# To be selected or not to be selected: A modeling and behavioral study of the mechanisms underlying stimulus-driven and top-down visual attention

Proefschrift
ter verkrijging van
de graad van Doctor aan de Universiteit Leiden,
op gezag van de Rector Magnificus prof.mr. P. F. van der Heijden,
volgens besluit van het College voor Promoties
te verdedigen op dinsdag 26 juni 2007
klokke 16.15 uur

door

**Gwendid T. van der Voort van der Kleij**

geboren te Leiderdorp
in 1977

## Promotiecommissie

Promotor:
Prof. dr. B. Hommel

Copromotor:
Dr. F. van der Velde

Referent:
Prof. dr. K. R. Ridderinkhof

Overige leden:
Prof. dr. A. H. C. van der Heijden
Prof. dr. J. L. Theeuwes
Dr. M. de Kamps
Dr. G. Wolters

# Contents

## List of abbreviations

| | |
|---|---|
| ANOVA | analysis of variance |
| AIT | anterior inferotemporal cortex |
| CIT | central inferotemporal cortex |
| CLAM | closed-loop attention model |
| CRF | classical receptive field |
| FEF | frontal eye field |
| FIT | feature integration theory |
| GSM | global saliency model |
| PFC | prefrontal cortex |
| PIT | posterior inferotemporal cortex |
| PP | posterior parietal cortex |
| RMSE | root mean squared error |
| RT | response time |
| SC | superior colliculus |
| SOA | stimulus onset asynchrony |
| V-PFC | ventral prefrontal cortex |
| VWM | visual working memory |
| WTA | winner-takes-all |

# Chapter 1 | Introduction

## Selective visual attention

The human visual system is limited in the amount of visual information that it can process at a time. If our environment would provide only a modest amount of visual information at a time, our visual system could just process it all. In reality, however, our environment projects an overdose of visual information to our eyes. To cope with this overload of visual information, our visual system selects only part of the available visual information at a time for further processing, and processes the rest of the visual information less extensively. This process is called *selective visual attention*.

## Stimulus-driven and top-down visual attention

Ideally, our visual system processes the visual information at a given time that helps us to act successfully in our environment. Most of the time (or maybe even all the time) our actions are influenced by knowledge, expectations and current goals. Hence, it would be helpful if our visual system selects visual information consistent with knowledge, expectations, and current goals, i.e., *top-down visual attention*.

For example, suppose that you are playing a tennis match. For that task it is very important to select and process the visual information related to the ball. Selection of the (visual information related to the) ball may be facilitated by knowledge that the ball has a round shape, a yellow color, or by expectations that the ball will be located in a specific section of the tennis court (in case you are returning the opponent's serve).

Nonetheless, it is important that our visual system also processes visual information that is not consistent with knowledge, expectations, and current goals. We need the flexibility to perceive and act upon novel or unexpected stimuli in our environment. For example, when preparing to serve in a tennis match, it is better to pause when a streaker suddenly enters the tennis court. Thus, it would be useful if our visual system selects visual information, independent of knowledge, expectations, and current goals as well, i.e., *stimulus-driven visual attention*.

Behavioral and neuroimaging studies on humans and neurophysiological studies on monkeys have provided evidence for both stimulus-driven and top-down visual attention (for an overview, see Corbetta & Shulman, 2002).

Numerous behavioral studies indicated that our visual system automatically selects an object that is distinguished by a unique feature from other objects (such as a large difference in color, orientation, or size) (e.g., Treisman & Gelade, 1980; for an overview, see Wolfe & Horowitz, 2004). Thus it appears that mechanisms of stimulus-driven visual attention make the location of an object with unique features more conspicuous or *salient* than the location of objects with common features (Cave, 1999; Itti & Koch, 2000; Koch & Ullman, 1985; Li, 2002; Wolfe, 1994). This phenomenon may be termed *global saliency* to make a distinction from other phenomena of stimulus-driven visual attention (e.g., an abrupt onset singleton) (see Chapter 7). Nonetheless, the terms stimulus-driven visual attention and (global) saliency are used interchangeably in this thesis, since no other phenomena of stimulus-driven visual attention are investigated.

Other studies showed that stimuli can be selected on the basis of information about location (i.e., space-based visual attention) (for an overview, see Yantis & Serences, 2003), nonspatial features (e.g., color, shape, and motion) (i.e., feature-based visual attention) (e.g., Bichot, Rossi, & Desimone, 2005; Chawla, Rees, & Friston, 1999; Martinez-Trujillo & Treue, 2004; Motter, 1994a, 1994b; Saenz, Buracas, & Boynton, 2002), and complex nonspatial features (i.e., object-based visual attention) (e.g., Chelazzi, Miller, Duncan, & Desimone, 1993; O'Craven, Downing, & Kanwisher, 1999) (see Chapter 6).

## Visual search

Selective visual attention is typically studied in visual search (for an overview, see Wolfe & Horowitz, 2004). In visual search studies, participants search for a target among a number of other items, the distracters. The number of distracters, the setsize, is typically varied, and the time (or accuracy) to indicate the presence or absence of the target is measured. If the response time is (relatively) independent of the number of distracters, it is concluded that the target can be efficiently searched (selected) among the distracters. If the response time increases with the number of distracters, it is concluded that the target cannot be efficiently searched among the distracters.

When stimulus-driven visual attention is studied in visual search, participants do not know the features of the target. The target is distinguished by a unique feature (or conjunction of features) from the distracters (e.g., a green target among blue distracters or a blue target among green distracters), and participants have to indicate whether a deviant item is present or not. Such a target is called a *singleton*.

Efficient search for a singleton among distracters can therefore be attributed to stimulus-driven visual attention (although the task instruction to search for a singleton may play a role as well (cf., Bacon & Egeth, 1994)).

When top-down visual attention is studied in visual search, participants do know one or more features of the target (e.g., the color). The target features are given during the task instructions or are cued before a session or trial. Efficient search for such a *cued-target* among distracters can be attributed to a combination of stimulus-driven and top-down visual attention.

After more than two decades of visual search studies and other studies, there is still a lot of discussion about which mechanisms underlie stimulus-driven visual attention and top-down visual attention, and how these mechanisms interact. We give an overview of several important findings of visual search studies, and of theories and models that are proposed to explain these findings, in Chapter 6.

Evidently, the ability to search for objects is tightly linked with the ability to recognize objects. One model that aims to integrate the mechanisms that underlie visual search and object recognition is the Closed-Loop Attention Model (CLAM) (Van der Velde, De Kamps, & Van der Voort van der Kleij, 2004). In CLAM, visual search arises from interaction between visual working memory in the prefrontal cortex, object recognition in the ventral pathway, and spatial selection in the dorsal pathway. CLAM strongly influenced the questions that are addressed in this thesis. Therefore, CLAM is discussed below. After that, an outline of the thesis is presented.

## CLAM

Figure 1 illustrates the overall connection structure of CLAM. Modeled after the basic architecture of the (visual) cortex, the model consists of four parts. The first part consists of the (lower) retinotopic areas of the visual cortex (e.g., V2-PIT). The second part consists of the networks in area AIT of the ventral pathway that process object identity (e.g., shape, color) (i.e., the feature maps). The third part consists of the networks in area PP of the dorsal pathway that process location information of objects in the visual field, and that transform this information into spatial coordinates for specific movements (e.g., eye, body, head, arm) (i.e., the spatial maps). The fourth part consists of visual working memory areas in the prefrontal cortex. The four parts are connected in a diamond structure, with reciprocal connections. In this way, the diamond connection structure of CLAM forms a closed loop.

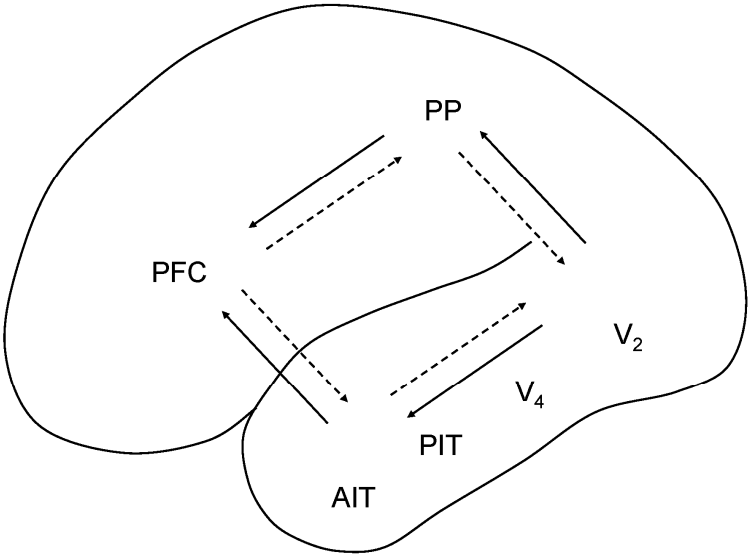


*Figure 1. The overall connection structure of CLAM. PFC = prefrontal cortex; AIT = anterior inferotemporal cortex; PIT = posterior inferotemporal cortex; PP = posterior parietal cortex.*

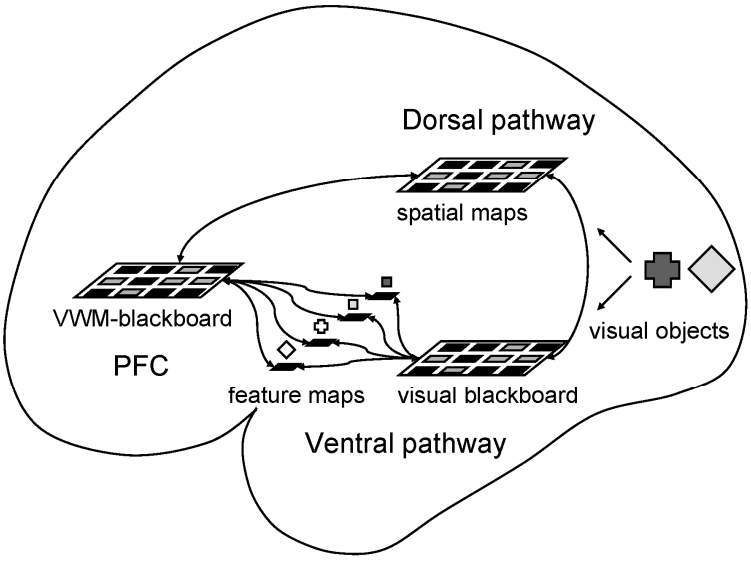


*Figure 2. The functional structure of CLAM.*

Figure 2 illustrates the functional structure of CLAM. Processing in CLAM starts in the retinotopic areas. The neurons in these areas have (relatively) small receptive fields and they typically encode conjunctions of elementary visual features. For instance, they encode elementary conjunctions of shape (e.g., orientation) with color, or conjunctions of shape with motion (e.g., an oriented bar moving in a particular direction). Because the areas are retinotopic, the neurons encode for location as well.

The ventral and dorsal pathways in CLAM emerge from the (lower) retinotopic areas. The ventral pathway transforms the retinotopic information into location invariant feature information about object identity. In Figure 2, the ventral pathway processes the feature information (i.e., shape, color) of a display that consists of a dark (blue) cross on the left and a light (yellow) diamond on the right. The dorsal pathway processes the spatial (location) information of the objects in this display. In CLAM, the ventral and dorsal pathway each consists of a combination of a feedforward network and a feedback network, which interact locally (Van der Velde & De Kamps, 2001).

Interaction between the ventral and dorsal pathway occurs in the retinotopic areas (e.g., V2-PIT). These areas function as a visual blackboard (Van der Velde & De Kamps, 2003) in which the features of an object (e.g., shape, color, location) can be related or bound. The notion of a *blackboard* derives from the fact that representations in these areas combine elementary feature information (e.g., shape, color) with location information. If one feature of an object (e.g., shape, color) is selected as a cue, the other features of the object (including its location) can be selected as well by means of an interaction process in the blackboard (i.e., feature-based or object-based visual attention). Likewise, the selection of the location of an object can be used to select the other features (e.g., shape, color) of the object by means of the interaction within the blackboard (i.e., space-based visual attention).

The ventral and dorsal pathway in CLAM also project (feedforward) to the prefrontal cortex (PFC). In the PFC, the features of a target object (or objects) are stored in a visual working memory (VWM) blackboard (Van der Velde & De Kamps, 2003). The VWM-blackboard in PFC is similar in nature to the visual blackboard in the visual cortex (e.g., on the level of retinotopic representation in PIT). It interacts with location invariant feature representations (e.g., shape, color) that are either located in the ventral pathway or in the PFC itself (or perhaps both). It also interacts with location representations that are either located in the PFC or in the

dorsal pathway (or both). The VWM-blackboard is used to bind the features (e.g., shape, color, location) of an object stored in visual working memory. The visual working memory in PFC projects back to the ventral and dorsal pathway, through the representations for features and location.

### Object-based visual attention in CLAM

Figure 3 illustrates the process of object-based (feature-based) visual attention in CLAM. A feature of a target object is stored in the VWM-blackboard. For instance, the shape of a cross (without a color) was presented earlier on the center of a display. Then, after a delay period, a display of two objects is presented, and the participant has to select the other features (e.g., color, location) of the cued object (i.e., the cross). In CLAM, the selection of the shape of a target object by a cue results in enhanced activation on the location of the target in the visual blackboard (V2-PIT). This enhanced activation results from the interaction between the feedforward network and the feedback network in the ventral pathway (Van der Velde & De Kamps, 2001). The feedforward network processes the identity of the objects in the display (e.g., shape, color). The feedback network in the ventral pathway carries the information of the cue back to the retinotopic areas (the visual blackboard). The cue-related activation in the feedback network is initiated by the information stored in the VWM-blackboard.

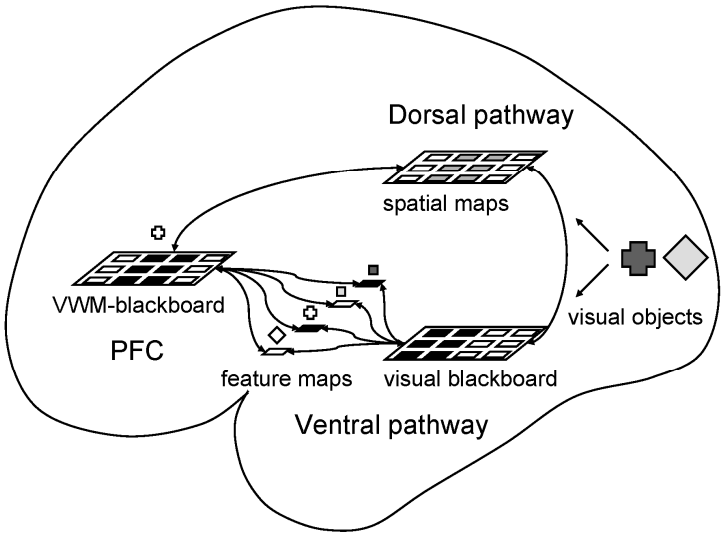


*Figure 3. An object-cue (i.e., the shape cross) in visual working memory initiates object selection in CLAM.*

### Space-based visual attention in CLAM

Figure 4 illustrates the process of space-based visual attention in CLAM. A spatial cue (without any identifiable shape) can be stored in the VWM-blackboard. This will result in an enhanced activation in the dorsal pathway that selects the location of one object (target) in a visual display. In turn, the selection of a location in the dorsal pathway will enhance activation on that location in the retinotopic areas (V2-PIT), which results in the selection of the shape and the color of the object on that location in the ventral pathway, in line with the notion of space-based visual attention.

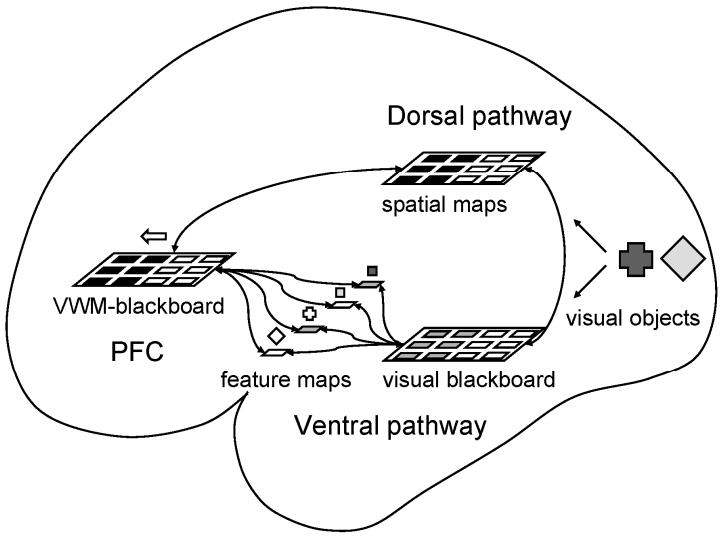


*Figure 4. A spatial cue (i.e., a symbolic cue such as an arrow indicating the left location) in visual working memory initiates object selection in CLAM.*

## Outline of the thesis

We have seen that CLAM provides an architecture that can account for object-based (feature-based) and space-based visual attention in visual search. In CLAM, top-down visual attention in visual search results from interaction between visual working memory in the prefrontal cortex, object recognition in the ventral pathway, and spatial selection in the dorsal pathway. Nonetheless, CLAM leaves many questions about the mechanisms of top-down visual attention in visual search open. Following the outline of CLAM (see Figure 5), several of these questions are addressed in this thesis by elaborating the visual working memory

in the prefrontal cortex and object recognition in the ventral pathway. In addition, this thesis explores mechanisms of stimulus-driven visual attention, and the interaction between mechanisms of stimulus-driven and top-down visual attention, by specifying spatial selection in the dorsal pathway, which was not made explicit in CLAM. The questions are investigated both by simulations and by behavioral experiments.

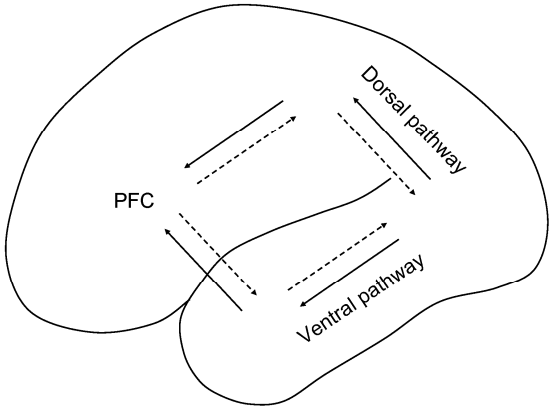


*Figure 5. Visual working memory in the prefrontal cortex, object recognition in the ventral pathway, and spatial selection in the dorsal pathway interact in CLAM.*

### Visual working memory in the prefrontal cortex

One assumption of CLAM is that objects that are maintained in visual working memory are represented in the VWM-blackboard in PFC. The VWM-blackboard in PFC binds the features of an object that is maintained in visual working memory, which are either located in the ventral and dorsal stream or in PFC itself (or both) (see Figure 6). Behavioral research suggested that the number of objects that can be maintained in visual working memory without interference (i.e., loss of information) is limited (to about four), but the number of object features (e.g., shape, color, location, motion, etc.) is unlimited for each of these objects (Vogel, Woodman, & Luck, 2001). *Chapter 2* investigates whether the architecture of VWM (Van der Velde & De Kamps, 2003) in CLAM can explain this finding. We varied the number of objects that are represented in the VWM-blackboard in PFC, and tested the model's ability to use information about the shape and location of an object to respectively bind the object's location and shape. The simulations indicated that our model cannot successfully bind the features of an object anymore as the VWM-

blackboard in PFC gets loaded with an increasing number of objects, which is in line with the behavioral findings.

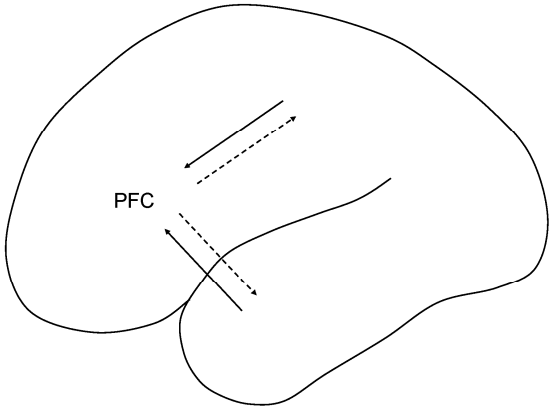


*Figure 6. The question addressed in Chapter 2 relates to visual working memory in the prefrontal cortex in CLAM.*

### Object recognition in the ventral pathway

The ventral pathway in CLAM is hypothesized to transform the retinotopic information into location invariant feature information about object identity (e.g., shape, color) (see Figure 7). What remains unclear, however, is how location invariant object recognition in the ventral pathway is attained. This question is addressed in *Chapter 3*.

Simulations explored whether location invariant object recognition in the ventral pathway can be attained by building up learning in the feedforward network. First, the feedforward network learns to identify simple features at all locations and therefore becomes selective for location invariant features. Next, the feedforward network in the ventral pathway learns to identify objects partly by learning new conjunctions of these location invariant features. Once the feedforward network is able to identify an object at a new location, all conditions for supervised learning of additional, location dependent features for the object are set. The learning in the feedforward network can be transferred to the feedback network, which is needed to localize an object at a new location. This learning scheme resulted in some degree of location invariance for object recognition in the ventral pathway in CLAM.

