speeded responses. The results suggest that detecting the presence of an animal can be done within a very short period of time in this image set, replicating ultra-rapid visual categorization (Fabre-Thorpe et al., 2001; Potter, 1976; Potter et al., 2014; Thorpe et al., 1996).

2.3.2.2 Feature localization

To quantify the localization precision, we measured the area of the 95% error ellipse of participants' localization responses relative to the normalized target centroids (Chen, Byrne, & Crawford, 2011). Figure 2.3B shows one participant's localization performance for different presentation durations and the corresponding error ellipses. The ellipse size indicates localization error; thus, a smaller ellipse means higher localization precision. Mean localization error was significantly modulated by presentation duration (Figure 2.3C; Greenhouse-Geisser correction ($\varepsilon = .755$); $F(2{,}27, 104.21) = 38.82$, $p < .001$) and by the *before-after* conditions ($F(1, 46) = 14.39$, $p < .001$). Post-hoc multiple comparisons with Bonferroni correction showed that the localization error was lower in the *before* condition than in the *after* condition ($M_{\text{diff}} = -47.01$, SE = 12.39, $p < .001$). The interaction between presentation duration and the *before-after* condition was not significant ($F(2.27, 104.21) = .27$, $p = .791$). As this null interaction indicated, localization performance patterns in both *before* and *after* conditions were almost equal across different presentation durations. In both conditions, localization was more precise when the presentation duration was extended from 20 ms to 100 ms (all ps < .001) but performance did not improve after 100 ms (all ps > .05). Therefore, similar to the animal detection results, feature localization seemed to be completed very rapidly and the current results do not support the idea that feedback processing improves localization precision.

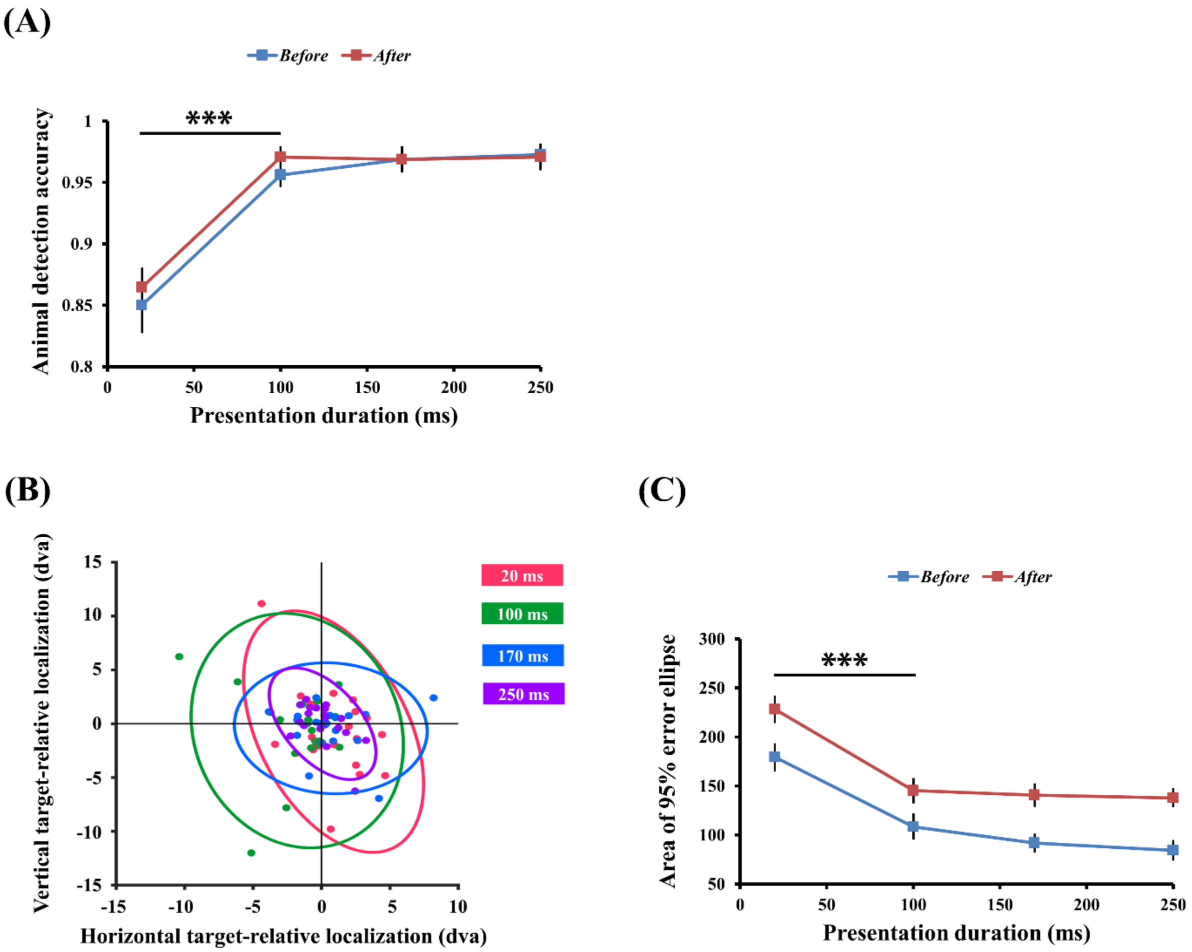


Figure 2.3. Experiment 1 results.

(A) Animal detection accuracy improved between 20 ms and 100 ms but it plateaued after 100 ms. This result did not vary by the timing of the task questions (before or after stimulus presentation). (B) An example of 95% localization error ellipses for each presentation duration. The centroids of the target features were normalized to (0, 0) and each colored dot indicates the participant's localization response relative to the target centroid. (C) Feature localization error was reduced between 20 ms and 100 ms but localization performance did not significantly change at longer presentation durations. Mean localization performance was better in the *before* condition but the performance pattern across different presentation durations was the same in both the *before* and *after* conditions. Error bars indicate SEM. *** $p < .001$

2.4 Experiment 2

Experiment 1 replicated prior rapid animal detection results and also suggested that feature localization may be accomplished within the feed-forward sweep. This seems to place our central hypothesis in doubt, namely, that feedback is required for localization. One possibility for this is that the animal-present images in Experiment 1 might have been too simple so that the result of this experiment did not represent detailed localization which demands full top-down processing (Rothenstein et al., 2008; Tsotsos et al., 2008). Each image contained only one animal that occupied most of the central area of the image, and its background was also very simple so that the target animal was well segregated from the background. Hence, participants might be able to extract the target animal's layout easily and guess the location of a target feature based on this gist rather than carefully localizing the feature by attention-demanding feature binding and distractor suppression. For instance, if one sees the contour of a bird's head then s/he could guess the relative location of a certain feature (e.g., beak) on its head. In other words, the cross-talk or interference within the visual hierarchy that would necessitate top-down feedback for correction is not so a great problem for images where there is no real visual conflict. A similar situation occurs in visual search where targets are well-differentiated from distractors (i.e., pop-out) and no focused attention is required. To control for this potential confounding factor, we conducted the same experiments with different stimuli. We used more complex animal-present images that contained more than one animal or (an) animal(s) with human(s), embedded in a complex background. In these images, the segregation between the target and background would not immediately occur, hence, more precise localization processes would be required. These new images should be more similar to the visual scenes we confront in our daily life. The masking stimuli in Experiment 2 were random polygons in random colors so that they were unrelated to the original images.

2.4.1 Materials and Methods

2.4.1.1 Participants

Twenty-four naïve participants (9 men, 15 women), between the ages of 17 and 34 years completed the experiment. They had normal or corrected-to-normal vision and their color vision was also intact. Informed written consent was obtained from all participants. Participants who were recruited from the Undergraduate Research Participants Pool of York University received course credit for their study participation while other participants were paid $15 CAD. The research was approved by York University's Human Participants Review Committee. According to the human research guidelines of York University, older minor age participants (e.g. 17 years old) do not require parental consent to participate in this research, as it fell under the category of minimal risk research.

2.4.1.2 Apparatus and stimuli

The apparatus was the same as in Experiment 1. We used 420 new images (210 animal-present, 210 animal-absent) collected from the Internet and selected from the MS COCO dataset (http://mscoco.org, Lin et al., 2014). The images subtended roughly 20° visual angle in width and 13° visual angle in height. In animal-present images, animal targets were mostly mammals and birds that live in groups or live with humans (Figure 2.4). Each image contained more than one animal or (an) animal(s) with human(s), and humans were again not categorized as animals. The size and location of the target animal randomly varied across different images, but it was not too small or too peripheral. The backgrounds of the images included natural landscapes, cityscapes, outdoor and indoor scenes. The animal-absent distractor images were also drawn from similar scene categories. Masking stimuli were created in MATLAB (The MathWorks, Inc.) with the Psychophysics Toolbox (Brainard, 1997; Pelli, 1997). They consisted of multiple random polygons that had different sizes, shapes, and colors.

Figure 2.4. Examples of animal-present images used in Experiment 2

2.4.1.3 Procedure

The experimental procedure was the same as in Experiment 1, except that the task question was always shown before stimulus presentation and we changed presentation durations. We did not test both the *before* and *after* conditions because in Experiment 1, these conditions did not affect the performance patterns across different presentation durations in the animal detection and feature localization tasks. When an image contained multiple animals, the task question was specific to one of these animals so that participants would not be confused (e.g., "Where is the muzzle of the black-and-white cow?"). An image could be presented for 25, 50, 75, 100, 170, 250, or 300 ms. Each participant completed 140 animal detection trials (70 animal-present and 70 animal-absent trials) and 140 feature localization trials (all animal-present), and these trials were randomly interleaved.

2.4.2 Results

2.4.2.1 Animal detection

Mean animal detection accuracy significantly varied depending on stimulus presentation duration (Figure 2.5A; Greenhouse-Geisser correction ($\varepsilon = .589$); $F(3.54, 81.32) = 65.963$, $p < .001$). Accuracy sharply improved between 25 ms and 100 ms ($F(3, 69) = 51.013$, $p < .001$), once again supporting the important role of feed-forward processing in object categorization. Nevertheless, performance also improved gradually between 100 ms and 300 ms (Greenhouse-Geisser correction ($\varepsilon = .69$); $F(2.07, 47.59) = 4.532$, $p = .015$). Across the two experiments, mean animal detection accuracy at the shortest presentation duration decreased from 84.33% (SD 11.32%, at 20 ms in the *before* condition) in Exp 1 to 68.33% (SD 10.98%, at 25 ms) in Exp 2. This indicated that rapid object categorization is slowed by increasing image complexity (Davenport & Potter, 2004; Joubert, Fize, Rousselet, & Fabre-Thorpe, 2008; Praß, Grimsen, König, & Fahle, 2013; Walker, Stafford, & Davis, 2008) and that top-down feedback processing could help segmenting the target object category in complex scenes.

2.4.2.2 Feature localization

The presentation duration significantly modulated localization error ($F(6, 138) = 36.148$, $p < .001$). Mean localization error was significantly reduced from 25 ms to 100 ms ($F(3, 69) = 16.772$, $p < .001$) and was continuously reduced afterwards (100 ms to 300 ms: $F(3, 69) = 11.393$, $p < .001$). The black line in Figure 2.5B shows these results. This decrement in localization error across longer presentation durations was not observed in Experiment 1 which suggests that precise feature localization among several distractors requires additional processing time, and thus, involves top-down feedback processing. Nevertheless, the present results might be driven by the addition of a longer presentation duration (300 ms). Therefore, we analyzed mean localization errors from 100 ms to 250 ms to match the results from Experiment 1. Even after excluding the localization error at 300 ms, localization errors significantly decreased as presentation duration increased above 100 ms ($F(2, 46) = 8.647$, $p = .001$). Also, the post-hoc comparisons showed that the decrement was significant when presentation duration was extended from 100 ms to 170 ms ($M_{diff} = 48.33$, SE = 13.55, $t(23) = 3.567$, $p = .002$) and from 100 ms to 250 ms ($M_{diff} = 53.11$, SE = 15.35, $t(23) = 3.459$, $p = .002$).

Our results suggest that recurrent processing contributes to improved localization accuracy that occurs beyond, and is complementary to, the localization afforded by a feed-forward process. This is a necessity, since a single neuron in human visual area TO, for example, can only localize to the size of its RF (about 20° of visual angle at 10° eccentricity (Kay et al., 2013)). The question is how a finer spatial location is determined at the level of description afforded by that layer of neurons. Indeed, the decrement in mean localization error was much greater within the 25-100 ms range ($M_{diff} = 105.72$) than within the 100-300 ms range ($M_{diff} = 72.82$). It is likely that initial feed-forward processing enables a fast but coarse location estimation, whereas subsequent feedback processing helps a slow but more precise localization process. If the visual input is relatively simple like the stimuli used in Experiment 1, which produces little conflict within feed-forward convergence neural communication, feature localization could be completed within the feed-forward sweep. In more complex visual scenes, feedback processes would be needed to provide fine localization. As feed-forward and feedback processing have different time courses, we might be able to reveal them by fitting our data in a piecewise linear manner, having 100 ms as the breakpoint (based on the results in Exp 1) and comparing it to the linear polynomial model.

The equation of the linear polynomial fit was

$$y = -0.5767x + 234.9$$

and the equations of the piecewise linear fit was

when $x < 100$ (ms)

$$y = -1.544x + 295.9$$

when $x \geq 100$ (ms)

$$y = -0.1764x + 138$$

To compare the goodness-of-fit of the piecewise linear model with that of the linear polynomial model (which assumes single process in localization), we computed the Akaike information criterion (AIC, Akaike, 1974) which penalizes model complexity. Smaller AIC values indicate better fit. The piecewise linear model (AIC = 43.02) explained the data better than the linear polynomial model (AIC = 50.12), and the red line in Figure 2.5B shows the result of the piecewise linear fitting. This supports the idea that both feed-forward and feedback processing contribute to feature localization, but each operates at different time points and with different effectiveness (e.g., slopes). It also suggests that feedback processing might be necessary specifically when a localization task requires attentional engagement (e.g., filtering out irrelevant distractors).

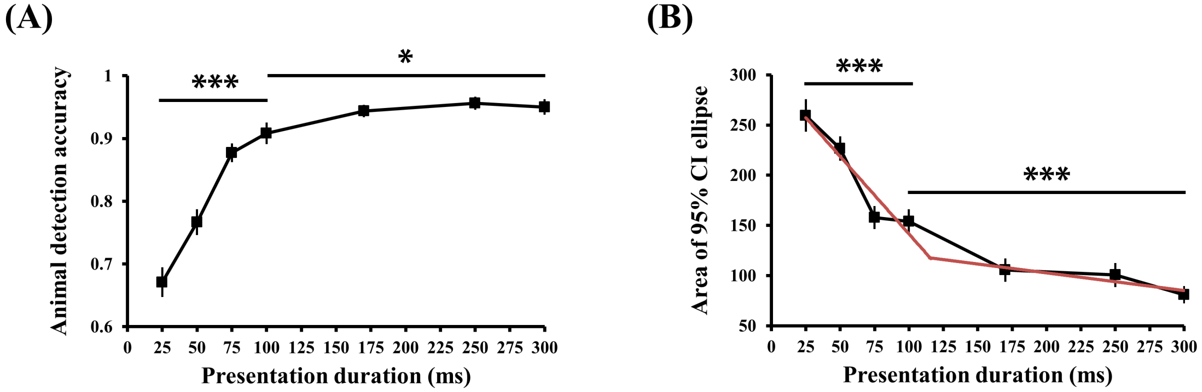


Figure 2.5. Experiment 2 results.

(A) Animal detection accuracy sharply improved from 25 ms to 100 ms and then it gradually improved as the stimuli were presented longer. (B) Black line: Feature localization error sharply decreased from 25 ms to 100 ms and it slowly but continuously decreased afterwards, suggesting that feedback processing enhanced localization performance. Red line: Piecewise linear fit to the average data. Error bars indicate SEM. * $p < .05$, *** $p < .001$

2.5 General discussion

The present study examined the roles of feed-forward and feedback visual processing in object categorization and localization. As many previous studies have demonstrated, the present study suggests that human vision can very rapidly determine the category of a certain object (i.e., animal) embedded in a visual scene, demonstrating that animal detection accuracy dramatically improved as stimulus presentation duration increased ~100 ms. However, animal detection accuracy at longer presentation durations was dependent on scene complexity. Performance did not significantly improve after 100 ms stimulus exposure for simple stimuli (Experiment 1), whereas it did improve at longer presentation durations when stimuli were more complex (Experiment 2). Therefore, rapid object categorization may not solely rely on feed-forward processing, but top-down feedback processing may be also involved when visual scenes need to be analyzed in more detail. Similarly, processing time for animal feature localization was dependent upon the complexity of visual scenes. Feature localization was as fast as animal detection when scenes were simple with a single animal. When scenes were cluttered, localization error was gradually reduced with increasing presentation times, indicating that feedback processing may be necessary for fine-grained localization.

The Selective Tuning model has claimed that different visual tasks require different processing strategies, such as feed-forward or feedback (Rothenstein et al., 2008; Tsotsos et al., 2008). While simple tasks, including object discrimination or categorization could be achieved within a single feed-forward pass, more complex tasks that require high-resolution information (e.g., precise localization) should involve top-down feedback processing, taking additional processing time. As previously mentioned, location information is blurred in high-order areas due to the integration of earlier neurons' RFs and because diverging feed-forward connections along the visual processing hierarchy cause cross-talk among visual signals. Hence, a top-down traversal that ameliorates this cross-talk must reach early visual areas where the finest location information is available (Buffalo et al., 2010). An important consequence of top-down localization is that when it selects target input connections, the remainder of the input (i.e., noise) within the same RF is suppressed, forming a suppressive surround around the target which eventually enhances the overall signal-to-noise ratio of the neuron (Boehler et al., 2009; Cutzu & Tsotsos, 2003; Hopf et al., 2006; Müller & Kleinschmidt, 2004; Müller et al., 2005; Tsotsos et al., 1995). Thus, the manifestation of the suppressive surround indirectly supports a top-down localization mechanism. Experimentally, it has been reported that a visual task that requires precise feature-location binding produces a suppressive surround but not a simple discrimination task. Furthermore, the effect of the surround suppression becomes evident around at 250 ms after stimulus onset (Boehler et al., 2009). This delay relative to the time course of the initial feed-forward sweep is consistent with the temporal range of top-down attentional modulation in early visual cortex (Boehler et al., 2008; Di Russo et al., 2003; Martínez et al., 1999, 2001; Mehta et al., 2000; Noesselt et al., 2002), implying that fine-grained localization completes within this feedback pass.

We used different scene complexities across the experiments and each led to different results in the categorization and localization tasks. In Experiment 1, target animals were centered and occupied large portions of images. Their backgrounds were also simple so that targets could be segregated from them easily. Due to these factors, competitive top-down selection of object information might not have been necessary for categorization and localization. Thus, both tasks were achieved within the feed-forward time range. In the localization task, it is possible that participants had localized the features based on the coarse layout of the animals and the relative spatial locations of the features without knowing their actual locations. ST suggested that this coarse level of location details is provided at intermediate layers of processing hierarchy, so the localization task in Experiment 1 is unlikely to necessitate a full top-down feedback traversal for more precise location information. It is also consistent with the findings that coarse location information is available in intermediate or higher-order object selective areas (Carlson et al., 2011; Cichy et al., 2011; DiCarlo & Maunsell, 2003; Grill-Spector et al., 1999; Hemond et al., 2007; MacEvoy & Epstein, 2007; Niemeier et al., 2005; Sayres & Grill-Spector, 2008; Schwarzlose et al., 2008; but see also Chakravarthi, Carlson, Chaffin, Turret, & VanRullen, 2014), allowing rapid access to this information after a single feed-forward sweep. On the other hand, more complex stimuli were used in Experiment 2 wherein animals were often presented with other distracting objects (e.g., humans), and the target animal was embedded in a complex and realistic background. These stimuli make target-background segregation much more difficult. As a result, task difficulty increased and both categorization and localization performance improved when additional processing time was provided, consistent with the contribution of feedback processing. Previous studies have reported that rapid object categorization is impaired when target objects are embedded in a complex background rather than in a simple background (Davenport & Potter, 2004; Joubert et al., 2008; Praß et al., 2013), and that categorization is attention-dependent when multiple foreground objects are presented together (Walker et al., 2008). The results of the present study are in line with these findings and they further suggest that object categorization in complex visual scenes requires time-consuming, top-down feedback processing. Similarly, feature localization in complex scenes requires a top-down, attention-mediated selective mechanism to overcome the crosstalk within the visual hierarchy and thus select the targets among various distractors, as ST claims. Experiment 2 would represent the nature of fine-grained localization with top-down feedback processing better than Experiment 1 because its stimuli are inherently attention-demanding, in the similar way that feature conjunction visual search tasks are more attention-demanding than feature pop-out tasks.

Moreover, they are much closer to the visual scenes we encounter in daily life, so they are more ecologically valid. Therefore, the results of Experiment 2 are likely to demonstrate the time course of precise localization more accurately.

Other theories prescribe computational decoding procedures that can take high level cortical representations as input and decode them to extract meaning, in particular, location information. For example, Hung et al. (2005) used a classifier-based readout technique to interpret the neural coding of selectivity and invariance at the IT population level. The activity of small neuronal populations over very short time intervals (as small as 12.5 ms) contained accurate and robust information about both object identity and category. Coarse information about position and scale could be read out over three positions. Isik and colleagues (2014) used neural decoding analysis to understand the timing of invariant object recognition in humans. They found that size and position-invariant visual information appears around 125 and 150 ms, respectively, and both develop in stages, with invariance to smaller transformations arising before invariance to larger transformations. They claimed that this supports a feed-forward hierarchical model of invariant object recognition where invariance increases at each successive visual area along the ventral stream. This is in contrast to the finding that showed how attention influences object position and identity information represented by the population of IT neurons when there is competition among objects (i.e., cluttered display) (Zhang et al., 2011). They found that before attention was employed, visual clutter significantly reduced the object information relative to when single object was presented. However, when attention was directed to a specific object, the amount of object information was restored to nearly the same level when the object was shown in isolation.

The difference between the results in these last two papers may be due to the different stimuli used, the latter requiring attention and the former not. We can conclude that although coarse location information is likely easily extracted after a single feed-forward pass for simple recognition tasks, more complex visual tasks that require image details (e.g., precise feature location) likely are not. Something more is needed for natural images and for tasks where more precision is required than simple coarse position (Buffalo et al., 2010). There are really two choices: 1) provide mechanisms that dynamically ameliorate the interference before interpretation; or, 2) provide mechanisms to correctly interpret corrupted representations. The methods just described are of the latter type. ST advocates for the former possibility and our experiments provide evidence to support this.

Another explanation could be that precise location simply emerges over time from the results of feed-forward processing, perhaps using some kind of evidence accumulation mechanism, and that no top-down process is at play at all. If this were true, there would be no recurrent suppressive surround or any kind of backwards activations throughout the cortex, as has been observed in other studies (Boehler et al., 2009; Buffalo et al., 2010; Mihalas, Dong, von der Heydt, & Niebur, 2011; Noudoost, Chang, Steinmetz, & Moore, 2010). In the emergence explanation, brain imaging methods would only observe increasing activation over time within one area and this is not the case. This would simply produce a single linear process for the increase in localization precision with increasing time, whereas Experiment 2 supported dual processes, a feed-forward process as well as a recurrent process. One might argue that fast local recurrent feedback within the ventral pathway of the visual system could achieve tasks that require visual scrutiny since mounting evidence supports that local feedback may compensate for disrupted initial feed-forward signals (Muckli et al., 2015; O'Reilly et al., 2013; Wyatte, Curran,

& O’Reilly, 2012a; Wyatte et al., 2014). For example, a recent MEG-fMRI study (Mohsenzadeh et al., 2018) demonstrated recurrent activity from IT to early visual cortex during object categorization that they did not attribute to attentional modulation due to the relatively early emergence of the feedback signal compared to the typically reported attention-related top-down signal (Boehler et al., 2008, 2009; Wyatte et al., 2014). However, the visual task they used is a simple face detection task that probably does not require precise feature binding nor localization, and the authors also acknowledged that local recurrent activity is involuntary and independent of attentional modulation (Roland, 2010; Roland et al., 2006). Therefore, that study might not test the role of top-down feedback processing in object recognition as local recurrent activity is unlikely to complete attention-demanding recognition tasks on its own.

To conclude, the current study demonstrates that precise localization information seems to require subsequent top-down feedback processing. Other natural tasks such as comparison (are two objects the same or different?) or measurement (which object is furthest away?) may also require more detailed localization processes than are available after a single feed-forward pass. In natural environments, precise localization is critical for goal-directed behavior, such as reaching and pointing, to correctly select a target object and avoid other distracting objects (Baldauf & Deubel, 2010), supporting the need for recurrent processing. It remains an open question as to how localization is involved in interaction with real-world, 3-dimensional objects. Future studies using real-world objects or virtual reality would allow greater insight into how top-down localization operates in a natural context and expand our understanding of the relationship between visual attention and action.

CHAPTER 3. The attentional suppressive surround: Eccentricity, location-based and feature-based effects and interactions

(Adapted from Yoo, Tsotsos, & Fallah (2018b))

3.1 Abstract

The Selective Tuning model of visual attention (Tsotsos, 1990) has proposed that the focus of attention is surrounded by an inhibitory zone, defining a center-surround attentional distribution. This attentional suppressive surround inhibits irrelevant information which is located close to attended information in physical space (Cutzu & Tsotsos, 2003; Hopf et al., 2010) or in feature space (Bartsch et al., 2017; Störmer & Alvarez, 2014; Tombu & Tsotsos, 2008). In Experiment 1, we investigate the interaction between location-based and feature-based surround suppression and hypothesize that the attentional surround suppression would be maximized when spatially adjacent stimuli are also represented closely within a feature map. Our results demonstrate that perceptual discrimination is worst when two similar orientations are presented in proximity to each other, suggesting the interplay of the two surround suppression mechanisms. The Selective Tuning model also predicts that the size of the attentional suppressive surround is determined by the receptive field size of the neuron which optimally processes the attended information. The receptive field size of the processing neurons is tightly associated with stimulus size and eccentricity. Therefore, Experiment 2 tests the hypothesis that the size of the attentional suppressive surround would become larger as stimulus size and eccentricity increase, corresponding to an increase in the neuron's receptive field size. We show that stimulus eccentricity but not stimulus size modulates the size of the attentional suppressive

surround. These results are consistent for both low- and high-level features (e.g., orientation and human faces). Overall, the present study supports the existence of the attentional suppressive surround and reveals new properties of this selection mechanism – the interaction between the two surround suppression mechanisms and the factor that determines the size of a suppressive surround.

3.2 Introduction

"The darkest place is under the candlestick." means that people tend to overlook nearby things. Interestingly, this expression corresponds well to one aspect of the Selective Tuning model (ST) of visual attention – center-surround distribution of attention (Tsotsos, 1990); for a full specification see Tsotsos (2011); for review, see Carrasco (2011). The center-surround attentional distribution demonstrates the attentional focus accompanied by a suppressive surround. When visual information is spatially close to the attentional focus so it falls within the suppressive zone, its processing is inhibited, making the attended information more conspicuous while filtering out irrelevant distractors. However, visual information beyond the suppressive surround is not affected. ST posits that top-down winner-take-all (WTA) processes select the strongest inputs to neurons (winner) at each processing level, and then prune away losing input connections around the winner. Those pruned connections form a suppressive surround and it helps precise localization of the attended stimulus. However, the other connections located far from the attended stimulus that do not interfere with the processing of the stimulus remain unaffected.

ST has made several predictions regarding the characteristics of the attentional surround suppression. Here, we introduce some predictions that are relevant to the hypotheses and findings of the current study. 1) Location-based attentional surround suppression occurs when a visual task requires spatial localization of the attended stimulus for visual scrutiny and it involves top-down feedback signaling throughout the visual processing hierarchy (Boehler et al., 2009; Hopf et al., 2010). 2) ST predicts that the size of the attentional suppressive surround would be determined by the neuron's RF size that best represents the attended stimulus (Tsotsos, 2011). Given that neurons have scales of representation defined by their RF sizes, then there would be a best size for any given visual stimulus. A main element of ST is that the attentional surround suppression helps reduce or eliminate the interference from unattended stimuli within a RF. Combining these, ST predicts that the size of the attentional suppressive surround should be related to the RF size of the neuron that best represents the attended stimulus and that the RF size should be closest to the size of the attended stimulus and large enough to include the whole stimulus. Motivated by this prediction, Hopf and colleagues (2010) hypothesized that the size of the attentional suppressive surround would differ depending on the processing level of the attended feature in visual hierarchy due to different neurons' RF sizes across visual areas. That is, a simple feature represented at a low-level visual area would produce a narrower suppressive surround than a complex feature because of smaller RF size in the low-level area. They, however, observed equally large suppressive surrounds for color- and luminance-targets, suggesting the earliest processing level of an attended feature may not influence the size of the attentional suppressive surround. A more appropriate test might have been to compare with more abstract targets (such as a complex object that would require neural activations higher than V4) but this remains for future work. In addition, this experiment did not consider how RF sizes are represented throughout the visual hierarchy as well as across the visual field and as a result this experimental design did not test these factors. Since then, Kay et al. (2013) have elucidated this relationship and our current design appropriately adjusts the methodology. 3) ST proposed that an analogous suppressive mechanism operates in the feature domain as well (Tsotsos, 1990). For example, directing attention to a certain feature (e.g., color red) facilitates its processing globally in the visual field but suppresses other similar features (e.g., orange) that represented closely to the attended feature in the same feature map. Features more dissimilar (e.g., green) from the attended one are not affected. Supporting evidence for feature-based surround suppression is observed in several feature domains (Bartsch et al., 2017; Fang, Becker, & Liu, 2019; Ho et al., 2012; Kiyonaga & Egner, 2016; Loach et al., 2008; Störmer & Alvarez, 2014; Tombu & Tsotsos, 2008; Wang et al., 2015).

In the present study, we investigated the properties of the attentional surround suppression that remain open. We set two research questions based on ST's original predictions and subsequent empirical findings. The first is whether the mechanisms of location-based and feature-based surround suppression interact with each other. Secondly, we explored what determines the size of the attentional suppressive surround, looking for empirical evidence for ST's prediction that the size of the attentional surround suppression is associated with the attended neuron's RF size (Tsotsos, 2011). To answer these questions, we conducted two experiments using a target discrimination task that has been well established as an effective measurement of the attentional suppressive surround (Cutzu & Tsotsos, 2003). During the task, participants were asked to judge whether a cued target (reference target) and the second target (probe) were the same or different while the distance between them systematically varied. Here, an attentional cue guides attention to the location of the reference target and prioritizes its processing, forming a suppressive surround around the reference target. If a *nearby* probe falls within the suppressive surround, target discrimination performance is deteriorated due to the suppression. Indeed, Cutzu and Tsotsos showed that target discrimination accuracy was lowest when two targets were closest to each other, but it improved when the inter-target distance increased, indicating location-based attentional surround suppression.

In Experiment 1, we hypothesized that the suppressive effects would be demonstrated more strongly when an attended stimulus and a nearby stimulus are located closely both in physical space and in a shared feature dimension if they indeed interact. Participants performed a similar task to the one in Cutzu and Tsotsos, but the stimuli were oriented bars instead of letters. As we hypothesized, we found that the center-surround attentional profile changes depending on both spatial and feature (i.e., orientation) distances between the target stimuli, indicating an interaction between location-based and feature-based surround suppression. Experiment 2 addressed whether stimulus size and eccentricity, which activate populations of neurons with correspondingly different RF sizes, led to differences in the size of the suppressive surround. Although the idea of the processing level of attended features has been considered previously (Hopf et al., 2010), sufficient variation in RF sizes to enable the detection of differences was not included, something that could be done by varying stimulus eccentricity (Hubel & Wiesel, 1974; Smith et al., 2001; Wilson & Sherman, 1976). Kay et al. (2013) provide a thorough set of relationships among RF size, eccentricity and visual areas, demonstrating that the stimulus eccentricities used in Hopf et al. (2010) may have involved too small a variation in RF sizes to be an effective experimental probe. By varying stimulus eccentricity, we hypothesized that a stimulus presented in the periphery would produce a wider suppressive surround than a stimulus presented in the foveal region. In addition, when the size of a stimulus becomes larger, more neurons will be involved in representing each part of the stimulus. Hence, we predicted that increments of the stimulus size would consequently make the size of the attentional suppressive surround larger. The current results suggested that stimulus eccentricity but not stimulus size changes the size of the attentional suppressive surround. This supports the portion of the original prediction that the surround size depends on the neuron RF that represents the target.

3.3 Experiment 1

3.3.1 Materials and methods

3.3.1.1 Participants

Thirty-eight York University students (ages 17-34, 29 female) participated in Experiment 1. All participants were unaware of the purpose of the experiment. They had normal or corrected-to-normal vision and normal color vision (tested with Ishihara Color Plates). Informed written consent was obtained from all participants and participants recruited from Undergraduate Research Participants Pool (URPP) of York University received course credits for their study participation. The research was approved by York University’s Human Participants Review Committee.

3.3.1.2 Apparatus and stimuli

Participants performed the experiments in a dark room. The position of each participant’s head was stabilized by a head and chin rest (Headspot, UHCOtech) placed 57 cm from a CRT monitor (21” View Sonic G225f, 1280x1024, 85 Hz). Each participant wore an infrared eye tracker (Eyelink II, SR Research, 500 Hz) monitoring the left eye position. The stimuli were created using MATLAB (The Mathworks Corp.) and the Psychophysics Toolbox (Brainard, 1997; Pelli, 1997). Experimental control was maintained by Presentation (Neurobehavioral Systems). Data analyses were conducted using MATLAB and SPSS (IBM).

The stimuli were oriented bars and their size was 0.6 degree of visual angle (dva). Two sets of stimuli were created based on the amount of orientation tilt (smaller or larger than 45°). Each set contained 8 right-tilted (quantized into 5° bins) and vertically symmetrical 8 left-tilted oriented bars. Vertical and horizontal orientations were not used. The set size of a stimulus array was 11 so there were 5 different inter-target distances. Stimuli were presented on an invisible annulus subtending 4 dva eccentricity (in radius). The approximate center-to-center inter-target distances were 2.27, 4, 6.05, 7.28, and 7.92 dva.

3.3.1.3 Procedure

Participants were required to look at a fixation cross presented at the center of the display and pressed a button on a response pad when they were ready to start the task (Figure 3.1). Once they pressed the button and if their fixation was maintained, an attentional cue (i.e., yellow filled circle) was presented for 150 ms at a random location on an invisible annulus. Then, all stimuli were presented simultaneously and remained on the screen for 100 ms. As mentioned earlier, the spatial surround suppression manifests at about 250 ms after stimulus onset (Boehler et al., 2009), and our timing choices were motivated by that result. Two of them were surrounded by yellow rings, indicating that they were the targets of the trial. The first target was always presented where the attentional cue was presented (100% valid), and the second target was presented randomly at one of the other locations. Task-irrelevant distractors filled the remaining locations. Target and distractor orientations were randomly selected within a stimulus set and stimulus selection was not tailored for each participant's performance. The stimulus array was followed by a 500 ms mask to remove sensory memory. Participants responded whether the two targets were the same or different by pressing buttons on the response pad. An experimental session consisted of 160 trials (10 second target locations x 2 trial types (same or different targets) x 8 repetitions, cue and first target locations were randomly selected).

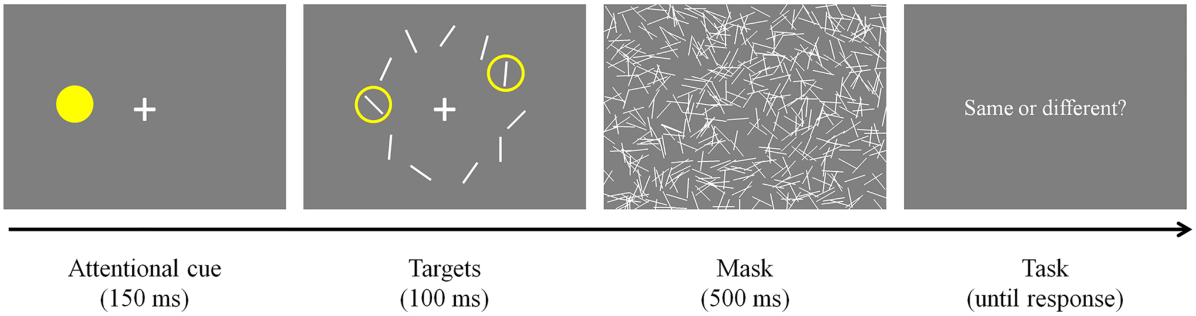


Figure 3.1. Procedure of Experiment 1.

After an attentional cue was presented for 150 ms, two target stimuli surrounded by yellow rings and distractors were presented for 100 ms. The stimulus array was masked for 500 ms to remove sensory memory. Participants responded whether the targets were the same or different by pressing buttons on a keyboard.

3.3.2 Results

Data were not included in the analysis if target discrimination accuracy (proportion correct) was lower than .65 in all inter-target distance conditions. Of the 38 participants, 11 participants were excluded through this procedure (mean accuracy (SD) = .496 (.037)). For the participants able to perform the task, a repeated measures analysis of variance (ANOVA) showed that target discrimination accuracy significantly changed across inter-target distances ($F(4, 104) = 3.536$, $p = .01$). Accuracy sharply decreased between the first and the second closest inter-target distances (2.27 vs. 4 dva: $M_{diff} = .067$, SE = .026, $t(26) = 2.551$, $p = .017$) but then, it tended to recover at the farthest inter-target distance (4 vs. 7.92 dva: $M_{diff} = -.039$, SE = .019, $t(26) = -2.026$, $p = .053$), showing a U-shaped profile (Figure 3.2A).

We also computed sensitivity, d', to measure participants' target discrimination performance independently of their response bias (Green & Swets, 1966; Macmillan & Creelman, 2005). d' significantly varied across inter-target distances ($F(4, 104) = 4.368$, $p = .003$; Figure 3.2B). In particular, d' also exhibited the U-shaped profile as we observed in simple accuracy analysis: d' diminished between the first and the second closest inter-target distances (2.27 vs. 4 dva: $M_{diff} = .46$, SE = .15, $t(26) = 3.01$, $p = .006$) and then, it recovered at the farthest inter-target distance (4 vs. 7.92 dva: $M_{diff} = -.23$, SE = .11, $t(26) = -2.142$, $p = .042$). Response bias systematically varied along d' (Greenhouse-Geisser correction ($\varepsilon = .668$): $F(2.67, 69.47) = 12.106$, $p < .001$) and participants' tendency to respond "same" is associated with greater sensitivity (Figure 3.2C). While finding differences between different stimuli is usually efficient (unless the differences are very subtle), it requires more thorough comparisons to confirm whether the same stimuli are indeed identical (Bamber, 1969; Belke & Meyer, 2002; Downing, 1971; Egeth, 1966; Farell, 1985; Hawkins, 1969). When attentional surround suppression is active so that information processing is constrained, it could be more difficult to make "same" judgment due to low sensitivity. The current response bias pattern is consistent with this idea, suggesting that surround suppression biases participants' decision making.

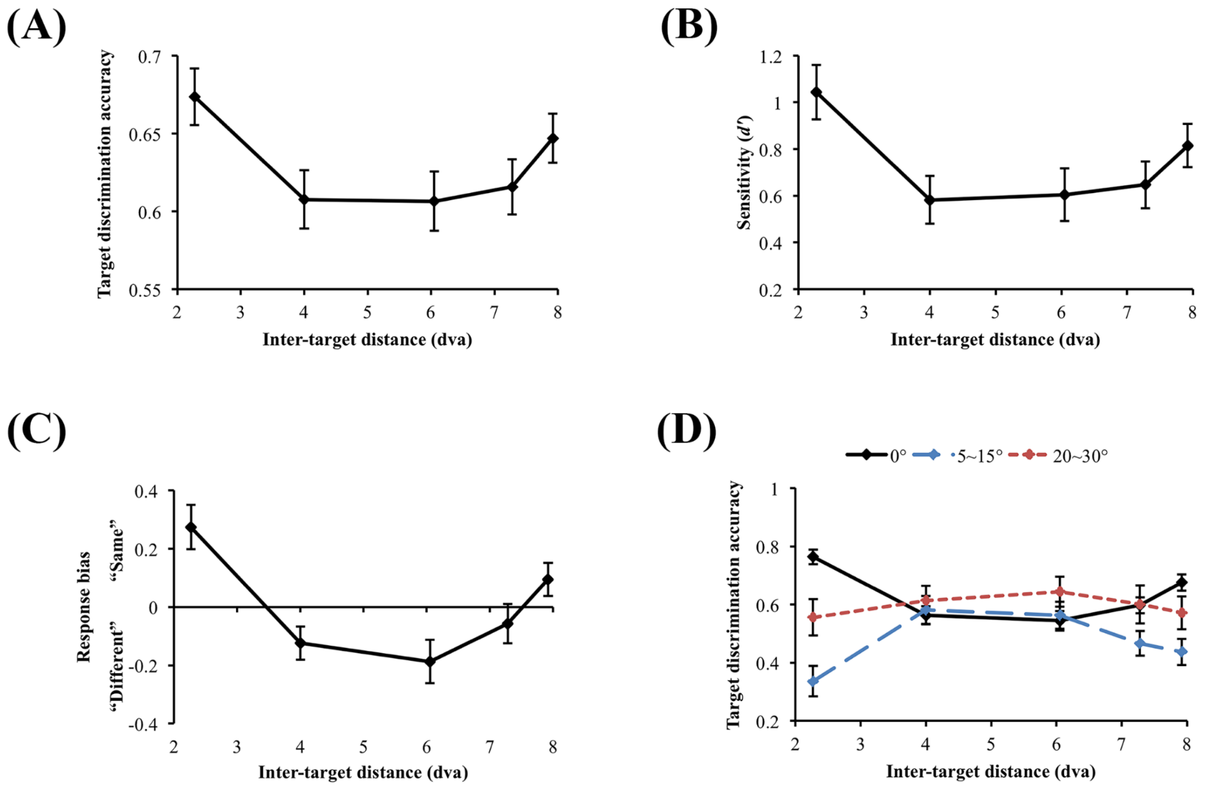


Figure 3.2. Experiment 1 results.

(A) Target discrimination accuracy (proportion correct) of Experiment 1. Accuracy was highest when the two targets were spatially closest and sharply decreased at the second closest distance. It eventually recovered at the farthest distance, demonstrating a U-shaped profile. (B) Participants' sensitivity for same-different discrimination (*d′*) also exhibited the comparable U-shaped profile. (C) Response bias changed along *d′*, demonstrating that "same" judgment, which requires thorough visual scrutiny, is associated with higher sensitivity. (D) Target discrimination accuracy varied depending on both orientation difference between the target stimuli and inter-target distance. When the target stimuli had the same orientation, accuracy was greatest at the closest inter-target distance, leading the U-shape profile. A typical attentional surround suppression profile (lowest accuracy at the closest inter-target distance) was observed when orientation difference between the targets was 5~15°. Surround suppression was not apparent when orientation difference became greater (20~30°). Error bars indicate SEM.

Nevertheless, the current discrimination accuracy and *d′* results do not match a typical profile of the attentional suppressive surround that previously reported (e.g., greatest suppression when two targets are closest to each other). We wondered if an interaction between location-based and feature-based surround suppression resulted in this unusual pattern. Hence, we broke down orientation differences between the target stimuli (0°, 5~15°, and 20~30° differences) and analyzed target discrimination accuracy for each orientation difference range (Figure 3.2D). Note that we analyzed accuracy because we cannot calculate *d′* after breaking down stimulus types. Target discrimination accuracy significantly varied depending on orientation difference between the target stimuli ($F(2, 46) = 12.639$, $p < .001$), demonstrating the lowest accuracy when orientation difference was small (5~15°) compared to the other conditions (all *p*s < .001). The main effect of inter-target distances was not significant ($F(4, 92) = .292$, $p = .882$). Importantly, there was a significant interaction between orientation difference and inter-target distance ($F(8, 184) = 4.857$, $p < .001$). At the closest inter-target distance (2.27 dva), target discrimination accuracy was highest when the target orientation was the same, and it was lowest when orientation difference was 5~15° ($M_{diff} = .428$, SE = .063, $t(26) = 6.798$, $p < .001$). The accuracy was intermediate when orientation difference was 20~30°. Participants' accuracy for 0° and 5~15° orientation differences diverged again at the farthest inter-target distance (7.92 dva: $M_{diff} = .239$, SE = .066, $t(26) = 3.59$, $p = .001$).

We also investigated how the profile of the attentional suppressive surround differs in each orientation difference condition. When the two targets had the same orientation, target discrimination accuracy significantly changed depending on inter-target distance (Greenhouse-Geisser correction ($\varepsilon = .739$): $F(2.96, 79.80) = 17.658$, $p < .001$). The accuracy was greatest at the closest inter-target distance and decreased at the second-closest inter-target distance (2.27 vs. 4 dva; $M_{diff} = 20.14\%$, SE = 3.03%, $t(26) = 6.65$, $p < .001$). Accuracy gradually increased towards the farthest distance from the second-closest distance ($F(3, 78) = 8.04$, $p < .001$). This result indicated that the U-shaped profile demonstrated in the previous analysis resulted from "same target" trials. A typical profile of attentional surround suppression was observed when the orientation difference between the targets was small (5~15°). Target discrimination accuracy significantly changed across inter-target distances ($F(4, 100) = 5.183$, $p =.001$), and it was lowest at the closest inter-target distance and then, improved at the second-closest inter-target distance (2.27 vs. 4 dva, $M_{diff} = -24.5\%$, SE = 7.07%, $t(26) = 3.467$, $p = .002$). Accuracy at the farthest distance tended to be higher than that at the closest distance (2.27 vs. 7.92 dva, $M_{diff} = -10.06\%$, SE = 5.17%, $t(26) = 1.944$, $p = .063$). Lastly, target discrimination accuracy did not differ across inter-target distances when targets' orientation difference was larger (20~30°: $F(4, 96) = 0.337$, $p =.852$). Consistent with our hypothesis, the current results suggest that attentional surround suppression is strongest when the stimuli are located closely to each other both in physical and feature spaces.

3.4 Experiments 2A & 2B

As aforementioned, one theoretical prediction within ST is that the size of the attentional suppressive surround is set by the neuron's RF size that best represents the attended stimulus (Tsotsos, 2011). At its core, this links a neuron's RF size with the size of the spatial attentional suppressive surround. The original rationale for the suppressive surround was to provide a mechanism to reduce the impact of context within a RF to allow a neuron to process "signal" rather than "noise" in a more direct manner via top-down manipulation. The effect of such a mechanism has been shown previously. For example, spatial and feature-based attention flexibly modulates neurons' RF profiles and sizes in the area MT (Anton-Erxleben et al., 2009; Niebergall, Khayat, Treue, & Martinez-Trujillo, 2011; Womelsdorf et al., 2006, 2008), and attentional load also affects the size and the spatial tuning of population RFs in early visual cortex (V1-V3, de Haas, Schwarzkopf, Anderson, & Rees, 2014). Anatomically, neurons' RFs become progressively larger as the level in the visual processing hierarchy and visual eccentricity increase (Kay et al., 2013). A previous study examined whether the spatial extent of the surround suppression differs depending on the processing level of an attended feature (Hopf et al., 2010). The results showed that the size of the attentional suppressive surround was identical for both luminance and color targets, suggesting that the ultimate top-down selection of these features could happen at the same level in visual hierarchy. However, the majority of neurons in early and intermediate visual areas are sensitive to both color and luminance (Bushnell, Harding, Kosai, Bair, & Pasupathy, 2011; Gouras & Kruger, 1979; Hubel & Livingstone, 1990; Johnson, Hawken, & Shapley, 2001; Lennie, Krauskopf, & Sclar, 1990; Thorell, de Valois, & Albrecht, 1984), hence, it is difficult to clearly separate the processing levels of these two features. Further, as mentioned earlier, it may be that the stimulus sizes and eccentricities did not make for sufficient experimental probes for this characteristic. In Experiment 2, we varied stimulus eccentricity, hypothesizing that a stimulus presented in the periphery would produce a wider suppressive surround than a stimulus presented in the foveal region due to increasing RF sizes toward the visual periphery (Hubel & Wiesel, 1974; Smith et al., 2001; Wilson & Sherman, 1976). We also hypothesized that manipulating the size of a stimulus would affect the size of the suppressive surround. Changes in the stimulus size influence the number of neurons that processes the stimulus within a visual processing level and then, the sum of these neurons' RFs would produce a corresponding suppressive surround. We used two types of stimuli (orientation and human faces in Experiment 2A and 2B, respectively) and three different stimulus size and eccentricity combinations for the target discrimination task.

3.4.1 Materials and Methods

3.4.1.1 Participants

Thirty-two (ages 18-34, 17 female) and thirty-one (ages 18-28, 17 female) York University students participated in Experiments 2A and 2B, respectively. All participants were unaware of the purpose of the experiment. They had normal or corrected-to-normal vision and normal color vision (tested with Ishihara Color Plates). Informed written consent was obtained from all participants and participants recruited from Undergraduate Research Participants Pool (URPP) of York University received course credits for their study participation. The research was approved by York University's Human Participants Review Committee.

3.4.1.2 Apparatus and stimuli

The apparatus was the same as in Experiment 1. The stimuli used in Experiment 2A were small (1 dva) and large (3 dva) oriented bars. Except for their sizes, the stimuli were equal to those used in Experiment 1. Human faces were used in Experiment 2B (TarrLab face database: www.face-place.org; Righi, Peissig, & Tarr, 2012). For human faces, only white male faces were used. All of them had short hair, the same viewpoints and facial expression (neutral), and did not wear accessories. The set size of a stimulus array was 11 so there were 5 different inter-target distances. Stimuli were presented on an invisible annulus at 7 dva eccentricity in the large and small (far) stimulus conditions, and at 4.8 dva eccentricity in the small (near) stimulus condition (in radius). The approximate center-to-center inter-stimulus distances were 3.94, 7.56, 10.58, 12.74, and 13.86 dva when stimuli were presented at 7 dva eccentricity, and 2.7, 5.2, 7.26, 8.74, and 9.5 dva when stimuli were presented at 4.8 dva eccentricity.

3.4.1.3 Procedure

The task was the same as in Experiment 1 except that one of the targets surrounded by a yellow ring was presented for 300 ms as an attentional cue (Figure 3.3). The ring had the same color and size as the yellow circle used in Experiment 1. We used this new attentional cue and lengthened its presentation duration to make participants focus more on the given stimulus so that they could better develop the attentional suppressive surround for both location and feature. Then, the second target and distractors were simultaneously presented for 200 ms. Since we assumed that the extremely short target presentation duration in Experiment 1 (100 ms) led to poor target discrimination accuracy, we used a longer target presentation duration in this experiment. Our pilot studies suggested that people needed much longer stimulus presentation duration specifically to perform face discrimination tasks. Half of the participants completed Experiment 2A (orientation condition) and the other half completed Experiment 2B (face condition). Within a stimulus type, there were three different display conditions (large, small (far), and small (near)). Each display condition consisted of 160 trials (10 second target locations x 2 trial types (same or different targets) x 8 repetitions, first target locations were randomly selected).

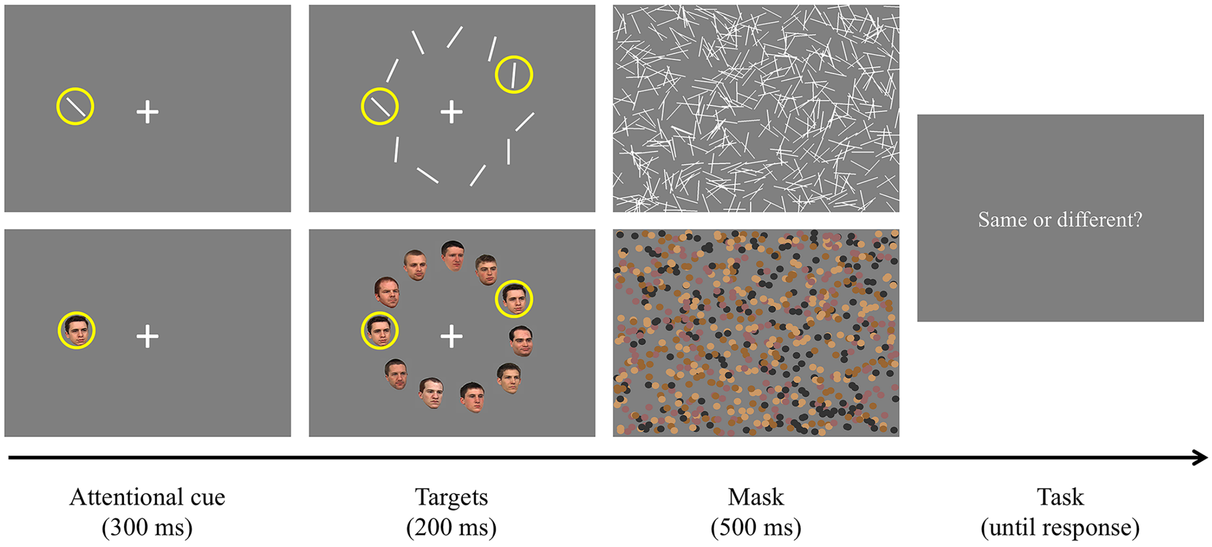


Figure 3.3. Procedure of Experiment 2.

The upper panel shows an example of Experiment 2A (orientation, different targets) and the lower panel shows an example of Experiment 2B (human face, same targets). Unlike Experiment 1, one of the target stimuli surrounded by a yellow ring was presented for 300 ms as the attentional cue. Later, the other stimuli were presented for 200 ms and another target stimulus was indicated by a yellow ring. They were followed by 500 ms masking stimuli. Participants pressed a predetermined key to report whether the targets were the same or different.

3.4.2 Results

3.4.2.1 Experiment 2A: Orientation

Data were not included in the analysis if target discrimination accuracy (proportion correct) was lower than .65 in all inter-target distance conditions in any of the stimulus display conditions. Through this process, data collected from 2 participants were excluded (mean accuracy (SD) = .454 (.034) and .569 (.146), respectively). Since we were mainly interested in how stimulus size and eccentricity affect the attentional profile, we did not report the results after breaking down target orientation differences as in Experiment 1. However, the attentional profile in each orientation difference condition was similar across the experiments (e.g., U-shaped profile resulted from same target orientation). First, we compared target discrimination accuracy between large and small (far) orientation conditions where the stimuli were presented at the same eccentricity (Figure 3.4A). The main effect of stimulus size was significant ($F(1, 29) = 13.794$, $p = .001$), indicating higher accuracy for large stimuli than small stimuli. Target discrimination accuracy was also significantly affected by inter-target distance ($F(4, 116) = 3.124$, $p = .018$). However, the interaction between stimulus size and inter-target distance was not significant ($F(4, 116) = 1.832$, $p = .127$), meaning that stimulus size did not change the attentional profile. We did not compare discrimination accuracy in the small (near) condition with accuracies in the other conditions because of different inter-target distances. When each stimulus size condition was analyzed separately, target discrimination accuracy for large orientation tended to vary across different inter-target distances ($F(4, 116) = 2.101$, $p = .085$). Accuracy for small (far) orientation significantly varied depending on inter-target distance ($F(4, 116) = 2.835$, $p = .028$). It decreased at the second closest inter-target distance (3.94 vs. 7.56 dva: $M_{diff} = .051$, SE = .023, $t(29) = 2.269$, $p = .031$) but it increased again at the farthest inter-target distance (7.56 vs. 13.86 dva: $M_{diff} = -.043$, SE = .021, $t(29) = .-2.076$, $p = .047$), showing the U-shaped profile as found in Experiment 1. Small orientations showed a similar profile when their eccentricity had decreased (small (near)). In the small (near) condition, target discrimination accuracy significantly varied across inter-target distance ($F(4, 116) = 8.442$, $p < .001$). Accuracy was significantly deteriorated at the second-closest inter-target distance (2.7 vs. 5.2 dva, $M_{diff} = .104$, SE = .019, $t(29) = 5.56$, p

< .001) but it recovered at the farthest distance (5.2 vs. 9.5 dva, M_{diff} = -.052, SE = .02, *t*(29) = -2.656, *p* = .013).

We performed a *d'* analysis (Figure 3.4B) and observed similar (but weaker) patterns as in the simple accuracy analysis. When large and small (far) orientation conditions were compared, large stimulus size contributed to higher *d'* (*F*(1, 29) = 14.493, *p* = .001) and *d'* tended to vary across inter-target distance (*F*(4, 116) = 2.322, *p* = .061). Again, the interaction between stimulus size and inter-target distance was not significant (*F*(4, 116) = 1.722, *p* = .15). Changes in inter-target distance did not strongly modulate *d'* in the large orientation condition (*F*(4, 116) = 1.681, *p* = .159). Nevertheless, *d'* in the small (far) and small (near) conditions varied across different inter-target distances (small (far): *F*(4, 116) = 2.401, *p* = .054 (marginal), small (near): *F*(4, 116) = 9.788, *p* < .001). In the latter two conditions, *d'* was attenuated at the second closest inter-target distance and then improved toward the farthest distance (all *p*s < .05), demonstrating U-shaped profiles. Participants' tendency to respond "same" was generally associated with greater *d'* (Figure 3.4C).