Nonetheless, it is unanswered whether location invariant object recognition relies on the detection of relatively simple features, or additionally on the detection of

more complex features. Efficient search is dependent on location invariant object recognition, as it requires that the target can reliably be identified among distracters (or that the distracters can reliably be identified along with the target and altogether discarded (Humphreys & Müller, 1993)), irrespective of the location of the target and distracters in the visual display. The question whether location invariant object recognition and efficient search rely on the detection of relatively simple features, or additionally on the detection of more complex features is addressed by three behavioral experiments in *Chapter 4*.

Wang, Cavanagh, and Green (1994) found that search for a digital 5 (digital 2) among digital 2's (digital 5's) is inefficient. The digital 2 and digital 5 differ only in the specific conjunctions of the same lines. Search for this target-distracter pair may be inefficient, because in general an object can only be recognized on the basis of relatively simple features (e.g., lines, edges). Alternatively, it is possible that an object can be recognized on the basis of more complex features (e.g., the global pattern), but only when an object is familiar enough. In this case, search for a digital 5 (digital 2) among digital 2's (digital 5's) may become efficient through training.

The first experiment in Chapter 4 investigates whether training could improve the stimulus familiarity and the search efficiency with the digital 2 and digital 5. We trained and measured stimulus familiarity independently of visual search efficiency, to study the relation between the increase of stimulus familiarity and the increase of search efficiency in a learning task. Search for a digital 5 (digital 2) among digital 2's (digital 5's) became more, but not fully, efficient through training. This suggests that intensive training does not enable objects to be recognized on the basis of more complex features, as required for efficient search. Instead, it appears that objects are (partially) recognized on the basis of relatively simple features, which are similar for the digital 2 and digital 5, confining the search efficiency.

The results further show that stimulus familiarity and search efficiency are partly dissociated. The stimulus familiarity (both of the target and the distracter) increased in our experiment, and visual search became more efficient as well. However, it was found that the search efficiency can be increased further without an effect on stimulus familiarity. Furthermore, the increase in search efficiency generalized substantially from trained to untrained locations (i.e., the effect of learning was largely location invariant).

The second and third experiments in Chapter 4 investigate whether the effect of learning persisted two months after training, and whether it transferred to other search tasks. It was found that the effect of learning was still (partly) present two months after training, and largely specific to the actual stimuli used.

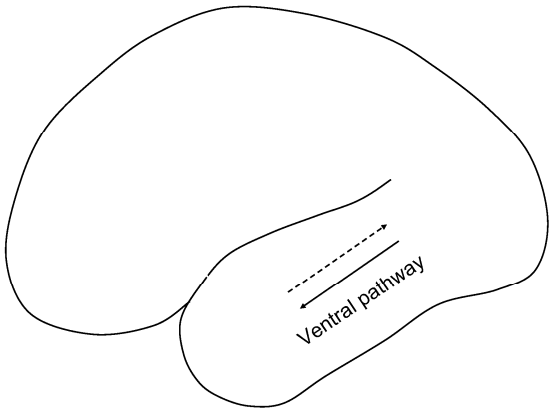


*Figure 7. The questions addressed in Chapters 3-4 relate to object recognition in the ventral pathway in CLAM.*

### Interaction between object recognition in the ventral pathway and spatial selection in the dorsal pathway

In Chapters 5-8, mechanisms of stimulus-driven visual attention and the interaction between mechanisms of stimulus-driven and top-down visual attention are studied by behavioral experiments and simulations.

Five behavioral experiments in *Chapter 5* explore whether the (global) saliency of objects gradually increases as fewer objects in the display share some characteristic, and the experiments explore the interaction of this gradual saliency with top-down visual attention (in the color dimension). In addition, the dynamics of gradual saliency and top-down visual attention over time are investigated.

Experiment 1 demonstrates that saliency is indeed gradual. Experiments 2-4 show that top-down visual attention makes the search for a target faster, even when the target is already located on a (gradually) salient location (e.g., the location of a color singleton). Experiment 5 indicates that colored elements activate the mechanisms responsible for saliency when they are presented for 50 ms, whereas

they enable the selection by top-down visual attention when they are presented for 100 ms.

*Chapter 6* presents an overview of several important findings of behavioral and neurophysiological studies in the realm of visual search, and of theories and models that are proposed to explain these findings. Two main questions that are addressed in this chapter are whether efficient search (which originally was attributed to mechanisms of stimulus-driven visual attention (Treisman & Gelade, 1980)) should be associated with processing in low cortical areas, and whether stimulus-driven visual attention is the result of bottom-up and horizontal processing, or alternatively of bottom-up, horizontal, and top-down processing. Several findings of the behavioral studies that we have reviewed suggest that efficient search cannot solely be attributed to processing in low cortical areas. The results of reviewed neurophysiological studies leave open whether stimulus-driven visual attention is the result of bottom-up and horizontal processing, or of bottom-up, horizontal, and top-down processing.

In *Chapter 7*, an explicit mechanism of global saliency is presented, the Global Saliency Model (GSM), and the interaction between the mechanisms of global saliency and top-down visual attention is specified. It is hypothesized that global saliency is the result of interaction between object recognition in the ventral pathway (Van der Velde & De Kamps, 2001) and spatial selection in the dorsal pathway (see Figure 8). Spatial selection in the dorsal pathway, which was not specified in CLAM, takes place in a number of interacting spatial maps. Consistent with the conclusions of the overview in Chapter 6, global saliency in GSM results from top-down processing in the ventral pathway, in addition to bottom-up and horizontal processing (in the ventral and dorsal pathway).

Simulations show that the model can explain several important findings in visual search, e.g., efficient search for a singleton among distracters (for an overview, see Wolfe & Horowitz, 2004) and the effects of target-distracter and distracter-distracter similarity (Duncan & Humphreys, 1989). In addition, it is shown that GSM can explain the findings of the behavioral experiments in Chapter 5.

Behavioral studies found that the response time to identify or match a target decreases with a larger distance between the target and an attended location (i.e., the location of a feature singleton) (e.g., Caputo & Guerra, 1998; Mounts, 2000). These results and other results have been interpreted as evidence that there is an *inhibitory annulus* around the focus of attention. *Chapter 8* investigates whether inhibition around the focus of attention might result from *pre-attentive lateral*

*inhibition*. Models of stimulus-driven visual attention usually assume that (pre-attentive) lateral inhibition between objects is stronger when objects share features with another (e.g., Itti & Koch, 2000; Wolfe, 1994). Hence, such a pre-attentive lateral inhibition account would predict that the inhibitory surround of attention grabbing distracter is stronger when a distracter shares features with the target than when it does not. The first behavioral experiment tested this prediction by manipulating the similarity between a target and distracter. No interaction was found. In fact, we found no evidence of an inhibitory surround if the target was also salient, even when a salient distracter grabbed attention. Moreover, in a second behavioral experiment it was found that a spatial cue, which grabbed attention, produces a facilitatory surround.

The results of our experiments suggest that the support for an inhibitory annulus around the focus of attention is less robust than it seemed, and that attention may instead facilitate the processing of stimuli near its focus. In line with GSM, it is proposed that salient objects inhibit surrounding objects (independent of whether they share features) not after grabbing attention, but pre-attentively through lateral inhibition.

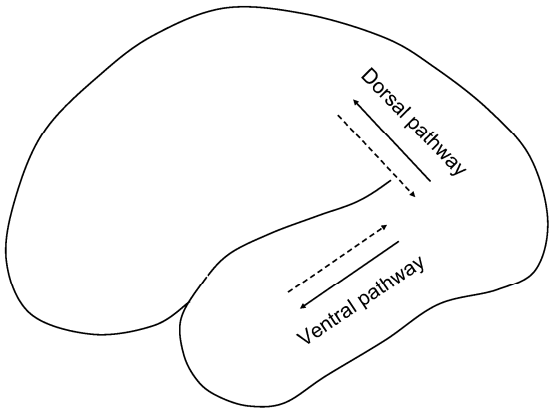


*Figure 8. The questions addressed in Chapters 5-8 relate to the interaction between object recognition in the ventral pathway and spatial selection in the dorsal pathway in CLAM.*

## Publications

Parts of Chapter 1 are included in a refereed publication, and Chapters 2, 3, 4, and 7 constitute refereed publications or are in preparation or submitted for refereed publication. To acknowledge the important contributions of the co-authors to

these publications, a list of references is presented here. Furthermore, I would like to mention that the study reported in Chapter 8 is done in collaboration with Martijn Meeter.

Chapter 1:
Van der Velde, F., De Kamps, M., & Van der Voort van der Kleij, G. T. (2004). CLAM: Closed-loop attention model for visual search. *Neurocomputing, 58-60,* 607-612.

Chapter 2:
Van der Voort van der Kleij, G.T., De Kamps, M., & Van der Velde, F. (2003). A neural model of binding and capacity in visual working memory. *Lecture Notes in Computer Science, 2714,* 771-778.

Van der Voort van der Kleij, G.T., De Kamps, M., & Van der Velde, F. (2004). Increasing number of objects impairs binding in visual working memory. *Neurocomputing, 58-60,* 599-605.

Chapter 3:
Van der Voort van der Kleij, G.T., Van der Velde, F., & De Kamps, M. (2005). Learning location invariance for object recognition and localization. *Lecture Notes in Computer Science, 3704,* 235-244.

Chapter 4:
Van der Voort van der Kleij, G.T., Van Winsen, R., & Van der Velde, F. (2006). Learning visual search: Dissociation between stimulus familiarity and search efficiency. Submitted to *Perception & Psychophysics*.

Chapter 7:
Van der Velde, F., Van der Voort van der Kleij, G. T., Haazebroek, P., & De Kamps, M. (in preparation). The Global Saliency Model.

# Chapter 2 | Increasing the number of objects impairs binding in visual working memory

The number of objects that can be maintained in visual working memory without interference is limited. We present simulations of a neural model of visual working memory in ventral prefrontal cortex that has this constraint as well. One layer in ventral PFC represents all objects in memory. These representations are used to bind the features (e.g., shape, location) of the objects. If there are too many objects, their representations interfere and therefore the quality of the representations degrades. Consequently, it becomes harder to bind the features for an object that is maintained in visual working memory.

## Introduction

Investigations (Vogel et al., 2001) have shown that humans have the ability to maintain a number of visual objects in visual working memory. A remarkable characteristic of this finding is that the number of objects that can be maintained in visual working memory without interference (i.e., loss of information) is limited (to about four), but the number of object features (e.g., shape, color, location, motion) is unlimited for each of the objects. We presented a model of visual working memory in prefrontal cortex (PFC) that theoretically can explain this characteristic (Van der Velde & De Kamps, 2003). A basic characteristic of this model is a *blackboard* that links different *processors* to one another. The processors in this case are networks for feature identification. The blackboard serves to bind the information processed in each of the specialized processors. Objects in visual working memory are represented in the blackboard. One layer in ventral PFC functions as the blackboard, containing representations that consist of conjunctions of identity information (e.g., shape, color) and location information. When too many objects are put in visual working memory, their representations in the blackboard interfere. Consequently, an object's representation in the blackboard muddles and the blackboard's performance to bind the features of an object degrades.

After getting deeper into this model of visual working memory, we present two simulations. One simulation explored how information about the shape of an object can be used to bind the object's location. Another simulation explored the

opposite binding route, i.e. how information about the location of an object can be used to bind the object's shape. The results reflect our expectations that the model is limited in the number of visual objects that it can maintain without interference complicating correct binding.

## Blackboard architecture of visual working memory in PFC

Our model of visual working memory in PFC is based on a neural blackboard architecture that is used in a simulation of object-based attention in the visual cortex (Van der Velde & De Kamps, 2001). We assume that the neural blackboard architecture is located in the ventral prefrontal cortex (V-PFC) (Van der Velde & De Kamps, 2003). This is in line with human neuroimaging studies and monkey studies (e.g., Wilson, Scalaidhe, & Goldman Rakic, 1993). Activation in V-PFC is sustained (reverberating) activation, characteristic of working memory activation in the cortex.

In the model (Figure 1A), the V-PFC has a layered structure with representations similar to the representations in the visual (temporal) cortex. First, the posterior inferotemporal cortex (PIT) connects to the blackboard. As in PIT itself, the representations in this layer of V-PFC consist of conjunctions of location and (partial) identity information (e.g., shape, color). The bottom layer of V-PFC is connected to higher-level areas in the visual cortex like the anterior inferotemporal cortex (AIT) and the posterior parietal cortex (PP), which process respectively the shape and location information of an object.

The connections from these higher-level areas to the bottom layer of V-PFC are similar to the connections in the feedback network of the visual cortex (Van der Velde & De Kamps, 2001). They associate all possible representations that are selective for an activated feature (e.g., shape, location). For example, if one shape is selected in AIT, then all representations in the bottom layer of V-PFC that are consistent with that shape (on every possible position) are activated. Note that these connections have a fan-out structure. Likewise, an attended location in PP activates all possible representations (e.g., for any shape) in the bottom layer of V-PFC on that location in (visual) space. The bottom layer of V-PFC thus represents the current focus of attention, whether this is based on location or (location-invariant) feature information. Consequently, interaction between the bottom layer of V-PFC and the blackboard can select the object representation that is consistent with the current attentional focus. The resulting activation in the select

layer of V-PFC can be used to bind the features of this object (Van der Velde & De Kamps, 2003).

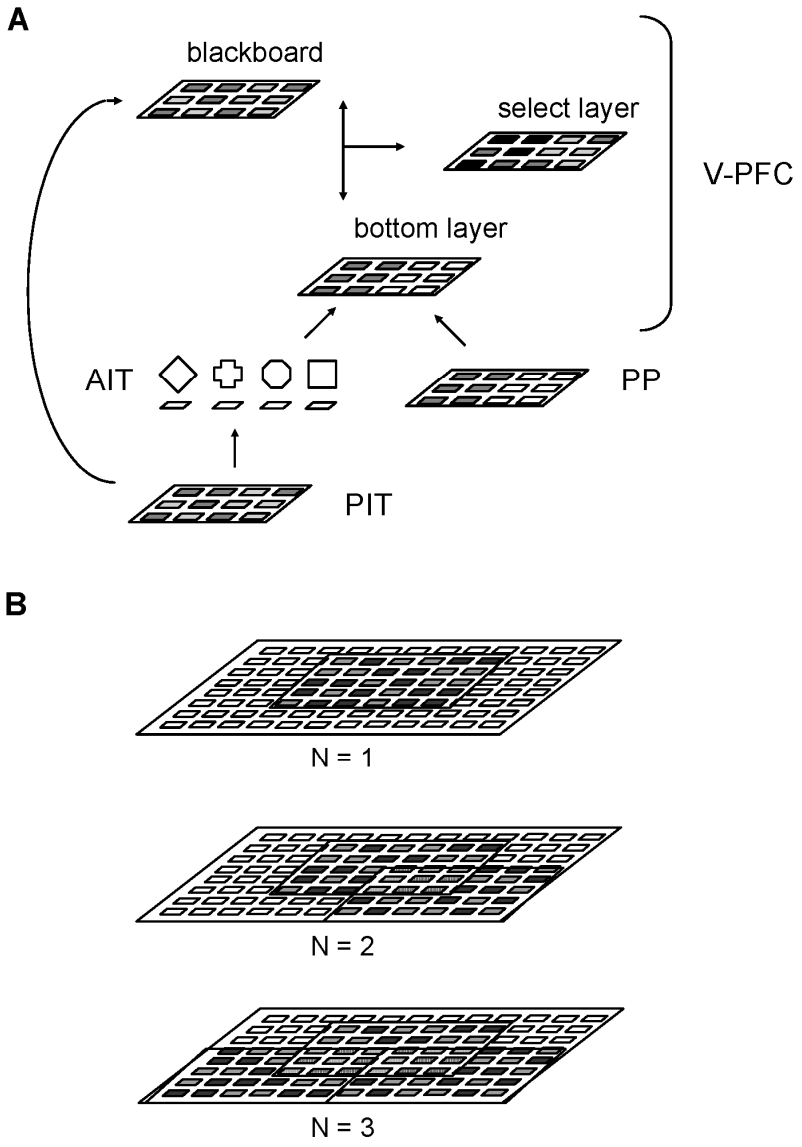


*Figure 1. (A) A blackboard architecture in the prefrontal cortex (PFC). PIT = posterior inferotemporal cortex; AIT = anterior inferotemporal cortex; PP = posterior parietal cortex; V-PFC = ventral prefrontal cortex. (B) Interference between object representations in the blackboard.*

## Feature binding in visual working memory

The nature of the representations in V-PFC and the connections with the higher-level areas in the visual cortex produces the behavioral findings described before. The blackboard architecture of V-PFC results in a binding of the feature representations of the objects maintained in visual working memory. Therefore, the features of an object can be retrieved (selected) in visual working memory as long as the representations of the objects stored in V-PFC do not interfere. However, when too many objects are present in a display, their representations in V-PFC will interfere, which results in loss of information (Figure 1B). As more objects are present in a display, the amount of interference increases, and it can be expected that the quality of the representation of an object in V-PFC becomes less. As a consequence, it becomes harder to correctly bind the feature representations of the objects that are maintained in visual working memory. V-PFC might end up binding wrong feature representations for an object that is attended to. Following simulations tested whether our model of the visual working memory shows this behavior.

## Simulations

For the simulations, we linked the V-PFC model with a (trained) neural network model of the ventral pathway in the visual cortex that is used in the simulation of object-based attention in the visual cortex (Van der Velde & De Kamps, 2001). This model consists of a feedforward network that includes the areas V1, V2, V4, PIT and AIT, and a feedback network that carries information about the identity of the objects to the lower areas in the visual cortex (V1 - PIT). The model shares the basic architecture and characteristics (i.e., the nature of the representations) of the visual cortex. The feedforward neural network was trained to identify 9 different objects on 9 possible positions (using backpropagation). After that, the feedback neural network was trained as well. Learning in the feedback network is based on the activity in the feedforward network that results when the feedforward network identifies an object. In the feedback network, the Hebbian learning rule is used so that the activation pattern in the feedforward network modifies the connections in the feedback network. In this way, the object selectivity in the feedforward network is transferred to the feedback network (Van der Velde & De Kamps, 2001). This was done successfully five times, each time resulting in slightly different connection weights between the layers, representing different instances of the model.

### Simulation 1: Binding the location by shape

This simulation explored the selection process in the V-PFC model that involves shape information. We expected that information about the shape of an object becomes less adequate to bind the object's location as the number of objects stored in visual working memory increases.

During simulations, displays consisting of N (different shaped) objects, with N ranging from 2 to 9, are presented to V1. For each N, 180 random displays are presented to each instance of the model. The objects, presented in separate, non-overlapping, positions, are processed in the visual cortex, and their PIT representations also activate the representations in the blackboard in V-PFC. The shape of one of the objects is selected (attended) in AIT (e.g., due to competition between all object shapes). The activation coding for this shape in AIT activates all representations in the bottom layer of V-PFC that are selective for that shape. As a result, the interaction between the bottom layer of V-PFC and the blackboard modulates the object representation in the select layer of V-PFC that is selective for the attended shape. Consequently, the activation in the select layer of V-PFC reflects the match between the representations in the blackboard and the bottom layer of V-PFC.

The artificial neurons can have activation values in the range -1 to 1. Positive and negative activation can be regarded as activity of separate populations of neurons (De Kamps & Van der Velde, 2001). Thus, negative activation in the bottom layer of V-PFC and negative activation in the blackboard is also a match. Therefore, we simulated the interaction between the blackboard and the bottom layer of V-PFC by computing the covariance between them. Note that these covariance values offer two kinds of information; the match (positive covariance) and the mismatch (negative covariance).

After every presentation of a display with N objects, the positive covariance for every possible position of an object in the select layer of V-PFC was computed. This positive covariance was then standardized by subtracting the mean positive covariance over all positions in the select layer of V-PFC from the positive covariance at a position in the select layer of V-PFC, and dividing this difference in positive covariance by the mean positive covariance over all positions in the select layer of V-PFC. The same was done for the negative covariance. We will further refer to this standardized positive and negative covariance as the *match* and *mismatch* respectively.

It may be clear that within every trial, one position in the select layer of V-PFC corresponds to the position of the attended object in the display, and N - 1 positions in this layer correspond to positions of objects in the display that are unattended. The rest of the positions (9 - N) in the select layer of V-PFC correspond to locations in the display where no object was presented.

Figure 2 shows the probability distribution over several amounts of match for positions in the select layer of V-PFC of attended objects and unattended objects separately. For each number of objects in visual working memory, data of all 5 instances of the neural network model are averaged over all relevant trials. Note that for successful binding to occur, the match should be high on the position of the attended object and low on positions of unattended objects. Only then the position of the attended object can be clearly distinguished from the positions of unattended objects in terms of match. As can be seen in Figure 2, this is the case if the number of objects held in visual working memory is low.