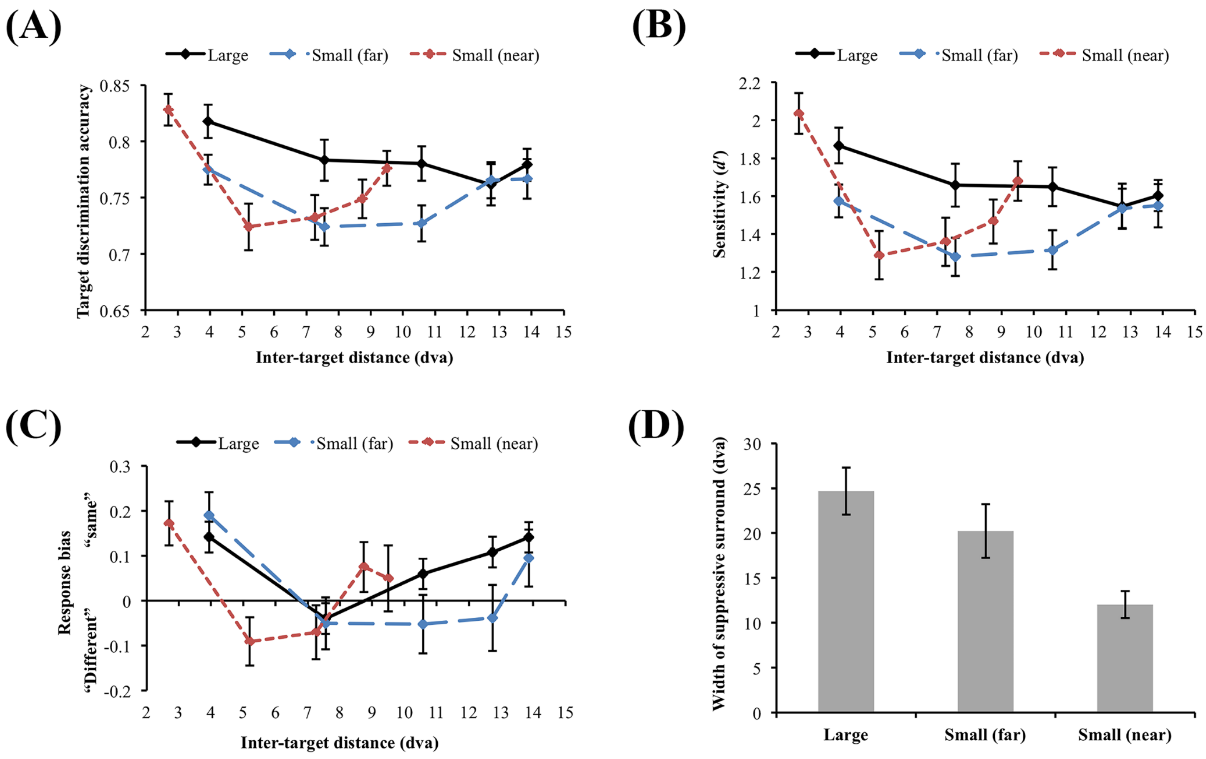


Figure 3.4. Experiment 2A results.

(A) Orientation target discrimination accuracy (proportion correct) in different stimulus size and eccentricity conditions. (B) The similar results were found when participants' sensitivity was analyzed. (C) Response bias varied along sensitivity, demonstrating an association between higher sensitivity and "same" response tendency. (D) Widths of the attentional suppressive surrounds. The sizes of the attentional suppressive surrounds in the large and small (far) orientation conditions did not statistically differ. Small (near) orientation condition produced a narrower suppressive surround compared to the other conditions. Error bars indicate SEM.

Next, we examined whether the size of the attentional suppressive surround is affected by stimulus eccentricity. To compare the size of the attentional suppressive surround across different display conditions, we used a curve fitting method because we cannot directly compare target discrimination accuracy through ANOVA when stimulus eccentricity differs (i.e., small (far) vs. small (near)). As shown in Experiment 1, we observed a U-shaped attentional profile again, hence, to represent attentional profile reflected in our data, we fitted a quadratic function (i.e., U-shaped) and measured its width as an indication of the size of the attentional suppressive surround. The quadratic function explained well the average target discrimination accuracy as a function of inter-target distance (goodness-of-fit is defined by adjusted R-squared, large = .770; small (far) = .820; small (near) = .972). Then, we fitted quadratic functions to individual data and measured its width (half-width at half-minimum). An outlier from the small (far) condition was excluded from the analysis (> 3 SD). Within the individual data, 56.66% of the data in the large condition, 73.33% in the small (far) condition, and 80% in the small (near) condition showed the U-shaped profile as the average data did, showing that there was individual variability in each condition, as indicated by the variability in widths (error bars) in Figure 3.4D. The results (Figure 3.4D) showed that the attentional suppressive surrounds in the large and small (far) conditions were not statistically different (M_{diff} = 4.71 dva, SE = 3.93 dva, $t(28) = 1.2$, $p = .240$), whereas the suppressive surround in the small (near) condition was narrower than those in the other conditions (large vs. small (near), M_{diff} = 12.65 dva, SE = 2.97 dva, $t(29) = 4.253$, $p < .001$; small (far) vs. small (near), M_{diff} = 8.49 dva, SE = 3.41 dva, $t(28) = 2.488$, $p = .019$).

3.4.2.2 Experiment 2B: Human faces

The same accuracy cut-off as in Experiment 2A was applied and one participant was excluded from the analysis through this process (mean accuracy (SD) = .61 (.191)). When the large and small (far) face conditions were compared, the main effects of stimulus size ($F(1, 29) = 6.535$, $p = .016$) and inter-target distance (Huynh-Feldt correction ($\varepsilon = .888$): $F(3.55, 102.97) =$

3.212, *p* = .02) were significant. However, their interaction was not significant (*F*(4, 116) = .871, *p* = .484), suggesting that stimulus size did not change the profile of the attentional suppressive surround. When each display condition was separately analyzed, target discrimination accuracy significantly varied depending on inter-target distance only in the large face condition (*F*(4, 116) = 2.594, *p* = .04) due to the accuracy peak at the second-farthest inter-target distance (8.74 dva). In contrast, accuracy remained unchanged across inter-target distances in the other conditions (small face (far): *F*(4, 116) = 1.337, *p* = .26), small face (near): *F*(4, 116) = .634, *p* = .639), thus, they did not suggest attentional surround suppression. Figure 3.5A shows the results of all conditions. *d'* analysis demonstrated the same results as we found in the simple accuracy analysis (Figure 3.5B). In the comparison between the large and small (far) face conditions, there was an advantage of large stimulus size (*p* = .01) but again, the interaction between stimulus size and inter-target distance was not significant. In addition, *d'* was modulated by inter-target distance only in the large face condition (*F*(4, 116) = 2.849, *p* = .027) but not in the other conditions. The patterns of response bias (Figure 3.5C), however, indicated that participants responded "different" more often when the inter-target distance was closer, but this tendency was gradually reversed as the inter-target distance increased (in all conditions *p*s < .001). We previously found an association between higher sensitivity and the response bias towards "same" judgments. Since face discrimination sensitivity did not vary across inter-target distances in this experiment, we questioned what caused the changes in the response bias. One possibility is that discrimination performance for same and different target faces could be very different and performance in one of these conditions might have affected the response bias. As we mentioned earlier, detecting any difference between different faces is usually easier than inspecting every feature on the same faces to confirm they are indeed the same, which requires more attention. If the assumption is correct, attentional surround suppression will be produced when target faces are the same, but it will be weak or even absent when target faces are different.

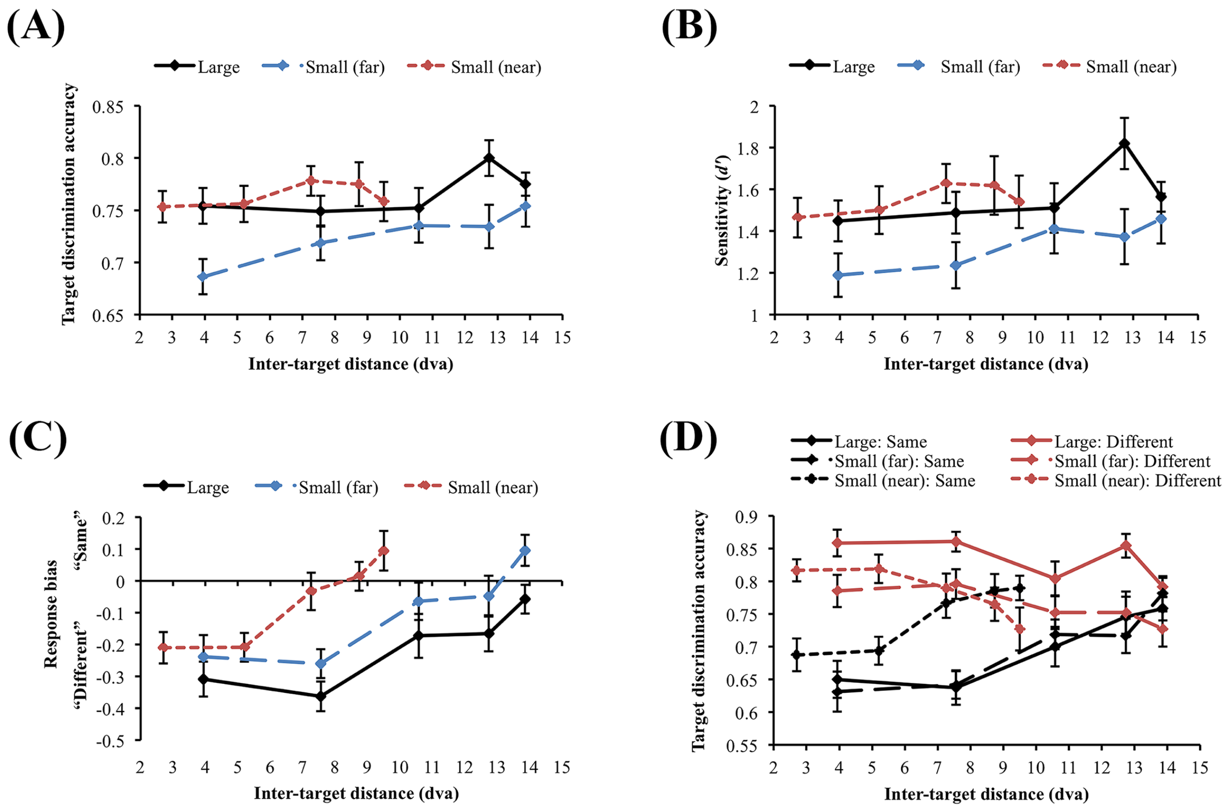


Figure 3.5. Experiment 2B results.

(A) Target discrimination accuracy and (B) sensitivity for face stimuli. No clear evidence for surround suppression was observed in both measures. (C) Participants tended to respond "same" more often as inter-target distance increased. (D) The identity of target faces affected target discrimination accuracy in different ways. Discrimination accuracy was greater when different faces were presented as the target stimuli (red lines), whereas it suffered when the target stimuli were the same face (black lines). Only same target faces produced attentional surround suppression. Error bars indicate SEM.

A repeated measures ANOVA was conducted to examine whether target identity (same vs. different) affected target discrimination accuracy (Figure 3.5D). First, we compared large and small (far) face conditions. The main effect of target identity on target discrimination accuracy was significant ($F(1, 29) = 26.135$, $p < .001$), demonstrating higher accuracy when target faces were different. It supports the idea that different faces are easier to discriminate than the same faces. Target discrimination accuracy was also affected by stimulus size ($F(1, 29) = 6.535$, $p = .016$), indicating better performance for larger stimuli. Inter-target distance significantly changed target discrimination accuracy (Huynh-Feldt correction ($\varepsilon = .89$), $F(3.55, 102.97) = 3.21$, $p = .02$). The two-way interaction between target identity and stimulus size was significant ($F(1, 29) = 5.58$, $p = .025$). Post-hoc tests with Bonferroni correction showed that accuracy for large faces was higher than that for small faces only when target faces were different ($M_{diff} = .071$, SE = .017, $p < .001$) but accuracy did not differ when target faces were the same ($M_{diff} = 0$, SE = .023, $p = .986$). The interaction between target identity and inter-target distance (Greenhouse-Geisser correction ($\varepsilon = .699$), $F(2.80, 81.10) = 15.417$, $p < .001$) was also significant. Target discrimination accuracy was higher in different face trials than in same face trials when inter-target distances were relatively shorter but accuracy in both same and different face trials became roughly equal at farther inter-target distances. Interaction between stimulus size and inter-target distance was not significant ($F(4, 116) = .871$, $p = .484$). Three-way interaction among target identity, stimulus size, and inter-target distance was not significant, either ($F(4, 116) = .100$, $p = .982$).

Within the same face trials, accuracy significantly varied across inter-target distances in all conditions (large: $F(4, 116) = 6.222$, $p < .001$; small (far): $F(4, 116) = 8.948$, $p < .001$; small (near): $F(4, 116) = 6.259$, $p < .001$). Importantly, accuracy gradually improved from the closest to the farthest inter-target distance (large: $M_{diff} = .108$, SE = .033, $t(29) = 3.247$, $p = .003$; small (far): $M_{diff} = .15$, SE = .031 , $t(29) = 4.826$, $p < .001$; small (near): $M_{diff} = .102$, SE = .029, $t(29) = 3.468$, $p = .002$). Consistent with our hypothesis, identifying the same faces required substantial attentional processing and produced attentional surround suppression. When the target faces were different, target discrimination accuracy was significantly affected by inter-target distance for large faces ($F(4, 116) = 3.828$, $p = .006$) and for small (near) faces ($F(4, 116) = 3.787$, $p = .006$) but not for small (far) faces ($F(4, 116) = 1.767$, $p = .140$). Contrary to the accuracy patterns in the same face trials, accuracy decreased between the closest and the farthest inter-target distances (large: $M_{diff} = -.067$, SE = .021, $t(29) = -3.247$, $p = .003$; small (far): $M_{diff} = -.058$, SE = .029 , $t(29) = 2.019$, $p = .053$ (marginal); small (near): $M_{diff} = -.09$, SE = .032 , $t(29) = 2.842$, $p = .008$). It indicated that there would be no attentional surround suppression when the task is relatively easy, so it does not demand much attentional processing.

Although attentional surround suppression was evident in all same face conditions, the attentional profiles quantitatively differed depending on stimulus eccentricity. In the small (near) condition, target discrimination accuracy initially improved from .69 to .77 as inter-target distance increased from 2.7 to 7.26 dva ($M_{diff} = .079$, SE = .026, $t(29) = 3.072$, $p = .005$) and then, it leveled off at farther distances (7.26 vs. 9.5 dva, $M_{diff} = .023$, SE = .027, $t(29) = .863$, $p = .395$). This plateau indicated the release from the surround suppression when the target stimuli were spatially well separated (Cutzu & Tsotsos, 2003). In contrast, target discrimination accuracy in the other conditions improved across the three farthest inter-target distances (10.58 vs. 13.86 dva, large: $M_{diff} = .058$, SE = .031, $t(29) = 1.891$, $p = .069$ (trend level), small (far): $M_{diff} = .063$, SE = .029, $t(29) = 2.186$, $p = .037$), meaning that surround suppression was still effective at this range. These results suggest that the size of the attentional suppressive surround is associated with stimulus eccentricity.

3.5 General discussion

The present study behaviorally examined the factors that manipulate the attentional suppressive surround, which is the hallmark of the Selective Tuning model. The results demonstrated that the profile of the attentional suppressive surround is dependent on the attended spatial location, on the attended feature, and on stimulus eccentricity.

3.5.1 Interaction between location-based and feature-based surround suppression

Location-based and feature-based surround suppression have been studied independently so far. In Experiment 1, we found that inter-target distances and target feature similarity shape the attentional suppressive surround together, demonstrating an interaction between the two suppressive mechanisms. Surround suppression was apparent when the target stimuli were close to each other spatially and in the feature map (i.e., when target orientations were similar). Then, it disappeared as the spatial and feature distances between the target stimuli increased. Although it is implicit in ST that such an interaction is possible, this experiment points to an explicit extension of ST by demonstrating the interaction between two forms of attentional surround suppression. As a next step, the nature of the interaction between the two suppressive mechanisms should be addressed, such as whether they additively or multiplicatively shape the attentional tuning profile.

Unlike the typical attentional surround suppression, judging the same orientation targets was best at the closest inter-target distance, suppressed at the intermediate distances, and then recovered at farther distances. This unexpected facilitation at the closest inter-target distance could result from perceptual grouping. As proposed by Gestalt psychologists, perceptually identical and spatially adjacent stimuli tend to be grouped together and facilitated together (Olson & Attneave, 1970; Schulz & Sanocki, 2003; Wagemans et al., 2012). However, this principle did not work on more complex features such as human faces (Experiment 2B).

One might ask if the current results could be explained by other attention models. In particular, the biased competition model (Desimone & Duncan, 1995; Reynolds et al., 1999) states that attention to one stimulus suppresses the unattended stimulus within a RF, thus it seems to predict similar location-based surround suppression. Both ST and biased competition models include attentional biases and competition among neural populations at their core. However, it should be noted that the idea of a suppressive surround due to top-down attentive localization processing was introduced earlier (Tsotsos, 1988, 1990) than most other models of attention, including the biased competition model (for review of attention models, see Carrasco (2011), and Tsotsos (2011)). They also differ from each other in terms of implementing suppressive mechanisms. The suppressive effects in the biased competition model result from attentional enhancement of the target stimulus, as Reynolds et al. (1999) say: "It is implemented here by increasing the efficacy of synapses projecting to the measured cell from the population activated by the attended stimulus. Increasing the strength of the signal from the attended stimulus population causes it to have a greater influence on the total mix of excitation and inhibition. Consequently, the response of the cell is driven toward the response that would be elicited if the attended stimulus were presented alone." Unlike the biased competition model, ST does not assume attentional gain mechanisms, proposing that attentional enhancement is a consequence of an explicit top-down suppressive mechanism. This top-down suppression involves a competitive process that inhibits irrelevant signals around the attended signal which in turn, leads to surround suppression. ST specifically formulates this process, showing how the potentially suppressed signals can be determined. In addition, the attentional surround suppression is an obligatory mechanism that happens even without distractors around the target stimulus (Boehler et al., 2011), whereas suppression in the biased competition requires multiple competing stimuli. Another fundamental difference between ST and biased competition models is that while the biased competition model focuses on attentional manifestations in earlier visual areas (V2/V4, Reynolds et al., 1999), ST includes attentional modulation across the full visual hierarchy which targets the higher-order areas first and then moves down to the lower level as shown in many studies (e.g., Boehler et al., 2009; Buffalo et al., 2010; Lauritzen et al., 2009). The biased competition model does not currently include critical components related to attentional surround suppression as it instead focuses on selection within a neuron's receptive field, while ST does predict a center-surround attentional profile (see Rothenstein & Tsotsos (2014) for the detailed comparisons between ST and other attention models).

When it comes to feature-based surround suppression, biased competition and ST predict different outcomes. While ST predicts a center-surround attentional profile of feature-based attention (Bartsch et al., 2017; Ho et al., 2012; Kiyonaga & Egner, 2016; Loach et al., 2008; Störmer & Alvarez, 2014; Tombu & Tsotsos, 2008; Wang et al., 2015), biased competition predicts a monotonic attentional modulation depending on feature similarity between the attended and unattended features (Martinez-Trujillo & Treue, 2004; Treue & Martinez-Trujillo, 1999). For instance, one recent biased competition attention model (Beuth & Hamker, 2015) implements feature-based attention as a top-down amplification signal. In their simulation, however, feature-based suppression varied as a linear function of feature similarity, which is inconsistent with the findings of current feature-based surround suppression. Another study proposed an attention model based on feedback connections and mutual inhibition within a visual area, which is related to ST (Miconi & VanRullen, 2016). A critical difference between their model and ST is that, as the authors explicitly pointed out, this model does not assume winner-take-all processes that prune away irrelevant connections around the attended input. Thus, their model does not support spatial surround suppression, let alone feature surround suppression, which are two key tenets of ST tested in the current study. Overall, the key factor of feature-based surround suppression predicted by ST has been supported empirically by the current study. Other attentional models may be modified in the future to incorporate that function.

3.5.2 Attentional suppressive surround is affected by stimulus eccentricity but not by stimulus size

Increments in stimulus size would activate more neurons within a visual processing level, each representing a fragment of the stimulus. We hypothesized as the sum of these neurons' RFs increases, it would also enlarge the spatial extent of the attentional suppressive surround. In Experiment 2, however, stimulus size did not change the size of the attentional suppressive surround for both orientation and human faces. One possibility is that because we presented stimuli at the peripheral visual field, the neuron's RF size at this eccentricity was already large enough to represent them as whole objects and due to large RF size, the neuron might not be sensitive to small changes in stimulus size. Alternatively, multiple neurons that represent each segment of the object could send converging input to the neuron in the next level, having a larger RF, thus, the size of the suppressive surround could be determined at a much higher level than initially expected. Since we cannot specify the level of stimulus processing in the behavioral study, these possibilities remain open. Future studies would be required to identify the exact level of stimulus processing and measure the size of the suppressive surround accordingly.

Although the level of stimulus processing is not clearly addressed, RF size differences across visual eccentricity are manifested throughout the visual hierarchy (Kay et al., 2013; Winawer, Horiguchi, Sayres, Amano, & Wandell, 2010). This means that the attended feature, independent of its processing level, would produce different suppressive surrounds in size depending on its eccentricity. Consistent with this idea, the present results showed that stimuli presented closer to the fovea produced a narrower suppressive surround when stimulus size was constant. One might argue that lower visibility at farther visual eccentricity could affect the current results. Nevertheless, both eccentricity conditions (near vs. far) showed roughly equal levels of target discrimination accuracy. Specifically, target discrimination accuracy in both eccentricity conditions is comparable at the farthest inter-target distance where stimulus comparison at the "far" eccentricity condition could have been worst due to low visibility (mean accuracy in the near vs. far conditions, orientation: 77.6% vs. 76.67%, face (same): 78.96% vs. 78.13%). Hence, visibility differences across visual eccentricity do not seem to strongly affect the results. One limitation of our experimental design is that it confounds spacing and eccentricity of the stimuli, so it might not capture the pure eccentricity effect on the attentional surround suppression. Future studies could resolve this issue it by carefully matching low-level parameters in different experimental conditions.

3.5.3 Attentional surround suppression reflects top-down selection refinement

ST proposes that attentional surround suppression occurs as a result of top-down selection processes that prune irrelevant connections that do not contribute to the representation of the attended stimulus. Consequently, this top-down propagation enhances spatial resolution of the attended stimulus and enables more precise localization by narrowing down the pass-zone throughout the visual processing hierarchy. To be more precise, the resolution is enhanced because interference (or entanglement) from the context where the attended stimulus is embedded is reduced. It leads to a novel hypothesis that attentional surround suppression would not be produced if a task can be performed without the need of such enhanced resolution (in location or in feature dimensions). Boehler and colleagues (2009) measured MEG responses during an orientation discrimination task which requires attentional focusing, and during a simple color discrimination task that can be done pre-attentively. Consistent with the hypothesis, only the orientation discrimination task elicited attentional surround suppression. In our Experiment 1, surround suppression did not happen when target orientations were sufficiently different and similarly, in Experiment 2B, we observed attentional surround suppression only when the target faces were identical but not when they were different. In both cases, one must examine the same or similar stimuli carefully to confirm their identities and it requires thorough attentional focusing. On the other hand, discrimination of easily distinguishable stimuli demands relatively lesser attentional focusing. For face stimuli, even though we controlled some features across different faces (e.g., races, emotion, perspectives, accessories, etc), there could be still distinctive features on the faces that made the task much easier. These results indicate that the top-down selection processes result in attentional surround suppression as ST suggests.

3.6 Conclusions

After the Selective Tuning model proposed the existence of attentional surround suppression, many studies have reported the properties of this phenomenon using various methods. The current study provides new findings, showing that location-based and feature-based attentional surround suppression operate simultaneously to precisely demarcate the stimulus-of-interest from irrelevant distractors. It also demonstrates that the spatial extent of the suppressive surround varies by stimulus eccentricity, representing a correlation between neurons' RF size and a suppressive surround. It would be worthwhile to explore whether the present results can be replicated by varying the types of visual stimuli or even perceptual modalities. In that way, we could generalize these findings across different cognitive domains and show whether the center-surround distribution of attention is an overarching mechanism that mediates information processing in the human brain. In addition, investigating the characteristics of attentional surround suppression will have a significant impact on practical applications as well, such as UI layout development.

CHAPTER 4. Feature-based attentional surround suppression in motion processing: Psychophysics and neurophysiology

4.1 Abstract

Attention to a certain visual feature inhibits the processing of features nearby in the same feature map, leaving further features unaffected (Tsotsos, 1990; Tsotsos et al., 1995). Here, we investigated this “feature-based surround suppression” in motion processing as well as the neurophysiological mechanisms underlying it. First, we used random dot patterns (RDPs) and measured motion repulsion in human participants (Marshak & Sekuler, 1979) in two conditions, while they equally attended to two superimposed RDPs (divided attention), and while they attended to one RDP direction and ignored the other (focused attention). Focusing attention to one motion direction reduced motion repulsion relative to dividing attention. Importantly, this reduction was stronger when the two RDPs directions were 30~40 deg apart than when they were 10 deg (very similar directions) and 50 deg apart (very dissimilar directions). This feature-based surround suppression also interacted with bottom-up factors such as motion speed and color of the attended motion surface. Second, we recorded the activity of direction-selective neurons in areas MT/MST to two RDPs positioned inside their receptive fields. One RDP always moved in the neurons preferred direction (preferred pattern) while the other could move in one of 12 different directions spaced every 30 deg (tuning pattern). When animals attended to the tuning pattern or ignored both patterns (fixation condition), neurons showed Gaussian tuning curves as a function of the difference between the neuron’s preferred direction and the direction of the tuning pattern (Martinez-Trujillo & Treue, 2004; Treue & Martinez-Trujillo, 1999). However, when the animals attended to the preferred pattern, tuning curves were better described by the

difference of two Gaussian functions, showing the minimum activity half way between the preferred and anti-preferred directions. Our results show the existence of a feature-based surround suppression mechanism at both psychophysical and neurophysiological levels. This is best modeled by a difference-of-Gaussians formulation, incompatible with linear models of attentional modulation. Such a mechanism reflects the complex dynamics of cortical circuits that can be triggered by feature-based attention.

4.2 Introduction

Attention to a given feature facilitates and prioritizes its processing while impairing the processing of other unattended features. This feature-based attention has been corroborated by a substantial number of studies with various methodologies (Anllo-Vento & Hillyard, 1996; Hillyard & Anllo-Vento, 1998; Martinez-Trujillo & Treue, 2004; Maunsell & Treue, 2006; Sàenz et al., 2002, 2003; Treue & Martinez-Trujillo, 1999). Among those experimental works, some researchers focused on the processing fate of unattended visual features. One proposal is that the suppressive effect of feature-based attention on the unattended features is greater for features similar to the attended feature relative to features that are dissimilar, exhibiting a center-surround attentional modulation.

This center-surround attentional modulation, or feature-based surround suppression, was predicted by the Selective Tuning model of visual attention (ST, Tsotsos, 1990, 2011; Tsotsos et al., 1995). Empirical studies on human vision have supported the existence of the feature-based surround suppression in several feature domains (e.g., orientation (Tombu & Tsotsos, 2008), color (Bartsch et al., 2017; Fang et al., 2019; Störmer & Alvarez, 2014), action (Loach et al., 2008), motion direction (Ho et al., 2012; Wang et al., 2015), visual working memory (Kiyonaga

& Egner, 2016)). Recent studies on surround suppression in motion direction discrimination used an attentional cueing paradigm in which participants attended to the cued direction and chose a target dot pattern that contained coherent motion (Ho et al., 2012; Wang et al., 2015). They found that participants' performance decreased as the direction offset between the cue and target stimuli became greater, but performance gradually "rebounded" when the offset was larger than 90 deg, indicating that feature-based surround suppression occurred in the motion direction domain. However, the neurophysiological mechanism of feature-based surround suppression in the motion direction domain is unclear. Furthermore, feature-based surround suppression could be also influenced by bottom-up factors, such as motion speed which modulate signal strength just as selective attention does. The present study aims to demonstrate feature-based surround suppression in the motion direction domain using a different experimental paradigm than the previous studies and examines the effects of top-down attention and bottom-up motion factors on feature-based surround suppression. It also investigates the neurophysiological mechanism that gives rise to feature-based surround suppression by measuring activity of direction-sensitive cells in non-human primates.