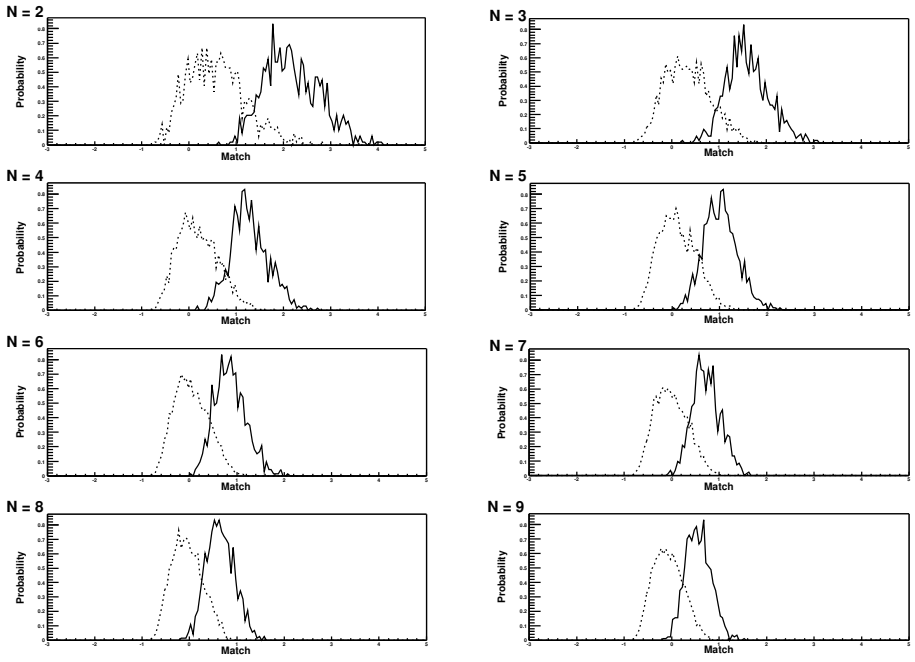


*Figure 2. Probability distribution of match for positions of attended objects (solid line) and positions of unattended objects (dashed line) in the select layer of V-PFC as a function of the number of objects in visual working memory (see the text for explanation). Y-axis: probability. X-axis: match, from negative (left) to positive (right).*

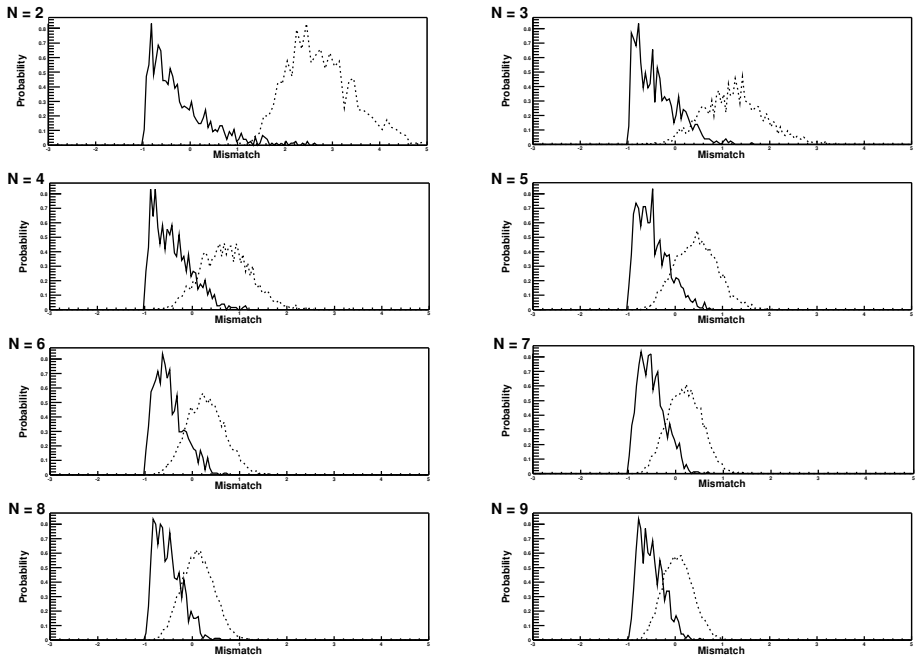


*Figure 3. Probability distribution of mismatch for positions of attended objects (solid line) and positions of unattended objects (dashed line) in the select layer of V-PFC as a function of the number of objects in visual working memory (see the text for explanation). Y-axis: probability. X-axis: mismatch, from negative (left) to positive (right).*

Figure 3 shows the probability distribution over several amounts of mismatch for positions in the select layer of V-PFC of attended objects and unattended objects separately. Again, for each number of objects in visual working memory, data of all 5 instances of the neural network model are averaged over all relevant trials. Note that for successful binding to occur, the mismatch should be low on the position of the attended object and high on positions of unattended objects. Only then the position of the attended object can be clearly distinguished from the positions of unattended objects in terms of mismatch. Again, as can be seen in Figure 3, this is the case if the number of objects held in visual working memory is low.

However, Figures 2 and 3 show that the probability distribution of match and mismatch for the positions of attended objects and for the positions of unattended objects start to overlap more and more as the number of objects in visual working memory increases. This means that the position of the attended object cannot be

reliably selected on the basis of positive or negative covariance. As the load on the visual working memory gets higher, positions of unattended objects will more frequently be selected instead. In other words, the binding process starts to break down.

The mean amount of match for positions of attended objects, positions of unattended objects and positions with no object is presented in Figure 4B together with the root mean squared error (RMSE). Picking the position of the attended object instead of a position of an unattended or empty position on the basis of match information clearly becomes very hard as the number of objects in visual working memory increases. Does mismatch information enable us to point out the correct position of an attended object when the number of objects stored in visual working memory increases? The answer is given in Figure 4A, and appears to be negative. The distinction between attended and unattended objects gets lost here as well. Filling up the visual working memory makes the level of mismatch that can be detected in the select layer of V-PFC on the position of the attended object more and more similar to the level of mismatch on other positions. Thus, based on mismatch information, binding begins to fail as well.

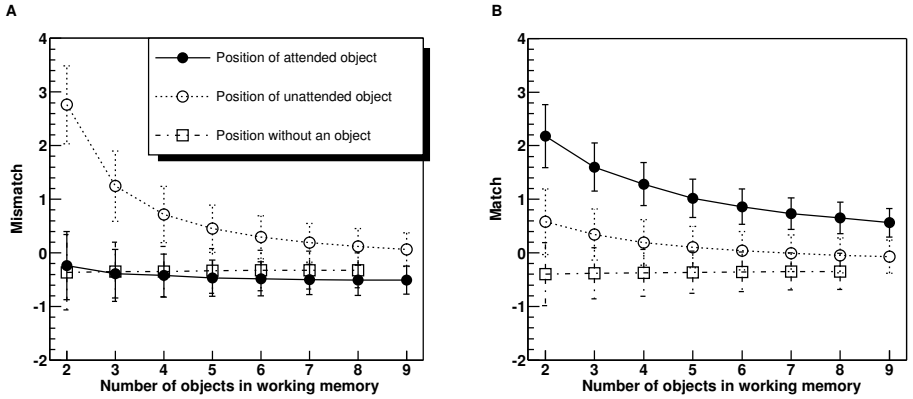


*Figure 4. (A) Mismatch (mean and RMSE) for positions of attended objects (solid line), positions of unattended objects (dot-dot line), and positions without an object (dash-dot line) in the select layer of V-PFC as a function of the number of objects in visual working memory (see the text for explanation). (B) Idem, but then for match.*

### Simulation 2: Binding the shape by location

This simulation explored the selection process in the V-PFC model that involves location information. We expected that information about the location of an object becomes less adequate to bind the object's shape as the number of objects stored in visual working memory increases.

During simulations, displays consisting of N (different shaped) objects, with N ranging from 2 to 9, are presented to V1. For each N, 90 random displays are presented to each instance of the model. The objects, presented in separate, non-overlapping, positions, are processed in the visual cortex, and their PIT representations also activate the representations in the blackboard in V-PFC. The location of one of the objects is selected (attended) in PP (e.g., due to competition between all object locations). The activation coding for this location in PP activates its corresponding location in the bottom layer of V-PFC. As a result, the interaction between the bottom layer of V-PFC and the blackboard modulates the object representation in the select layer of V-PFC at the attended location. The activation in the select layer of V-PFC is processed further by AIT to identify the object's shape.

For simplicity, the activity in PP that represents a certain location after competition between all object locations, its one-to-one connections to the bottom layer of V-PFC, and the interaction between the blackboard and the bottom layer of V-PFC are simulated altogether in one step by modulating the object representation in the blackboard at the attended location. To implement the last step regarding the binding of the object's shape, the blackboard layer served as input to area AIT, which is trained to identify shape information. A winner-takes-all mechanism in AIT selects the identified shape.

The nature of attentional modulation is being debated. The model does not include a clear perspective on this part. Instead, we have taken a more pragmatic stand to simulate, approximately, two competing hypotheses. Attention may either increase the sensitivity for attended features by providing an extra input to neurons representing those, or may boost the response strength for attended features without changing the sensitivity to them (Treue, 2001). We will refer to the former mechanism as *additive* and to the latter as *multiplicative*. Logically, though this is not simulated here, attention may involve a combination of both mechanisms as well.

Hence, location information modulated the representation in the blackboard in two qualitatively different ways during separate runs. In *multiplicative runs*, the

activity of neurons representing the attended location in the blackboard was multiplied by a certain factor. Alternatively, in *additive runs*, these neurons were given extra input, and new activation values were accordingly computed. To ensure results that are sufficiently robust, multiplicative and additive runs were done with a varying modulation strength from respectively 1 to 2 and 0 to 0.5, with a similar step size of 0.05. In additive runs, the range of extra input was chosen to balance apparent levels of sensory input.

Figure 5 shows the probability of successful binding over the number of objects in visual working memory and modulation strength, for both additive and multiplicative runs. For each number of objects in visual working memory, data of all 5 instances of the neural network model are averaged over all relevant trials. Note that a modulation strength of 0 in the additive runs and of 1 in the multiplicative runs actually means that there is no selection by location information at all. Hence, the proportion of correct binding for each N should equal chance level. Figure 5 indeed reflects this fact. Interestingly, we see that a slight increase in modulation strength immediately improves binding. Nevertheless, there appears to be a limit in the benefit of increasing the modulation strength. This makes sense as modulated neurons reach their maximum firing rate at some point.

Moreover, modulation strength also affects unattended, overlapping object representations. Both for additive and multiplicative runs, binding is better when the number of objects held in visual working memory is low, even for quite high values of modulation strength. In other words, as the number of objects increases, the model becomes less reliable to select an object's shape based on its location information. Hence, the binding process starts breaking down. Comparing the additive and multiplicative runs, we see that the latter show slightly better binding (i.e., boosting the output of neurons enables better binding than increasing the input). This makes sense as multiplication amplifies the representation in the blackboard without affecting its structure, while adding does modify the structure of the representation to some extent.

So far we have assumed that the representation in the blackboard is identical to the one in PIT. However, this is not likely to be true. It is possible that the representation in the blackboard is reduced compared to PIT. New simulations explored the binding power of the model given a sparse and reduced representation in the blackboard. Before the location information of one object

modulated the activity in the blackboard, a competition mechanism in the blackboard reduced its representation and made it sparse.

Subtracting an inhibitory input from each neuron's input, which allows 30 percent of the neurons to be active, and computing new activation values, implemented this competition process. In additive runs, the modulation strength now ranged from 0 to 0.3 to balance lower sensory input.

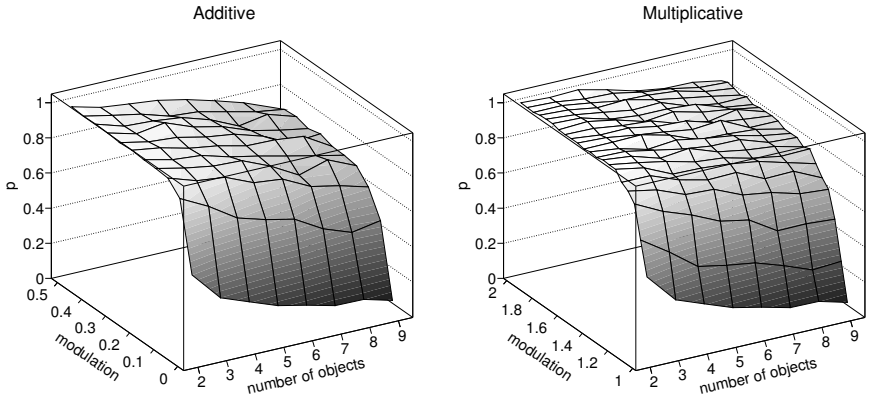


*Figure 5. Proportion of correct binding as a function of the number of objects in visual working memory and modulation strength. See the text for explanation.*

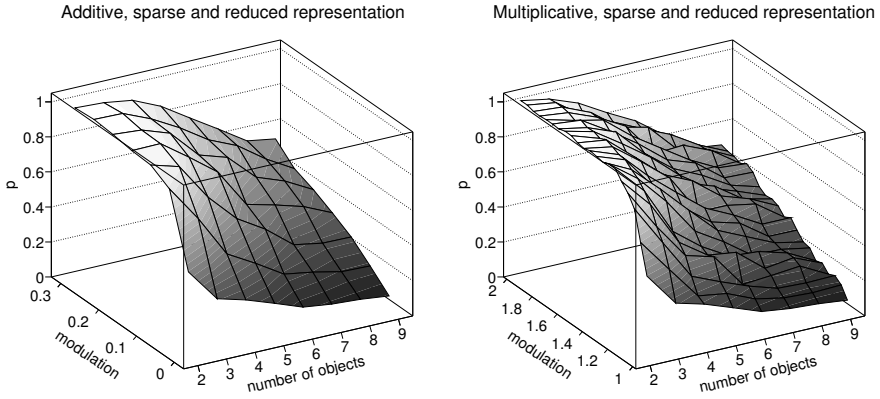


*Figure 6. Proportion of correct binding as a function of the number of objects in visual working memory and modulation strength, given a sparse and reduced representation in the blackboard. See the text for explanation.*

Figure 6 shows the probability of successful binding over the number of objects in visual working memory and modulation strength, for these runs. We see that even when the representation in the blackboard is sparse and reduced compared to the one in PIT, it can still bind the shape to the location of an object considerably when the number of objects in visual working memory is low. As expected, for higher number of objects the binding impairment already seen in former runs is amplified, as a higher number of objects leads to more competition and thus to a more reduced and sparse representation in the blackboard.

## Discussion

The simulations point out that the model of visual working memory that we presented is limited in the number of objects that it can maintain in memory without interference (i.e., loss of information). Our model cannot successfully bind the features (e.g., location, shape) of an attended object anymore as it gets loaded with more objects. This is in accordance with behavioral findings about visual working memory (Vogel et al., 2001). Naturally, our simulations are of a qualitative nature. The fact that there is a limit in the number of objects that people can maintain in visual working memory is (probably) inherent to its architecture. The model that we presented shares this characteristic. When exactly the limit in visual working memory is reached will depend on other factors as well, like the level of alertness and the contrast of the objects with the background.

Our model predicts that this limit is also partly dependent on the distance between objects in a display. Another prediction from our model is that the resolution of spatial attention is comparably limited in other tasks than visual working memory. Selection by location information is dependent on the amount of interference between object representations in the ventral pathway of the visual cortex. Note that it does not matter whether spatial attention (also) acts upon areas with a higher spatial resolution (e.g., V1 or V2), when areas like V4 and PIT, due to their conjunction representations, are still used to bind object's features. Selecting an object by a more centered focus (e.g., a Gaussian) of its location may overcome some interference between object representations. However, it also risks ignoring important information.

# Chapter 3 | Learning location invariance for object recognition and localization

A visual system not only needs to recognize a stimulus, it also needs to find the location of the stimulus. In this chapter, we present a neural network model that is able to generalize its ability to identify objects to new locations in its visual field. The model consists of a feedforward network for object identification and a feedback network for object localization. The feedforward network first learns to identify simple features at all locations and therefore becomes selective for location invariant features. This network subsequently learns to identify objects partly by learning new conjunctions of these location invariant features. Once the feedforward network is able to identify an object at a new location, all conditions for supervised learning of additional, location dependent features for the object are set. The learning in the feedforward network can be transferred to the feedback network, which is needed to localize an object at a new location.

## Introduction

Imagine yourself walking through the wilderness. It is very important that you recognize the company of a predator, wherever the predator appears in your visual field. Location invariant recognition enables us to associate meaningful information (here: danger) with what we see, independent of where we see it. Hence location invariance is a very important feature of our visual system.

Nonetheless, location invariant recognition also implies a loss of location information about the object we have identified. Yet, information about where something is in our environment is also essential in order to react in a goal-directed manner upon what is out there.

Van der Velde and De Kamps (2001) have previously proposed a neural network model of visual object-based attention, in which the identity of an object is used to select its location among other objects. This model consists of a feedforward network that identifies (the shape of) objects that are present in its visual field. In addition, the model also consists of a feedback network that has the same connection structure as the feedforward network, but with reciprocal connections. The feedback network is trained with the activation in the feedforward network as input (Van der Velde & De Kamps, 2001). By using a Hebbian learning procedure,

the selectivity in the feedforward network is transferred to the feedback network. We argue that this is a very natural and simple way to keep the feedback network continuously up to date with ongoing learning in the feedforward network.

How does this architecture allow the step to go from implicitly knowing what to knowing where? Suppose the feedforward network identifies a circle in its visual field. The feedback network carries back information about the identity of this shape to the lower (retinotopic) areas of the model. In these areas, the feedback activation produced by the circle interacts with feedforward activation produced by the circle. The interaction between the feedforward network and the feedback network (in local microcircuits) results in a selective activation at locations in the retinotopic areas of the model that correspond to the location of the circle. This activation can be used to direct spatial attention to the location of the target (Van der Velde & De Kamps, 2001).

Previous research has focused on location invariant recognition in feedforward neural networks (Fukushima, 2004; Riesenhuber & Poggio, 2000). Several models are proposed, in which information processing is routed in a bottom-up manner to a salient location rather than to other locations (e.g., Itti & Koch, 2000). The goal of this chapter is to explore the complementary task of finding, in a top-down manner, the location of what is recognized in a location invariant manner in the visual field. The model of Amit and Mascaro can perform this task (Amit & Mascaro, 2003). They assume a replica module with multiple copies of the local feature input that gives (gated) input to a centralized module that learns to identify objects completely independent of location, and vice versa. We provide an alternative mechanism for location invariant object recognition, by which cells in the feedforward network not only become selective for location invariant features, but also for location dependent features. Next, we explore how learning such location invariant object recognition in the feedforward network transfers to location invariant learning in the feedback network in our neural network model. This transfer is necessary in order to find something at a new location.

We have built up learning in the feedforward network in such a way that it initially learns to identify simple features (e.g., oriented lines, edges) at all possible locations. After that, the feedforward network learns to identify objects at some possible locations. The rationale behind this learning procedure is that learning to recognize an object may then partly involve abstracting new conjunctions of known, location invariant features. This enables the feedforward network to generalize its ability to identify an object at trained locations to new locations. A

simulation of the network confirmed this line of thought. This simulation is first presented in this chapter.

The second simulation presented here investigated how the ability of the feedforward network to recognize an object at a new location relates to finding an object at a new location, given the fact that learning in the feedforward network is built up in successive stages. The simulation demonstrates that recognizing an object at a new location does not automatically lead to finding that new location of the object. However, we show that the recognition of an object at a new location facilitates efficient, supervised learning of additional location dependent features in the feedforward network. Once the improved selectivity for the object at that location in the feedforward network is transferred to the feedback network, the interaction between the feedforward network and the feedback network does enable the selection of the new location of the object.

## Network architecture

For the simulations we used a similar neural network model of (the ventral pathway in) the visual cortex as was used in the simulation of object-based attention in the visual cortex (Van der Velde & De Kamps, 2001). It basically consists of a feedforward network that includes the areas V1, V2, V4, the posterior inferotemporal cortex (PIT), the central inferotemporal cortex (CIT) and the anterior inferotemporal cortex (AIT), and of a feedback network that carries information about the identity of the object to the lower retinotopic areas in the visual cortex (V2 - PIT). The model shares the basic architecture and characteristics of the visual cortex. First, the receptive field's size of cells in an area increases, while climbing up the visual processing hierarchy. Second, the connections between cells in the network are determined so that the retinotopic organization is maintained throughout area V1 to area PIT. Yet, the high-level areas CIT and AIT have input connections from all cells in the previous area. Cells in CIT and AIT thus receive information covering the whole visual field (all positions). Every two successive areas are interconnected. For example, area AIT only receives input from area CIT.

Figure 1 illustrates the architecture of the network schematically. From area V1 to area PIT, cells are arranged in a two-dimensional array that makes up the visual field. The number of layers in an area defines the number of cells per retinotopic position (e.g., two from area V2 to area PIT). Multiple layers within an area are not interconnected. Each layer in V1 codes for line segments of one of four different

orientations (vertical, horizontal, left diagonal, and right diagonal). The input is set in area V1 by activating cells in the four layers of cells. Area AIT functions as the output layer of the network.