In the first experiment, we measured feature-based surround suppression in the motion direction domain using motion repulsion. Motion repulsion is an overestimation of the directional difference between the two superimposed motion surfaces (Marshak & Sekuler, 1979) due to mutual inhibition of neurons that preferentially respond to different motion directions (Hiris & Blake, 1996; Marshak & Sekuler, 1979; Mather & Moulden, 1980; Perry & Fallah, 2014; Wilson & Kim, 1994). Studies demonstrated that feature-based attention modulates motion repulsion (Chen, Meng, Matthews, & Qian, 2005; Tzvetanov, Womelsdorf, Niebergall, & Treue, 2006). In particular, motion repulsion became weaker when attention was selectively directed to one of the two superimposed motion directions than when they were viewed equally, suggesting that feature-based attention reduced motion repulsion by suppressing mutual inhibition (Chen et al., 2005).

The magnitude of motion repulsion also depends on bottom-up factors (Braddick, Wishart, & Curran, 2002; Curran & Benton, 2003; Perry, Tahiri, & Fallah, 2014). For instance, motion repulsion is reduced when motion speed increases probably because faster motion speed enhances motion signal strength, thus, improving perceptual judgment (c.f., Palmer, Huk, & Shadlen, 2005). Based on these findings, we measured motion repulsion under different attention and speed conditions while the directional difference between two motion surfaces varied. We hypothesized that motion repulsion would be reduced to a greater degree when one of the two *similar* directions is attended than when they are dissimilar because of feature-based surround suppression. We also anticipated that this decrement of motion repulsion would be more apparent when motion speed is faster, indicating an interaction between feature-based attention and motion speed. As hypothesized, our results supported the feature-based surround suppression in motion processing and the interaction between top-down attention and bottom-up factors.

Our second experiment focused on the neurophysiological mechanisms of feature-based surround suppression. We measured the tuning curves of direction-selective neurons in monkey's middle temporal (MT) and medial superior temporal (MST) visual areas by placing two moving random dot patterns (RDPs) within a neuron's receptive field (RF) – one RDP always moved in a neuron's preferred direction and the other moved in one of twelve directions. We instructed animals to direct attention either to one or the other pattern. We observed that when a neuron's preferred direction was attended, the tuning curve of the neuron exhibited a center-surround profile: Neuronal response to direction at about 90 deg away from the preferred one was more inhibited than to direction 180 deg from the preferred one (anti-preferred).

4.3 Experiment 1: Psychophysics

4.3.1 Methods and materials

4.3.1.1 Participants

Fourteen naïve participants (5 men, 9 women), between the ages of 20 and 30 years completed the experiment. They had normal or corrected-to-normal visual acuity, and normal color vision. Written informed consent was obtained from all participants and they were paid for their participation ($30 CAD). The research was approved by York University's Human Participants Review Committee.

4.3.1.2 Apparatus and stimuli

Experiments were conducted in a dark room. Participants sat 57 cm from a CRT monitor (21" View Sonic G225f, 1280x1024, 85 Hz) and their heads were stabilized on a head and chin rest (Headspot, UHCOtech, Houston, TX). Participants wore an infrared eye tracker (Eyelink II, SR Research, 500 Hz, Mississauga, ON, Canada) monitoring the left eye position. Random dot patterns (RDPs) were created through MATLAB (MathWorks, Natick, MA) and the Psychophysics Toolbox (Brainard, 1997; Pelli, 1997). Experimental control was maintained by Presentation (Neurobehavioral Systems, Berkeley, CA).

An annular RDP consisted of two superimposed motion surfaces (RDP size = 15 dva (degree in visual angle) in diameter, inner aperture size = 6 dva in diameter, dot size = 0.15 dva, 75 dots per surface). The directions of the motion surfaces changed every trial and the dots in each motion surface moved in the same direction (100% coherent). Directional difference between the two surfaces systematically varied by 10~50 deg (10 deg step). Motion speed was either 3 deg/sec or 6 deg/sec and both motion surfaces moved in the same speed. Dots in one motion surface were red (luminance: 24.67 cd/m^2) and those in the other surface were green (24.64 cd/m^2) to make participants easily segregate them, without affecting direction repulsion (Perry & Fallah, 2012).

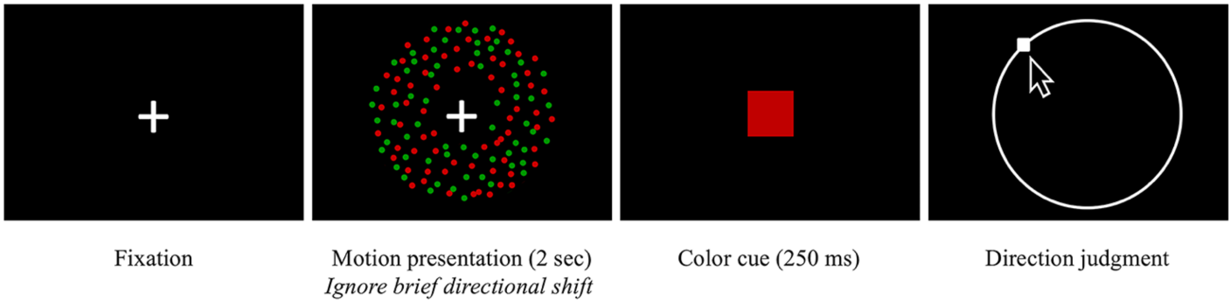
4.3.1.3 Procedure

Figure 4.1 illustrates the experimental procedure. There were two conditions – divided and focused attention. In the divided attention condition, typical motion repulsion was measured. The RDPs were presented for 2 sec once participants fixated a white cross centered on a screen for 200 ms. Participants had to maintain the fixation until the RDPs disappeared, otherwise an error message was presented, and the trial was randomly interleaved in the remaining trials. Participants were asked to view both motion surfaces equally to perceive their directions. During motion presentation, a brief directional shift (100 ms) on either motion surface could randomly occur in 80% of trials, and then it went back to the original direction. The amount of shift was randomly selected from the range between 30 and 40 deg when motion speed was 3 deg/sec, and between 20 and 30 deg when motion speed was 6 deg/sec to equalize the perceptual strength of directional shifts across different motion speed conditions. A shift could occur from 650 to 1100 ms after RDPs onset. Participants were asked to ignore this directional shift while they viewed the RDPs. After motion presentation, a color cue (either red or green) appeared for 250 ms to indicate which motion surface was the target in that trial. Each color cue was presented equally throughout the experiment. Participants reported the motion direction of the target surface by clicking along a white circular outline.

In the focused attention condition, after maintaining central fixation for 200 ms, a color cue appeared before the RDPs were presented to indicate which motion surface should be attended. Participants were required to attend only to the cued motion surface (target) while ignoring the other surface. To make sure whether they selectively attended to the target surface, they clicked the right mouse button within a 1 sec after the onset of a brief directional shift. The directional shift could occur only on the target surface. If there was no shift, participants did not respond and waited until they viewed the white circular outline. If participants missed the shift, responded too late, or made a false alarm, an error message was presented, and the trial was discarded. They reported the motion direction of the target surface only when the attention task was successful performed.

The attention conditions were blocked, and participants performed both conditions twice in a random order. At the beginning of each attention condition, participants were given 10 practice trials whose data were not used and then, they performed 100 trials as the main experiment (200 trials for each attention condition, in total). There was a mandatory break time after every 25 trials and participants could have extra break time if they wanted.

(A) Divided attention



(B) Focused attention

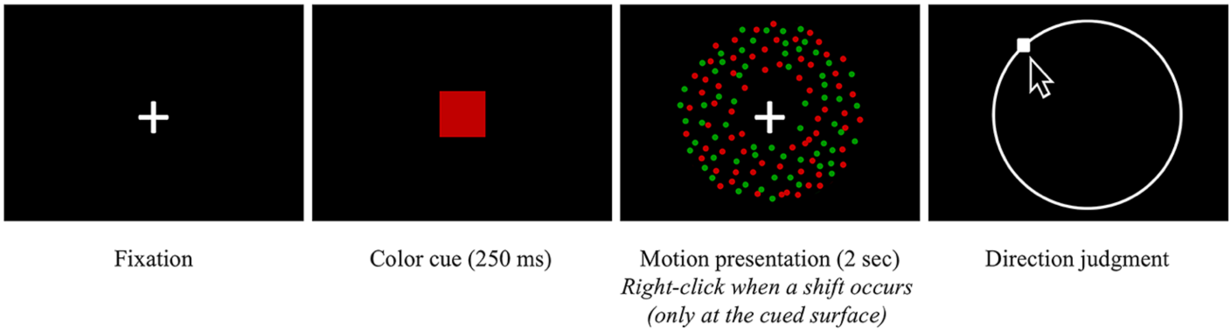


Figure 4.1. Experimental conditions and procedure of Experiment 1.

(A) In the divided attention condition, participants attended to both motion surfaces equally, ignoring any directional shift. A color cue was presented after RDPs disappeared and then, participants reported a direction of the cued motion surface by clicking along a white circular outline. (B) In the focused attention condition, a color cue indicated which motion surface participants had to attend. Participants were asked to detect a directional shift on the cued surface if occurred, and then reported motion directions of this surface as in the divided attention condition.

4.3.1.4 Data analysis

We first sorted participants' direction judgment as Perry and Fallah (2012) did to reduce variability in data. A correct direction judgment should fall within a range that extended from halfway between the two motion directions to 45 deg away from the motion direction of the target surface. Then, motion repulsion, defined by the difference between the reported (perceived) and the actual motion direction, was calculated. Since participants performed an additional attention task in the focused attention condition, only trials in which both directional shift detection and motion direction judgment were successful were included in the analysis. Motion repulsions in the two attention conditions were compared to quantify the attentional modulation.

4.3.2 Results

First, we analyzed how feature-based attention and motion speed influenced participants' direction judgment accuracy – i.e., whether reported directions fell within a valid response range (Figure 4.2A, see Data analysis for details). Direction judgment accuracy in the focused attention condition was measured only if the attentional task (detecting a brief directional shift) was successfully performed. An analysis of variance (ANOVA) demonstrated significant main effects of the attention conditions ($F(1, 13) = 6.683$, $p = .023$) and motion speed ($F(1, 13) = 22.443$, $p < .001$) on directional judgment accuracy. Post-hoc pairwise comparisons with Bonferroni correction showed that direction judgment was more accurate in the focused attention condition than in the divided attention condition ($M_{diff} = 4.59\%$, SE = 1.78%). Since participants tracked only one motion direction in the focused attention condition, direction judgments had been easier. This pattern was significant when motion speed was 3 deg/sec ($M_{diff} = 5.55\%$, SE = 2.44%, $t(13) = 2.269$, $p = .041$), and it was at a trend level when motion speed was 6 deg/sec ($M_{diff} = 3.78\%$, SE = 7.03%, $t(13) = 2.014$, $p = .065$). Direction judgment accuracy was higher when motion speed was faster (M_{diff} = 9.03%, SE = 1.91%), and this result was evident in both attention conditions (divided: M_{diff} = 9.85% SE = 1.69%, *t*(13) = 5.822, *p* < .001, focused: M_{diff} = 8.09%, SE = 2.73%, *t*(13) = 2.97, *p* = .011). Color of the target motion surface did not influence directional judgment (*F*(1, 13) = .051, *p* = .825). All interactions between the variables were not significant.

Figure 4.2B shows the mean accuracy of directional shift detection in the focused attention condition. The mean accuracy was 83.14% (SD 13.45%) and 82.93% (9.79%) when motion speed was 3 deg/sec and 6 deg/sec, respectively. They did not statistically differ (*p* = .915), indicating that difficulty of the attentional task was well controlled across different speed conditions. When it was broken down by the color of the attended motion surface (target), the main effect of the target surface color on directional shift detection was significant (*F*(1, 13) = 13.995, *p* = .002). Detecting directional shifts was better when the target motion surface was in red than when it was in green (M_{diff} = 3.21%, SE = .86%). The performance advantage in the red motion surfaces likely represents strong attentional guidance of red (Pomerleau, Fortier-Gauthier, Corriveau, Dell'Acqua, & Jolicœur, 2014; Tchernikov & Fallah, 2010). This difference was at a trend level when motion speed was 3 deg/sec (M_{diff} = 2.86%, SE = 1.40%, *t*(13) = 2.038, *p* = .062) and was significant when motion speed was 6 deg/sec (M_{diff} = 3.57%, SE = 1.28%, *t*(13) = 2.794, *p* = .015). Interaction between motion speed and the color of the target motion surface was not significant (*F*(1, 13) = .12, *p* = .734).

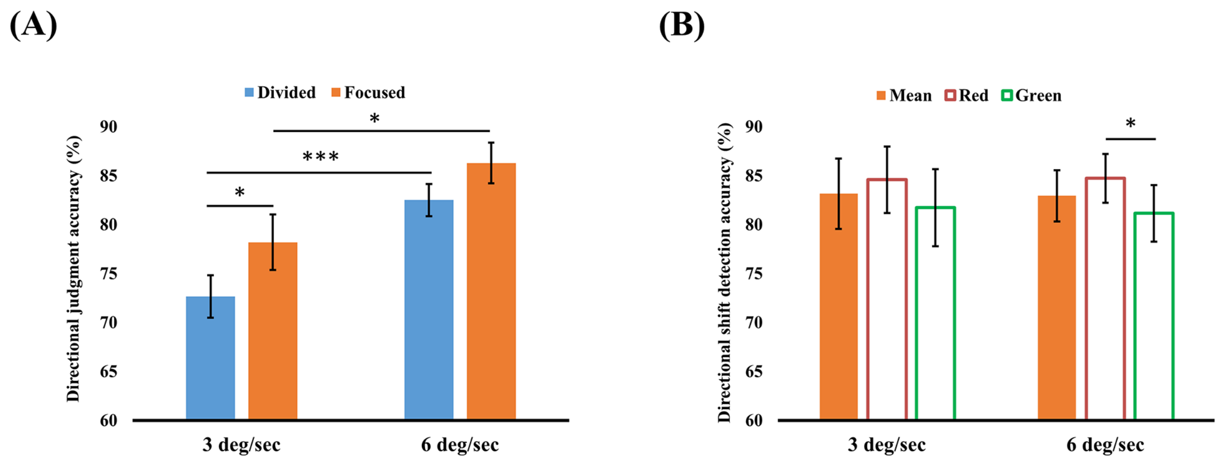


Figure 4.2. Direction judgment accuracy.

(A) Direction judgment was more accurate in the focused attention condition, and when motion speed was faster. (B) Detection of directional shifts (only in the focused attention condition) was not affected by motion speed. However, direction discrimination was more accurate when the color of the attended motion surface (target) was in red. Error bars indicate SEM. * $p < .05$, *** $p <. 001$

We further analyzed motion repulsion of the trials in which participants correctly discriminated motion direction. When motion speed was 3 deg/sec (Figure 4.3A), the main effect of the attention condition on motion repulsion was not significant ($F(1, 13) = 2.224$, $p = .16$), whereas motion repulsion significantly varied depending on the directional difference between the two motion surfaces ($F(4, 52) = 2.715$, $p = .04$). Importantly, motion repulsion was influenced by the interaction between the attention condition and directional difference ($F(4, 52) = 3.958$, $p = .007$). Motion repulsion was significantly reduced in the focused attention condition when the directional difference was 30 deg ($M_{diff} = 4.40$ deg, SE = 1.41 deg, $p = .008$) and this effect was marginal when directional difference was 40 deg ($M_{diff} = 2.81$ deg, SE = 1.34 deg, p

= .056). Reduction of motion repulsion at around 30~40 deg difference suggests that feature-based surround suppression played a role by inhibiting an unattended motion direction that represented closely to the attended motion direction.

The amount of motion repulsion was different depending on the color of the target motion surface ($F(1, 13) = 28.589$, $p < .001$; Figure 4.3B). Motion repulsion was smaller when the target motion surface was green than it was red ($M_{diff} = 7.07$ deg, SE = 1.32 deg). Although the attention condition did not affect motion repulsion ($F(1, 13) = 2.07$, $p = .174$), there was a significant interaction between the target motion surface color and the attention condition ($F(1, 13) = 9.199$, $p = .01$). We conducted an ANOVA separately for each target surface color to investigate the effects of colors more specifically. When the target motion surface was *red*, motion repulsion was not affected by the attention conditions ($F(1, 13) = 1.188$, $p = .296$) but by the directional difference between the two motion surfaces ($F(4, 52) = 7.522$, $p < .001$). The interaction between these two variables was not significant ($F(4, 52) = .8$, $p = .531$). Hence, there was no evidence for reduction of motion repulsion by feature-based attention and feature-based surround suppression. On the other hand, when the target motion surface was *green*, the main effect of the attention conditions was significant ($F(1, 13) = 22, 631$, $p < .001$), indicating that feature-based attention reduced motion repulsion ($M_{diff} = 3.96$ deg, SE = .83 deg) as previously reported (Chen et al., 2005). Motion repulsion was not modulated by directional difference ($F(4, 52) = .802$, $p = .529$) but the interaction between the attention conditions and directional difference was significant ($F(4, 52) = 6.699$, $p < .001$). Motion repulsion significantly decreased in the focused attention condition when directional difference was around at 20~40 deg (20 deg: $p = .051$ (marginal), 30 deg: $p < .001$, and 40 deg: $p = .005$). This indicates that the feature-based surround suppression in the 3 deg/sec condition was mainly derived from the trials where the target motion surface was green.

Change in motion speed from 3 deg/sec to 6 deg/sec affected motion repulsion ($F(1, 13) = 54.343, p < .001$) – faster motion speed (6 deg/sec) reduced motion repulsion ($M_{diff} = 3.47$ deg, SE = .47 deg) as previously reported (Braddick et al., 2002; Curran & Benton, 2003; Perry et al., 2014). When motion speed was 6 deg/sec (Figure 4.3C), the attention condition significantly modulated motion repulsion ($F(1, 13) = 19.557, p = .001$), demonstrating that motion repulsion was generally smaller in the focused attention condition ($M_{diff} = 3.18$ deg, SE = .72 deg). Directional difference ($F(4, 52) = 1.652, p = .175$) and the interaction between the attention condition and directional difference ($F(4, 52) = 1.049, p = .391$) did not significantly affect motion repulsion. Motion repulsion was significantly smaller in the focused attention condition at all directional differences, except at 10 deg difference, and the difference in motion repulsion (divided - focused) did not statistically vary across directional differences. This result made it difficult to specify the effect of the feature-based surround suppression (i.e., greater reduction of motion repulsion when motion directions are similar). We broke down the data by the color of the target motion surface to see whether there were unexpected interactions between the surface color and the other variables that gave rise to the present result.

Motion repulsion was modulated by the color of the target surface ($F(1, 13) = 14.491, p = .002$), showing smaller motion repulsion in *green* surfaces ($M_{diff} = 3.07$ deg, SE = .76 deg; Figure 4.3D). In addition, attention condition significantly affected motion repulsion ($F(1, 13) = 16.507, p = .001$). Motion repulsion was smaller in the focused attention condition than in the divided attention condition ($M_{diff} = 3.07$ deg, SE = .76 deg). The interaction between the target surface color and the attention condition was not significant ($F(1, 13) = .865, p = .369$). Unlike when motion speed was 3 deg/sec, the attention condition modulated motion repulsion even when the color of the target motion surface was red ($F(1, 13) = 5.177$, $p = .04$). Post-hoc pairwise comparisons showed that motion repulsion was reduced in the focused attention condition (M_{diff} = 2.22 deg, SE = .97 deg) and it was significant when the directions of RDKs differed by 40~50 deg (40 deg: $p = .036$, 50 deg: $p = .02$). When green was the color of the target motion surface, feature-based attention also significantly reduced motion repulsion ($F(1, 13) = 8.148$, $p = .014$; divided vs. focused: M_{diff} = 3.93 deg, SE = 1.38 deg). The reduction was at the trend level at 10~20 deg difference (10 deg: $p = .053$, 20 deg: $p = .085$) and became significant at 30~40 deg difference (30 deg: $p = .008$, 40 deg: $p = .021$). Motion repulsion patterns for green motion surfaces were consistent across different motion speeds, whereas feature-based attention played a role only if motion speed was faster when the target surface was red. Therefore, the almost universal decrement in motion repulsion in the 6 deg/sec condition might have resulted from the interaction between the color of the target motion surface and motion speed.

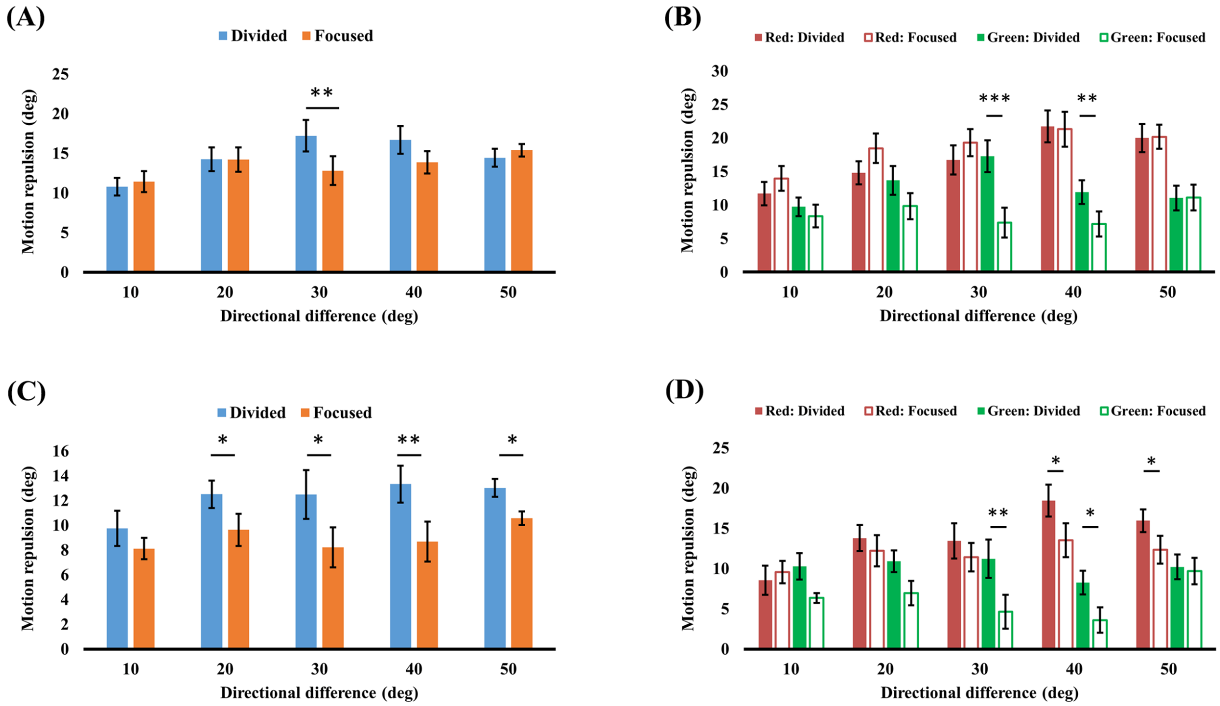


Figure 4.3. Motion repulsion in different experimental conditions.

When motion speed was 3 deg/sec, (A) the amount of motion repulsion between the two attention conditions significantly differed when directional difference was 30 deg. The same effect was marginal when directional difference was 40 deg ($p = .056$). (B) Attentional modulation on motion repulsion was evident only when the color of the target motion surface was green. When motion speed was 6 deg/sec, (C) Motion repulsion was reduced in the focused attention condition and it was true across all directional differences, except at 10 deg difference. (D) Unlike when motion speed was slower, attentional modulation on motion repulsion was observed regardless of the color of the target motion surface. Error bars indicate SEM. * $p < .05$, ** $p < .01$, *** $p < .001$

4.4 Experiment 2: Neurophysiology

4.4.1 Methods and materials

4.4.1.1 Apparatus and stimuli

We recorded the responses of direction-selective neurons in areas MT and MST of two male macaque monkeys in different task conditions. After initial training, a head post, a scleral search coil (Judge, Richmond, & Chu, 1980) to monitor eye position (Robinson, 1963), and a recording chamber were implanted in each animal. A custom computer program running on an Apple Macintosh PowerPC controlled the stimulus presentations, monitored eye position and behavioral responses during the experiments, and recorded the behavioral and neuronal data. The experiments reported in this study were conducted according to local and national rules and regulations and were approved by the Regierungspraesidium Tuebingen (Germany).

Random dot patterns (RDPs) consisted of small bright dots with a density of 5 dots/dva^2 within a stationary circular virtual aperture on a dark computer monitor. The luminance of the dots was 55 cd/m^2 and the viewing distance was 57 cm. The diameter of the aperture varied from about 1° to 12° depending on the size of each neuron's RF so that the stimulus did not exceed the boundaries of the classical RF. Movement of the dots was created by displacement of each dot by the appropriate amount at the monitor refresh rate of 75 Hz. We measure the speed and direction tuning of the neuron by online display of the tuning curves and chose the speed at which the neuron produced the strongest response (Treue & Martinez-Trujillo, 1999). In every trial we presented two RDPs of equal size inside the neurons RF. One pattern always moved in the neuron’s preferred direction, estimated by online display of the responses to a single RDP inside the RF moving in different directions (see Treue & Martinez-Trujillo (1999)). The other pattern could move in one of twelve different directions spaced every 30 deg.

4.4.1.2 Recordings

Extracellular recordings from the left hemisphere were conducted using tungsten microelectrodes (impedance 0.5–2 mΩ, Microprobe and FHC). Electrodes were lowered through a recording chamber implanted on top of the parietal bone until reaching the approximate location of MT/MST. Single units were isolated with a window discriminator, and eccentricity, direction-selectivity, and position of the electrode within the recorded area were determined. We recorded only from those neurons showing clear direction-selectivity during initial mapping (see Treue & Martinez-Trujillo (1999) for methods).

4.4.1.3 Procedure

The animals were trained to attend to one of the RDPs (target) and ignore the other RDP (distractor), while directing gaze to a fixation point. Figure 4.4A illustrates a sample stimulus display of the experiment. One of the RDPs always moved in the neurons preferred direction (preferred pattern) and the other could move in one of 12 different directions from trial to trial (tuning pattern, in steps of 30 deg). There were three different experimental conditions depending on which feature on the display was attended (Figure 4.4B). The animals were cued to attend to 1) the preferred pattern (attend-preferred condition), or 2) tuning pattern (attend-tuning condition). The target appeared for a short time period and then disappeared to re-appear together with the distracter. The task for the animals was to sustain attention to the attended pattern and release a lever within a 450 ms response window after the pattern changed direction. The animals had to ignore changes in the direction of the unattended pattern, which happened in 50% of the trials. In this task, the animals scored hits in more than 85% of the trials (animal F

86% and animal M 87%), indicating that they indeed allocated attention to the attended pattern and ignored the unattended pattern. 3) In a third condition (fixation condition), the animals attended to the fixation point and detected its color change while ignoring both RDPs in the periphery.

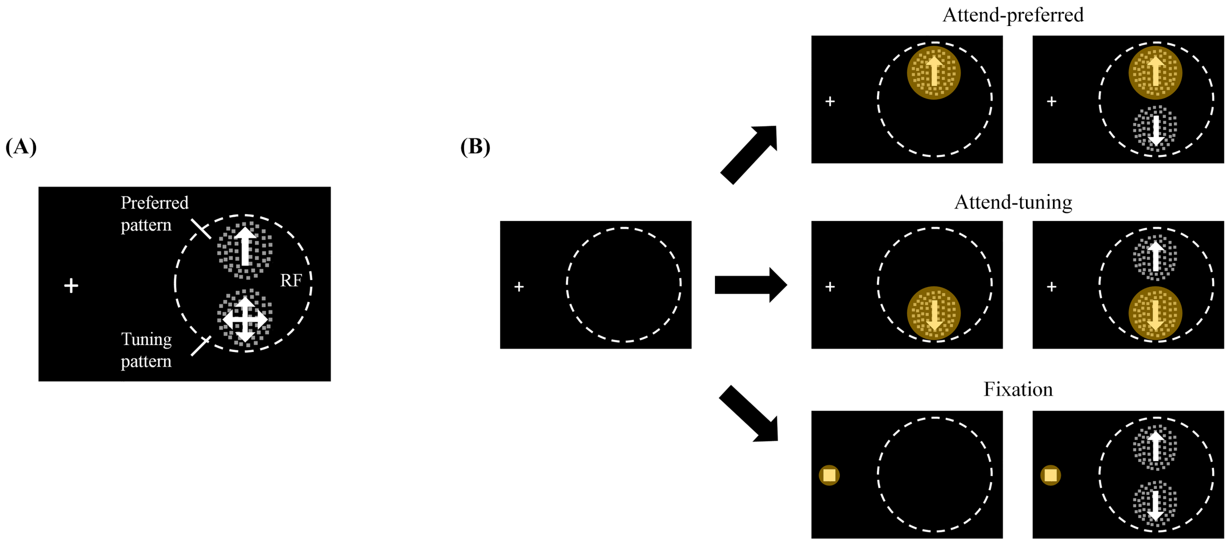


Figure 4.4. Stimuli and conditions in Experiment 2

(A) Experiment 2 stimulus configuration. The preferred pattern which always moved to the neuron's preferred direction and the tuning pattern which could move to one of 12 directions were presented within a neuron's RF. The fixation point was presented on the left side of the display. (B) Three attention conditions. While an animal looked at the fixation point, it was cued to attend to either the preferred (attend-preferred) or tuning (attend-tuning) patterns and detected a directional shift on the attended pattern, or to the fixation point that changed into a small square and the animal detected a change in the square color while ignoring the RDPs (fixation condition).