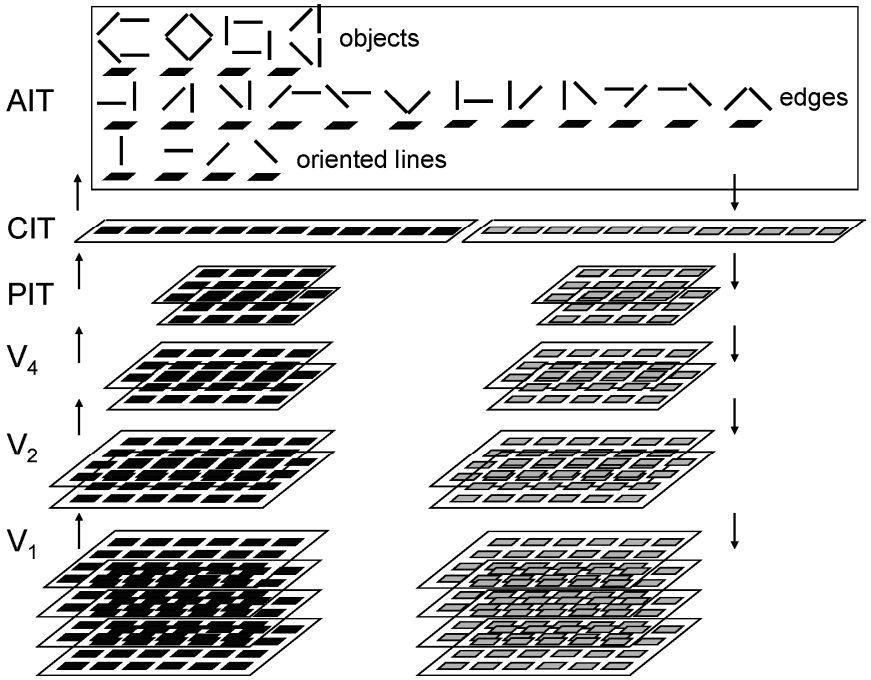


*Figure 1. The architecture of the network. The symbols above the cells in layer AIT show the features that the cells were trained to identify.*

## Simulating location invariant object identification

The network was trained with backpropagation in three successive stages. In the first stage, the network learned to identify oriented line segments (having the length of two cells in the input layer) presented at any position within the network's visual field. In the second stage, the network was trained to identify edges consisting of various combinations of the oriented line segments (see Figure 1) at any position within the network's visual field. In order to avoid (potential) catastrophic interference, the oriented line segments learned in the previous stage were also included in the training. Note that the nature of the collection of edges (two different combinations of each identical set of line segments) forces the network to abstract local relation information at a low level in order to identify the

edges correctly. Hence, throughout these two stages of supervised training, the network learned to identify features of increasing complexity. In the final stage, the network was trained to identify objects (see Figure 1) consisting of line segments and of one or more trained edges. Importantly, the network was only exposed to the objects at four possible locations (see Figure 2A). Again, the training set also incorporated features that were previously learned (at all locations).

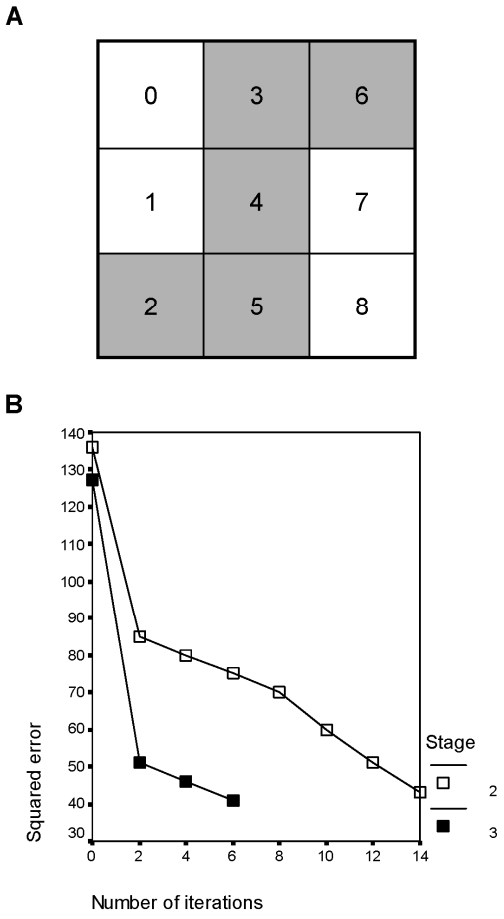


*Figure 2. (A) The nine possible locations in the visual field where objects were presented during testing. The network was exposed to objects at four locations during training (white). Before testing, the objects had never been presented at the five other (gray) locations. (B) Squared error of the network's output over the number of epochs during training, for the second (2) and third (3) learning stage.*

The first two training stages were chosen to generate a network, in which cells in V4 and PIT are selective for a variety of simple and more complex features like the cells in comparable areas of the monkey brain (Tanaka, 1996). The training in the first and second stages offered the network an opportunity to draw on formerly constructed selectivity while encoding new, more complex information in the third stage (i.e., bootstrapping). Note that the exact features that cells in the network learn to abstract are not set in advance, but develop as a result of learning. Furthermore, representation in the network is distributed, due to the connection structure of the network (Van der Velde & De Kamps, 2001).

Cells in CIT have input connections that cover the whole visual field. In principle, during training these cells could become selective only for features that appear in a subset of the visual field. However, the number of cells in area CIT was not sufficient to allow such a specialization for location information. In order to identify the oriented lines and edges at all locations, the cells in CIT learned to abstract features largely independent of location information.

Interestingly, if cells in area CIT are selective for features largely independent of location information after the first two training stages, then the network may subsequently learn to identify the objects partly by learning new conjunctions of such location invariant features. In other words, the network could shape the selectivity of some cells by building upon the location invariant selectivity of cells that are already present. Such a mechanism would give the network the ability to generalize the identification of the objects to locations where the objects have never been presented before.

## Results of location invariant object identification

We trained the feedforward neural network according to the training scheme described above. This was done successfully five times, each time resulting in slightly different connection weights between the areas in the network.

Figure 2B shows the squared error of the network's output over the number of passes that the network has gone through the training set, both for the second and the third stage of training. The data for only one network are displayed in the graph, but these data are well representative for other instances of the network. As can be seen in Figure 2B, the network very quickly learns to identify the objects in the third stage, once it has learned to identify the oriented lines and the edges in the previous stage.

After the training, the network's response was tested for each of the four objects presented at nine possible locations. Four of the locations were identical to the locations at which the objects appeared during training. In contrast, the objects were never presented before at the other five locations (see Figure 2A). Given the connection structure of the network, more cells in the network receive input from an object when it is presented in the center of its visual field than when it is presented in a more peripheral location. Therefore, locations where objects appeared during training and new locations are chosen in such a way that on average the same number of cells in the network respond to an object at each kind of location (i.e., trained or new), apart from the center location.

Each panel in Figure 3 shows the activation value of one cell in area AIT after the processing of its selective object and the other objects, at each location. Each cell clearly responds selectively to the object that it has been trained to identify.

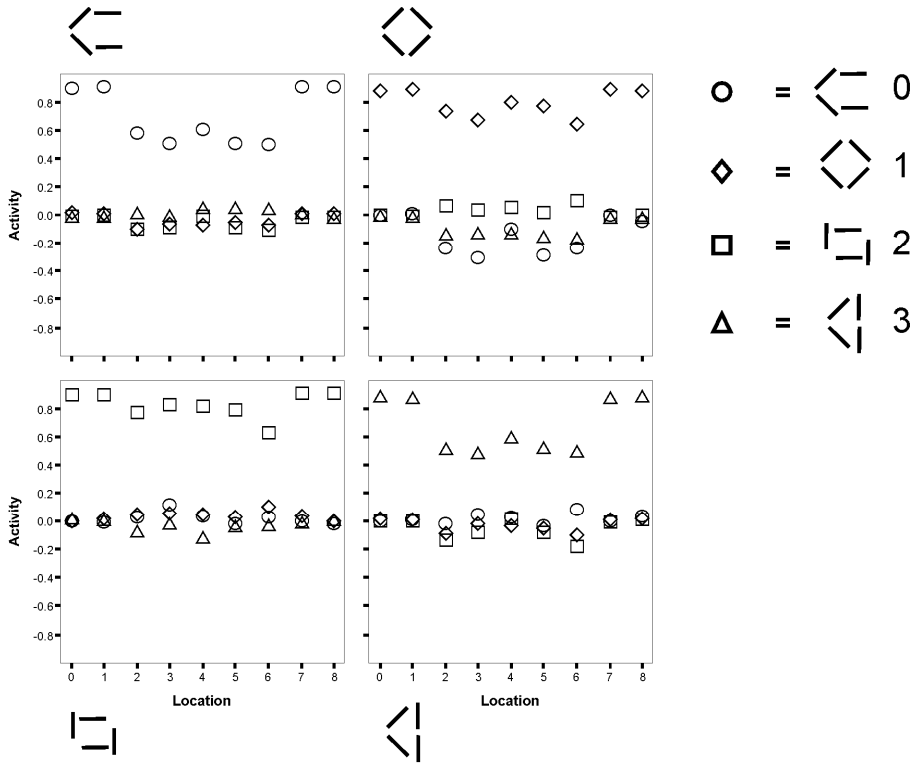


*Figure 3. Each panel shows the activation values of one cell in area AIT trained to identify the object drawn above or under the graph, after presentation of each of the 4 objects at both trained (i.e., locations 0, 1, 7, and 8) and untrained (i.e., locations 2, 3, 4, 5, and 6) locations.*

Moreover, each cell is optimally active when its preferred object appears at one of the trained locations, but it is also active, although to a lesser extend, when its preferred object appears at a new location. Particularly, the diamond and the square (object 1 and 2) are identified most strongly at new locations. The reduced response for a preferred object at new locations compared to trained locations shows that the network partly encodes location dependent features for the objects. This possibly takes place lower in the processing hierarchy of the network. However, the network is clearly able to generalize its identification of objects to new locations. This shows that the network also abstracts new conjunctions of known location invariant features in addition to location dependent features.

## Simulating location invariant top-down visual search

In the second simulation the model performed a top-down visual search task. In this task, a cue is presented first. After that, the target object, matching the cue, appears in the visual field with three distracters (see Figure 4A). The location of the cued object then has to be selected. The network was tested on this visual search task repeatedly with each of the four objects presented as the target. For each target object, 180 random search displays are presented (set as input) to the network. In the model the task is simulated as follows.

In the simulation, a cue selectively activates a cell in area AIT of the feedback network. Top-down activation in the feedback network results in the activation of all other cells in lower areas of the feedback network that are selective for features of that object. Next, the cued object and the other objects are set as input at random, non-overlapping locations in the visual field of the feedforward network. The feedforward network of the model processes all the objects simultaneously. After that, the interaction between the processing in the feedforward network and in the feedback network is simulated by computing the covariance between the activation of cells in the feedforward network and the activation of cells in the feedback network (Van der Velde & De Kamps, 2001).

For each object, the covariance values of all the cells selective for the object in area PIT are summed up. To normalize the sum for each object, the sum of covariance values for an object is divided by the number of cells, which are selective for the object. The group of cells selective for one of the presented objects that has the highest level of normalized covariance indicates the location selected for the target. Note that area PIT still has a retinotopic organization and that cells in this area thus are also partly selective for location information.

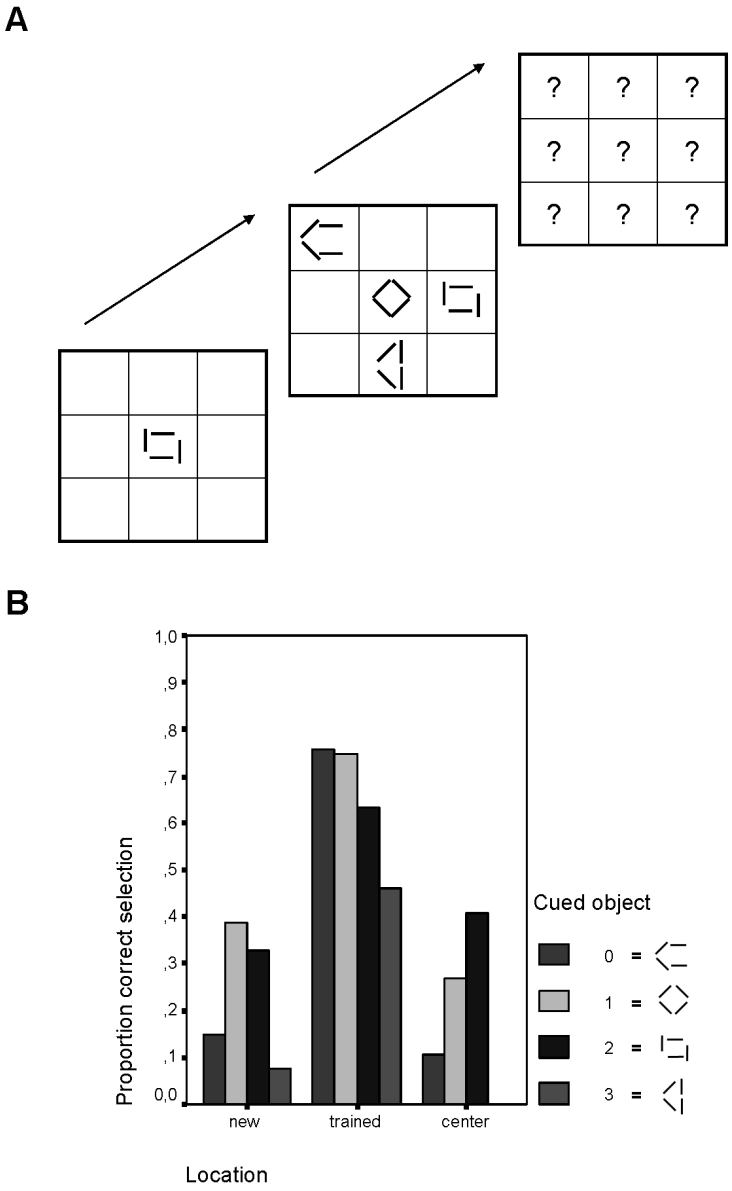


*Figure 4. (A) The top-down visual search task. A cue first indicates the target object (left) and after that the target object is presented between other objects (middle). The model then has to select the location of the target object (right). (B) The proportion of correct selections of the target's location for each of the objects as the target, when the target is presented at the new locations, the trained locations, or the (new) center location.*

## Results of location invariant top-down visual search

Figure 4B illustrates how the (partly) location invariant object identification displayed by the feedforward network (see Figure 3) relates to the model's ability to find the location of an object between other objects. For each of the four objects as the target, the proportion of correct selections of the target's location in the visual field is depicted separately for the trained locations, the new locations, and the (new) center location of the target. The data are averaged over five instances of the model. As can be seen in Figure 4B, the network is better in finding the target's location when its location is one of the locations at which the network is trained to identify the target, than when its location is one of the locations at which the network is not trained to identify the target. Apparently, the network's ability to generalize its identification of an object to new locations does not transfer automatically to the task of finding the location of an object between other objects.

Part of the reason probably lies in the quality of the feedback connections that are the basis for top-down attentional selection in the model. The connections in the feedback network are trained in a Hebbian manner on all the activation patterns in the feedforward network during training (Van der Velde & De Kamps, 2001). As a result, cells in the feedback network that are selective for trained locations code more elaborate information about an object than cells that are selective for new locations (see Figure 3). That is, at trained locations, cells in the feedback network are selective for both location invariant features and for location dependent features, just like cells in the feedforward network. Instead, at new locations, cells in the feedback network are at most selective for location invariant features.

Furthermore, to retrieve information about the location of an object at new locations, the reduced object selectivity in the feedback network has to interact with the activation in the feedforward network, which is also less selective for an object at new locations than for an object at trained locations. Hence, the limitations in the feedback encoding of an object at new locations and the limitations in the feedforward encoding of an object at new locations aggravate each other.

Despite this multiplicative effect of a less elaborated encoding of an object at new locations, we would still expect the network to select the location of the target in a visual search task somewhat above chance level (i.e., proportion correct selection = .25). Figure 4B points out that this is, on average, not the case in our simulation. It is possible that cells in the network that respond to multiple objects present in the

visual field (i.e., cells with large receptive fields), degrade the already basic, generalized feedforward encoding of the target at a new location too much for the model to put its top-down selection mechanism into effective use (Van der Voort van der Kleij, De Kamps, & Van der Velde, 2003). Nevertheless, the network selects object 1 and 2 at new locations between other objects above chance level. Note that these two objects are precisely the objects, which the feedforward network already identified most strongly at new locations (see Figure 3).

## Bridging the gap between recognition and localization

In summary, even when the network recognizes an object at a new location, this does not mean that it can immediately find the location of that object. Obviously, in real life it is very important that we rapidly learn to bridge this gap. What is the mechanism that may constitute that bridge?

The first simulation demonstrates that an object at a new location can be identified. All requirements for supervised learning are therefore present; an object is present at a new location and it is recognized. Figure 2B shows that, in supervised learning, the feedforward network can learn to abstract additional location dependent features of objects relatively fast. As a result the feedforward network becomes more selective for the object at that new location. This increased selectivity of the feedforward network transfers to the feedback network by means of the Hebbian learning in the feedback network (Van der Velde & De Kamps, 2001). After this, the interaction between the feedforward network and the feedback network will enable the localization of the object.

A similar result has emerged in a study, in which participants had to search for a triangle of a particular orientation between triangles of another orientation (Sigman & Gilbert, 2000). The ability of the participants to identify the target between the other objects improved dramatically over several days of training, but this learning was localized to a particular region of the visual field, namely the area used for training. This result might indicate that representations of the trained object are build separately for different positions across the cortical area (Sigman & Gilbert, 2000).

It is crucial for the mechanism that we propose that the feedforward network learns in a build up manner, in which more complex features can partly be learned from more simple, location invariant, features. This allows the network to generalize its ability to identify an object to new locations and triggers more

elaborated, location dependent learning that allows the network to find the object at new locations as well.

## Discussion

Our neural network model predicts that the generalization to new locations by the visual system is more restricted when we have to find an object between other objects than when we have to recognize an object. In line with the second simulation, and with the study of Sigman and Gilbert (2000), we hypothesize that when we search for an object between other objects, the abstraction of new location dependent features of an object may be essential to make the search more reliable. It might also speed up the search process.

We speculate that a visual system can rapidly abstract additional, location dependent features that are needed to reliably find an object at new locations, once it recognizes an object to some extent. Learning new, location dependent features proceeds in parallel to learning new conjunctions of known location invariant features. It possibly takes place mostly lower in the visual processing hierarchy. Our suggestions relate to Ahissar and Hochstein's (2004) Reverse Hierarchy Theory (RHT), although RHT specifically focuses on perceptual learning, and asserts that visual perceptual learning gradually progresses backwards from high-level areas to the input levels of the visual system.

A visual system may generalize its recognition of an object to new locations, when it learns to identify the object partly by means of new conjunctions of location invariant features for which cells of the system are already selective. A simulation demonstrated this principle in our neural network model. Such learning may take place higher up the visual processing hierarchy. Our neural network model learned to recognize objects at multiple locations before testing its ability to generalize recognition to new locations. Yet, the neural network model may have shown comparable location invariant object recognition with fewer trained locations. Nevertheless, it is very likely that we learn to recognize an object at multiple locations, even during a single observation, due to movement of the object or ourselves (e.g., eye-movements, head movements).

The neural network model localizes objects in disjoint windows, like some other models of visual search (Amit & Mascaro, 2003). In the future, the selection of one of multiple disjoint windows may be substituted by a winner-takes-all process, which selects the location with the highest activation in the retinotopic areas of

the model after the interaction between the feedforward and the feedback network (see GSM in Chapter 7).

The neural network model is not yet very robust to clutter. Scaling up its size and changing training to include a larger number of features and objects, will make its cells selective for a larger collection of both location dependent and location invariant features. In addition, providing multiple examples of an object with a realistic amount of within-object variability will strengthen the need to learn the most informative features for discriminating between that object and other objects (Amit & Mascaro, 2003). Together these extensions could result in sparser object representations, helping the neural network model to cope with clutter.

# Chapter 4 | Learning visual search: A dissociation between stimulus familiarity and search efficiency

Previous studies have shown that stimulus familiarity has an effect on visual search efficiency. However, stimulus familiarity was either not tested in these studies, or it was tested in a way that was (partly) confounded with search efficiency itself. In this study, we tested stimulus familiarity independently of visual search efficiency, to compare the increase of stimulus familiarity with the increase of search efficiency in a learning task. The results show that stimulus familiarity and search efficiency are partly dissociated. Stimulus familiarity increases search efficiency, but search efficiency can be increased further without an effect on stimulus familiarity. The effects of learning generalized substantially from trained to untrained locations. Furthermore, the effects of learning were still (partly) present two months after training, and were largely specific to the actual stimuli used.

## Introduction

In a visual search task, participants have to search for a target item among a variable number of distracters. Depending on the combination of the target and distracters, the response time may be (relatively) independent of the number of distracters, or increase with the number of distracters. Search is labeled efficient when the response time is (relatively) independent of the number of distracters and inefficient when the response time increases with the number of distracters. In this chapter, we investigate the relation between stimulus familiarity and search efficiency. In particular, we investigate the relation between learning stimulus familiarity and learning visual search efficiency.

Several studies have investigated the effect of stimulus familiarity on visual search, with sometimes conflicting results. Wang, Cavanagh, and Green (1994) asked participants to search for a target among distracters in the four different conditions of target and distracter familiarity. They compared the search efficiency across the resulting unfamiliar target–unfamiliar distracters (U-U), familiar target–unfamiliar distracters (F-U), unfamiliar target–familiar distracters (U-F), and familiar target–familiar distracters (F-F) conditions. Search was efficient only in the U-F condition. Wang et al. (1994) proposed that unfamiliar

items elicit more activation than familiar items, and consequently attract more attention. According to this hypothesis search is efficient in the U-F condition, because the unfamiliar target is processed before the familiar distracters, whereas the target is processed just as the distracters or even after the distracters in the other conditions (in the absence of an effective attentional set). Hence, Wang et al. (1994) suggested that a difference in familiarity between the target and distracters determines search efficiency.

However, in Wang et al.'s (1994) experiment, the F-U and U-F conditions were studied with one set of items (i.e., N / Z versus mirrored N / Z), while the F-F condition was studied with another set of items (i.e., digital 2 versus digital 5). Wang et al.'s (1994) result that search for a target among familiar distracters is efficient only when the target itself is not familiar may thus also be attributed to stimulus differences (Malinowski & Hübner, 2001; Shen & Reingold, 2001).

Malinowski and Hübner (2001) and Shen and Reingold (2001) investigated the effect of target and distracter familiarity with one set of items for all conditions. They circumvented stimulus differences between conditions by comparing search performance between two groups of participants, which differed in familiarity with the items. In Malinowski and Hübner's (2001) study, Slavic participants were familiar with both N and mirrored N (each serving as target and distracter), whereas the German participants were only familiar with N. In line with the results from Wang et al. (1994), search was not efficient when the target was familiar and the distracters unfamiliar. However, the results further showed that search was not only efficient among familiar distracters when the target was unfamiliar, but also when it was familiar. Shen and Reingold (2001) presented Chinese and English participants two Chinese characters and their 180° rotated forms. The Chinese characters and their rotated forms differed only in the relative position of the components (i.e., a rectangle and a plus sign). The results from the Chinese participants indicated that for both the familiar and the unfamiliar targets, search was more efficient (but not efficient) among familiar distracters than among unfamiliar distracters. The familiarity of the target did not alter the search efficiency. English participants showed no difference in search efficiency between any of the four (U-U) conditions. Both studies provide evidence that the familiarity of the distracters, rather than a difference in familiarity between the target and distracters, determines search efficiency (Malinowski & Hübner, 2001; Shen & Reingold, 2001). More specific, search is (more) efficient when the distracters are familiar and (more) inefficient when the distracters are unfamiliar.

In the studies of Wang et al. (1994), Malinowski and Hübner (2001) and Shen and Reingold (2001), stimuli were used that were assumed (on different grounds) to be either familiar of unfamiliar. But a test of stimulus familiarity, independent of visual search efficiency itself, was not used in these experiments. In the case of Malinowski and Hübner's (2001) study, one can assume that Slavic participants are more familiar with the mirrored N than German participants. Likewise, in the case of Shen and Reingold's (2001) study, one can assume that Chinese participants are more familiar with Chinese characters than English participants. In the study of Wang et al. (1994), though, stimulus familiarity is less clear. For example, in the critical F-F condition, Wang et al. assumed that the digital 2 and digital 5 are familiar stimuli. However, the digital 2 and digital 5 are particular visual exemplars of the categories (concepts) 2 and 5. The concepts 2 and 5 are familiar, but that does not imply that the particular visual exemplars digital 2 and digital 5 are equally familiar. Identification of objects at the categorical level can be faster than the identification of exemplars (Rosch, Mervis, Gray, Johnson, & Boyes-Braem, 1976). Indeed, we show here that the familiarity of the digital 2 and digital 5 can be improved significantly by training, which indicates that these stimuli were not familiar in Wang et al.'s study. In turn, this undermines their conclusion that search is not efficient in the F-F condition.

The importance of testing stimulus familiarity independently of search efficiency can be further illustrated with the study of Mruczek and Sheinberg (2005). They investigated the effect of stimulus familiarity on visual search by training participants on a set of natural images. In this way, stimulus familiarity was controlled in the experiment. That is, stimulus familiarity increased in the course of the experiment, so that its effect on search efficiency could be investigated. A main conclusion of the study was that distracter familiarity improves visual search efficiency, in line with the conclusions of Malinowski and Hübner (2001) and Shen and Reingold (2001).

However, Mruczek and Sheinberg (2005) investigated stimulus familiarity in two different ways. The increase in familiarity of the targets was investigated by measuring the response time (RT) for target identification. Targets became more familiar in the course of the experiment, because the RT of their identification decreased. Yet, the familiarity of the distracters was not investigated in this way. Instead, distracter familiarity was trained and investigated with a visual search task. The increased efficiency of this task, observed in the course of the experiment, was taken as a measure of the increase in distracter familiarity. The

difficulty of this approach is that stimulus (distracter) familiarity is no longer an independent variable, i.e. independent from search efficiency. As a result, the conclusion that distracter familiarity improves search efficiency is based on a confounding of distracter familiarity with search efficiency. Search efficiency is the operational definition of distracter familiarity in this experiment, thus the conclusion in effect states that increased search efficiency improves search efficiency.

The results of Mruczek and Sheinberg (2005) do show that search efficiency can be trained. However, to investigate whether stimulus familiarity per se influences search efficiency, a confounding between stimulus familiarity and search efficiency has to be avoided. That is, to investigate the effect of stimulus familiarity on search efficiency, both factors have to be investigated and measured separately. Likewise, learning stimulus familiarity has to be separated from learning search efficiency, to investigate the effect of the one on the other.

In this study, we investigated the relation between stimulus familiarity and search efficiency. As stimuli we used the digital 2 and digital 5 used by Wang et al. (1994). Wang et al. (1994) assumed that the digital 2 and digital 5 are familiar stimuli, but as noted above, they could have confused the familiarity of the concepts 2 and 5 with the (visual) familiarity of the exemplars digital 2 and digital 5. Malinowski and Hübner (2001) also suggested that the digital 2 and digital 5 are rather atypical versions of the numbers, and that such a deviation from the standard impairs search performance. Thus, we investigated whether training could improve stimulus familiarity and search efficiency with the digital 2 and digital 5. To disentangle stimulus familiarity from search efficiency, we trained and measured stimulus familiarity and search efficiency separately. In this way, we could investigate the effect of stimulus familiarity on search efficiency and vice versa.

To study stimulus familiarity (Experiment 1), we used an identification task in which one stimulus was presented (either the digital 2 or the digital 5). The task of the participant was to identify the stimulus as fast as possible. The RTs in the identification task are a measure of the familiarity of the stimuli, as in the case of the target in Mruczek and Sheinberg's (2005) study. To study learning of stimulus familiarity, participants performed the identification task during a number of consecutive days (> 5760 trials). A decrease in RT during the training phase can be seen as an increase of stimulus familiarity (as in the target case of Mruczek & Sheinberg, 2005).

To study search efficiency (Experiment 1), participants were first tested on a search for the digital 2 among digital 5's and on a search for the digital 5 among digital 2's. Then, participants were trained on one of these two search tasks (> 5760 trials). After training, participants were again tested on a search for the digital 2 among digital 5's and on a search for the digital 5 among digital 2's (participants always searched for a known target).

Thus, participants were trained on only one search task. This allows us to investigate the relation between stimulus familiarity and search efficiency. First, with one search task only, the distracter is presented more often than the target during the training phase. This could influence the familiarity of the two stimuli, that is, the distracter could become more familiar than the target. If so, there will be a difference in RT between the distracter and the target in the identification task (in favor of the distracter). Second, the increase of the familiarity of the stimuli during training could influence search efficiency, even for the search combination that was not trained. In particular, if both stimuli (the digital 2 and digital 5) are equally familiar, and if search efficiency depends only on stimulus familiarity (i.e., the target and distracter familiarity), there should be no difference in search efficiency between the trained search task and the untrained search task.

A further question addressed with our first experiment was the specificity of learning for location. Two previous studies have found effects of learning conjunction search that were highly specific for trained locations (Sigman & Gilbert, 2000; Treisman, Vieira, & Hayes, 1992). We tested the effect of location by comparing the effect of learning at trained and untrained locations within the visual field. Again, if search efficiency depends only on stimulus familiarity, irrespective of the trained locations, and the digital 2 and digital 5 are equally familiar, there should be no difference in search efficiency between the trained locations and the untrained locations.

Finally, we investigated whether the effect of learning search efficiency persisted two months after training (Experiment 2), and whether it transferred to other search tasks (Experiment 3).

## Experiment 1

To investigate the effect of learning on search efficiency, one search task was presented to participants during training. Participants thus searched either for the digital 2 among digital 5's, or for the digital 5 among digital 2's during the training phase. Before and after training, we tested the performance on both

search tasks. To investigate the effect of learning on stimulus familiarity, participants performed an identification task. In this task, participants viewed a single item display, and they had to identify whether the item was the digital 2 or the digital 5. The identification task was alternated with the search task during training.

Furthermore, during training, the items were briefly presented at a subset of locations within the visual field. The presentation time (150 ms) was chosen to prevent voluntary eye-movements, to minimize exposure at other locations within the visual field. We tested the performance on both search tasks before and after training at the same subset of locations (*trained locations*), and at another subset of locations within the visual field (*untrained locations*). This allowed us to determine to what extent learning was location-specific.

Before training, performance on both search tasks and at both subsets of locations should be equivalent. Hence, we compared search performance in five conditions (see Table 1). The first condition comprised both search tasks at both subsets of locations before training. In the second and third condition, the untrained search task was presented after training, respectively at untrained and at trained locations. In the fourth and fifth condition, the trained search task was presented after training, respectively at untrained and at trained locations. For brevity, we will leave out the specification "after training" for the last four conditions in the remainder of the text.

*Table 1*

*Combinations of search task and locations before and after training, and how they map onto the five conditions in Experiments 1 and 2 (1 = Both search tasks, both subsets of locations, before training; 2 = Untrained search task, untrained locations, after training; 3 = Untrained search task, trained locations, after training; 4 = Trained search task, untrained locations, after training; 5 = Trained search task, trained locations, after training)*

| Search task | Locations | | | |
|---|---|---|---|---|
| | BT, untrained locations | BT, trained locations | AT, untrained locations | AT, trained locations |
| BT, untrained search task | 1 | 1 | | |
| BT, trained search task | 1 | 1 | | |
| AT, untrained search task | | | 2 | 3 |
| AT, trained search task | | | 4 | 5 |

*Note.* BT = before training; AT = after training.

## Method

*Participants*

Eight participants with normal or corrected-to-normal vision voluntarily took part in the experiment. They were paid for their participation. Six participants were 20-23 years old and two participants were 49-50 years old.

*Stimuli*

Stimuli were presented on 17" Targa TM 1769-A monitors, with a resolution of 1024 to 768 pixels and a refresh rate of 100 Hz. See Figures 1A and 1B for examples of the stimulus display. The items were the same digital 2 and digital 5 as in Wang et al.'s (1994) study. They subtended 0.6° horizontally and 0.9° vertically. Both in the search task and in the identification task, the items appeared randomly at 12 of 24 possible locations on two virtual presentation circles. The small presentation circle contained 8 possible locations, and the large presentation circle 16. The diameter of the small and large presentation circle was about 7° and 14° respectively. The items appeared either in the first and third quadrant of the presentation circles, or in the second and fourth quadrant of the presentation circles. In the search task, the target was either present or absent, and the setsize varied from 1 to 12 items. In the identification task, either one digital 2 or one digital 5 was presented (each one an equal number of times). The computer-generated stimuli were black and appeared on a white background. A quarter of the participants were randomly assigned to each of the four combinations of search task and locations for training.

*Procedure*

Participants were seated in a dimly lit room at approximately 60 cm of the screen. Each trial began with the presentation of a fixation cross at the center of the screen for 600 ms, which remained visible in the stimulus display. Immediately thereafter, the stimulus display appeared. During training, the stimulus display was visible for 150 ms. Before and after training, the stimulus display remained present until a response was given. In the search and identification task participants were asked to indicate respectively whether the target was absent or present and whether the item was a digital 2 or digital 5, by pressing one of two keyboard buttons. Participants were requested to respond as quickly as possible without making mistakes. A black word ("wrong") was flashed for 400 ms

following errors. The response was followed by an interval of 200 ms until the onset of the fixation cross for the following trial.

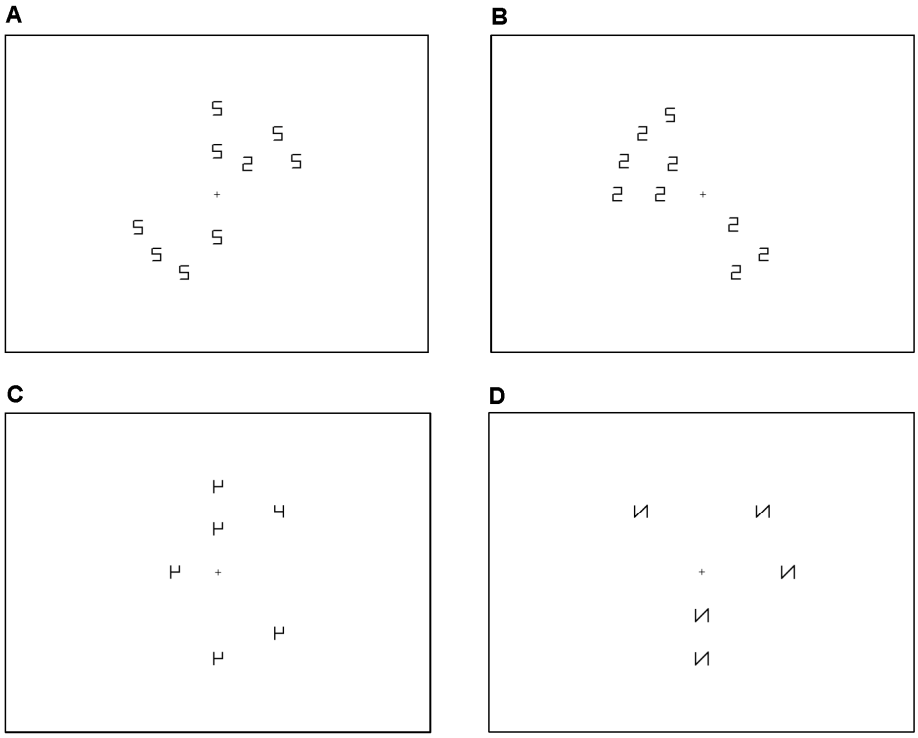


*Figure 1. (A) Example of a stimulus display in Experiment 1, for search for the digital 2 among digital 5's at one subset of locations. (B) Example of a stimulus display in Experiment 1, for search for the digital 5 among digital 2's at another subset of locations. (C) Example of a stimulus display in Experiment 3, for search for the digital 4 among mirrored digital 4's. (D) Example of a stimulus display in Experiment 3, for search for the N among mirrored N's.*

Each participant served in one session of about 2 hour before training, six sessions of about 1.5 hour during training, and one session of about 2 hour after training. Before and after training, a session consisted of 44 blocks of 48 trials (11 blocks for each combination of search task and locations), preceded by 48 practice trials. One cycle of 4 blocks was repeated 11 times. Within a cycle, there were two blocks for one search task, followed by two for the other, with the same order of locations in each pair. Before each block, the search target was displayed on the screen until

participants pressed a key. There were 24 presentation combinations in each block (target absent vs. target present × 12 setsizes), and each combination was repeated two times in a block in random order. After each block, participants received feedback about their average response time and their accuracy in the last block, and a comparison to the previous block. At the same time, they were encouraged to take a break.

During training, a session consisted of 40 blocks of 48 trials, preceded by 12 practice trials. One cycle of 20 blocks was repeated twice. Within a cycle, there were 10 blocks for one specific combination of search task and locations (trained search task, trained locations), followed by 10 blocks for the identification task. Participants performed the six training sessions within two weeks, and no more than one session was scheduled per day.

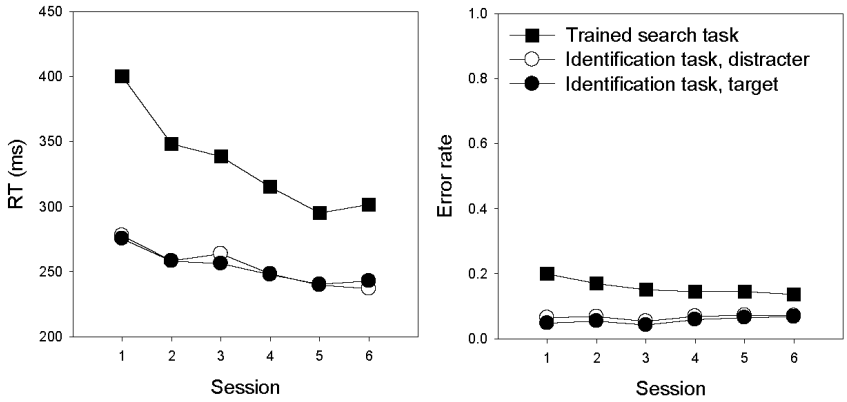


*Figure 2. Response time and error rate as a function of training session, for the trained search task, and for the identification task separately for the target and distracter item.*

## Results: Training sessions

Figure 2 shows the response time (RT) and the error rate as a function of training session for the identification task. The two stimuli are defined as targets and distracters, in terms of their use in the trained search task. A repeated measures analysis of variance (ANOVA) was performed on the RTs and the error rates separately, treating the two stimuli (i.e., targets and distracters) and training session (i.e., 1, 2, 3, 4, 5, and 6) as within-subject variables. There was only a main effect of training session for the RTs [$F(5, 35) = 11.72$, $p < .001$], indicating that the

stimulus familiarity of the targets and distracters improved in the course of the experiment. Figure 2 also shows the RT and the error rate as a function of training session for the trained search task. The results show that search efficiency of the trained search task improved in the course of the experiment. The increase of search efficiency could have affected stimulus familiarity. In particular, the distracters outnumber the targets in the training phase of search efficiency. This could have induced a difference in familiarity between the target and the distracter. However, the target was identified just as fast and accurate as the distracter in the identification task [RTs, $F(1, 7) = 0.012$, $p = .917$; error rates, $F(1, 7) = 1.24$, $p = .302$] (see Figure 2).

### Results: Comparison before and after training

Figure 3A shows the RT and the error rate as a function of condition, setsize, and target presence. For each participant, response times (RTs) that were more than 2.5 standard deviations above or below the mean RT of each combination of search task, locations, target presence and setsize, before and after training, were eliminated. This removed 2.25% of the trials. Analyses of RTs and search slopes are done over correct trials.

RTs and error rates before training were analyzed with an ANOVA, with search task, locations, and target presence as within-subject variables. As expected, for the RTs, there was only a main effect of target presence [$F(1, 7) = 30.70$, $p = .001$]. The main effect of search task [$F(1, 7) = 0.14$, $p = .715$], the main effect of locations [$F(1, 7) = 0.85$, $p = .387$], the Search Task × Locations interaction [$F(1, 7) = 0.01$, $p = .919$], the Search Task × Target Presence interaction [$F(1, 7) = 0.55$, $p = .484$], the Location × Target Presence [$F(1, 7) = 3.46$, $p = .105$], and the Search Task × Locations × Target Presence interaction [$F(1, 7) = 0.04$, $p = .842$] were all not significant. Likewise, for the error rates, there was only a main effect of target presence [$F(1, 7) = 101.79$, $p < .001$]. The main effect of search task [$F(1, 7) = 0.39$, $p = .552$], the main effect of locations [$F(1, 7) = 0.11$, $p = .755$], the Search Task × Locations interaction [$F(1, 7) = 0.25$, $p = .636$], the Search Task × Target Presence interaction [$F(1, 7) = 0.06$, $p = .820$], the Location × Target Presence [$F(1, 7) = 0.03$, $p = .868$], and the Search Task × Locations × Target Presence interaction [$F(1, 7) = 3.90$, $p = .089$] were all not significant. This allowed us to collapse over the four combinations of search task and locations before training, in line with our proposed condition scheme. We will refer to the five combinations of search task and locations before and after training (see Table 1) as the factor condition.