4.4.1.4 Data analysis

We recorded neuronal responses in the three attention conditions. Neurons were recorded from if they showed putative direction selectivity explored online using a moving bar (Martinez-Trujillo & Treue, 2004; Treue & Martinez-Trujillo, 1999). We computed average firing rates during the interval from 200 to 1200 milliseconds after the onset of the two patterns, as a function of the tuning pattern's direction relative to the direction of the preferred pattern. The responses of each neuron were normalized to the response in the attend-preferred condition and then averaged across neurons. The responses of both MT and MST neurons were pooled since the direction selectivity and tuning curve profiles were very similar between the two areas. Repeated measures analysis of variance (ANOVA) and paired samples t-tests on the average normalized neuronal response were conducted to examine how neuronal responses varied depending on experimental conditions.

We fitted two different models, the single Gaussian and the difference of Gaussians (DoG) to neuronal responses, using the MATLAB curve fitting toolbox (Mathwork Inc., USA). We used the Akaike Information Criterion (AIC, Akaike, 1974) to assess the relative quality of each model fit which penalizes model complexity (lower AIC value indicates better fit).

4.4.2 Results

We recorded the responses of 78 direction-selective neurons in areas MT and MST of two male macaque monkeys to moving RDPs in different task conditions. Figure 4.5A shows the average normalized neuronal response of each attention condition. A repeated measures ANOVA showed that the neuronal response was significantly modulated by the attention conditions (*F*(2,

122) = 47.058, *p* < .001). Post-hoc multiple comparisons with Bonferroni correction demonstrated that the response was greatest in the attend-preferred condition than in the other conditions (all *ps* < .001). On the other hand, no significant difference was observed between the attend-tuning and the fixation conditions (*p* = 1). The main effect of the directional difference between the preferred and the tuning patterns was significant (*F*(11, 671) = 45.768, *p* < .001). Neuronal response was greatest when both preferred and tuning patterns moved in the same direction (0 deg difference, all *p*s < .001) and then, it gradually decreased as directional difference between the two motion patterns increased. The interaction between attention conditions and directional difference was also significant (*F*(22, 1342) = 13.561, *p* < .001). To shed light on the nature of this interaction, we examined the profile of feature-based attentional modulation under different attention conditions. We collapsed neuronal responses when the absolute directional difference between the preferred and the tuning patterns was the same (e.g., +30 deg and -30 deg) because the profile of the average normalized neuronal response was almost symmetrical in all attention conditions (Figure 4.5B).

Neuronal responses in the fixation condition demonstrated a neuron's default feature tuning when attention was directed to the fixation point, not to RDPs. The response was maximal when the tuning pattern moved in the neuron's preferred direction because two patterns moving in the neuron's preferred direction were shown inside the RF. Here is important to consider the two patterns were positioned within the excitatory RF and did not extend into the purely inhibitory surround. This was assessed during initial mapping of the RF (see methods). The response was minimal when the tuning pattern moved in the neuron's anti-preferred direction. There was a monotonic decrement in response as a function of the difference between the neuron's preferred direction and the direction of the tuning pattern.

Neuronal response in the attend-tuning condition also monotonically decreased as the directions between the preferred and the tuning patterns became dissimilar. In this condition, neuronal responses were greater than those in the fixation condition when the direction of the tuning pattern was closer to the neuron's preferred direction (at 0 and 30 deg differences, all *p*s < .001). On the other hand, responses in the attend-tuning condition were lower than those in the fixation condition when the directions of the tuning pattern became similar to the neuron's anti-preferred direction (150 deg difference, M_{diff} = -.06, SE = .03, *t*(76) = -1.942, *p* = .056 (marginal); 180 deg difference, M_{diff} = -.07, SE = .04, *t*(76) = -2.101, *p* = .039). In the attend-preferred condition, the maximal neuronal response was also observed when the directions of the preferred and tuning patterns were the same. However, responses were lowest when the tuning pattern moved in directions ~90 deg away from the preferred direction, not when the tuning pattern moved in the anti-preferred direction. The minimum neuronal response was significantly lower than the maximum response at 0 deg difference (90 vs. 0 deg, M_{diff} = -.22, SE = .02, *t*(76) = -8.78, *p* < .001) and it was also lower than the response for the anti-preferred direction (90 vs. 180 deg, M_{diff} = -.04, SE = .02, *t*(76) = -2.272, *p* = .026). This result indicates that in the attend-preferred condition a suppressive surround was formed around the neuron's preferred direction.

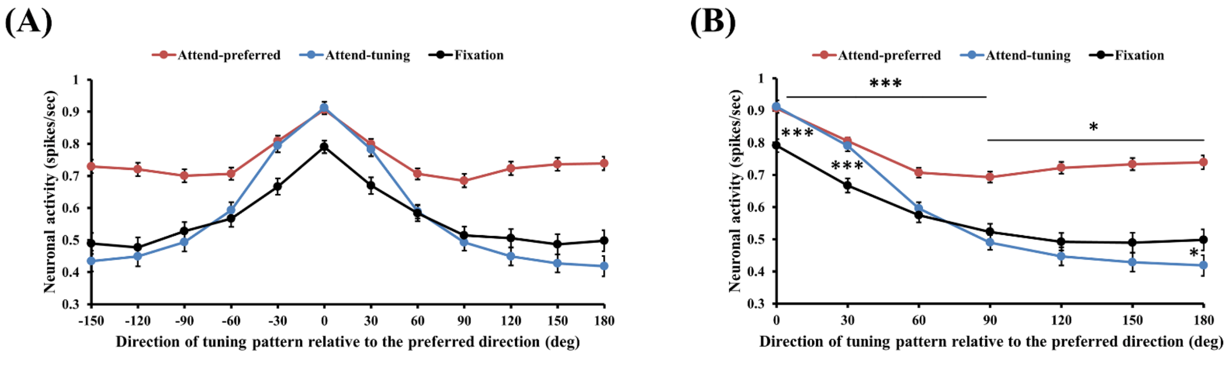


Figure 4.5. Average normalized neuronal response.

(A) Average normalized neuronal response in each attention condition. (B) Normalized neuronal responses averaged across the same (absolute) directional difference. Responses in the attend-tuning and the fixation conditions monotonically decreased as the directions between the preferred and tuning patterns became dissimilar. Response in the attend-tuning condition was larger than that in the fixation condition when the direction of the tuning pattern was closer to the neuron's preferred direction (0~30 deg). This pattern was reversed when the direction of the tuning pattern was similar to the neuron's anti-preferred direction. In the attend-preferred condition, however, response was lowest when the direction of the tuning pattern was approximately 90 deg away from the preferred direction. Then, the response increased as directional difference increased (~180 deg difference), supporting the feature-based surround suppression. Error bars indicate SEM. * $p < .05$, ** $p < .01$, *** $p < .001$

We further evaluated the neuronal response profiles with non-linear curve fitting and model comparisons. First, we fitted the single Gaussian and the DoG models to the average normalized neuronal response (population response) in each attention condition. The difference between these two models is that the single Gaussian cannot explain feature-based surround suppression, while the DoG can. In the attend-preferred condition (Figure 4.6A), the DoG (adjusted $R^2 = .990$) model explained the population response better than the single Gaussian model (adjusted $R^2 = .886$), suggesting that the neuronal response fits feature-based surround suppression. In the other attention conditions (Figure 4.6B and 4.6C), both models could explain the response profiles similarly well (all adjusted R^2, attend-tuning condition: single Gaussian (.994), DoG (.999); fixation condition: single Gaussian (.957), DoG (.987)). The results support that feature-based surround suppression was only apparent in the attend-preferred condition.

We also compared different model fits at the individual neuron level. Our curve fitting algorithm sometimes failed to fit a given individual neuronal response because the fit did not converge due to variability in the data, or because some data points were missing. We included individual neurons in the analysis only if both models successfully fitted their responses and then, compared different model fits of the same neuronal response (i.e., pairwise comparison): 69 neurons in the attend-preferred condition, 53 neurons in the attend-tuning condition, and 59 neurons in the fixation condition were included in the analysis. As a goodness-of-fit measure, we computed the AIC for each model fit (see data analysis). In the attend-preferred condition (Figure 4.6D), the AIC was greater for the single Gaussian model than the DoG model (M_{diff} = 2.22, SE = 1.03, *t*(68) = 2.148, *p* = .035), meaning that the latter model explained the data better. Conversely, the single Gaussian model fitted the data better in the attend-tuning condition (Figure 4.6E), demonstrating a smaller AIC value than the DoG model did (M_{diff} = -2.25, SE = .92, *t*(52) = -2.448, *p* = .018). In the fixation condition (Figure 4.6F), there was no significant difference in AIC between the two models (M_{diff} = .53, SE = .88, *t*(58) = .605, *p* = .548).

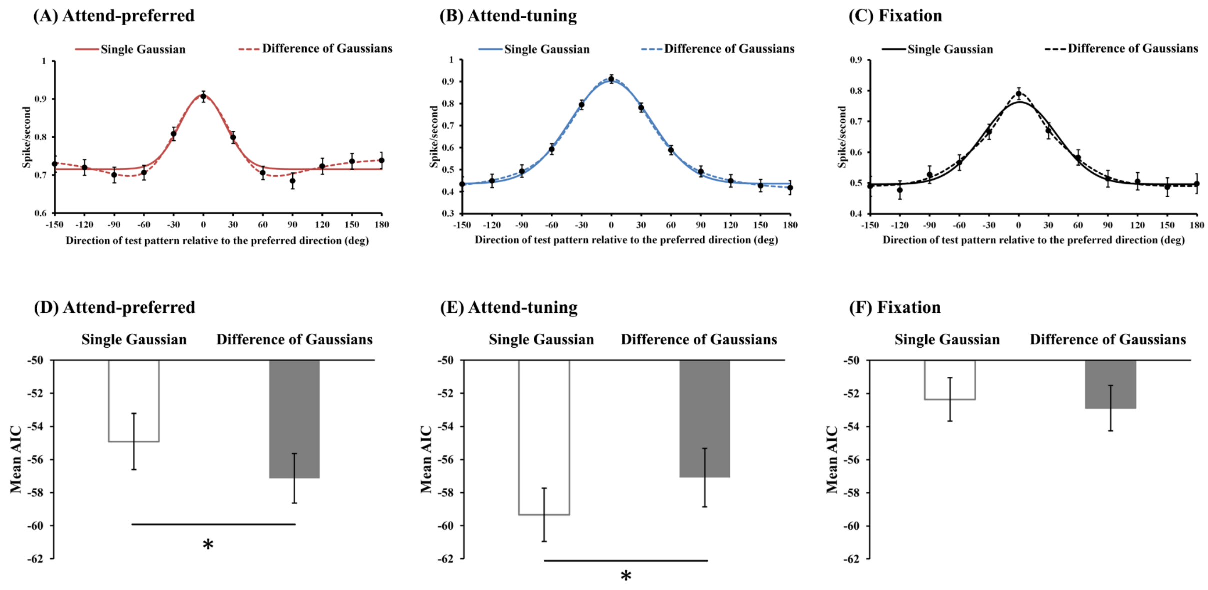


Figure 4.6. Model comparisons at the population and individual neuron levels.

(Upper panels) Different model fits to the average normalized neuronal response in the (A) attend-preferred, (B) attend-tuning, and (C) fixation conditions. Solid curve indicates the single Gaussian model fit and dashed curve indicates the difference of Gaussians (DoG) model fit. DoG model fitted better the response profile in the attend-preferred condition. Both models explained well the response profiles in the other conditions. (Lower panels) Akaike Information Criterion (AIC) for each model fit was measured at individual neuron level. Smaller AIC value means better model fit. White and gray bars indicate the mean AIC of the single Gaussian and the DoG models, respectively. (D) DoG model fitted the responses better in the attend-preferred condition. (E) Single Gaussian model fitted the responses better in the attend-tuning condition. (F) No significant difference between the two models was observed in the fixation condition. Error bars indicate SEM. * $p < .05$

4.5 General discussion

The present study demonstrated feature-based surround suppression in the motion direction domain at both behavioral and neurophysiological levels. In the behavioral study (Experiment 1), we observed feature-based surround suppression reduces motion repulsion when the two superimposed random dot motions moved in similar directions. At the neurophysiological level (Experiment 2), we showed that feature-based attention modulates the motion direction tuning curves of MT/MST neurons producing a similar center-surround effect. These finding confirm a main prediction of the Selective Tuning model (Tsotsos, 1990, 2011; Tsotsos et al., 1995).

Previous studies have found the attentional surround suppression when directing attention to a spatial location, demonstrating a ring of inhibition around the attended location (Cutzu & Tsotsos, 2003; Hopf et al., 2006, 2010). This non-linear effect of spatial attention is not predicted by models in which the attentional enhancement of attended locations degrades monotonically as a function of the distance from the attentional focus (Downing, 1988; Laberge & Brown, 1986; LaBerge & Brown, 1989). The current findings, along with the previous reports (Bartsch et al., 2017; Fang et al., 2019; Ho et al., 2012; Kiyonaga & Egner, 2016; Loach et al., 2008; Störmer & Alvarez, 2014; Tombu & Tsotsos, 2008; Wang et al., 2015), indicate the same surround suppressive mechanism does exist in feature space.

We also questioned whether top-down feature-based attention would interact with bottom-up motion factors. The results demonstrated that the feature-based attentional modulation varied depending on motion speed and the color of the target motion surface. With slower motion speed (3 deg/sec), the feature-based surround suppression of motion direction was apparent only when the color of the target motion surface was green. When motion speed was faster (6 deg/sec), however, feature-based surround suppression was observed for both green and red target motion surfaces, implying the interplay between top-down feature-based attention and bottom-up factors. The present results appear to be inconsistent with the previous report that the color of motion surfaces does not influence motion repulsion (Perry & Fallah, 2012). However, in their study, motion repulsion was not measured separately by the motion surface color, hence, the effect of the surface color might have not been observed. Greater motion repulsion in red surfaces suggests that these surfaces are more strongly affected by inhibition. Such a result is consistent with a recent study that showed red facilitated response inhibition compared to green (Blizzard, Fierro-Rojas, & Fallah, 2017). When the speed of the surfaces was faster, producing more motion strength, it is likely that the balance in mutual inhibition between the two colored surfaces was more even, which then allowed for feature-based attentional modulation and surround suppression to be seen on the red surface.

In the second experiment, we examined the neurophysiological mechanisms of feature-based surround suppression in monkey MT/MST neurons. We had a critical condition in the present study which has not been addressed in the previous literature (Martinez-Trujillo & Treue, 2004; Treue & Martinez-Trujillo, 1999). In this condition, a direction tuning curve was measured while attention was always directed to a neuron’s preferred motion direction (attend-preferred). The minimum neuronal response was observed when the direction of the tuning pattern was 60~90 deg away from the preferred direction, and then it rebounded toward the anti-preferred direction. This response dip indicates that there was a suppressive surround around the attended feature (i.e., preferred direction) so that the nearby unattended features were inhibited. Furthermore, this direction tuning curve was explained better by the difference-of-Gaussians model which represents the surround suppression than by the single Gaussian model. Together, the present finding agrees with a prediction of the Selective Tuning model on feature-based surround suppression.

One puzzling finding in this experiment was that the feature distance where surround suppression occurred was different from the one observed in Experiment 1, using behavioral measurements. Feature-based surround suppression in our behavioral study happened when the directional difference between two motions was smaller (30~40 deg) than in the neurophysiological study (60~90 deg). The previous behavioral studies reported broader feature-based suppressive surrounds in the motion direction domain than those in the other feature domains (maximum suppression around at around 90 deg difference, Ho et al., 2012; Wang et al., 2015) and this range is similar to what we found in Experiment 2, probably due to the broad direction tuning curve of MT neurons (Albright, 1984; Britten & Newsome, 1998; Snowden, Treue, & Andersen, 1992). There are several possible explanations for the discrepancy between the present two experiments. It may be that a narrower surround suppression profile in our Experiment 1 could be due to the nature of motion repulsion. Motion repulsion is typically attenuated as the two superimposed motions move in more dissimilar directions, thus, one would not find the feature-based surround suppression when the directional difference becomes greater. In addition, we used colored, superimposed RDPs in Experiment 1, while we used RDPs of the same color and spatially separated in Experiment 2. These differences in experimental designs are likely to influence the profile of feature-based surround suppression.

One interesting question for future studies is what would be the exact neurophysiological mechanisms that allows for feature-based surround suppression. There is no specific answer to this question, but one clear prediction is that it must occur via inhibitory interactions between neurons encoding the attended and unattended features. It may be possible that attention triggers inhibitory interactions between populations of neurons encoding the different features of stimuli that are visually available via specific interneuron types. A prediction is that the strength of such interactions will follow the non-linearity we observed in our study. One appealing idea is that such non-linearity may be due to the specific reach of horizontal inhibitory projections of the different interneuron types that vary as a function of the distance from the cell bodies. For example, one can conceive that the maximum strength projections may reach groups of principal cells tuned for directions that are intermediate between the preferred and anti-preferred direction. This hypothesis, however, needs to be tested in future studies.

4.6 Conclusions

To conclude, we found evidence for feature-based surround suppression in the motion direction domain at both behavioral and neuronal levels. Behaviorally, feature-based surround suppression interacted with other bottom-up factors such as motion speed and the color of the target motion surface. Direction-selective neurons demonstrated a center-surround tuning curve when attention was directed to their preferred direction. The current results, as well as the previous findings on feature-based surround suppression (Bartsch et al., 2017; Fang et al., 2019; Ho et al., 2012; Kiyonaga & Egner, 2016; Loach et al., 2008; Störmer & Alvarez, 2014; Tombu & Tsotsos, 2008; Wang et al., 2015) support the prediction of the Selective Tuning model. The consistent findings of feature-based surround suppression across different feature domains suggest that this suppressive mechanism is an underlying principle of visual selection. As ST describes, visual selection is the result of a top-down, layer-by-layer, attentive competitive process (Tsotsos, 2011). It necessarily interacts with any bottom-up sensory processes and local processes due to horizontal projections within and between cortical columns. The final observed neural modulation reveals the effect of this interaction with the size of the surround determined by the spatial layout of feature coding within an area and the constraints on the length of the horizontal projections. A more detailed investigation of these characteristics is warranted.

CHAPTER 5. General discussion

5.1 Summary of findings

The Selective Tuning model incorporates the major elements of human visual attention and has made important predictions that have been strongly supported by several experimental works. One of its hallmarks is the attentional surround suppression which is dependent on top-down feedback (recurrent) processing. While feedback signals trace back to the connections that represent the attended visual input throughout the visual processing hierarchy, the attentional surround suppression reduces interferences from the irrelevant connections around the attentional focus. This top-down process with the surround suppression enhances the resolution of the attended visual input as the feedback signal propagates from higher to lower levels in the visual hierarchy (note that finer visual information, in terms of scale or feature, is available in lower cortical areas). This property is also reflected in the attentional binding strategies of ST, proposing that the length of the feedback traversal should be correlated with the level of visual details that a task needs. Each step of the feedback traversal in the visual hierarchy imposes a temporal delay which indicates the processing time required to complete a visual task. A substantial number of studies reported that the attentional surround suppression mechanism exists in the spatial domain (Bahcall & Kowler, 1999; Boehler et al., 2009; Caputo & Guerra, 1998; Cutzu & Tsotsos, 2003; Hopf et al., 2006, 2010; Mounts, 2000a, 2000b; Müller & Kleinschmidt, 2004) and also in the various feature domains (Bartsch et al., 2017; Ho et al., 2012; Kiyonaga & Egner, 2016; Loach et al., 2008; Störmer & Alvarez, 2014; Tombu & Tsotsos, 2008; Wang et al., 2015). Furthermore, it has been experimentally demonstrated that attention-mediated, top-down feedback processing in the visual system gives rise to the attentional surround suppression (Bartsch et al., 2017; Boehler et al., 2009). This empirical evidence supports the biological validity of ST. Nevertheless, there are still more research questions and the predictions that remain open. Here are some future questions and how ST currently deals with these questions.

- Are there structural/functional mappings between the brain regions/networks and the components of ST? ST is relatively less specific on how its elements correspond to the brain areas than the other models that specifically target certain brain areas and their neuronal activity. ST also does not explicitly address where attentional signals come from. Further research on attentional networks in the brain may be required to elaborate the models of visual attention.
- It is possible to learn to attend? Capability of selective attention differs depending on developmental stages, prior experiences, and the integrity of brain functions. ST currently does not have components that realize learning; however, it has potential to implement a simple learning strategy that leads to adaptive behavior (Gottlieb, 2012). Recently, there was an attempt to incorporate ST's top-down processing in learning algorithms for object localization (Biparva & Tsotsos, 2017).
- How is the result of attentional selection relayed to the subsequent processing stages, such as decision making, motor preparation or memory? How do these functions interact with each other in the brain? The Selective Tuning Attention References (STAR) model, an extension of ST, combines the visual hierarchy with other components necessary for visual task execution, including long-term, working memory and fixation controller

(Tsotsos & Kruijne, 2014). The development of the STAR framework would potentially represent information transmission.

- How does attention to multiple stimuli change the attentional profile? Studies have reported that the separation between stimuli which modulates competitive interactions would impact divided attention (Alvarez & Cavanagh, 2005; Alvarez, Gill, & Cavanagh, 2012; Franconeri, Alvarez, & Cavanagh, 2013; Scalf & Beck, 2010). The profile of divided attention could be tested once a proper stimulus array is determined. The same idea could be tested in the feature level: How does attention to multiple features (within a feature domain or across different feature domains) manipulate the attentional profile?

Three studies in this dissertation were conducted to investigate the unexamined predictions of ST and provide better understanding of visual attention. The first set of experiments (Chapter 2) tested the idea that fine-grained visual localization requires top-down feedback signals to earlier visual areas (Rothenstein et al., 2008; Tsotsos, 2011; Tsotsos et al., 2008). According to the attentional binding strategies of ST, the feed-forward pass fulfills simple visual tasks that do not require visual scrutiny (e.g., simple detection, categorization), thus these tasks can be performed very rapidly (Convergence Binding). Experimentally, feed-forward effects on rapid object categorization have been demonstrated (e.g., Potter, 1976; Serre et al., 2007; Thorpe et al., 1996). In contrast, top-down feedback processing to earlier areas (Full Recurrence Binding) should be deployed when a task requires high-resolution information and it takes longer time to allow feedback traversals after the feed-forward sweep. The present study investigated whether different binding processes are involved in object categorization and feature localization by controlling stimulus presentation durations. In the first experiment, relatively simple natural scenes were presented as stimuli. Consistent with the previous findings, object categorization was almost perfect even if the stimulus exposure was very short. Feature localization also completed very rapidly, suggesting that localization in simple images could be done based on coarse location information and that it may not necessarily require a full top-down feedback traversal. More complex and ecologically valid images were used in the second experiment. Here, both object categorization and feature localization performance improved as stimulus presentation duration increased (at least ~300 ms). The results suggest that when a task requires visual scrutiny with attention, such as searching for elements in cluttered scenes, it takes more processing time so that the feedback signal can access the early visual areas where the fine details are available.

The second study (Chapter 3) investigated the potential factors that influence the attentional suppressive surround. Location-based and feature-based surround suppression have been studied independently so far. If the two suppressive mechanisms interact, the attentional surround suppression would be maximized when an attended stimulus and an unattended stimulus are adjacent both in physical space and in a shared feature dimension. Participants performed a target discrimination task (Cutzu & Tsotsos, 2003) while the spatial distance and orientation difference between the two targets (cued item and probe) varied. Consistent with the initial hypothesis, target discrimination performance was worst when two similar orientated bars were presented at nearby locations and this suppressive effect was reduced when spatial and feature distances became greater. These results suggest that the location-based and feature-based surround suppression interact depending on spatial and feature distances between stimuli. In the second experiment, the potential determinants of the size of the location-based suppressive surround were examined with the target discrimination task. ST predicted that the size of the suppressive surround would be associated with the size of a neuron's RF that best represents the task-relevant stimulus (Tsotsos, 2011). In the present experiment, stimulus size and eccentricity were manipulated. The number of neurons that represent a stimulus could vary depending on the size of this stimulus, and the size of the suppressive surround would be affected by the sum of these neurons' RF sizes. It is also well established that the RF size progressively increases toward the visual periphery (Hubel & Wiesel, 1974; Kay et al., 2013; Smith et al., 2001), potentially affecting the suppressive surround. The results showed that stimulus eccentricity modulated the size of the suppressive surround, supporting the relationship between the attended neuron's RF size and the size of the suppressive surround. In contrast, stimulus size did not affect the size of the suppressive surround. One possible explanation is that the stimulus might have been represented by a single large neuron (due to stimulus eccentricity or the processing level in the visual hierarchy), not by multiple small neurons, which may not be sensitive to changes in stimulus size.

The last study of this dissertation (Chapter 4) investigated feature-based surround suppression in motion processing. In addition to supporting the existence of this suppressive mechanism in the motion direction domain, the current study explored the interplay between feature-based attention and bottom-up factors and the underlying neurophysiological mechanism of feature-based surround suppression. In the behavioral experiment, motion repulsion (Marshak & Sekuler, 1979) was measured when attention was directed to one of the two superimposed motion directions and when attention was equally divided to the two directions. The hypothesis was that motion repulsion would be reduced to a greater degree when one of the two similar directions is attended than when the directions are dissimilar because feature-based surround suppression would strongly filter out an interference from the unattended direction in the former.

The result supported this hypothesis by demonstrating decreased motion repulsion when the difference between the attended and the unattended directions was 30~40 deg, and feature-based attention interacted with bottom-up factors, such as motion speed and the color of the motion surfaces. In the neurophysiological experiment, direction tuning curves of monkeys' MT/MST neurons were measured while they attended to different motion stimuli presented within the RF. When monkeys attended to each direction of the tuning pattern or to a fixation point, ignoring motion stimuli in the RF, neuronal activity was greatest when the tuning pattern moved to the neuron's preferred direction whereas the activity was lowest when the pattern moved to the neuron's anti-preferred direction, exhibiting a single Gaussian-like tuning curve. However, the direction tuning curve showed a center-surround profile (difference-of-Gaussians) when monkeys attended to a neuron's preferred direction. That is, the minimum neuronal activity was observed when the direction of the tuning curve was half way between the preferred and anti-preferred directions but then it recovered as the direction of the tuning pattern became more dissimilar from the preferred one. Therefore, both behavioral and neurophysiological experiments provided converging evidence for feature-based surround suppression in the motion direction domain.

5.2 Limitations

Each study in this dissertation was carefully designed based on previously established methods, however, there are still some limitations. The general limitation of the present studies is that they mostly rely on human behavioral data to draw conclusions. Although the current results are consistent with what ST predicted, they do not clarify the underlying mechanisms that give rise to the behavioral consequences. For example, the first study (Chapter 2) suggested that feature localization in cluttered scenes requires top-down feedback processing, but using behavioral results, we could not figure out which areas and what circuitry are involved or determine the exact timing of the feedback signal. Combining different neuroimaging methods would be helpful to overcome this limitation and provide further insights of the visual system.

Regarding the second set of experiments (Chapter 3), the interaction between location-based and feature-based surround suppression would have been demonstrated better if the experimental conditions were more thoroughly controlled. The same-different discrimination task consists of the equal number of same and different target trials to prevent participants' response bias. Due to this fact, the results on the same target trials have greater statistical power and more strongly represented than those on the different target trials (note that the different target trials were further broken down during the analysis based on the feature distance). Feature distance could be manipulated better in a future study as now the feature distance range where surround suppression occurs is known from the present study and the other related studies that investigated surround suppression in different feature domains. In addition, manipulating stimulus size and eccentricity unavoidably confounds other variables such as stimulus spacing and visibility. Stimulus size affects the spacing between the stimuli, and to avoid spatial overlap or visual crowding (Pelli, Palomares, & Majaj, 2004) due to inadequate spacing, stimuli eccentricity should be adjusted accordingly. Therefore, it is difficult to control each factor independently to the others and the results could be unintentionally affected by the other factors.