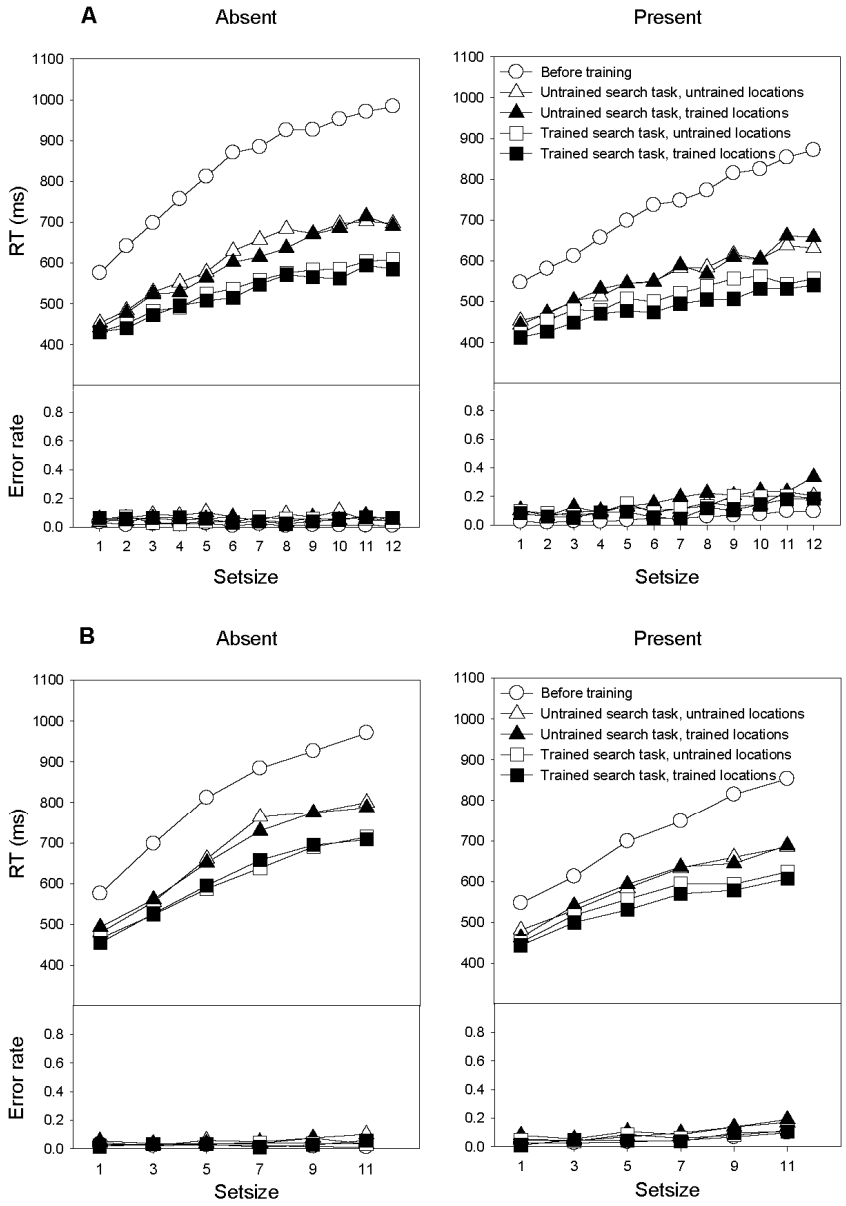


*Figure 3. (A) Response time and error rate as a function of setsize, for the condition before training and the four conditions after training, separately for target absent and present trials. (B) Response time and error rate as a function of setsize, for the condition before training and the four conditions two months after training, separately for target absent and present trials.*

*Response times*

RTs were submitted to an ANOVA with condition and target presence as within-subject variables.[1] The analysis revealed a main effect of condition [$F(4, 28) = 65.17$, $p < .001$ (Greenhouse-Geisser)]. Responses were faster after training than before training [condition 1 versus 2, $t(7) = 6.10$, $p < .001$; condition 1 versus 3, $t(7) = 6.81$, $p < .001$; condition 1 versus 4, $t(7) = 12.34$, $p < .001$; and condition 1 versus 5, $t(7) = 13.34$, $p < .001$]. Responses to the trained search task were faster than to the untrained search task [condition 2 versus 4, $t(7) = 4.27$, $p = .004$; condition 2 versus 5, $t(7) = 4.54$, $p = .003$; condition 3 versus 4, $t(7) = 4.28$, $p = .004$; and condition 3 versus 5, $t(7) = 5.17$, $p = .001$]. Moreover, responses to the trained search task at trained locations were fastest [condition 4 versus 5, $t(7) = 3.50$, $p = .010$]. As expected, responses were slower in target absent than in target present trials [$F(1, 7) = 20.15$, $p = .003$ (Greenhouse-Geisser)].

In addition, condition interacted significantly with target presence [$F(4, 28) = 15.52$, $p < .001$]. Planned comparisons revealed that responses in target absent trials were slowed down less (in comparison to target present trials) after training than before training [condition 1 versus 2, $t(7) = 3.41$, $p = .011$; condition 1 versus 3, $t(7) = 4.55$, $p = .003$; condition 1 versus 4, $t(7) = 6.25$, $p < .001$; and condition 1 versus 5, $t(7) = 4.74$, $p = .002$]. Responses in target absent trials were also slowed down less for the trained search task than for the untrained search task, at untrained locations [condition 2 versus 4, $t(7) = 2.76$, $p = .028$].

*Search slopes*

For each participant, we computed the linear regression of RT on setsize, separately for each condition and for target absent and target present trials. The search slopes found for each participant were submitted to an ANOVA with condition and target presence as within-subject variables. Search slopes differed across conditions [$F(4, 28) = 22.30$, $p < .001$ (Greenhouse-Geisser)]. Search slopes were shallower after training than before training [condition 1 versus 2, $t(7) = 3.72$, $p = .007$; condition 1 versus 3, $t(7) = 2.98$, $p = .021$; condition 1 versus 4, $t(7) = 6.94$, $p < .001$; and condition 1 versus 5, $t(7) = 6.71$, $p < .001$]. In addition, search slopes were shallower for the trained search task than for the untrained search task [condition 2 versus 4, $t(7) = 4.56$, $p = .003$; condition 2 versus 5, $t(7) = 5.08$, $p = .001$; condition 3 versus 4, $t(7) = 3.53$, $p = .010$; and condition 3 versus 5, $t(7) = 3.66$, $p = .008$]. Finally, search slopes tended to be steeper in target absent than in target

present trials, but this effect was only marginally significant [$F(1, 7) = 5.25$, $p = .056$].

*Error rates*

The mean error rate was 6.51%. Error rates were submitted to an ANOVA with condition, target presence and setsize as within-subject variables. The analysis revealed a main effect of condition [$F(4, 28) = 12.01$, $p = .001$ (Greenhouse-Geisser)]. The error rate was higher after training than before training [condition 1 versus 2, $t(7) = -5.13$, $p = .001$; condition 1 versus 3, $t(7) = -4.24$, $p = .004$; condition 1 versus 4, $t(7) = -3.08$, $p = .018$; and condition 1 versus 5, $t(7) = -3.28$, $p = .014$]. Participants made less errors to the trained search task at trained locations than to the untrained search task at trained or untrained locations [condition 2 versus 5, $t(7) = 6.02$, $p = .001$; and condition 3 versus 5, $t(7) = 4.46$, $p = .003$].

Furthermore, participants made more errors in target present than in target absent trials [$F(1, 7) = 67.23$, $p < .001$]. This difference in error rate tended to be larger after training than before training [condition 1 versus 2, $t(7) = 2.22$, $p = .062$; condition 1 versus 3, $t(7) = 4.82$, $p = .002$; and condition 1 versus 4, $t(7) = 2.77$, $p = .028$], except for the trained search task at trained locations [condition 1 versus 5, $t(7) = 1.45$, $p = .19$; condition 3 versus 5, $t(7) = -4.19$, $p = .004$; and condition 4 versus 5, $t(7) = -2.31$, $p = .054$], as indicated by a significant Condition × Target Presence interaction [$F(4, 28) = 7.20$, $p < .001$]. This may reflect that the bias for absent responses slightly increases through training, except for the trained search task at trained locations.

In addition, the error rate increased with an increasing setsize [$F(11, 77) = 10.49$, $p < .001$ (Greenhouse-Geisser)]. This effect was more pronounced for target present trials than for target absent trials [$F(11, 77) = 11.50$, $p < .001$ (Greenhouse-Geisser)]. Importantly, condition did not interact with setsize [$F(44, 308) = 1.44$, $p = .233$ (Greenhouse-Geisser)].

## Discussion

The results show that the stimulus familiarity of the digital 2 and digital 5 increased during training. Thus, contrary to Wang et al.'s (1994) assumption, the digital 2 and digital 5 were not familiar stimuli in their experiment. Therefore, the F-F condition in their experiment was in fact an U-U condition. Therefore, the lack of search efficiency obtained with the F-F condition in Wang et al.'s

experiment cannot be seen as evidence for the notion that efficient search occurs only in the U-F condition.

The results also show that search for the digital 2 (digital 5) among digital 5's (digital 2's) can become more efficient through training. Yet, search for the digital 2 (digital 5) among digital 5's (digital 2's) did not become as efficient as in Wang et al's (1994) U-F condition, despite intensive practice (> 5760 search trials). The average search slope in Wang et al's (1994) U-F condition, averaged over target absent and target present trials, was 5 ms / item for the mirrored N-N target-distracter pair and 11.5 ms / item for the mirrored Z-Z target-distracter pair. In comparison, the average search slope decreased from 33 ms / item before training to 13 ms / item for the trained search task at trained locations in our experiment.

Hence, visual search became more efficient in our experiment, and stimulus familiarity (both of the target and the distracter) increased as well. Nonetheless, we obtained a dissociation between stimulus familiarity and search efficiency in our experiment. Stimulus familiarity, and the increase of stimulus familiarity during learning, was the same for targets and distracters. If stimulus familiarity (either of the target, the distracters, or both) is the only contributing factor to search efficiency, search efficiency should be similar for the trained and the untrained search task, because both tasks consisted of equally familiar stimuli. Yet, search efficiency increased more in the trained search task than in the untrained search task. Naturally, the distracters outnumbered the target in the trained search task (averaged over all setsizes). Even though this had no effect on the familiarity of the distracters (as compared to the familiarity of the target), the training of target and distracters in the trained search task had an additional effect on search efficiency.

The additional increase in search efficiency of the trained search task could perhaps have resulted from a locality effect, in particular for the distracters in the trained search task. If training a search task affects the local representation of the distracters, there would have to be a difference between the trained locations and the untrained locations for the trained search displays. Because the trained search task is not trained at the untrained locations, the difference between the trained and untrained search tasks would have to disappear at the untrained locations.

However, improvement in search performance specific to the trained search task was not limited to trained locations. Also at untrained locations, performance on the trained search task was clearly faster and more efficient than performance on the untrained search task. In fact, the average search slope for the trained search

task was not shallower at trained locations (13 ms / item) than at new (14 ms / item) locations. However, responses to the trained search task were faster at trained locations (504 ms) than at untrained locations (523 ms). Thus, the improvement in search performance specific to the trained search task generalizes greatly from trained to untrained locations, except for a small benefit in response time at trained locations, as compared with untrained locations.

Finally, Table 2 reveals that the search slope varied enormously among participants before and after training (for the trained search task at trained locations). Participants also differed considerably in the decrease in search slope through training.

*Table 2*

*Search slope (in milliseconds per item) for each participant before training and after training for the trained search task at trained locations*

| Condition | Participant | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| BT | 43.9 | 26.2 | 22.4 | 45.4 | 36.2 | 24.2 | 34.2 | 34.7 |
| AT, trained search task, trained locations | 5.7 | 8.3 | 10.5 | 34.6 | 11.5 | 5.1 | 11.6 | 16.5 |

*Note.* BT = before training; AT = after training.

## Experiment 2

In this experiment we investigated whether the effect of learning in Experiment 1, such as the difference between the trained search task and the untrained search task, persisted over two months. Therefore, we tested the performance on each combination of search task (trained or untrained) and location (trained or untrained) two months after training, and compared it with the performance before training.[2]

### Method

The same 8 participants as those in Experiment 1 voluntarily took part in the experiment. The stimuli, design and procedure were equal to those before and after training in Experiment 1, except for the fact that only the odd numbered setsize conditions from Experiment 1 (i.e., 1, 3, 5, 7, 9, and 11) were included. This change was made to reduce the duration of the second experiment, as the participants already had served about thirteen hours in the first experiment. The

setsize thus varied from 1 to 11. The experiment was divided into 24 blocks of 48 trials (6 blocks for each combination of search task and locations), preceded by 48 practice trials. One cycle of 4 blocks was repeated 6 times. Within a cycle, there were two blocks for one search task, followed by two for the other, with the same order of locations in each pair. There were 12 presentation combinations in each block (target absent vs. target present × 6 setsizes), and each combination was repeated four times in a block in random order.

As in Experiment 1, we mapped each combination of search task and locations onto the conditions 2-5 (see Table 1). For brevity, we will leave out the specification "two months after training" for these conditions in the remainder of this section.

## Results

Figure 3B plots the RT and the error rate as a function of condition, setsize and target presence. For each participant, RTs that were more than 2.5 standard deviations above or below the mean RT of each combination of search task, locations, target presence and setsize, before and two months after training, were eliminated. This removed 2.37% of the trials. Analyses of RTs and search slopes are done over correct trials.

### *Response times*

RTs were submitted to an ANOVA with condition and target presence as within-subject variables. The analysis revealed a main effect of condition [$F(4, 28) = 18.36$, $p = .001$ (Greenhouse-Geisser)]. Responses were faster after training than before training [condition 1 versus 2, $t(7) = 3.26$, $p = .014$; condition 1 versus 3, $t(7) = 3.52$, $p = .010$; condition 1 versus 4, $t(7) = 6.49$, $p < .001$; and condition 1 versus 5, $t(7) = 6.12$, $p < .001$]. Responses to the trained search task were faster than to the untrained search task [condition 2 versus 4, $t(7) = 2.66$, $p = .033$; condition 2 versus 5, $t(7) = 2.85$, $p = .025$; condition 3 versus 4, $t(7) = 2.82$, $p = .026$; and condition 3 versus 5, $t(7) = 3.13$, $p = .017$]. As expected, responses were slower in target absent than in target present trials [$F(1, 7) = 20.27$, $p = .003$ (Greenhouse-Geisser)].

In addition, condition interacted significantly with target presence [$F(4, 28) = 5.11$, $p = .003$]. Planned comparisons revealed that responses in target absent trials were slowed down less (in comparison to target present trials) for the trained search task than before training [condition 1 versus 4, $t(7) = 4.69$, $p = .002$; and condition 1 versus 5, $t(7) = 2.57$, $p = .037$]. Responses in target absent trials were

slowed down least for the trained search task at untrained locations [condition 2 versus 4, $t(7) = 3.03$, $p = .019$; condition 3 versus 4, $t(7) = 2.20$, $p = .063$; and condition 5 versus 4, $t(7) = 2.47$, $p = .043$].

*Search slopes*

For each participant, we computed the linear regression of RT on setsize, separately for each condition and for target absent and target present trials. The search slopes found for each participant were submitted to an ANOVA with condition and target presence as within-subject variables. Search slopes differed across conditions [$F(4, 28) = 6.50$, $p < .001$]. Search slopes for the trained search task were shallower than before training [condition 1 versus 4, $t(7) = 4.54$, $p = .003$; and condition 1 versus 5, $t(7) = 3.67$, $p = .008$]. In addition, search slopes tended to be shallower for the trained search task than for the untrained search task, but this was only marginally significant [condition 2 versus 4, $t(7) = 2.21$, $p = .063$; condition 2 versus 5, $t(7) = 1.98$, $p = .088$; condition 3 versus 4, $t(7) = 2.22$, $p = .062$; and condition 3 versus 5, $t(7) = 2.11$, $p = .073$]. Finally, search slopes were steeper in target absent than in target present trials [$F(1, 7) = 11.65$, $p = .011$].

*Error rates*

The mean error rate was 4.93%. Error rates were submitted to an ANOVA with condition, target presence and setsize as within-subject variables. The analysis revealed a main effect of condition [$F(4, 28) = 6.54$, $p = .001$]. The error rate was higher for the untrained search task than before training [condition 1 versus 2, $t(7) = -2.84$, $p = .025$; and condition 1 versus 3, $t(7) = -2.61$, $p = .035$]. Participants made less errors to the trained search task than to the untrained search task [condition 2 versus 4, $t(7) = 2.65$, $p = .033$; condition 2 versus 5, $t(7) = 3.56$, $p = .009$; condition 3 versus 4, $t(7) = 2.35$, $p = .051$; and condition 3 versus 5, $t(7) = 3.70$, $p = .008$]. Furthermore, participants made more errors in target present than in target absent trials [$F(1, 7) = 26.98$, $p = .001$ (Greenhouse-Geisser)].

In addition, the error rate increased with an increasing setsize [$F(5, 35) = 12.45$, $p = .001$ (Greenhouse-Geisser)]. This effect was more pronounced for target present trials than for target absent trials [$F(5, 35) = 4.64$, $p = .022$ (Greenhouse-Geisser)].

## Discussion

We still found an effect of training on the search tasks two months after training. The performance on the trained search task was faster and more efficient than

before training (see Figure 4). Moreover, performance was faster, less erroneous, and slightly more efficient on the trained search task than on the untrained search task. Furthermore, there was no benefit in response time anymore for the trained search task at trained locations, as compared with untrained locations. Hence, two months after training, there was no longer any effect of learning that was specific for trained locations.

## Experiment 3

The digital 2 and digital 5 differ only in the global pattern (i.e., the specific conjunction of the same lines). Having established that an effect of training visual search for the digital 2 (digital 5) among digital 5's (digital 2's) persisted two months after training, we tested in a third experiment whether learning in this task transferred to another task in which the target and distracters differ only in the global pattern (digital 4 among mirrored digital 4's), and/or to a task in which the target and distracters differ in a visual feature, i.e., the orientation of the oblique, (N among mirrored N's). Leonards, Rettenbach, Nase, and Sireteanu (2002, Experiment 5), also investigated whether learning a task in which the target and distracters differ only in the global pattern transfers to a task in which the target and distracters additionally differ in a visual feature. They found no transfer of learning between the two tasks.

### Method

*Participants*

Six of the participants of Experiment 1 and 14 naïve participants with normal or corrected-to-normal vision voluntarily took part in the experiment.

*Stimuli*

Two target-distracter pairs were used: the digital 4 as target among mirrored digital 4's as distracters, and the N as target among mirrored N's as distracters (see Figures 1C and 1D). The digital 4 and mirrored digital 4 subtended $0.6^{o} \times 0.9^{o}$ and the N and mirrored N $0.9^{o} \times 0.9^{o}$. The setsize varied from 1 to 6, and items appeared randomly at all 24 possible locations on the two virtual presentation circles.

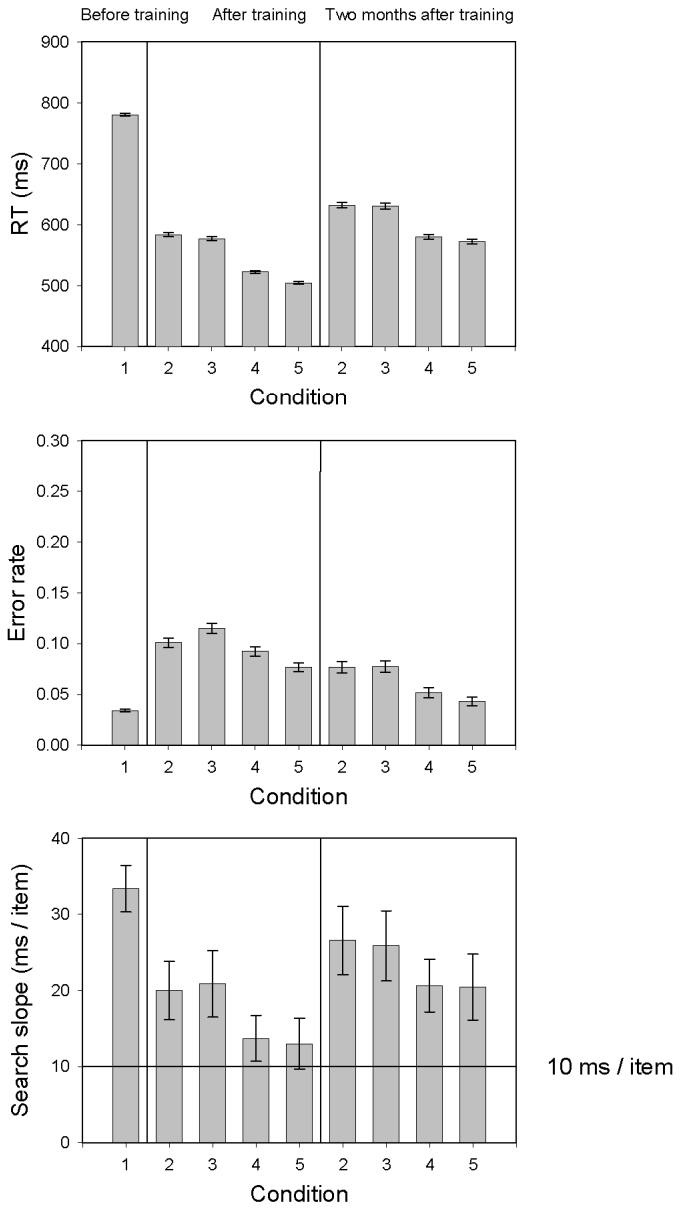


*Figure 4. Response time, error rate and search slope for the condition before training, the four conditions after training, and the four conditions two months after training. The error bars show the standard error of the mean.*

*Procedure*

The procedure was identical to the one before and after training in Experiment 1, except for the number of trials. The experiment consisted of 12 blocks of 48 trials, preceded by 24 practice trials. One cycle of 2 blocks was repeated 6 times. Within a cycle, there was one block for one target-distracter pair, followed by one for the other. Before each block, the search target was displayed on the screen until participants pressed a key. There were 12 presentation combinations in each block (target absent vs. target present × 6 setsizes), and each combination was repeated four times in a block in random order.

## Results

Figure 5 shows the RT and the error rate as a function of experience (naïve participants versus trained participants), target presence and setsize, for each target-distracter pair. For each participant, RTs that were more than 2.5 standard deviations above or below the mean RT of each combination of target-distracter pair, target presence and setsize, were eliminated. This removed 2.20% of the trials. Analyses of RTs and search slopes are done over correct trials.

*Response times*

RTs were submitted to an ANOVA with target-distracter pair and target presence as within-subject variables, and experience as a between-subject variable. Responses to the N-mirrored N pair were slower than to the digital 4-mirrored digital 4 pair [$F(1, 18) = 13.94$, $p = .002$ (Greenhouse-Geisser)]. As expected, responses were slower in target absent than in target present trials [$F(1, 18) = 43.23$, $p < .001$ (Greenhouse-Geisser)].

*Search slopes*

For each participant, we computed the linear regression of RT on setsize, separately for each target-distracter pair and for target absent and target present trials. The search slopes found for each participant were submitted to an ANOVA with target-distracter pair and target presence as within-subject variables, and experience as a between-subject variable. Search slopes were steeper for the N-mirrored N pair than for the digital 4-mirrored digital 4 pair [$F(1, 18) = 7.17$, $p = .015$ (Greenhouse-Geisser)]. Furthermore, search slopes were steeper in target absent than in target present trials [$F(1, 18) = 12.95$, $p = .002$ (Greenhouse-Geisser)].

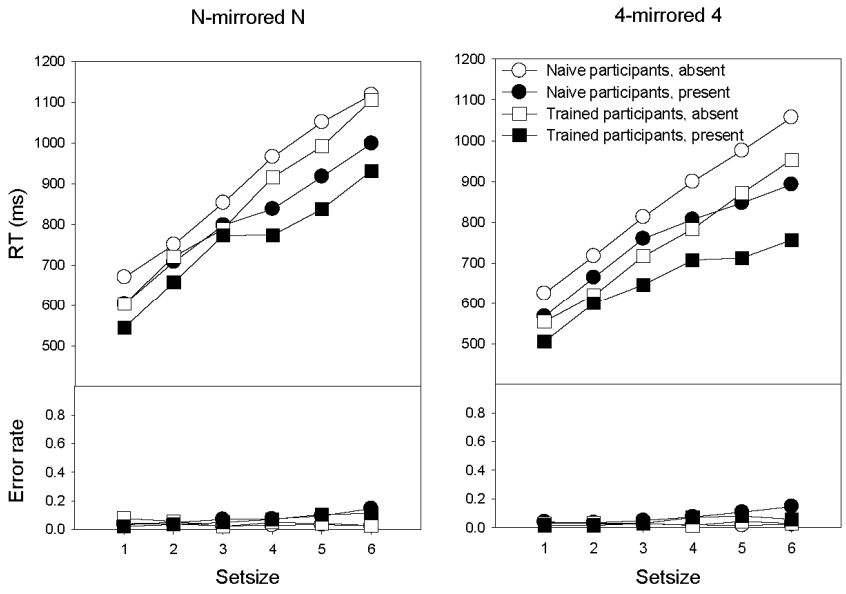


*Figure 5. Response time and error rate as a function of experience (trained participants / naïve participants), target presence and absence, and setsize, when searching for N among mirrored N's, and 4 among mirrored 4's.*

### *Error rates*

The mean error rate was 5.04%. Error rates were submitted to an ANOVA with target-distracter pair, target presence and setsize as within-subject variables, and experience as a between-subject variable. Participants made more errors to the N-mirrored N pair than to the digital 4-mirrored digital 4 pair [$F(1, 18) = 4.46$, $p = .049$ (Greenhouse-Geisser)]. Also, participants made more errors in target present than in target absent trials [$F(1, 18) = 19.69$, $p < .001$ (Greenhouse-Geisser)]. Furthermore, the error rate increased with an increasing setsize [$F(5, 90) = 6.82$, $p < .001$ (Greenhouse-Geisser)]. This effect was more pronounced for target present trials than for target absent trials [$F(5, 90) = 10.55$, $p < .001$ (Greenhouse-Geisser)].

## Discussion

Figure 5 suggests that trained participants searched faster for N among mirrored N's and for digital 4 among mirrored digital 4's than naïve participants, especially for larger setsizes. However, the search performance on these tasks did not differ

reliably between naïve participants and trained participants, irrespective of the fact that the latter group of participants had extensive practice with the task of searching for digital 2 (digital 5) among digital 5's (digital 2's). Thus, there was no significant benefit of learning a search task in which the target and distracters differ only in the global pattern (i.e., the specific conjunction of the same lines) on another search task in which the target and distracters differ only in the global pattern, or on a search task in which the target and distracters differ in a visual feature. Hence, we replicated the finding of Leonards et al. (2002), who found no transfer of learning a task in which the target and distracters differ only in the global pattern to a task in which the target and distracters in addition differ in a visual feature. The results further suggest that learning in Experiment 1 is confided to the actual stimuli used. Thus, the results of this experiment indicate that the results of Experiment 1 are based on visual learning.

## General discussion

Previous studies provide evidence that stimulus familiarity affects visual search efficiency. Wang et al. (1994) concluded on the basis of their experiment that visual search is efficient when the target is unfamiliar and the distracters are familiar (i.e., the U-F condition). In contrast, Malinowski and Hübner (2001) and Shen and Reingold (2001) presented evidence that visual search is efficient when the distracters are familiar, regardless of target familiarity. The difference between these studies is thus the condition in which the target and the distracters are familiar (i.e., the F-F condition). Wang et al. observed inefficient search in this condition, in contrast with the (more) efficient search observed by Malinowski and Hübner (2001) and Shen and Reingold (2001). However, in their F-F condition, Wang et al. (1994) used the digital 2 and digital 5 as stimuli, assuming that they are familiar. Here, we showed that the digital 2 and digital 5 can be made more familiar by training, which suggests that they were not familiar in Wang et al.'s study. Thus, the F-F condition in their study was in effect an U-U condition. As a result, the conclusion of Wang et al. that efficient search does not occur in the F-F condition is unfounded.

Yet, the results of Experiment 1 also demonstrate that, even after extensive training, search for the digital 2 (digital 5) among digital 5's (digital 2's) did not become as efficient as in Wang et al.'s (1994) U-F condition. Perhaps the similarity between the digital 2 and digital 5, both consisting of the same lines, could also have limited the search efficiency in Wang et al.'s (1994) F-F condition. This

suggestion is consistent with results of Shen and Reingold (2001, Experiment 1), who tested search efficiency for several item pairs (i.e., capital letters and digital numbers versus their mirrored forms) in the U-F and the F-U condition. The difference in search efficiency between the U-F condition and the F-U condition, as well as the search efficiency in each condition, were larger when the two items differed in a low-level feature (or a global orientation cue) than when they differed only in the arrangements of the same lines.

The results of the present study emphasize the importance of controlling for the level of stimulus familiarity when studying the effect of familiarity in visual search. In previous studies, stimulus familiarity was not measured, but assumed to exist. One exception is the study of Mruczek and Sheinberg (2005). They trained participants on a set of natural images, to be used as targets or distracters in a search task. In Mruczek and Sheinberg's (2005) study, target familiarity was tested by measuring RTs in an identification task. However, Mruczek and Sheinberg (2005) measured distracter familiarity by means of a search task, in which increased search efficiency was taken as a measure of increased familiarity. As a result, distracter familiarity was confounded with search efficiency in this study, undermining its conclusion that increased distracter familiarity results in more efficient search.

Therefore, to investigate the effect of stimulus familiarity on search efficiency, and in particular to investigate the relation between learning stimulus familiarity and learning search efficiency, a test of stimulus familiarity is needed that is independent of search efficiency itself. To this end, we studied and tested stimulus familiarity (both of the target and distracters) using the RTs in an identification task.

We trained participants on the identification task of the digital 2 and digital 5, together with one of the two possible search tasks with these stimuli. Thus, participants were trained to search for the digital 2 (or digital 5) among digital 5's (or digital 2's). Participants were tested on both search tasks (before and after training). The results of the identification task show that the stimuli became more familiar in the course of the experiments, and that there was no difference in the familiarity of the stimuli, despite the fact that the distracter stimulus was presented more often in the trained search task than the target stimulus. The results of both the trained and the untrained search task show that search for the digital 2 (digital 5) among digital 5's (digital 2's) became more efficient through training.[3]

Although the search efficiency thus increased with stimulus familiarity in our experiment, stimulus familiarity and search efficiency were also (partly) dissociated in our experiment. Stimulus familiarity and the increase of stimulus familiarity during learning were the same for targets and distracters. Thus, if stimulus familiarity is the only contributing factor to search efficiency, search efficiency should be similar for the trained and the untrained search task, given that both tasks consisted of equally familiar stimuli. Yet, search efficiency increased more in the trained search task than in the untrained search task.

As we noted above, the additional increase in search efficiency of the trained search task could perhaps have resulted from a locality effect, in particular for the distracters in the trained search task. This suggestion is in line with the structure of the visual cortex. Stimuli are processed and represented through a hierarchy of areas, beginning in the lower areas, in which the neurons have small receptive fields, and ending in the higher areas, in which neurons have large receptive fields. Training of stimulus familiarity could in particular have an effect on the higher areas, because stimulus identity is represented in these areas. However, training a search task could also affect the processing and representation in the lower areas. In particular, the representation and processing of the distracters could be affected in these areas, for example, due to an increased interaction between these distracter representations. This interaction would not necessarily affect the higher areas in the visual cortex, and would therefore not influence the familiarity of the distracters, but it could influence the process of searching a target among the distracters.

If training a search task affects the local representation of the distracters, there would have to be a difference between the trained locations and the untrained locations for the trained search displays. Furthermore, because the trained search task is not trained at the untrained locations, the difference between the trained and untrained search tasks would have to disappear at the untrained locations.

Nonetheless, the improvement in performance specific to the trained search task was not limited to trained locations. Also at untrained locations, performance on the trained search task was clearly faster and more efficient than performance on the untrained search task. This difference remained two months after training. Furthermore, although responses to the trained search task were faster at trained locations than at untrained locations directly after training, this difference disappeared two months after training. Thus, the improvement in search

performance specific to the trained search task generalizes from trained to untrained locations, and this transfer is sustainable over time.

The (partial) dissociation between stimulus familiarity and search efficiency observed in our experiments shows that search efficiency does not only depend on the familiarity of the distracters (Malinowski & Hübner, 2001; Shen & Reingold, 2001; Mruczek & Sheinberg, 2005) or on the difference in familiarity between the target and the distracters (Wang et al., 1994). Apparently, learning the distracters as a group also affects search efficiency, even though it does not result in an increased familiarity of the distracter stimulus as compared to the familiarity of the target stimulus. Previous research has shown that learning results in better grouping of the distracters, and that this better grouping of the distracters facilitates faster target detection (Karni & Sagi, 1991; Treisman, 1982).

In our experiments, distracter grouping also transferred to untrained locations. In contrast, Treisman et al. (1992, Experiment 3) found effects of learning conjunction search that were highly specific for trained locations. Participants learned to search for four targets defined by a conjunction of a color and shape (i.e., a letter) among distracters. Two of the four targets were presented more often at one possible display location (non-overlapping) than at the other seven possible display locations, the consistent targets. In the course of training (about 4500 trials), a large benefit emerged for the consistent targets in their frequent location, and an increasing cost when they appeared in the infrequent locations. However, in Treisman et al.'s (1992) third experiment targets were defined by a conjunction of a color and shape. We have proposed before that binding of a color and shape requires interaction between bottom-up and top-down processing at lower retinotopic visual areas (Van der Velde & De Kamps, 2001; Van der Velde, De Kamps, & Van der Voort van der Kleij, 2004). As receptive fields within lower visual areas are relatively small, neuronal modification at this level may result in highly location-specific learning.

Sigman and Gilbert (2000) also found effects of learning conjunction search that were highly specific for trained locations. They trained participants to detect a triangle of a particular orientation among triangles of other orientations. After training, search for the trained target was efficient within the training region, but not outside the trained region of the visual field. However, in Sigman and Gilbert's (2000) experiment, the untrained region was more eccentric than the trained region. As search efficiency is shown to decrease with a smaller search items' size / eccentricity ratio (Humphreys, Quinlan, & Riddoch, 1989), the higher

search efficiency within the trained region than within the untrained region in Sigman and Gilbert's (2000) study may be attributed to the difference in eccentricity. Moreover, Sigman and Gilbert's (2000) task required processing at lower visual areas, in which the spatial resolution is high, while processing at higher visual areas sufficed for our task (c.f., Ahissar & Hochstein, 1997, 2004). The stimulus array and setsize were respectively smaller and larger in Sigman and Gilbert's (2000) task (stimulus array, $4.2^{o} \times 4.2^{o}$; setsize, 24) than in our task (stimulus array, diameter $7^{o}$-$14^{o}$; setsize, 1-12). As a result, neuronal modification may have resulted in stronger location-specific learning in Sigman and Gilbert's (2000) experiment than in our experiment.

The results of Experiment 3 suggest that the results of Experiment 1 (and 2) are based on visual learning. Thus, it seems that in Experiment 1 a grouping of distracters was learned with a representation at a high level of the visual hierarchy (perhaps comparable to a Gestalt pattern). In this way, the recognition of the pattern of distracters could transfer to other untrained locations. Learning the distracters as a pattern had an effect on search efficiency over and above the effect produced by the increase of stimulus familiarity, but it did not affect stimulus familiarity itself. The persistence of the results two months after training suggests that pattern learning of distracters is stable over time.

# Chapter 5 | Interaction between gradual saliency and top-down visual attention within the color dimension

Models of visual attention suggest that stimulus-driven and top-down attentional mechanisms together select locations for attention by respectively favoring locations with unique features, explaining pop-out, and locations with designated target features. Here, we investigated whether the (stimulus-driven) saliency of elements gradually increases as fewer elements in the display share some characteristic, and the interaction of this gradual saliency with top-down visual attention (for color). Experiment 1 demonstrates that saliency is gradual, while the benefit of shifting attention to elements from a minority colored set was restricted. Experiments 2-4 show that top-down visual attention decreases the response time when the target is already salient. Experiment 5 shows that colored elements already activate the mechanisms responsible for saliency when they are presented for 50 ms, whereas they enable the selection by top-down visual attention when they are presented for 100 ms.

## Introduction

In pop-out visual search, participants are shown displays composed of multiple elements on a background. All of the elements share the same features (e.g., color, shape, size, etc), the *distracters*, but one element, the *singleton*, differs in the value of one of these features. The number of distracters has a minimal effect on the time needed to detect the singleton. In other words, the search slope is (almost) zero, and search is very efficient. There are many studies showing (almost) zero search slopes to detect the presence or the absence of a singleton embedded between distracters (for an overview, see Wolfe & Horowitz, 2004).

Pop-out is considered as a bottom-up effect in most models of visual attention (Cave, 1999; Itti & Koch, 2000; Li, 2002; Wolfe, 1994). Bottom-up processing refers to the processing of a stimulus from lower-level areas to higher-level areas in the visual processing hierarchy. It is driven by the stimulus and top-down knowledge does not play a role. In these models, the relative uniqueness of each element with respect to its context, the *saliency*, is first computed through bottom-up processing. Then, spatial attention is directed either automatically (Cave, 1999; Itti & Koch, 2000; Li, 2002; Wolfe, 1994) or voluntary (Treisman & Sato, 1990) to

the most salient location. In pop-out visual search, this is the location of the singleton, which is unique compared to the other elements.

The question arises as to whether saliency is an all-or-none phenomenon, or whether elements become increasingly salient as fewer and fewer elements in the display share a characteristic. Our hypothesis is that elements become increasingly salient as fewer and fewer elements in the display share a characteristic. We will refer to this as *gradual saliency*.

Studies of conjunction search have provided indirect evidence for gradual saliency (Sobel & Cave, 2002; Zohary & Hochstein, 1989). In conjunction search, a target is (usually) defined by a conjunction of two features, and the distracters share one of the two target features. Zohary and Hochstein (1989), for example, had participants search for a red vertical element, whereas the distracters were green vertical elements and red horizontal elements. They varied the proportions of the two types of distracters, and showed that the search for the target proceeded through the smallest group of distracters; *smaller-group search*. Sobel and Cave (2002) have replicated this finding in several variations of this task. Smaller-group search indicates that rarer elements are searched earlier or faster than more common elements. Thus, smaller-group search in conjunction search is consistent with the existence of gradual saliency.

However, in conjunction search it is very advantageous to determine the smaller group of distracters and to search this group of distracters, because the target is always present within the smaller group of distracters (except for target absent trials). The present study was designed to evaluate the existence of gradual saliency, without the incentive to search the smaller group of elements.

Toward this end we developed a method, in which the target that has to be searched for (i.e., an oriented line) is superimposed on colored elements, which are not relevant for the task at hand. The colored elements are divided in two sets, each with a particular color. The proportion of these sets is varied, but the overall amount of elements (i.e., given by the combination of both subsets) remains the same. As a result we obtain two sets of differently colored elements, with different proportions. We compare the time to identify the target on the smallest of these sets with the time needed to identify the target on the largest of these sets. Gradual saliency predicts that search for the target will be faster on the smallest of these sets compared to the largest of these sets.

Fixing the total number of colored elements in the display has the advantage of having an equally strong global transient generated by the onset of colored

elements, independently of the proportions of the two sets of differently colored elements. Our method bears some similarity with the distance method (Turatto, Galfano, Gardini, & Mascetti, 2004), in which the number of colored elements is held constant and a target is superimposed with a varying distance from the irrelevant color singleton. An important difference with the distance method is that in our approach the target can appear on a whole range of gradually salient locations, instead of on the classic set of a singleton and a no singleton location.

Furthermore, our method provides the possibility to study the effect of top-down visual attention on (gradual) saliency. Because the colored elements are not searched for (i.e., they are not the target), we can investigate whether top-down visual attention for one of these colored elements influences the search for the target. A number of studies have sought to examine the role of top-down visual attention for elements that share one particular feature with the target in conjunction search, such as limiting search to the group of elements sharing the target's color for a target defined by a conjunction of color and orientation. Furthermore, gradual saliency was either implicitly present due to the proportion between the two types of distracters, or explicitly present due to varying proportions between the two types of distracters. Such studies concluded that top-down visual attention for elements that share one particular feature with the target (i.e., restricting search to a single set of distracters) is automatic in conjunction visual search (Bacon & Egeth, 1997; Egeth, Virzi, & Garbart, 1984), or can be induced by instructions (Kaptein, Theeuwes, & Van der Heijden, 1995).

However, as already argued by Cave and Wolfe (1990) and by Sobel and Cave (2002), the previously mentioned studies suffer from limitations concerning the attribution of findings to either top-down visual attention or to smaller-group search. For example, the presence of top-down visual attention for one target feature in Egeth et al.'s (1984) study can alternatively be explained by smaller-group search, as there were relatively few distracters of the set of distracters that participants were instructed to restrict their search to, compared to the other set of distracters. In Kaptein et al.'s (1995) experiment, smaller-group search was made partly ineffective by making participants search for a target defined by a conjunction of a color and a orientation, of which the orientation was difficult to discriminate from the orientation of the distracters ($0^{o}$ versus $20^{o}$). In Bacon and Egeth's (1997) study, smaller-group search was not efficient due to an unequal distribution of distracter types over all the trials. For this reason, participants

might have relied more strongly on top-down visual attention for the target feature that less frequently dominated the search displays.

Sobel and Cave (2002) tested the respective roles of smaller-group search and top-down visual attention in conjunction search more directly, by manipulating the proportions of the two distracter types. Furthermore, in separate experiments, they manipulated the discriminability of the defining features of the target, the density of the display elements, and the use of explicit instructions to restrict search to one set of distracters. Sobel and Cave (2002) found that participants largely relied on smaller-group search and little or none on top-down visual attention for elements that share one particular feature with the target, as long as both target features were easily discriminable from the distracting features and the display was dense enough with search elements. The results of this study suggest that top-down visual attention for one particular target feature does not guide visual search when easily discriminable stimuli allow guidance by saliency (whether driven by bottom-up or top-down grouping factors).

The experiments of Sobel and Cave (2002) were designed to explore the balance between smaller-group search and top-down visual attention for elements that share one particular feature with the target. Although very interesting, these experiments are not ideal to unravel the interaction between gradual saliency and top-down visual attention. The reason is that in a conjunction search task, participants always have to search for a target that is defined by a combination of two features. Hence, participants probably always adopt an attentional set encompassing *both defining features* to some extent, despite explicit instructions to restrict search specifically to one set of distracters.

In our method we investigate the interaction between gradual saliency and top-down visual attention in a design, in which the latter is manipulated on top of the search for the target. This is accomplished by defining our manipulation of top-down visual attention and gradual saliency in another dimension (i.e., color) than top-down visual attention for the target (i.e., orientation). As has already been mentioned, our design also reduces the incentive to search the smaller group of elements with a particular color compared to conjunction search.