In the last study (Chapter 4), the experimental paradigms between the behavioral and the neurophysiological studies were different because these two experiments were initially conducted independently from each other. This difference might have influenced the profile of feature-based surround suppression, producing a narrower suppressive surround in the behavioral study which reflects the nature of motion repulsion (Marshak & Sekuler, 1979). Furthermore, the stimulus configuration in the neurophysiological study – presenting two spatially separated motion patterns within a RF – involves spatial attention, not purely isolating feature-based attention. Both studies ultimately support the feature-based surround suppression in the motion direction domain, a unified experimental paradigm would have demonstrated more congruent results.

5.3 Why are different attentional profiles observed?

Although substantial evidence supports the presence of a center-surround attentional profile, this differs from long-standing, classic profiles for spatial attention such as spotlight or gradient model that do not propose a inhibitory zone around the attentional focus and also did not test its possibility (Eriksen & St. James, 1986; Eriksen & Yeh, 1985; LaBerge & Brown, 1989; Posner et al., 1980). It is important to address this difference. A potential factor that could influence an attentional profile is the type of visual tasks used in different studies. The studies that supported spotlight or simple gradient of attention used relatively easy visual tasks, such as detection of stimulus onset and binary target discrimination (Downing, 1988; Handy, Kingstone, & Mangun, 1996; Henderson & Macquistan, 1993; Hughes & Zimba, 1985; Posner et al., 1980), not necessarily requiring visual scrutiny nor feature binding. In contrast, as previously emphasized, a critical condition that induces the attentional surround suppression is a top-down selection process. The attentional surround suppression has been typically found with visual tasks that required fine discrimination and localization of target stimuli (Bahcall & Kowler, 1999; Boehler et al., 2009; Caputo & Guerra, 1998; Cave & Zimmerman, 1997; Cutzu & Tsotsos, 2003; Hopf et al., 2006; Mounts, 2000a, 2000b; Skelton & Eriksen, 1976). A psychophysical study analyzed reaction time in a simple feature detection task and in a more attention-demanding object individuation task when the spatial separation between the targets varied (McCarley & Mounts, 2007). The results showed that reaction time in the feature detection task was not affected by target separation, whereas reaction time was significantly longer when target separation was smaller in the object individuation task. It suggests that attentional surround suppression occurs only if a visual task requires higher resolution or disambiguation of stimulus' properties. Boehler et al. (2009) also reported comparable findings by measuring MEG signals during orientation and color discrimination tasks. They found that attentional surround suppression was apparent only in the orientation discrimination task which needs fine details and localization of a target but not in the color discrimination task. Furthermore, this is also consistent with the findings in the current dissertation. Surround suppression was observed when people make same-different judgments on the same/similar orientation and the same face targets but there was no evidence of surround suppression for dissimilar orientation and different face targets (Chapter 3). While the former needs more detailed comparisons of visual features to make decisions, the latter does not demand effortful visual analyses. To conclude, attentional profiles could vary depending on specific computational processes required to achieve given visual tasks and it may resolve the controversies about the distribution of attention.

5.4 Future directions

There are several possible future studies that could develop and expand the present findings. Attentional selection in vision is often followed by goal-directed behavior such as reaching and grasping of the selected object. The localization task in Chapter 2 could be applied in a more realistic situation to examine the interaction between vision and action. In particular, one might use real-world objects that we typically use in our everyday life or a virtual reality environment. Real-world objects are more memorable (Snow, Skiba, Coleman, & Berryhill, 2014) and draw more attention (Gomez, Skiba, & Snow, 2018) than pictures. Furthermore, even infants show strong preferences for real objects (Gerhard, Culham, & Schwarzer, 2016). Therefore, localization in a realistic situation could be qualitatively different from performing the same task using 2D stimuli, and it would be able to demonstrate the genuine impact of top-down feedback processing on executing goal-directed behavior (e.g., preparation of a proper hand posture, speed of movement, on-line correction of motor responses).

The second study provided supporting evidence for the interaction between location-based and feature-based surround suppression but only at the phenomenal level. Future studies could elucidate the nature of the relationship between these two suppressive effects. For example, one could investigate the form of their interaction (whether it is additive, multiplicative, or mixture of these two). It would be also worth studying how feature-based surround suppression is modulated by the interplay between different features. So far, studies on feature-based surround suppression have searched for this suppressive phenomenon within a feature domain. However, it should be addressed whether the similar suppressive effects exert when different types of features are attended together (feature conjunction) because visual stimuli we are confronted in the real world typically consist of multiple features. This future study would grant more ecological validity to the surround suppression mechanisms. Potential results are many. 1) Surround suppression could occur only for the attended feature conjunction but not for individual features of this conjunction. 2) Surround suppression could occur for both feature conjunction and its individual features but it would be relatively stronger for the attended feature conjunction.

3) The magnitude of surround suppression could be equal for both feature conjunction and its individual features. Understanding the relationship between different types of attentional surround suppression would also provide further insight into practical application, such as designing user-interface layout and displaying items effectively to maximize user experience and performance.

Another research question is whether the center-surround profile of attention is found in other perceptual modalities and cognitive domains. Loach and colleagues (2008) previously demonstrated the effect of the surround suppression on action affordance, suggesting that surround suppression is not just confined to the visual domain. Moreover, it is reported that representation in visual working memory can influence perception in a center-surround manner (Kiyonaga & Egner, 2016). If the surround suppression is an overarching selective mechanism in the brain, future studies should be able to find the equivalent suppressive effects in different modalities and cognitive domains, such as in the auditory domain where tonotopic organization is apparent (Romani, Williamson, & Kaufman, 1982; Wessinger, Buonocore, Kussmaul, & Mangun, 1997).

Attentional surround suppression could function as a diagnostic tool which evaluates whether one's attentional system is intact. The ability to selectively allocate attention and filter out irrelevant information matures as development progresses and research supports that top-down feedback processes are strengthened in the late stage of development (Amso & Scerif, 2015; Farrant & Uddin, 2015; Gaspelin, Margett-Jordan, & Ruthruff, 2015). Motivated by these findings, a recent study measured the attentional surround suppression across different age groups and demonstrated that this suppressive mechanism does not appear in childhood, but it is manifested in pre-adolescence and adulthood (Wong-Kee-You, Tsotsos, & Adler, 2019). This result is consistent with late development of feedback cortical connectivity and it also fits ST's notion that the attentional surround suppression results from top-down, attention-mediated feedback processing. The same idea can be applied to the patient groups with attention-related deficits, such as autism spectrum disorder (ASD). Compared to the control group, ASD group shows weaker surround suppression (Ronconi et al., 2018), suggesting inefficiency in inhibiting irrelevant visual information that ASD typically demonstrates (Adams & Jarrold, 2012; Burack, 1994; Remington, Swettenham, Campbell, & Coleman, 2009; Rubenstein & Merzenich, 2003). Testing the attentional surround suppression on multiple patient groups would shed light on the nature of the attentional network of each disorder and suggest which attentional function should be targeted to relieve symptoms.

5.5 Final thoughts

Selective attention is a crucial mechanism which prioritizes task-relevant information among a myriad of perceptual inputs. Although people would agree with this general function of selective attention, how attention exerts its effects and the underlying mechanisms are still a matter of controversy. Motivated by the proposals of the Selective Tuning model, this dissertation experimentally demonstrates the role of top-down feedback processing in fine-grained attentional selection and the characteristics of the attentional surround suppression, the consequence of top-down feedback in both physical and feature spaces. The experiments in this dissertation suggest further strategies to amplify attentional resolution that the visual system utilizes to demarcate relevant information from distracting one. A future question is what other strategies the visual system would use to achieve successful information selection and whether the Selective Tuning model could predict/incorporate them. For example, a recent study reported that attention to one of two similar feature representations protects the attended representation from distortion, whereas the unattended representation is greatly repulsed away from its veridical feature (Bae & Luck, 2017). The similar results were found with the electrophysiological method as well (Chapman, Geweke, & Störmer, 2019). Such findings provide novel details of the selection process that attention calibrates attended and unattended information in a different manner. Furthermore, proper implementation of experimental findings in a model would help develop machine/computer vision and lead to the next predictions on the visual system at the same time. I hope the present research potentially could contribute to the improvements in several fields of vision science by the synergy between experimental works and modelling.

References

Adams, N. C., & Jarrold, C. (2012). Inhibition in autism: Children with autism have difficulty inhibiting irrelevant distractors but not prepotent responses. *Journal of Autism and Developmental Disorders*, *42*(6), 1052–1063. https://doi.org/10.1007/s10803-011-1345-3

Adesnik, H., Bruns, W., Taniguchi, H., Huang, Z. J., & Scanziani, M. (2012). A neural circuit for spatial summation in visual cortex. *Nature*, *490*(7419), 226–231. https://doi.org/10.1038/nature11526

Ahissar, M., & Hochstein, S. (1997). Task difficulty and the specificity of perceptual learning. *Nature*, *387*(6631), 401–406. https://doi.org/10.1038/387401a0

Ahissar, M., & Hochstein, S. (2000). The spread of attention and learning in feature search: Effects of target distribution and task difficulty. *Vision Research*, *40*(10–12), 1349–1364. https://doi.org/10.1016/S0042-6989(00)00002-X

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, *19*(6), 716–723. https://doi.org/10.1109/TAC.1974.1100705

Albright, T. D. (1984). Direction and orientation selectivity of neurons in visual area MT of the macaque. *Journal of Neurophysiology*, *52*(6), 1106–1130. https://doi.org/10.1152/jn.1984.52.6.1106

Alink, A., Schwiedrzik, C. M., Kohler, A., Singer, W., & Muckli, L. (2010). Stimulus predictability reduces responses in primary visual cortex. *Journal of Neuroscience*, *30*(8), 2960–2966. https://doi.org/10.1523/JNEUROSCI.3730-10.2010

Altmann, C. F., Bülthoff, H. H., & Kourtzi, Z. (2003). Perceptual organization of local elements into global shapes in the human visual cortex. *Current Biology*, *13*(4), 342–349. https://doi.org/10.1016/S0960-9822(03)00052-6

Alvarez, G. A., & Cavanagh, P. (2005). Independent resources for attentional tracking in the left and right visual hemifields. *Psychological Science*, *16*(8), 637–643. https://doi.org/10.1111/j.1467-9280.2005.01587.x

Alvarez, G. A., Gill, J., & Cavanagh, P. (2012). Anatomical constraints on attention: Hemifield independence is a signature of multifocal spatial selection. *Journal of Vision*, *12*(5), 9. https://doi.org/10.1167/12.5.9

Amso, D., & Scerif, G. (2015). The attentive brain: Insights from developmental cognitive neuroscience. *Nature Reviews Neuroscience*, *16*(10), 606–619. https://doi.org/10.1038/nrn4025

Anllo-Vento, L., & Hillyard, S. A. (1996). Selective attention to the color and direction of moving stimuli: Electrophysiological correlates of hierarchical feature selection. *Perception and Psychophysics*, *58*(2), 191–206. https://doi.org/10.3758/BF03211875

Anton-Erxleben, K., Stephan, V. M., & Treue, S. (2009). Attention reshapes center-surround receptive field structure in macaque cortical area MT. *Cerebral Cortex*, *19*(10), 2466–2478. https://doi.org/10.1093/cercor/bhp002

Auckland, M. E., Cave, K. R., & Donnelly, N. (2007). Nontarget objects can influence perceptual processes during object recognition. *Psychonomic Bulletin & Review*, *14*(2), 332–337. https://doi.org/10.3758/BF03194073

Awh, E., Jonides, J., & Reuter-Lorenz, P. A. (1998). Rehearsal in spatial working memory. *Journal of Experimental Psychology: Human Perception and Performance*, *24*(3), 780–790. https://doi.org/10.1037/0096-1523.24.3.780

Awh, E., & Pashler, H. (2000). Evidence for split attentional foci. *Journal of Experimental Psychology: Human Perception and Performance*, *26*(2), 834–846.

https://doi.org/10.1037/0096-1523.26.2.834

Bae, G. Y., & Luck, S. J. (2017). Interactions between visual working memory representations. *Attention, Perception, and Psychophysics*, *79*(8), 2376–2395. https://doi.org/10.3758/s13414-017-1404-8

Bahcall, D. O., & Kowler, E. (1999). Attentional interference at small spatial separations. *Vision Research*, *39*(1), 71–86. https://doi.org/10.1016/S0042-6989(98)00090-X

Baldauf, D., & Deubel, H. (2010). Attentional landscapes in reaching and grasping. *Vision Research*, *50*(11), 999–1013. https://doi.org/10.1016/j.visres.2010.02.008

Bamber, D. (1969). Reaction times and error rates for “same”-“different” judgments of multidimensional stimull. *Perception & Psychophysics*, *6*(3), 169–174. https://doi.org/10.3758/BF03210087

Bar, M. (2004). Visual objects in context. *Nature Reviews Neuroscience*, *5*(8), 617–629. https://doi.org/10.1038/nrn1476

Bar, M., Kassam, K. S., Ghuman, A. S., Boshyan, J., Schmid, A. M., Dale, A. M., … Halgren, E. (2006). Top-down facilitation of visual recognition. *Proceedings of the National Academy of Sciences*, *103*(2), 449–454. https://doi.org/10.1073/pnas.0507062103

Barceló, F., Suwazono, S., & Knight, R. T. (2000). Prefrontal modulation of visual processing in humans. *Nature Neuroscience*, *3*(4), 399–403. https://doi.org/10.1038/73975

Bartsch, M. V., Loewe, K., Merkel, C., Heinze, H.-J., Schoenfeld, M. A., Tsotsos, J. K., & Hopf, J.-M. (2017). Attention to color sharpens neural population tuning via feedback processing in the human visual cortex hierarchy. *The Journal of Neuroscience*, *37*(43), 0666–17. https://doi.org/10.1523/JNEUROSCI.0666-17.2017

Belke, E., & Meyer, A. S. (2002). Tracking the time course of multidimensional stimulus discrimination: Analyses of viewing patterns and processing times during “same”-"different" decisions. *European Journal of Cognitive Psychology*, *14*(2), 237–266. https://doi.org/10.1080/09541440143000050

Betz, T., Kietzmann, T. C., Wilming, N., & König, P. (2010). Investigating task-dependent top-down effects on overt visual attention. *Journal of Vision*, *10*(3), 1–14. https://doi.org/10.1167/10.3.15

Beuth, F., & Hamker, F. H. (2015). A mechanistic cortical microcircuit of attention for amplification, normalization and suppression. *Vision Research*, *116*, 241–257. https://doi.org/10.1016/j.visres.2015.04.004

Biederman, I., Mezzanotte, R. J., & Rabinowitz, J. C. (1982). Scene perception: Detecting and judging objects undergoing relational violations. *Cognitive Psychology*, *14*(2), 143–177. https://doi.org/10.1016/0010-0285(82)90007-X

Biparva, M., & Tsotsos, J. K. (2017). STNet: Selective tuning of convolutional networks for object localization. *Proceedings of the IEEE International Conference on Computer Vision*, 2715–2723. https://doi.org/10.1109/ICCVW.2017.319

Blizzard, S., Fierro-Rojas, A., & Fallah, M. (2017). Response inhibition is facilitated by a change to red over green in the stop signal paradigm. *Frontiers in Human Neuroscience*, *10*, 655. https://doi.org/10.3389/fnhum.2016.00655

Boehler, C. N., Schoenfeld, M. A., Heinze, H.-J., & Hopf, J.-M. (2008). Rapid recurrent processing gates awareness in primary visual cortex. *Proceedings of the National Academy of Sciences*, *105*(25), 8742–8747. https://doi.org/10.1073/pnas.0801999105

Boehler, C. N., Tsotsos, J. K., Schoenfeld, M. A., Heinze, H.-J., & Hopf, J.-M. (2011). Neural mechanisms of surround attenuation and distractor competition in visual search. *Journal of*

Neuroscience, *31*(14), 5213–5224. https://doi.org/10.1523/JNEUROSCI.6406-10.2011

Boehler, C. N., Tsotsos, J. K., Schoenfeld, M. A., Heinze, H. J., & Hopf, J. M. (2009). The center-surround profile of the focus of attention arises from recurrent processing in visual cortex. *Cerebral Cortex*, *19*(4), 982–991. https://doi.org/10.1093/cercor/bhn139

Bouvier, S., & Treisman, A. (2010). Visual feature binding requires reentry. *Psychological Science*, *21*(2), 200–204. https://doi.org/10.1177/0956797609357858

Boyce, S. J., Pollatsek, A., & Rayner, K. (1989). Effect of background information on object identification. *Journal of Experimental Psychology: Human Perception and Performance*, *15*(3), 556–566. https://doi.org/10.1037/0096-1523.15.3.556

Boynton, G. M. (2005). Attention and visual perception. *Current Opinion in Neurobiology*, *15*(4), 465–469. https://doi.org/10.1016/j.conb.2005.06.009

Braddick, O. J., Wishart, K. A., & Curran, W. (2002). Directional performance in motion transparency. *Vision Research*, *42*(10), 1237–1248. https://doi.org/10.1016/S0042-6989(02)00018-4

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, *10*(4), 433–436. https://doi.org/10.1163/156856897X00357

Braun, J., Koch, C., & Davis, J. L. (2001). *Visual attention and cortical circuits*. Cambridge: MIT press.

Britten, K. H., & Newsome, W. T. (1998). Tuning bandwidths for near-threshold stimuli in area MT. *Journal of Neurophysiology*, *80*(2), 762–770. https://doi.org/10.1152/jn.1998.80.2.762

Buffalo, E. A., Fries, P., Landman, R., Liang, H., & Desimone, R. (2010). A backward progression of attentional effects in the ventral stream. *Proceedings of the National Academy of Sciences*, *107*(1), 361–365. https://doi.org/10.1073/pnas.0907658106

Bullier, J. (2001). Integrated model of visual processing. *Brain Research Reviews*, *36*(2–3), 96–107. https://doi.org/10.1016/S0165-0173(01)00085-6

Burack, J. A. (1994). Selective attention deficits in persons with autism: Preliminary evidence of an inefficient attentional lens. *Journal of Abnormal Psychology*, *103*(3), 535–543. https://doi.org/10.1037/0021-843X.103.3.535

Bushnell, B. N., Harding, P. J., Kosai, Y., Bair, W., & Pasupathy, A. (2011). Equiluminance cells in visual cortical area V4. *Journal of Neuroscience*, *31*(35), 12398–12412. https://doi.org/10.1523/JNEUROSCI.1890-11.2011

Caputo, G., & Guerra, S. (1998). Attentional selection by distractor suppression. *Vision Research*, *38*(5), 669–689.

Carlson, T., Hogendoorn, H., Fonteijn, H., & Verstraten, F. A. J. (2011). Spatial coding and invariance in object-selective cortex. *Cortex*, *47*(1), 14–22. https://doi.org/10.1016/j.cortex.2009.08.015

Carrasco, M. (2011). Visual attention: The past 25 years. *Vision Research*, *51*(13), 1484–1525. https://doi.org/10.1016/j.visres.2011.04.012

Cave, K. R., & Zimmerman, J. M. (1997). Flexibility in spatial attention before and after practice. *Psychological Science*, *8*(5), 399–403. https://doi.org/10.1111/j.1467-9280.1997.tb00433.x

Chakravarthi, R., Carlson, T. A., Chaffin, J., Turret, J., & VanRullen, R. (2014). The temporal evolution of coarse location coding of objects: Evidence for feedback. *Journal of Cognitive Neuroscience*, *26*(10), 2370–2384. https://doi.org/10.1162/jocn_a_00644

Chapman, A. F., Geweke, F., & Störmer, V. S. (2019). Feature-based attention resolves differences in target-distractor similarity through multiple mechanisms. *Journal of Vision*, *19*(10), 45a-45a. https://doi.org/10.1167/19.10.45a

Chelazzi, L., Duncan, J., Miller, E. K., & Desimone, R. (1998). Responses of neurons in inferior temporal cortex during memory-guided visual search. *Journal of Neurophysiology*, *80*(6), 2918–2940. https://doi.org/10.1152/jn.1998.80.6.2918

Chen, M., Yan, Y., Gong, X., Gilbert, C. D., Liang, H., & Li, W. (2014). Incremental integration of global contours through interplay between visual cortical areas. *Neuron*, *82*(3), 682–694. https://doi.org/10.1016/j.neuron.2014.03.023

Chen, Y., Byrne, P., & Crawford, J. D. (2011). Time course of allocentric decay, egocentric decay, and allocentric-to-egocentric conversion in memory-guided reach. *Neuropsychologia*, *49*(1), 49–60. https://doi.org/10.1016/j.neuropsychologia.2010.10.031

Chen, Y., Meng, X., Matthews, N., & Qian, N. (2005). Effects of attention on motion repulsion. *Vision Research*, *45*(10), 1329–1339. https://doi.org/10.1016/j.visres.2004.11.005

Chun, M. M. (2000). Contextual cueing of visual attention. *Trends in Cognitive Sciences*, *4*(5), 170–178. https://doi.org/10.1016/S1364-6613(00)01476-5

Chun, M. M., & Jiang, Y. (1998). Contextual Cueing: Implicit Learning and Memory of Visual Context Guides Spatial Attention. *Cognitive Psychology*, *36*(1), 28–71. https://doi.org/10.1006/cogp.1998.0681

Chun, M. M., & Jiang, Y. (2003). Implicit, long-term spatial contextual memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*(2), 224–234. https://doi.org/10.1037/0278-7393.29.2.224

Cichy, R. M., Chen, Y., & Haynes, J.-D. (2011). Encoding the identity and location of objects in human LOC. *NeuroImage*, *54*(3), 2297–2307. https://doi.org/10.1016/j.neuroimage.2010.09.044

Cichy, R. M., Pantazis, D., & Oliva, A. (2014). Resolving human object recognition in space and time. *Nature Neuroscience*, *17*(3), 455–462. https://doi.org/10.1038/nn.3635

Crouzet, S. M., Kirchner, H., & Thorpe, S. J. (2010). Fast saccades toward faces: Face detection in just 100 ms. *Journal of Vision*, *10*(4), 1–17. https://doi.org/10.1167/10.4.16

Curran, W., & Benton, C. P. (2003). Speed tuning of direction repulsion describes an inverted U-function. *Vision Research*, *43*(17), 1847–1853. https://doi.org/10.1016/S0042-6989(03)00302-X

Cutzu, F., & Tsotsos, J. K. (2003). The selective tuning model of attention: Psychophysical evidence for a suppressive annulus around an attended item. *Vision Research*, *43*(2), 205–219. https://doi.org/10.1016/S0042-6989(02)00491-1

Davenport, J. L., & Potter, M. C. (2004). Scene consistency in object and background perception. *Psychological Science*, *15*(8), 559–564. https://doi.org/10.1111/j.0956-7976.2004.00719.x

de Haas, B., Schwarzkopf, D. S., Anderson, E. J., & Rees, G. (2014). Perceptual load affects spatial tuning of neuronal populations in human early visual cortex. *Current Biology*, *24*(2), R66–R67. https://doi.org/10.1016/j.cub.2013.11.061

Deco, G., & Zihl, J. (2001). A neurodynamical model of visual attention: Feedback enhancement of spatial resolution in a hierarchical system. *Journal of Computational Neuroscience*, *10*(3), 231–253. https://doi.org/10.1023/A:1011233530729

Delorme, A., Richard, G., & Fabre-Thorpe, M. (2000). Ultra-rapid categorisation of natural scenes does not rely on colour cues: A study in monkeys and humans. *Vision Research*, *40*(16), 2187–2200. https://doi.org/10.1016/S0042-6989(00)00083-3

Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual*

Review of Neuroscience, *18*(1), 193–222. https://doi.org/10.1146/annurev.neuro.18.1.193

Di Lollo, V., Enns, J. T., & Rensink, R. A. (2000). Competition for consciousness among visual events: The psychophysics of reentrant visual processes. *Journal of Experimental Psychology: General*, *129*(4), 481–507. https://doi.org/10.1037/0096-3445.129.4.481

Di Russo, F., Martinez, A., & Hillyard, S. A. (2003). Source analysis of event-related cortical activity during visuo-spatial attention. *Cerebral Cortex*, *13*(5), 486–499. https://doi.org/10.1093/cercor/13.5.486

DiCarlo, J. J., & Maunsell, J. H. R. (2003). Anterior inferotemporal neurons of monkeys engaged in object recognition can be highly sensitive to object retinal position. *Journal of Neurophysiology*, *89*(6), 3264–3278. https://doi.org/10.1152/jn.00358.2002

Downing, B. D. (1971). Response probabilities and “same-different” reaction times. *Perception & Psychophysics*, *9*(2), 213–215.

Downing, C. J. (1988). Expectancy and visual-spatial attention: Effects on perceptual quality. *Journal of Experimental Psychology: Human Perception and Performance*, *14*(2), 188–202. https://doi.org/10.1037/0096-1523.14.2.188

Downing, P. E. (2000). Interactions between visual working memory and selective attention. *Psychological Science*, *11*(6), 467–473. https://doi.org/10.1111/1467-9280.00290

Egeth, H. E. (1966). Parallel versus serial processes in multidimensional stimulus discrimination. *Perception & Psychophysics*, *1*(4), 245–252. https://doi.org/10.3758/BF03207389

Egly, R., & Homa, D. (1984). Sensitization of the visual field. *Journal of Experimental Psychology: Human Perception and Performance*, *10*(6), 778–793. https://doi.org/10.1037/0096-1523.10.6.778

Eimer, M., & Grubert, A. (2014). Spatial attention can be allocated rapidly and in parallel to new visual objects. *Current Biology*, *24*(2), 193–198. https://doi.org/10.1016/j.cub.2013.12.001

Epstein, R., & Kanwisher, N. (1998). A cortical representation of the local visual environment. *Nature*, *392*(6676), 598–601. https://doi.org/10.1038/33402

Eriksen, C. W., & St. James, J. D. (1986). Visual attention within and around the field of focal attention: A zoom lens model. *Perception & Psychophysics*, *40*(4), 225–240. https://doi.org/10.3758/BF03211502

Eriksen, C. W., & Yeh, Y.-Y. (1985). Allocation of attention in the visual field. *Journal of Experimental Psychology: Human Perception and Performance*, *11*(5), 583–597. https://doi.org/10.1037/0096-1523.11.5.583

Esterman, M., & Yantis, S. (2010). Perceptual expectation evokes category-selective cortical activity. *Cerebral Cortex*, *20*(5), 1245–1253. https://doi.org/10.1093/cercor/bhp188

Evans, K. K., & Treisman, A. (2005). Perception of objects in natural scenes: Is it really attention free? *Journal of Experimental Psychology: Human Perception and Performance*, *31*(6), 1476–1492. https://doi.org/10.1037/0096-1523.31.6.1476

Fabre-Thorpe, M., Delorme, A., Marlot, C., & Thorpe, S. (2001). A limit to the speed of processing in ultra-rapid visual categorization of novel natural scenes. *Journal of Cognitive Neuroscience*, *13*(2), 171–180. https://doi.org/10.1162/089892901564234

Fabre-Thorpe, M., Richard, G., & Thorpe, S. J. (1998). Rapid categorization of natural images by rhesus monkeys. *NeuroReport*, *9*(2), 303–308. https://doi.org/10.1097/00001756-199801260-00023

Fahrenfort, J. J., Scholte, H. S., & Lamme, V. A. F. (2007). Masking disrupts reentrant processing in human visual cortex. *Journal of Cognitive Neuroscience*, *19*(9), 1488–1497. https://doi.org/10.1162/jocn.2007.19.9.1488

Fang, M. W. H., Becker, M. W., & Liu, T. (2019). Attention to colors induces surround suppression at category boundaries. *Scientific Reports*, *9*(1), 1–13. https://doi.org/10.1038/s41598-018-37610-7

Farell, B. (1985). “Same”–"different" judgments: A review of current controversies in perceptual comparisons. *Psychological Bulletin*, *98*(3), 419–456. https://doi.org/10.1037/0033-2909.98.3.419

Farrant, K., & Uddin, L. Q. (2015). Asymmetric development of dorsal and ventral attention networks in the human brain. *Developmental Cognitive Neuroscience*, *12*, 165–174. https://doi.org/10.1016/j.dcn.2015.02.001

Felleman, D. J., & Van Essen, D. C. (1991). Distributed hierarchical processing in the primate cerebral cortex. *Cerebral Cortex*, *1*(1), 1–47. https://doi.org/10.1093/cercor/1.1.1

Foxe, J., & Simpson, G. (2002). Flow of activation from V1 to frontal cortex in humans. *Experimental Brain Research*, *142*(1), 139–150. https://doi.org/10.1007/s00221-001-0906-7

Franconeri, S. L., Alvarez, G. A., & Cavanagh, P. (2013). Flexible cognitive resources: Competitive content maps for attention and memory. *Trends in Cognitive Sciences*, *17*(3), 134–141. https://doi.org/10.1016/j.tics.2013.01.010

Friedman, A. (1979). Framing pictures: The role of knowledge in automatized encoding and memory for gist. *Journal of Experimental Psychology: General*, *108*(3), 316–355. https://doi.org/10.1037/0096-3445.108.3.316

Friston, K., & Kiebel, S. (2009). Predictive coding under the free-energy principle. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1521), 1211–1221. https://doi.org/10.1098/rstb.2008.0300

Fukushima, K. (1986). A neural network model for selective attention in visual pattern recognition. *Biological Cybernetics*, *55*(1), 5–15.

Gaspelin, N., Margett-Jordan, T., & Ruthruff, E. (2015). Susceptible to distraction: Children lack top-down control over spatial attention capture. *Psychonomic Bulletin & Review*, *22*(2), 461–468. https://doi.org/10.3758/s13423-014-0708-0

Gerhard, T. M., Culham, J. C., & Schwarzer, G. (2016). Distinct visual processing of real objects and pictures of those objects in 7- to 9-month-old infants. *Frontiers in Psychology*, *7*(827). https://doi.org/10.3389/fpsyg.2016.00827

Ghose, G. M., & Maunsell, J. (1999). Specialized representations in visual cortex: A role for binding? *Neuron*, *24*(1), 79–85. https://doi.org/10.1016/S0896-6273(00)80823-5

Gilbert, C. D., & Li, W. (2013). Top-down influences on visual processing. *Nature Reviews Neuroscience*, *14*(5), 350–363. https://doi.org/10.1038/nrn3476

Gomez, M. A., Skiba, R. M., & Snow, J. C. (2018). Graspable objects grab attention more than images do. *Psychological Science*, *29*(2), 206–218. https://doi.org/10.1177/0956797617730599

Gottlieb, J. (2012). Attention, learning, and the value of information. *Neuron*, *76*(2), 281–295. https://doi.org/10.1016/j.neuron.2012.09.034

Gouras, P., & Kruger, J. (1979). Responses of cells in foveal visual cortex of the monkey to pure color contrast. *Journal of Neurophysiology*, *42*(3), 850–860.

Gray, C. M. (1999). The temporal correlation hypothesis of visual feature integration. *Neuron*, *24*(1), 31–47. https://doi.org/10.1016/S0896-6273(00)80820-X

Gray, C. M., König, P., Engel, A. K., & Singer, W. (1989). Oscillatory responses in cat visual cortex exhibit inter-columnar synchronization which reflects global stimulus properties. *Nature*, *338*(6213), 334–337. https://doi.org/10.1038/338334a0

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

Grill-Spector, K., & Kanwisher, N. (2005). Visual recognition. *Psychological Science*, *16*(2), 152–160. https://doi.org/10.1111/j.0956-7976.2005.00796.x

Grill-Spector, K., Kushnir, T., Edelman, S., Avidan, G., Itzchak, Y., & Malach, R. (1999). Differential processing of objects under various viewing conditions in the human lateral occipital complex. *Neuron*, *24*(1), 187–203. https://doi.org/10.1016/S0896-6273(00)80832-6

Haider, B., Krause, M. R., Duque, A., Yu, Y., Touryan, J., Mazer, J. A., & McCormick, D. A. (2010). Synaptic and network mechanisms of sparse and reliable visual cortical activity during nonclassical receptive field stimulation. *Neuron*, *65*(1), 107–121. https://doi.org/10.1016/j.neuron.2009.12.005

Hamker, F. H. (2005). The reentry hypothesis: The putative interaction of the frontal eye field, ventrolateral prefrontal cortex, and areas V4, IT for attention and eye movement. *Cerebral Cortex*, *15*(4), 431–447. https://doi.org/10.1093/cercor/bhh146

Handy, T. C., Kingstone, A., & Mangun, G. R. (1996). Spatial distribution of visual attention: Perceptual sensitivity and response latency. *Perception & Psychophysics*, *58*(4), 613–627. https://doi.org/10.3758/BF03213094

Hawkins, H. L. (1969). Parallel processing in complex visual discrimination. *Perception & Psychophysics*, *5*(1), 56–64.

Haxby, J. V, Gobbini, M. I., Furey, M. L., Ishai, A., Schouten, J. L., & Pietrini, P. (2001). Distrubuted and overlapping representations of face and objects in ventral temporal cortex. *Science*, *293*(5539), 2425–2430. https://doi.org/10.1126/science.1063736

Hemond, C. C., Kanwisher, N. G., & Op de Beeck, H. P. (2007). A preference for contralateral stimuli in human object- and face-selective cortex. *PLoS ONE*, *2*(6), e574. https://doi.org/10.1371/journal.pone.0000574

Henderson, J. M., & Macquistan, A. D. (1993). The spatial distribution of attention following an exogenous cue. *Perception & Psychophysics*, *53*(2), 221–230. https://doi.org/10.3758/BF03211732

Hillyard, S. A., & Anllo-Vento, L. (1998). Event-related brain potentials in the study of visual selective attention. *Proceedings of the National Academy of Sciences*, *95*(3), 781–787. https://doi.org/10.1073/pnas.95.3.781

Hiris, E., & Blake, R. (1996). Direction repulsion in motion transparency. *Visual Neuroscience*, *13*(1), 187–197.

Ho, T. C., Brown, S., Abuyo, N. A., Ku, E.-H. J., & Serences, J. T. (2012). Perceptual consequences of feature-based attentional enhancement and suppression. *Journal of Vision*, *12*(8), 15. https://doi.org/10.1167/12.8.15

Hochstein, S., & Ahissar, M. (2002). View from the top: Hierarchies and reverse hierarchies in the visual system. *Neuron*, *36*(5), 791–804. https://doi.org/10.1016/S0896-6273(02)01091-7

Hopf, J.-M., Boehler, C. N., Luck, S. J., Tsotsos, J. K., Heinze, H. J., & Schoenfeld, M. A. (2006). Direct neurophysiological evidence for spatial suppression surrounding the focus of attention in vision. *Proceedings of the National Academy of Sciences*, *103*(4), 1053–1058. https://doi.org/10.1073/pnas.0507746103

Hopf, J.-M., Boehler, C. N., Schoenfeld, M. A., Heinze, H.-J., & Tsotsos, J. K. (2010). The spatial profile of the focus of attention in visual search: Insights from MEG recordings. *Vision Research*, *50*(14), 1312–1320. https://doi.org/10.1016/j.visres.2010.01.015

Hopf, J.-M., Boelmans, K., Schoenfeld, M. A., Luck, S. J., & Heinze, H.-J. (2004). Attention to features precedes attention to locations in visual search: Evidence from electromagnetic brain responses in humans. *Journal of Neuroscience*, *24*(8), 1822–1832. https://doi.org/10.1523/JNEUROSCI.3564-03.2004

Hubel, D. H., & Livingstone, M. S. (1990). Color and contrast sensitivity in the lateral geniculate body and primary visual cortex of the macaque monkey. *Journal of Neuroscience*, *10*(7), 2223–2237.

Hubel, D. H., & Wiesel, T. N. (1968). Receptive fields and functional architecture of monkey striate cortex. *The Journal of Physiology*, *195*(1), 215–243. https://doi.org/10.1113/jphysiol.1968.sp008455

Hubel, D. H., & Wiesel, T. N. (1974). Uniformity of monkey striate cortex: A parallel relationship between field size, scatter, and magnification factor. *The Journal of Comparative Neurology*, *158*(3), 295–305. https://doi.org/10.1002/cne.901580305

Hughes, H. C., & Zimba, L. D. (1985). Spatial maps of directed visual attention. *Journal of Experimental Psychology: Human Perception and Performance*, *11*(4), 409–430. https://doi.org/10.1037/0096-1523.11.4.409

Hung, C. P., Kreiman, G., Poggio, T., & DiCarlo, J. J. (2005). Fast readout of object identity from macaque inferior temporal cortex. *Science*, *310*(5749), 863–866. https://doi.org/10.1126/science.1117593

Hupé, J. M., James, A. C., Payne, B. R., Lomber, S. G., Girard, P., & Bullier, J. (1998). Cortical feedback improves discrimination between figure and background by V1, V2 and V3 neurons. *Nature*, *394*(6695), 784–787. https://doi.org/10.1038/29537

Isik, L., Meyers, E. M., Leibo, J. Z., & Poggio, T. (2014). The dynamics of invariant object recognition in the human visual system. *Journal of Neurophysiology*, *111*(1), 91–102. https://doi.org/10.1152/jn.00394.2013

Ito, M., Westheimer, G., & Gilbert, C. D. (1998). Attention and perceptual learning modulate contextual influences on visual perception. *Neuron*, *20*(6), 1191–1197. https://doi.org/10.1016/S0896-6273(00)80499-7

Itti, L., & Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research*, *40*(10–12), 1489–1506. https://doi.org/10.1016/S0042-6989(99)00163-7

Itti, L., Koch, C., & Niebur, E. (1998). A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *20*(11), 1254–1259. https://doi.org/10.1109/34.730558

Jehee, J. F. M., Lamme, V. A. F., & Roelfsema, P. R. (2007). Boundary assignment in a recurrent network architecture. *Vision Research*, *47*(9), 1153–1165. https://doi.org/10.1016/j.visres.2006.12.018

Johnson, E. N., Hawken, M. J., & Shapley, R. (2001). The spatial transformation of color in the primary visual cortex of the macaque monkey. *Nature Neuroscience*, *4*(4), 409–416. https://doi.org/10.1038/86061

Johnson, J. S., & Olshausen, B. A. (2003). Timecourse of neural signatures of object recognition. *Journal of Vision*, *3*(7), 4. https://doi.org/10.1167/3.7.4

Joubert, O. R., Fize, D., Rousselet, G. A., & Fabre-Thorpe, M. (2008). Early interference of context congruence on object processing in rapid visual categorization of natural scenes. *Journal of Vision*, *8*(13), 11. https://doi.org/10.1167/8.13.11

Juan, C. H., & Walsh, V. (2003). Feedback to V1: A reverse hierarchy in vision. *Experimental*

Brain Research, *150*(2), 259–263. https://doi.org/10.1007/s00221-003-1478-5

Judge, S. J., Richmond, B. J., & Chu, F. C. (1980). Implantation of magnetic search coils for measurement of eye position: An improved method. *Vision Research*, *20*(6), 535–538. https://doi.org/10.1016/0042-6989(80)90128-5

Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: A module in human extrastriate cortex specialized for face perception. *The Journal of Neuroscience*, *17*(11), 4302–4311. https://doi.org/10.1523/JNEUROSCI.17-11-04302.1997

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. https://doi.org/10.1016/S0896-6273(00)80734-5

Kay, K. N., Winawer, J., Mezer, A., & Wandell, B. A. (2013). Compressive spatial summation in human visual cortex. *Journal of Neurophysiology*, *110*(2), 481–494. https://doi.org/10.1152/jn.00105.2013

Keysers, C., Xiao, D. K., Földiák, P., & Perrett, D. I. (2001). The speed of sight. *Journal of Cognitive Neuroscience*, *13*(1), 90–101. https://doi.org/10.1162/089892901564199

Kim, M.-S., & Cave, K. R. (1999). Top-down and bottom-up attentional control: On the nature of interference from a salient distractor. *Perception & Psychophysics*, *61*(6), 1009–1023. https://doi.org/10.3758/BF03207609

Kirchner, H., & Thorpe, S. J. (2006). Ultra-rapid object detection with saccadic eye movements: Visual processing speed revisited. *Vision Research*, *46*(11), 1762–1776. https://doi.org/10.1016/j.visres.2005.10.002

Kiyonaga, A., & Egner, T. (2016). Center-surround inhibition in working memory. *Current Biology*, *26*(1), 64–68. https://doi.org/10.1016/j.cub.2015.11.013

Koch, C., & Ullman, S. (1987). Shifts in selective visual attention: Towards the underlying neural circuitry. In L. M. Vaina (Ed.), *Matters of Intelligence* (pp. 115–141). https://doi.org/10.1007/978-94-009-3833-5_5

Koivisto, M., Mäntylä, T., & Silvanto, J. (2010). The role of early visual cortex (V1/V2) in conscious and unconscious visual perception. *NeuroImage*, *51*(2), 828–834. https://doi.org/10.1016/j.neuroimage.2010.02.042

Koivisto, M., Railo, H., Revonsuo, A., Vanni, S., & Salminen-Vaparanta, N. (2011). Recurrent processing in V1/V2 contributes to categorization of natural scenes. *Journal of Neuroscience*, *31*(7), 2488–2492. https://doi.org/10.1523/jneurosci.3074-10.2011

Koivisto, M., & Silvanto, J. (2011). Relationship between visual binding, reentry and awareness. *Consciousness and Cognition*, *20*(4), 1293–1303. https://doi.org/10.1016/j.concog.2011.02.008

Koivisto, M., & Silvanto, J. (2012). Visual feature binding: The critical time windows of V1/V2 and parietal activity. *NeuroImage*, *59*(2), 1608–1614. https://doi.org/10.1016/j.neuroimage.2011.08.089

Kok, P., Rahnev, D., Jehee, J. F. M., Lau, H. C., & de Lange, F. P. (2012). Attention reverses the effect of prediction in silencing sensory signals. *Cerebral Cortex*, *22*(9), 2197–2206. https://doi.org/10.1093/cercor/bhr310

Kveraga, K., Boshyan, J., & Bar, M. (2007). Magnocellular projections as the trigger of top-down facilitation in recognition. *Journal of Neuroscience*, *27*(48), 13232–13240. https://doi.org/10.1523/JNEUROSCI.3481-07.2007

LaBerge, D., Auclair, L., & Sieroff, E. (2000). Preparatory attention: Experiment and theory. *Consciousness and Cognition*, *9*(3), 396–434. https://doi.org/10.1006/ccog.1999.0429

Laberge, D., & Brown, V. (1986). Variations in size of the visual field in which targets are presented: An attentional range effect. *Perception & Psychophysics*, *40*(3), 188–200. https://doi.org/10.3758/BF03203016

LaBerge, D., & Brown, V. (1989). Theory of attentional operations in shape identification. *Psychological Review*, *96*(1), 101–124. https://doi.org/10.1037/0033-295X.96.1.101

Lamme, V. A. F., Zipser, K., & Spekreijse, H. (2002). Masking interrupts figure-ground signals in V1. *Journal of Cognitive Neuroscience*, *14*(7), 1044–1053. https://doi.org/10.1162/089892902320474490

Lamme, V. A., Supèr, H., & Spekreijse, H. (1998). Feedforward, horizontal, and feedback processing in the visual cortex. *Current Opinion in Neurobiology*, *8*(4), 529–535. https://doi.org/10.1016/S0959-4388(98)80042-1

Lauritzen, T. Z., D'Esposito, M., Heeger, D. J., & Silver, M. A. (2009). Top-down flow of visual spatial attention signals from parietal to occipital cortex. *Journal of Vision*, *9*(13), 18. https://doi.org/10.1167/9.13.18

Layher, G., Giese, M. A., & Neumann, H. (2014a). Learning representations of animated motion sequences-A neural model. *Topics in Cognitive Science*, *6*(1), 170–182. https://doi.org/10.1111/tops.12075

Layher, G., Schrodt, F., Butz, M. V., & Neumann, H. (2014b). Adaptive learning in a compartmental model of visual cortex-how feedback enables stable category learning and refinement. *Frontiers in Psychology*, *5*, 1287. https://doi.org/10.3389/fpsyg.2014.01287

Layton, O. W., Mingolla, E., & Yazdanbakhsh, A. (2014). Neural dynamics of feedforward and feedback processing in figure-ground segregation. *Frontiers in Psychology*, *5*, 972. https://doi.org/10.3389/fpsyg.2014.00972

Lee, T. S., & Mumford, D. (2003). Hierarchical Bayesian inference in the visual cortex. *Journal of the Optical Society of America A*, *20*(7), 1434–1448. https://doi.org/10.1364/JOSAA.20.001434

Lennie, P., Krauskopf, J., & Sclar, G. (1990). Chromatic mechanisms in striate cortex of macaque. *The Journal of Neuroscience*, *10*(2), 649–669. https://doi.org/10.1523/JNEUROSCI.10-02-00649.1990

Li, W., Piëch, V., & Gilbert, C. D. (2006). Contour saliency in primary visual cortex. *Neuron*, *50*(6), 951–962. https://doi.org/10.1016/j.neuron.2006.04.035

Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., … Zitnick, C. L. (2014). Microsoft COCO: Common objects in context. In *ECCV* (pp. 740–755). https://doi.org/10.1007/978-3-319-10602-1_48

Linkenkaer-Hansen, K., Palva, J. M., Sams, M., Hietanen, J. K., Aronen, H. J., & Ilmoniemi, R. J. (1998). Face-selective processing in human extrastriate cortex around 120 ms after stimulus onset revealed by magneto- and electroencephalography. *Neuroscience Letters*, *253*(3), 147–150. https://doi.org/10.1016/S0304-3940(98)00586-2

Liu, H., Agam, Y., Madsen, J. R., & Kreiman, G. (2009). Timing, timing, timing: Fast decoding of object information from intracranial field potentials in human visual cortex. *Neuron*, *62*(2), 281–290. https://doi.org/10.1016/j.neuron.2009.02.025

Liu, J., Harris, A., & Kanwisher, N. (2002). Stages of processing in face perception: An MEG study. *Nature Neuroscience*, *5*(9), 910–916. https://doi.org/10.1038/nn909

Loach, D., Frischen, A., Bruce, N., & Tsotsos, J. K. (2008). An attentional mechanism for selecting appropriate actions afforded by graspable objects. *Psychological Science*, *19*(12), 1253–1257. https://doi.org/10.1111/j.1467-9280.2008.02234.x

MacEvoy, S. P., & Epstein, R. A. (2007). Position selectivity in scene- and object-responsive occipitotemporal regions. *Journal of Neurophysiology*, *98*(4), 2089–2098. https://doi.org/10.1152/jn.00438.2007

Macmillan, N. A., & Creelman, C. D. (2005). Detection Theory: A User's Guide. In *Detection theory: A user's guide (2nd ed.).* https://doi.org/10.1017/CBO9781107415324.004

Marshak, W., & Sekuler, R. (1979). Mutual repulsion between moving visual targets. *Science*, *205*(4413), 1399–1401. https://doi.org/10.1126/science.472756

Martinez-Trujillo, J. C., & Treue, S. (2004). Feature-based attention increases the selectivity of population responses in primate visual cortex. *Current Biology*, *14*(9), 744–751. https://doi.org/10.1016/j.cub.2004.04.028

Martínez, A., Anllo-Vento, L., Sereno, M. I., Frank, L. R., Buxton, R. B., Dubowitz, D. J., … Hillyard, S. A. (1999). Involvement of striate and extrastriate visual cortical areas in spatial attention. *Nature Neuroscience*, *2*(4), 364–369. https://doi.org/10.1038/7274

Martínez, A., Di Russo, F., Anllo-Vento, L., Sereno, M. I., Buxton, R. B., & Hillyard, S. A. (2001). Putting spatial attention on the map: Timing and localization of stimulus selection processes in striate and extrastriate visual areas. *Vision Research*, *41*(10–11), 1437–1457. https://doi.org/10.1016/S0042-6989(00)00267-4

Mather, G., & Moulden, B. (1980). A simultaneous shift in apparent direction: further evidence for a "distribution-shift" model of direction coding. *Quarterly Journal of Experimental Psychology*, *32*(2), 325–333. https://doi.org/10.1080/14640748008401168

Maunsell, J. H. R., & Cook, E. P. (2002). The role of attention in visual processing. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *357*(1424), 1063–1072. https://doi.org/10.1098/rstb.2002.1107

Maunsell, J. H. R., & Treue, S. (2006). Feature-based attention in visual cortex. *Trends in Neurosciences*, *29*(6), 317–322. https://doi.org/10.1016/j.tins.2006.04.001

McCarley, J. S., & Mounts, J. R. W. (2007). Localized attentional interference affects object individuation, not feature detection. *Perception*, *36*(1), 17–32. https://doi.org/10.1068/p5598

McMains, S. A., & Somers, D. C. (2004). Multiple spotlights of attentional selection in human visual cortex. *Neuron*, *42*(4), 677–686. https://doi.org/10.1016/S0896-6273(04)00263-6

Mehta, A. D., Ulbert, I., & Schroeder, C. E. (2000). Intermodal selective attention in monkeys. I: distribution and timing of effects across visual areas. *Cerebral Cortex*, *10*(4), 343–358. https://doi.org/10.1093/cercor/10.4.343

Miconi, T., & VanRullen, R. (2016). A feedback model of attention explains the diverse effects of attention on neural firing rates and receptive field structure. *PLOS Computational Biology*, *12*(2), e1004770. https://doi.org/10.1371/journal.pcbi.1004770

Mihalas, S., Dong, Y., von der Heydt, R., & Niebur, E. (2011). Mechanisms of perceptual organization provide auto-zoom and auto-localization for attention to objects. *Proceedings of the National Academy of Sciences*, *108*(18), 7583–7588.

Mohsenzadeh, Y., Qin, S., Cichy, R. M., & Pantazis, D. (2018). Ultra-rapid serial visual presentation reveals dynamics of feedforward and feedback processes in the ventral visual pathway. *ELife*, *7*, e36329. https://doi.org/10.7554/eLife.36329

Moores, E., Laiti, L., & Chelazzi, L. (2003). Associative knowledge controls deployment of visual selective attention. *Nature Neuroscience*, *6*(2), 182–189. https://doi.org/10.1038/nn996

Moran, J., & Desimone, R. (1985). Selective attention gates visual processing in the extrastriate cortex. *Science*, *229*(4715), 782–784. https://doi.org/10.1126/science.4023713

Morishima, Y., Akaishi, R., Yamada, Y., Okuda, J., Toma, K., & Sakai, K. (2009). Task-specific signal transmission from prefrontal cortex in visual selective attention. *Nature Neuroscience*, *12*(1), 85–91. https://doi.org/10.1038/nn.2237

Mouchetant-Rostaing, Y., Giard, M. H., Delpuech, C., Echallier, J. F., & Pernier, J. (2000). Early signs of visual categorization for biological and non-biological stimuli in humans. *NeuroReport*, *11*(11), 2521–2525. https://doi.org/10.1097/00001756-200008030-00035

Mounts, J. R. W. (2000a). Attentional capture by abrupt onsets and feature singletons produces inhibitory surrounds. *Perception and Psychophysics*, *62*(7), 1485–1493. https://doi.org/10.3758/BF03212148

Mounts, J. R. W. (2000b). Evidence for suppressive mechanisms in attentional selection: Feature singletons produce inhibitory surrounds. *Perception and Psychophysics*, *62*(5), 969–983. https://doi.org/10.3758/BF03212082

Muckli, L., De Martino, F., Vizioli, L., Petro, L. S., Smith, F. W., Ugurbil, K., … Yacoub, E. (2015). Contextual feedback to superficial layers of V1. *Current Biology*, *25*(20), 2690–2695. https://doi.org/10.1016/j.cub.2015.08.057

Müller, M. M., & Hübner, R. (2002). Can the spotlight of attention be shaped like a doughnut? Evidence from steady-state visual evoked Potential. *Psychological Science*, *13*(2), 119–124. https://doi.org/10.1111/1467-9280.00422

Müller, M. M., Malinowski, P., Gruber, T., & Hillyard, S. A. (2003). Sustained division of the attentional spotlight. *Nature*, *424*(6946), 309–312. https://doi.org/10.1038/nature01812

Müller, N. G., & Kleinschmidt, A. (2004). The attentional “spotlight’s” penumbra: Center-surround modulation in striate cortex. *NeuroReport*, *15*(6), 977–980.

https://doi.org/10.1097/00001756-200404290-00009

Müller, N. G., Mollenhauer, M., Rösler, A., & Kleinschmidt, A. (2005). The attentional field has a Mexican hat distribution. *Vision Research*, *45*(9), 1129–1137. https://doi.org/10.1016/j.visres.2004.11.003

Mumford, D. (1991). On the computational architecture of the neocortex. *Biological Cybernetics*, *65*(2), 135–145. https://doi.org/10.1007/BF00202389

Navalpakkam, V., & Itti, L. (2006). An integrated model of top-down and bottom-up attention for optimizing detection speed. *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*, *2*, 2049–2056. https://doi.org/10.1109/CVPR.2006.54

Niebergall, R., Khayat, P. S., Treue, S., & Martinez-Trujillo, J. C. (2011). Expansion of MT neurons excitatory receptive fields during covert attentive tracking. *Journal of Neuroscience*, *31*(43), 15499–15510. https://doi.org/10.1523/JNEUROSCI.2822-11.2011

Niemeier, M., Goltz, H. C., Kuchinad, A., Tweed, D. B., & Vilis, T. (2005). A contralateral preference in the lateral occipital area: Sensory and attentional mechanisms. *Cerebral Cortex*, *15*(3), 325–331. https://doi.org/10.1093/cercor/bhh134

Noesselt, T., Hillyard, S. A., Woldorff, M. G., Schoenfeld, A., Hagner, T., Jäncke, L., … Heinze, H.-J. (2002). Delayed striate cortical activation during spatial attention. *Neuron*, *35*(3), 575–587. https://doi.org/10.1016/S0896-6273(02)00781-X

Noudoost, B., Chang, M. H., Steinmetz, N. A., & Moore, T. (2010). Top-down control of visual attention. *Current Opinion in Neurobiology*, *20*(2), 183–190. https://doi.org/10.1016/j.conb.2010.02.003

O'Reilly, R. C., Wyatte, D., Herd, S., Mingus, B., & Jilk, D. J. (2013). Recurrent processing during object recognition. *Frontiers in Psychology*, *4*, 124. https://doi.org/10.3389/fpsyg.2013.00124

Olson, R. K., & Attneave, F. (1970). What variables produce similarity grouping? *The American Journal of Psychology*, *83*(1), 1. https://doi.org/10.2307/1420852

Ozeki, H., Finn, I. M., Schaffer, E. S., Miller, K. D., & Ferster, D. (2009). Inhibitory stabilization of the cortical network underlies visual surround suppression. *Neuron*, *62*(4), 578–592. https://doi.org/10.1016/j.neuron.2009.03.028

Palmer, tephen E. (1975). The effects of contextual scenes on the identification of objects. *Memory & Cognition*, *3*(5), 519–526. https://doi.org/10.3758/BF03197524

Palmer, J., Huk, A. C., & Shadlen, M. N. (2005). The effect of stimulus strength on the speed and accuracy of a perceptual decision. *Journal of Vision*, *5*(5), 376–404. https://doi.org/10.1167/5.5.1

Pascual-Leone, A., & Walsh, V. (2001). Fast backprojections from the motion to the primary visual area necessary for visual awareness. *Science*, *292*(5516), 510–512. https://doi.org/10.1126/science.1057099

Pasupathy, A., & Connor, C. E. (2002). Population coding of shape in area V4. *Nature Neuroscience*, *5*(12), 1332–1338. https://doi.org/10.1038/nn972

Peelen, M. V., & Kastner, S. (2011). A neural basis for real-world visual search in human occipitotemporal cortex. *Proceedings of the National Academy of Sciences*, *108*(29), 12125–12130. https://doi.org/10.1073/pnas.1101042108

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, *10*(4), 437–442. https://doi.org/10.1163/156856897X00366

Pelli, D. G., Palomares, M., & Majaj, N. J. (2004). Crowding is unlike ordinary masking: Distinguishing feature integration from detection. *Journal of Vision*, *4*(12), 12. https://doi.org/10.1167/4.12.12

Perry, C. J., & Fallah, M. (2012). Color improves speed of processing but not perception in a motion illusion. *Frontiers in Psychology*, *3*, 92. https://doi.org/10.3389/fpsyg.2012.00092

Perry, C. J., & Fallah, M. (2014). Feature integration and object representations along the dorsal stream visual hierarchy. *Frontiers in Computational Neuroscience*, *8*, 84. https://doi.org/10.3389/fncom.2014.00084

Perry, C. J., Tahiri, A., & Fallah, M. (2014). Feature integration within and across visual streams occurs at different visual processing stages. *Journal of Vision*, *14*(2), 10. https://doi.org/10.1167/14.2.10

Peyrin, C., Michel, C. M., Schwartz, S., Thut, G., Seghier, M., Landis, T., … Vuilleumier, P. (2010). The neural substrates and timing of top-down processes during coarse-to-fine categorization of visual scenes: A combined fMRI and ERP study. *Journal of Cognitive Neuroscience*, *22*(12), 2768–2780. https://doi.org/10.1162/jocn.2010.21424

Pomerleau, V. J., Fortier-Gauthier, U., Corriveau, I., Dell'Acqua, R., & Jolicœur, P. (2014). Colour-specific differences in attentional deployment for equiluminant pop-out colours: Evidence from lateralised potentials. *International Journal of Psychophysiology*, *91*(3), 194–205. https://doi.org/10.1016/j.ijpsycho.2013.10.016

Poort, J., Raudies, F., Wannig, A., Lamme, V. A. F., Neumann, H., & Roelfsema, P. R. (2012). The role of attention in figure-ground segregation in areas V1 and V4 of the visual cortex. *Neuron*, *75*(1), 143–156. https://doi.org/10.1016/j.neuron.2012.04.032

Posner, M. I. (1980). Orienting of attention. *Quarterly Journal of Experimental Psychology*, *32*(1), 3–25. https://doi.org/10.1080/00335558008248231

Posner, M. I., Snyder, C. R., & Davidson, B. J. (1980). Attention and the detection of signals. *Journal of Experimental Psychology: General*, *109*(2), 160–174. https://doi.org/10.1037/0096-3445.109.2.160

Potter, M. C. (1976). Short-term conceptual memory for pictures. *Journal of Experimental Psychology. Human Learning and Memory*, *2*(5), 509–522. https://doi.org/10.1037/0278-7393.2.5.509

Potter, M. C., Wyble, B., Hagmann, C. E., & McCourt, E. S. (2014). Detecting meaning in RSVP at 13 ms per picture. *Attention, Perception, and Psychophysics*, *76*(2), 270–279. https://doi.org/10.3758/s13414-013-0605-z

Praß, M., Grimsen, C., König, M., & Fahle, M. (2013). Ultra rapid object categorization: Effects of level, animacy and context. *PLoS ONE*, *8*(6), e68051. https://doi.org/10.1371/journal.pone.0068051

Puckett, A. M., & DeYoe, E. A. (2015). The attentional field revealed by single-voxel modeling of fMRI time courses. *Journal of Neuroscience*, *35*(12), 5030–5042. https://doi.org/10.1523/JNEUROSCI.3754-14.2015

Puri, A. M., Wojciulik, E., & Ranganath, C. (2009). Category expectation modulates baseline and stimulus-evoked activity in human inferotemporal cortex. *Brain Research*, *1301*, 89–99. https://doi.org/10.1016/j.brainres.2009.08.085

Quian Quiroga, R., Reddy, L., Kreiman, G., Koch, C., & Fried, I. (2005). Invariant visual representation by single neurons in the human brain. *Nature*, *435*(7045), 1102–1107. https://doi.org/10.1038/nature03687

Rao, R. P. N., & Ballard, D. H. (1999). Predictive coding in the visual cortex: A functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience*, *2*(1), 79–87. https://doi.org/10.1038/4580

Remington, A., Swettenham, J., Campbell, R., & Coleman, M. (2009). Selective attention and perceptual load in autism spectrum disorder. *Psychological Science*, *20*(11), 1388–1393. https://doi.org/10.1111/j.1467-9280.2009.02454.x

Reynolds, J. H., & Chelazzi, L. (2004). Attentional modulation of visual processing. *Annual Review of Neuroscience*, *27*(1), 611–647. https://doi.org/10.1146/annurev.neuro.26.041002.131039

Reynolds, J. H., Chelazzi, L., & Desimone, R. (1999). Competitive mechanisms subserve attention in macaque areas V2 and V4. *The Journal of Neuroscience*, *19*(5), 1736–1753. https://doi.org/10.1523/JNEUROSCI.19-05-01736.1999

Reynolds, J. H., & Heeger, D. J. (2009). The normalization model of attention. *Neuron*, *61*(2), 168–185. https://doi.org/10.1016/j.neuron.2009.01.002

Riesenhuber, M., & Poggio, T. (1999). Hierarchical models of object recognition in cortex. *Nature Neuroscience*, *2*(11), 1019–1025. https://doi.org/10.1038/14819

Riesenhuber, M., & Poggio, T. (2002). Neural mechanisms of object recognition. *Current Opinion in Neurobiology*, *12*(2), 162–168. https://doi.org/10.1016/S0959-4388(02)00304-5

Righi, G., Peissig, J. J., & Tarr, M. J. (2012). Recognizing disguised faces. *Visual Cognition*, *20*(2), 143–169. https://doi.org/10.1080/13506285.2012.654624

Ro, T., Breitmeyer, B., Burton, P., Singhal, N. S., & Lane, D. (2003). Feedback contributions to visual awareness in human occipital cortex. *Current Biology*, *13*(12), 1038–1041. https://doi.org/10.1016/S0960-9822(03)00337-3

Robinson, D. A. (1963). A method of neasuring eye movemnent using a scieral search coil in a magnetic field. *IEEE Transactions on Bio-Medical Electronics*, *10*(4), 137–145. https://doi.org/10.1109/TBMEL.1963.4322822

Roelfsema, P. R. (2006). Cortical algorithms for perceptual grouping. *Annual Review of Neuroscience*, *29*(1), 203–227. https://doi.org/10.1146/annurev.neuro.29.051605.112939

Roelfsema, P. R., & Houtkamp, R. (2011). Incremental grouping of image elements in vision. *Attention, Perception, & Psychophysics*, *73*(8), 2542–2572. https://doi.org/10.3758/s13414-011-0200-0

Roelfsema, P. R., Lamme, V. A. F., & Spekreijse, H. (1998). Object-based attention in the primary visual cortex of the macaque monkey. *Nature*, *395*(6700), 376–381. https://doi.org/10.1038/26475

Roelfsema, P. R., Lamme, V. A. F., Spekreijse, H., & Bosch, H. (2002). Figure-ground segregation in a recurrent network architecture. *Journal of Cognitive Neuroscience*, *14*(4), 525–537. https://doi.org/10.1162/08989290260045756

Roland, P. E. (2010). Six principles of visual cortical dynamics. *Frontiers in Systems Neuroscience*, *4*, 28. https://doi.org/10.3389/fnsys.2010.00028

Roland, P. E., Hanazawa, A., Undeman, C., Eriksson, D., Tompa, T., Nakamura, H., … Ahmed, B. (2006). Cortical feedback depolarization waves: A mechanism of top-down influence on early visual areas. *Proceedings of the National Academy of Sciences*, *103*(33), 12586–12591. https://doi.org/10.1073/pnas.0604925103

Romani, G., Williamson, S., & Kaufman, L. (1982). Tonotopic organization of the human auditory cortex. *Science*, *216*(4552), 1339–1340. https://doi.org/10.1126/science.7079770

Ronconi, L., Gori, S., Federici, A., Devita, M., Carna, S., Sali, M. E., … Facoetti, A. (2018). Weak surround suppression of the attentional focus characterizes visual selection in the ventral stream in autism. *NeuroImage: Clinical*, *18*, 912–922. https://doi.org/10.1016/j.nicl.2018.02.014

Rothenstein, A. L., Rodríguez-Sánchez, A. J., Simine, E., & Tsotsos, J. K. (2008). Visual feature binding within the Selective Tuning attention framework. *International Journal of Pattern Recognition and Artificial Intelligence*, *22*(5), 861–881. https://doi.org/10.1142/S0218001408006648

Rothenstein, A. L., & Tsotsos, J. K. (2014). Attentional Modulation and Selection – An Integrated Approach. *PLoS ONE*, *9*(6), e99681. https://doi.org/10.1371/journal.pone.0099681

Rousselet, G. A., Fabre-Thorpe, M., & Thorpe, S. J. (2002). Parallel processing in high-level categorization of natural images. *Nature Neuroscience*, *5*(7), 629–630. https://doi.org/10.1038/nn866

Rousselet, G. A., Macé, M. J.-M., & Fabre-Thorpe, M. (2003). Is it an animal? Is it a human face? Fast processing in upright and inverted natural scenes. *Journal of Vision*, *3*(6), 440–455. https://doi.org/10.1167/3.6.5

Rubenstein, J. L. R., & Merzenich, M. M. (2003). Model of autism: increased ratio of excitation/inhibition in key neural systems. *Genes, Brain and Behavior*, *2*(5), 255–267. https://doi.org/10.1034/j.1601-183X.2003.00037.x

Saalmann, Y. B., Pigarev, I. N., & Vidyasagar, T. R. (2007). Neural mechanisms of visual attention: How top-down feedback highlights relevant locations. *Science*, *316*(5831), 1612–1615. https://doi.org/10.1126/science.1139140

Sàenz, M., Buracas, G. T., & Boynton, G. M. (2002). Global effects of feature-based attention in human visual cortex. *Nature Neuroscience*, *5*(7), 631–632. https://doi.org/10.1038/nn876

Sàenz, M., Buraĉas, G. T., & Boynton, G. M. (2003). Global feature-based attention for motion and color. *Vision Research*, *43*(6), 629–637. https://doi.org/10.1016/S0042-6989(02)00595-3

Salin, P.-A., & Bullier, J. (1995). Corticocortical connections in the visual system: Structure and function. *Physiological Reviews*, *75*(1), 107–154. https://doi.org/10.1152/physrev.1995.75.1.107

Sayres, R., & Grill-Spector, K. (2008). Relating retinotopic and object-selective responses in human lateral occipital cortex. *Journal of Neurophysiology*, *100*(1), 249–267. https://doi.org/10.1152/jn.01383.2007

Scalf, P. E., & Beck, D. M. (2010). Competition in visual cortex impedes attention to multiple items. *Journal of Neuroscience*, *30*(1), 161–169. https://doi.org/10.1523/JNEUROSCI.4207-09.2010

Schall, J. D., Sato, T. R., Thompson, K. G., Vaughn, A. A., & Juan, C.-H. (2004). Effects of search efficiency on surround suppression during visual selection in frontal eye field. *Journal of Neurophysiology*, *91*(6), 2765–2769. https://doi.org/10.1152/jn.00780.2003

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. https://doi.org/10.1017/S004857729897010X

Schulz, M. F., & Sanocki, T. (2003). Time course of perceptual grouping by color. *Psychological Science*, *14*(1), 26–30. https://doi.org/10.1111/1467-9280.01414

Schwarzlose, R. F., Swisher, J. D., Dang, S., & Kanwisher, N. (2008). The distribution of category and location information across object-selective regions in human visual cortex. *Proceedings of the National Academy of Sciences*, *105*(11), 4447–4452.

https://doi.org/10.1073/pnas.0800431105

Schyns, P. G., & Oliva, A. (1994). From blobs to boundary edges: Evidence for time- and spatial-scale-dependent scene recognition. *Psychological Science*, *5*(4), 195–200. https://doi.org/10.1111/j.1467-9280.1994.tb00500.x

Seghier, M. L., & Vuilleumier, P. (2006). Functional neuroimaging findings on the human perception of illusory contours. *Neuroscience & Biobehavioral Reviews*, *30*(5), 595–612. https://doi.org/10.1016/j.neubiorev.2005.11.002

Self, M. W., Lorteije, J. A. M., Vangeneugden, J., van Beest, E. H., Grigore, M. E., Levelt, C. N., … Roelfsema, P. R. (2014). Orientation-tuned surround suppression in mouse visual cortex. *Journal of Neuroscience*, *34*(28), 9290–9304. https://doi.org/10.1523/JNEUROSCI.5051-13.2014

Self, M. W., van Kerkoerle, T., Supèr, H., & Roelfsema, P. R. (2013). Distinct roles of the cortical layers of area V1 in figure-ground segregation. *Current Biology*, *23*(21), 2121–2129. https://doi.org/10.1016/j.cub.2013.09.013

Serre, T., Oliva, A., & Poggio, T. (2007). A feedforward architecture accounts for rapid categorization. *Proceedings of the National Academy of Sciences*, *104*(15), 6424–6429. https://doi.org/10.1073/pnas.0700622104

Shpaner, M., Molholm, S., Forde, E., & Foxe, J. J. (2013). Disambiguating the roles of area V1 and the lateral occipital complex (LOC) in contour integration. *NeuroImage*, *69*, 146–156. https://doi.org/10.1016/j.neuroimage.2012.11.023

Shulman, G. L., Remington, R. W., & McLean, J. P. (1979). Moving attention through visual space. *Journal of Experimental Psychology: Human Perception and Performance*, *5*(3), 522–526. https://doi.org/10.1037/0096-1523.5.3.522

Sigman, M., & Gilbert, C. D. (2000). Learning to find a shape. *Nature Neuroscience*, *3*(3), 264–269. https://doi.org/10.1038/72979

Silvanto, J., Cowey, A., Lavie, N., & Walsh, V. (2005). Striate cortex (V1) activity gates awareness of motion. *Nature Neuroscience*, *8*(2), 143–144. https://doi.org/10.1038/nn1379

Silverstein, D. N. (2015). A computational investigation of feedforward and feedback processing in metacontrast backward masking. *Frontiers in Psychology*, *6*, 6. https://doi.org/10.3389/fpsyg.2015.00006

Singer, W. (1999). Neuronal synchrony: A versatile code for the definition of relations? *Neuron*, *24*(1), 49–65. https://doi.org/10.1016/S0896-6273(00)80821-1

Skelton, J. M., & Eriksen, C. W. (1976). Spatial characteristics of selective attention in letter matching. *Bulletin of the Psychonomic Society*, *7*(2), 136–138.

Smith, A. T., Singh, K. D., Williams, A. L., & Greenlee, M. W. (2001). Estimating receptive field size from fMRI data in human striate and extrastriate visual cortex. *Cerebral Cortex*, *11*(12), 1182–1190. https://doi.org/10.1093/cercor/11.12.1182

Snow, J. C., Skiba, R. M., Coleman, T. L., & Berryhill, M. E. (2014). Real-world objects are more memorable than photographs of objects. *Frontiers in Human Neuroscience*, *8*, 837. https://doi.org/10.3389/fnhum.2014.00837

Snowden, R. J., Treue, S., & Andersen, R. A. (1992). The response of neurons in areas V1 and MT of the alert rhesus monkey to moving random dot patterns. *Experimental Brain Research*, *88*(2), 389–400. https://doi.org/10.1007/BF02259114

Spratling, M. W. (2010). Predictive coding as a model of response properties in cortical area V1. *Journal of Neuroscience*, *30*(9), 3531–3543. https://doi.org/10.1523/JNEUROSCI.4911-09.2010

Spratling, M. W., & Johnson, M. H. (2004). A feedback model of visual attention. *Journal of Cognitive Neuroscience*, *16*(2), 219–237. https://doi.org/10.1162/089892904322984526

Stokes, M., Thompson, R., Cusack, R., & Duncan, J. (2009). Top-down activation of shape-specific population codes in visual cortex during mental imagery. *Journal of Neuroscience*, *29*(5), 1565–1572. https://doi.org/10.1523/JNEUROSCI.4657-08.2009

Störmer, V. S., & Alvarez, G. A. (2014). Feature-based attention elicits surround suppression in feature space. *Current Biology*, *24*(17), 1985–1988. https://doi.org/10.1016/j.cub.2014.07.030

Summerfield, C., Trittschuh, E. H., Monti, J. M., Mesulam, M. M., & Egner, T. (2008). Neural repetition suppression reflects fulfilled perceptual expectations. *Nature Neuroscience*, *11*(9), 1004–1006. https://doi.org/10.1038/nn.2163

Summerfield, J. J., Lepsien, J., Gitelman, D. R., Mesulam, M. M., & Nobre, A. C. (2006). Orienting attention based on long-term memory experience. *Neuron*, *49*(6), 905–916. https://doi.org/10.1016/j.neuron.2006.01.021

Tchernikov, I., & Fallah, M. (2010). A color hierarchy for automatic target selection. *PLoS ONE*, *5*(2), e9338. https://doi.org/10.1371/journal.pone.0009338

Thorell, L. G., de Valois, R. L., & Albrecht, D. G. (1984). Spatial mapping of monkey VI cells with pure color and luminance stimuli. *Vision Research*, *24*(7), 751–769. https://doi.org/10.1016/0042-6989(84)90216-5

Thorpe, S., Fize, D., & Marlot, C. (1996). Speed of processing in the human visual system. *Nature*, *381*(6582), 520–522. https://doi.org/10.1038/381520a0

Tombu, M., & Tsotsos, J. K. (2008). Attending to orientation results in an inhibitory surround in orientation space. *Perception & Psychophysics*, *70*(1), 30–35.

https://doi.org/10.3758/PP.70.1.30

Tomita, H., Ohbayashi, M., Nakahara, K., Hasegawa, I., & Miyashita, Y. (1999). Top-down signal from prefrontal cortex in executive control of memory retrieval. *Nature*, *401*(6754), 699–703. https://doi.org/10.1038/44372

Torralba, A., Oliva, A., Castelhano, M. S., & Henderson, J. M. (2006). Contextual guidance of eye movements and attention in real-world scenes: The role of global features in object search. *Psychological Review*, *113*(4), 766–786. https://doi.org/10.1037/0033-295X.113.4.766

Treisman, A. (1985). Preattentive processing in vision. *Computer Vision, Graphics and Image Processing*, *31*(2), 156–177. https://doi.org/10.1016/S0734-189X(85)80004-9

Treisman, A. (1996). The binding problem. *Current Opinion in Neurobiology*, *6*(2), 171–178. https://doi.org/10.1016/S0959-4388(96)80070-5

Treisman, A. (1998). Feature binding, attention and object perception. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *353*(1373), 1295–1306. https://doi.org/10.1098/rstb.1998.0284

Treisman, A., & Gormican, S. (1988). Feature analysis in early vision: Evidence from search asymmetries. *Psychological Review*, *95*(1), 15–48. https://doi.org/10.1037//0033-295X.95.1.15

Treisman, A. M. (1969). Strategies and models of selective attention. *Psychological Review*, *76*(3), 282–299. https://doi.org/10.1037/h0027242

Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, *12*(1), 97–136. https://doi.org/10.1016/0010-0285(80)90005-5

Treisman, A., & Schmidt, H. (1982). Illusory conjunctions in the perception of objects. *Cognitive*

Psychology, *14*(1), 107–141. https://doi.org/10.1016/0010-0285(82)90006-8

Treisman, A., & Souther, J. (1985). Search asymmetry: A diagnostic for preattentive processing of separable features. *Journal of Experimental Psychology: General*, *114*(3), 285–310. https://doi.org/10.1037/0096-3445.114.3.285

Treue, S., & Martinez-Trujillo, J. C. (1999). Feature-based attention influences motion processing gain in macaque visual cortex. *Nature*, *399*(6736), 575–579. https://doi.org/10.1038/21176

Tsotsos, J. K. (1988). A “complexity level” analysis of immediate vision. *International Journal of Computer Vision*, *1*(4), 303–320. https://doi.org/10.1007/BF00133569

Tsotsos, J. K. (1990). Analyzing vision at the complexity level. *Behavioral and Brain Sciences*, *13*(3), 423–445. https://doi.org/10.1017/S0140525X00079577

Tsotsos, J. K. (1991). Is complexity theory appropriate for analyzing biological systems? *Behavioral and Brain Sciences*, *14*(4), 770–773. https://doi.org/10.1017/S0140525X00072484

Tsotsos, J. K. (1992). On the relative complexity of active vs. passive visual search. *International Journal of Computer Vision*, *7*(2), 127–141. https://doi.org/10.1007/BF00128132

Tsotsos, J. K. (2011). *A computational perspective on visual attention*. Cambridge, MA: MIT Press.

Tsotsos, J. K., Culhane, S. M., Kei Wai, W. Y., Lai, Y., Davis, N., & Nuflo, F. (1995). Modeling visual attention via selective tuning. *Artificial Intelligence*, *78*(1–2), 507–545. https://doi.org/10.1016/0004-3702(95)00025-9

Tsotsos, J. K., & Kruijne, W. (2014). Cognitive programs: Software for attention’s executive.

Frontiers in Psychology, *5*, 1260. https://doi.org/10.3389/fpsyg.2014.01260

Tsotsos, J. K., Liu, Y., Martinez-Trujillo, J. C., Pomplun, M., Simine, E., & Zhou, K. (2005). Attending to visual motion. *Computer Vision and Image Understanding*, *100*(1-2 SPEC. ISS.), 3–40. https://doi.org/10.1016/j.cviu.2004.10.011

Tsotsos, J. K., Rodríguez-Sánchez, A. J., Rothenstein, A. L., & Simine, E. (2008). The different stages of visual recognition need different attentional binding strategies. *Brain Research*, *1225*, 119–132. https://doi.org/10.1016/j.brainres.2008.05.038

Tzvetanov, T., Womelsdorf, T., Niebergall, R., & Treue, S. (2006). Feature-based attention influences contextual interactions during motion repulsion. *Vision Research*, *46*(21), 3651–3658. https://doi.org/10.1016/j.visres.2006.05.018

Ullman, S. (1995). Sequence seeking and counter streams: A computational model for bidirectional information flow in the visual cortex. *Cerebral Cortex*, *5*(1), 1–11. https://doi.org/10.1093/cercor/5.1.1

Vanduffel, W., Tootell, R. B. H. ., & Orban, G. A. (2000). Attention-dependent suppression of metabolic activity in the early stages of the macaque visual system. *Cerebral Cortex*, *10*(2), 109–126. https://doi.org/10.1093/cercor/10.2.109

VanRullen, R., & Koch, C. (2003). Visual selective behavior can be triggered by a feed-forward process. *Journal of Cognitive Neuroscience*, *15*(2), 209–217. https://doi.org/10.1162/089892903321208141

VanRullen, R., & Thorpe, S. J. (2001). Is it a bird? Is it a plane? Ultra-rapid visual categorisation of natural and artifactual objects. *Perception*, *30*(6), 655–668. https://doi.org/10.1068/p3029

Vidyasagar, T. R. (1998). Gating of neuronal responses in macaque primary visual cortex by an attentional spotlight. *NeuroReport*, *9*(9), 1947–1952. https://doi.org/10.1097/00001756-

199806220-00006

Volberg, G., & Greenlee, M. W. (2014). Brain networks supporting perceptual grouping and contour selection. *Frontiers in Psychology*, *5*, 264. https://doi.org/10.3389/fpsyg.2014.00264

von der Malsburg, C. (1999). The what and why of binding: The modeler's perspective. *Neuron*, *24*(1), 95–104. https://doi.org/10.1016/S0896-6273(00)80825-9

Wagemans, J., Elder, J. H., Kubovy, M., Palmer, S. E., Peterson, M. A., Singh, M., & von der Heydt, R. (2012). A century of Gestalt psychology in visual perception: I. Perceptual grouping and figure–ground organization. *Psychological Bulletin*, *138*(6), 1172–1217. https://doi.org/10.1037/a0029333

Walker, S., Stafford, P., & Davis, G. (2008). Ultra-rapid categorization requires visual attention: Scenes with multiple foreground objects. *Journal of Vision*, *8*(4), 21. https://doi.org/10.1167/8.4.21

Wang, Y., Miller, J., & Liu, T. (2015). Suppression effects in feature-based attention. *Journal of Vision*, *15*(5), 15. https://doi.org/10.1167/15.5.15.doi

Watanabe, T., Harner, A. M., Miyauchi, S., Sasaki, Y., Nielsen, M., Palomo, D., & Mukai, I. (1998). Task-dependent influences of attention on the activation of human primary visual cortex. *Proceedings of the National Academy of Sciences*, *95*(19), 11489–11492. https://doi.org/10.1073/pnas.95.19.11489

Wessinger, C. M., Buonocore, M. H., Kussmaul, C. L., & Mangun, G. R. (1997). Tonotopy in human auditory cortex examined with functional magnetic resonance imaging. *Human Brain Mapping*, *5*(1), 18–25. https://doi.org/10.1002/(SICI)1097-0193(1997)5:1<18::AID-HBM3>3.0.CO;2-Q

Wilson, H. R., & Kim, J. (1994). A model for motion coherence and transparency. *Visual Neuroscience*, *11*(6), 1205–1220. https://doi.org/10.1017/S0952523800007008

Wilson, J. R., & Sherman, S. M. (1976). Receptive-field characteristics of neurons in cat striate cortex: Changes with visual field eccentricity. *Journal of Neurophysiology*, *39*(3), 512–533. https://doi.org/10.1152/jn.1976.39.3.512

Winawer, J., Horiguchi, H., Sayres, R. A., Amano, K., & Wandell, B. A. (2010). Mapping hV4 and ventral occipital cortex: The venous eclipse. *Journal of Vision*, *10*(5), 1. https://doi.org/10.1167/10.5.1

Wolfe, J. M. (1994). Guided Search 2.0 A revised model of visual search. *Psychonomic Bulletin & Review*, *1*(2), 202–238. https://doi.org/10.3758/BF03200774

Wolfe, J. M., Butcher, S. J., Lee, C., & Hyle, M. (2003). Changing your mind: On the contributions of top-down and bottom-up guidance in visual search for feature singletons. *Journal of Experimental Psychology: Human Perception and Performance*, *29*(2), 483–502. https://doi.org/10.1037/0096-1523.29.2.483

Wolfe, J. M., & Cave, K. R. (1999). The psychophysical evidence for a binding problem in human vision. *Neuron*, *24*(1), 11–17. https://doi.org/10.1016/S0896-6273(00)80818-1

Wolfe, J. M., Horowitz, T. S., & Michod, K. O. (2007). Is visual attention required for robust picture memory? *Vision Research*, *47*(7), 955–964. https://doi.org/10.1016/j.visres.2006.11.025

Womelsdorf, T., Anton-Erxleben, K., Pieper, F., & Treue, S. (2006). Dynamic shifts of visual receptive fields in cortical area MT by spatial attention. *Nature Neuroscience*, *9*(9), 1156–1160. https://doi.org/10.1038/nn1748

Womelsdorf, T., Anton-Erxleben, K., & Treue, S. (2008). Receptive field shift and shrinkage in macaque middle temporal area through attentional gain modulation. *Journal of Neuroscience*, *28*(36), 8934–8944. https://doi.org/10.1523/JNEUROSCI.4030-07.2008

Wong-Kee-You, A. M. B., Tsotsos, J. K., & Adler, S. A. (2019). Development of spatial suppression surrounding the focus of visual attention. *Journal of Vision*, *19*(7), 9. https://doi.org/10.1167/19.7.9

Wyatte, D., Curran, T., & O'Reilly, R. (2012a). The limits of feedforward vision: Recurrent processing promotes robust object recognition when objects are degraded. *Journal of Cognitive Neuroscience*, *24*(11), 2248–2261. https://doi.org/10.1162/jocn_a_00282

Wyatte, D., Herd, S., Mingus, B., & O'Reilly, R. (2012b). The role of competitive inhibition and top-down feedback in binding during object recognition. *Frontiers in Psychology*, *3*, 182. https://doi.org/10.3389/fpsyg.2012.00182

Wyatte, D., Jilk, D. J., & O'Reilly, R. C. (2014). Early recurrent feedback facilitates visual object recognition under challenging conditions. *Frontiers in Psychology*, *5*, 674. https://doi.org/10.3389/fpsyg.2014.00674

Yoo, S.-A., Tsotsos, J. K., & Fallah, M. (2018b). The attentional suppressive surround: Eccentricity, location-based and feature-based effects and interactions. *Frontiers in Neuroscience*, *12*, 710. https://doi.org/10.3389/fnins.2018.00710

Yoo, S.-A., Tsotsos, J. K., & Fallah, M. (2019). Feed-forward visual processing suffices for coarse localization but fine-grained localization in an attention-demanding context needs feedback processing. *PLOS ONE*, *14*(9), e0223166. https://doi.org/10.1371/journal.pone.0223166

Young, M., & Yamane, S. (1992). Sparse population coding of faces in the inferotemporal cortex. *Science*, *256*(5061), 1327–1331. https://doi.org/10.1126/science.1598577

Zeki, S. M. (1978). Uniformity and diversity of structure and function in rhesus monkey prestriate visual cortex. *The Journal of Physiology*, *277*(1), 273–290. https://doi.org/10.1113/jphysiol.1978.sp012272

Zhang, Y., Meyers, E. M., Bichot, N. P., Serre, T., Poggio, T. A., & Desimone, R. (2011). Object decoding with attention in inferior temporal cortex. *Proceedings of the National Academy of Sciences*, *108*(21), 8850–8855. https://doi.org/10.1073/pnas.1100999108

Zipser, K., Lamme, V. A. F., & Schiller, P. H. (1996). Contextual modulation in primary visual cortex. *The Journal of Neuroscience*, *16*(22), 7376–7389. https://doi.org/10.1523/JNEUROSCI.16-22-07376.1996