## Experiment 1: Gradual saliency

In the experiments of Sobel and Cave (2002), the proportions of the two distracter types were varied in a conjunction search task, and participants showed smaller-group search with dense displays and with features that were highly discriminable

(see also Zohary & Hochstein, 1989). Interestingly, this indicates that saliency is gradual (i.e., not an all-or-none phenomenon). However, it is possible that participants searched the smaller group of elements, because the target was always present within this group of elements. The first experiment was designed to test whether elements become increasingly salient as fewer and fewer elements in the display share a characteristic, while the benefit of shifting attention to elements from a minority colored set was restricted. The target was superimposed on one of fifteen colored elements. A minority of the elements (the *minority colored set*) was colored in one particular color and the majority of the elements (the *majority colored set*) in a different color. The target was equally likely to appear on an element from a minority colored set or on an element from a majority colored set.

## Method

### *Participants*

Eighteen Leiden University undergraduate students with normal or corrected-to-normal vision voluntarily took part in the experiment. All participants reported to have normal color vision. They were either paid for their participation, or received credits to partially fulfill the requirements of a psychology class.

### *Stimuli*

Stimuli were presented on 17" Targa TM 1769-A monitors, with a resolution of 1024 to 768 pixels and a refresh rate of 100 Hz. Each trial began with the presentation of a fixation symbol for 600 ms. The fixation symbol was a gray "+" (each intersecting line measuring 1.1 degree of visual angle) located on the center of a black background. The fixation symbol was followed by a blank screen for 200 ms, after which the search display appeared.

The search display consisted of 15 colored disks randomly placed on two virtual presentation circles against a black background (see Figure 1). The small presentation circle contained 8 potential disk locations, and the large presentation circle 16. The diameter of the small and large presentation circle was about 7 and 14 degrees of visual angle respectively, while the disks measured 1.1 in degrees of visual angle. Each disk was either green or blue. The colors green, blue, and gray were made equiluminant.

One disk contained an oriented, black line, which measured approximately 0.5 degree of visual angle. The oriented line, the target, was tilted 45° to the left or 45° to the right.

Although the total number of disks in the search display was always fifteen, the ratio between the numbers of disks of each color was varied. Each search display was equally likely to contain 0, 1, 3, 5, 7, 8, 10, 12, 14, or 15 disks of one color with 15, 14, 12, 10, 8, 7, 5, 3, 1 or 0 disks of the other color.
Each potential disk location was equally likely to contain a colored disk, and the target was equally likely to be placed in one of 1, 3, 5, 7, 8, 10, 12, 14, or 15 identically colored disks. Furthermore, the location of the target in a disk on the small versus the large presentation circle was independently varied, implying that the location of the target in a disk was equally likely to be on the small or on the large presentation circle. The locations of the target in a disk on the small presentation circle and on the large presentation circle are the two levels of the factor eccentricity. Finally, the color of the disk that contained the target was equally likely to be green or blue in all conditions.

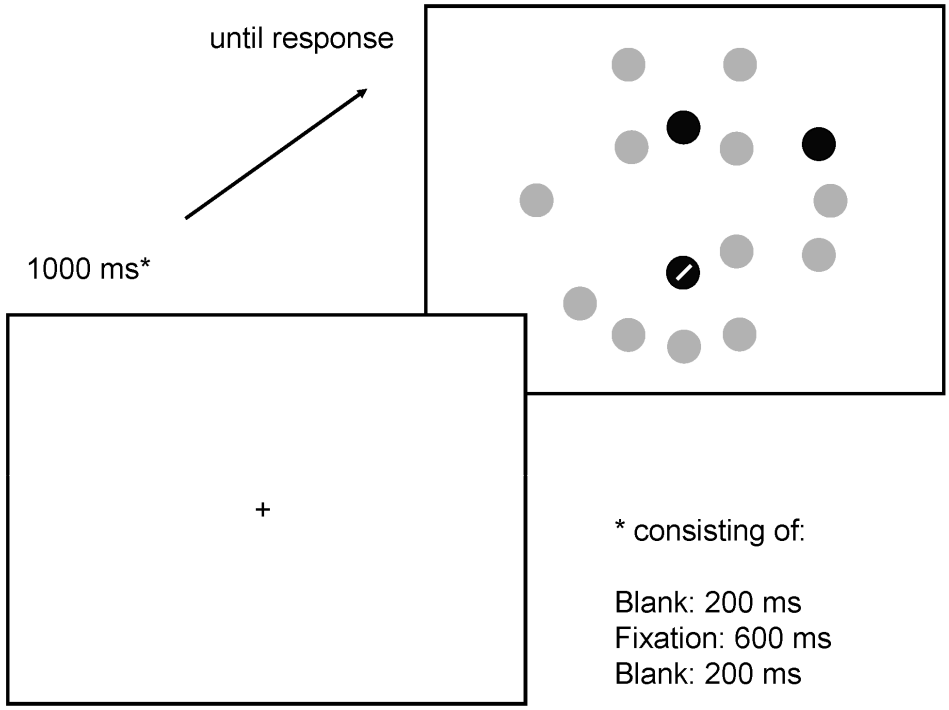


*Figure 1. Sequence of displays in Experiment 1. Gray denotes the color blue, black denotes the color green, and the white line denotes the black target.*

*Procedure*

Participants were seated in a dimly lit room at approximately 60 cm of the screen. They were instructed to respond to the diagonal line and to indicate whether it was tilted to the left or tilted to the right, by pressing one of two keyboard buttons. Participants were requested to respond as quickly as possible without making mistakes. A yellow word ("wrong") was flashed for 400 ms following errors. The search display remained visible until one of the buttons was pressed (i.e., self-terminated response). The response was followed by an interval of 200 ms until the onset of the fixation symbol for the following trial.

The experiment consisted of ten blocks of 36 trials, preceded by 24 practice trials. After each block, participants received feedback about their average response time (RT) and their accuracy in the last block, and a comparison to the previous block. Feedback also functioned as a self-paced break.

## Results

RTs that were faster than 200 ms or slower than 4000 ms were excluded from the analysis. This removed 0.08% of the trials. Figure 2A shows the RT and the error rate as a function of the number of identically colored elements, on one of which the target was superimposed, for all targets. Figure 2B plots the RT and the error rate as a function of the number of identically colored elements, on one of which the target was superimposed, separately for targets on the small and on the large presentation circle. For example, when the number of identical elements is one, the target was placed in one uniquely colored element. Similarly, when the number of identical elements is fifteen, all elements in the display had the same color and the target was placed in one of them.

RTs were submitted to an analysis of variance (ANOVA), with the number of identical elements and eccentricity as within-subject variables. The main effect of the number of identical elements, $F(8, 136) = 19.08$, $p < .001$ (Greenhouse-Geisser), the main effect of eccentricity, $F(1, 17) = 77.16$, $p < .001$, and the Number Of Identical Elements × Eccentricity interaction, $F(8, 136) = 6.82$, $p < .001$ (Greenhouse-Geisser), were all significant.

Planned comparisons between pairs of successive conditions showed that responses were faster in the condition in which the number of identical elements was 1 versus 3, $F(1, 17) = 31.95$, $p < .001$; 3 versus 5, $F(1, 17) = 11.58$, $p = .003$; 5 versus 7, $F(1, 17) = 14.46$, $p = .001$; and 15 versus 14, $F(1, 17) = 8.21$, $p = .011$.

An ANOVA of error rates with the number of identical elements and eccentricity as within-subject variables, revealed no significant effects. As can be seen in Figure 2B, errors are equally distributed over all conditions. The pattern of error rates discards the possibility of a speed-accuracy trade-off.

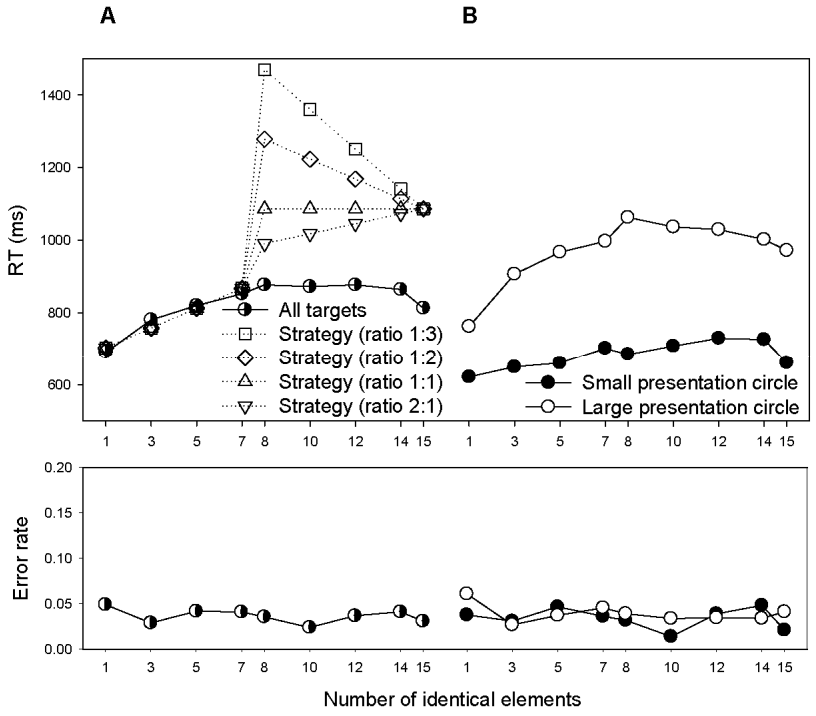


*Figure 2. (A) Response time (top) and error rate (bottom) in Experiment 1 as a function of the number of identical elements, on one of which the target was superimposed. Response times that are predicted by the strategy of voluntarily searching elements from the minority colored set before elements from the majority colored set are also shown, for ratios 1:1, 1: 2, 1:3, and 2:1 between the target present and the target absent search slope (see the text for explanation). (B) Response time (top) and error rate (bottom) in Experiment 1 as a function of the number of identical elements, on one of which the target was superimposed, separately for targets on the small and on the large presentation circle.*

## Discussion

We found evidence for gradual saliency, while the benefit of shifting attention to elements from a minority colored set was restricted. Responses for targets that are located on elements from a minority colored set are faster than for targets that are

located on elements from a majority colored set. More specific, responses are fastest for targets on color singletons, but there are also RT benefits for targets on elements from minority colored sets with more than one element (see Figure 2A). Responses for targets that are located on one of fourteen identically colored elements in the presence of one uniquely colored element are slower than for targets that are located on one of fifteen identically colored elements.

Taken together, elements from a minority colored set are either searched earlier or faster than elements from a majority colored set. Similar to the locations of color singletons, the locations of elements from a minority colored set with more than one element seem relatively salient.

One might argue that participants had an incentive to shift top-down visual attention more to elements from the minority colored set than to elements from the majority colored set, and that the results of Experiment 1 therefore do not reflect (gradual) saliency. The reason is that each individual element from the minority colored set has a higher probability that the target is placed on it than each individual element from the majority colored set, since the target is equally likely to be superimposed on an element from the minority colored set or an element from the majority colored set.

Suppose that participants indeed voluntarily searched elements from the minority colored set before elements from the majority colored set. In that case, the RTs in the conditions with 1 to 7 identical elements reflect the time that is needed to search through (on average) *half of the elements from the minority colored set* (target present search), whereas the RTs in the conditions with 8 to 15 identical elements reflect both the time that is needed to search through *all the elements from the minority colored set* (target absent search), and the time that is needed to search through (on average) *half of the elements from the majority colored set* (target present search). Hence, the RTs in the conditions with 1 to 7 identical elements indicate the search slope when the target is present (i.e., 27 ms / item). Based on the target present search slope, and the ratio between the target present and the target absent search slope, one can predict the RTs in the conditions with 8 to 15 identical elements, in which all the elements from the minority colored set are searched before elements from the majority colored set.

One generally assumes that the ratio between the target present and the target absent search slope is 1:2. Figure 2A shows the predicted RTs for this ratio and other ratios between the target present and the target absent search slope: 1:1, 1:3, and 2:1. For a ratio of 1:2 between the target present and the target absent search

slope, the predicted RT greatly increases between the condition with 7 identical elements and the condition with 8 identical elements. The reason is that (according to the strategy) in the condition with 7 identical elements on average half of the elements from the minority colored set are searched (i.e., 3.5), while in the condition with 8 identical elements all of the elements from the minority colored set are searched (i.e., 7), plus on average half of the elements from the majority colored set (i.e., 4). Likewise, for the ratios 1:1, 1:3, or (very unlikely) 2:1 between the target present and the target absent search slope, the predicted RT also greatly increases between the condition with 7 identical elements and the condition with 8 identical elements. However, the actual RT only slightly increases between the condition with 7 identical elements and the condition with 8 identical elements. Furthermore, in the condition with 15 identical elements, the predicted RT is much higher than the actual RT.

In conclusion, it is evident that the actual RTs do not fit the predicted pattern of RTs in the conditions with 8 to 15 identical elements, irrespective of the specific ratio between the target present and the target absent search slope. Thus, the strategy of voluntarily searching elements from the minority colored set before elements from the majority colored set does not explain the results of Experiment 1.

As expected, participants were faster to identify targets on the small presentation circle than targets on the large presentation circle (see Figure 2B). In addition, the RT benefit for targets appearing on elements from a minority colored set was larger for targets on the large presentation circle than for targets on the small presentation circle. One explanation for this finding is that elements are less strongly represented with a larger eccentricity from fixation (Parkhurst, Law, & Niebur, 2002). As a consequence, the benefit of an increase in saliency due to the relative uniqueness of an element with respect to its context may be larger for more eccentric stimuli. The faster responses for targets on the small presentation circle than for targets on the large presentation circle, and the stronger RT benefit of gradual saliency for targets on the large presentation circle than for targets on the small presentation circle, are replicated in the other experiments of this chapter. These findings will not be repeatedly discussed, because they are not the main interest of this chapter.

## Experiments 2A and 2B: Gradual saliency and top-down visual attention

In a second experiment we investigated the interaction between gradual saliency and top-down visual attention. Sobel and Cave (2002) independently varied the accessibility of smaller-group search and the presence of top-down visual attention for elements that share one particular feature with the target in conjunction search. Instructions to search through the elements that share one particular feature with the target had little effect when both the target features were easily discriminable. Participants only searched the subset of elements that share one particular feature with the target when the discrimination of the other target feature was difficult. This effect was strengthened by explicit instructions to search the easily discriminable feature.

In conjunction search, participants have to search for a target that is defined by a conjunction of two (or more) features. Hence, participants might always adopt an attentional set encompassing both (or all) defining features to some extent, despite explicit instructions to search the subset of elements that share one particular feature with the target and to ignore the other elements. To examine the interaction between gradual saliency and top-down visual attention, independently from top-down visual attention for the target, we defined the target in another dimension (i.e., orientation) than top-down visual attention and gradual saliency (i.e., color).

In Experiments 2A and 2B, top-down visual attention was either set by a color cue at the beginning of each trial, or was absent due to a neutral cue. In Experiment 2A, we used colored disks as cues (*explicit cues*), whereas we used words (*symbolic cues*) in Experiment 2B. Symbolic cues can mediate the processing of elements exclusively by top-down mechanisms. In addition, explicit cues might also prime the color (Theeuwes, Reimann, & Mortier, 2006).

### Method

*Participants*

A total of thirty-six participants from the same student population as described in Experiment 1 were tested (18 in Experiment 2A and 18 in Experiment 2B).

*Stimuli*

Stimuli were presented on the same apparatus as in Experiment 1. The stimuli were equal to those in Experiment 1, with the difference that all disks now

contained an oriented black line, which measured approximately 0.5 degree of visual angle. One of these lines was tilted 45° to the left or 45° to the right; the target. The other lines were homogeneously oriented, either horizontal or vertical. This modification makes the disk with the target more similar to the disks without the target. As a consequence, these stimuli might expose larger benefits from gradual saliency and from top-down visual attention than the stimuli in Experiment 1.

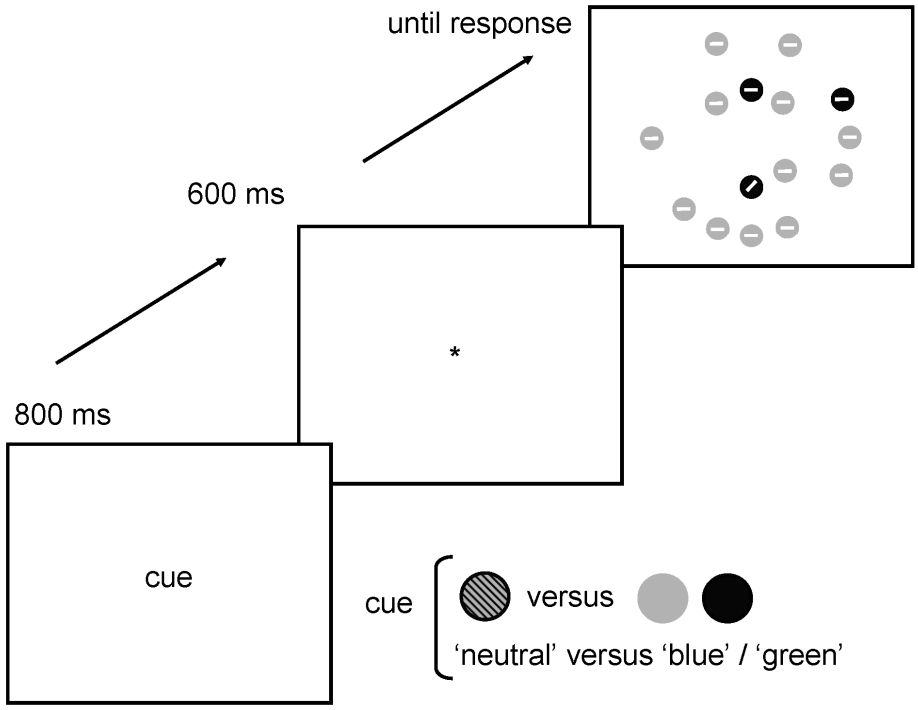


*Figure 3. Sequence of displays in Experiments 2A and 2B. Gray denotes the color blue, black denotes the color green, and the white lines denote the gray lines.*

Moreover, search displays were now preceded by a cue that was visible at the center of the display for 800 ms (see Figure 3). For the participants in Experiment 2A, the cue was either a colored (i.e., green or blue) disk, or a gray disk. It subtended approximately 2.2 degree of visual angle. In Experiment 2B, the cue was the word "green", "blue" or "neutral". In *color cue* trials, the cue preceding the search display indicated the color of the disk that would contain the target, whereas in *neutral cue*

trials the cue was neutral (e.g., gray) and therefore not informative. The cue was a color cue in half of the trials. In the rest of the trials the cue was neutral, and the probability that the target appeared in a green or in a blue disk was equal. The cue was followed by a fixation period for another 600 ms. The fixation symbol was a gray "*" and measured about 0.5 degree of visual angle.

*Procedure*

In addition to the instructions given to participants in Experiment 1, participants in Experiments 2A and 2B were also instructed to direct their attention to the green disks when the cue was green, and to direct their attention to the blue disks when the cue was blue. They were further informed that a neutral cue was equally likely to be followed by a target in a green or in a blue disk.

The response was followed by an interval of 600 ms until the onset of the cue for the following trial. The experiment consisted of eleven blocks of 36 trials, preceded by 24 practice trials. For the rest, the procedure was the same as in Experiment 1.

## Results

RTs that were faster than 200 ms or slower than 4000 ms were excluded from the analysis. This removed 0.31% of the trials in Experiment 2A and 0.93% of the trials in Experiment 2B. Figure 4A shows the RT and the error rate as a function of the number of identically colored elements and cueing for the explicit cues used in Experiment 2A. Figure 4B plots the RT and the error rate as a function of the number of identically colored elements and cueing for the symbolic cues used in Experiment 2B.

***Experiment 2A.*** RTs were submitted to an ANOVA with cueing, the number of identical elements, and eccentricity as within-subject variables. There were main effects of cueing, $F(1, 17) = 21.04$, $p < .001$; the number of identical elements, $F(8, 136) = 27.20$, $p < .001$ (Greenhouse-Geisser), and eccentricity, $F(1, 17) = 145.10$, $p < .001$. The Cueing × Eccentricity interaction, $F(1, 17) = 14.20$, $p = .002$ and the Number Of Identical Elements × Eccentricity interaction, $F(8, 136) = 7.78$, $p < .001$ (Greenhouse-Geisser) were also significant.

Planned comparisons between pairs of successive conditions showed that responses were faster in the condition in which the number of identical elements was 1 versus 3, $F(1, 17) = 59.20$, $p < .001$; 5 versus 7, $F(1, 17) = 13.11$, $p = .002$; 8

versus 10, $F(1, 17) = 4.51$, $p = .049$; and 15 versus 14, $F(1, 17) = 5.49$, $p = .032$ (3 versus 5, $F(1, 17) = 4.39$, $p = .051$).

The mean RT for each level of the number of identical elements was compared between the color cue and the neutral cue condition. Paired samples t-tests (two-tailed) revealed faster responses in the color cue condition than in the neutral cue condition for 1 identical element $t(17) = 5.09$, $p < .001$; 3 identical elements $t(17) = 3.25$, $p = .005$; 5 identical elements $t(17) = 3.58$, $p = .002$; 7 identical elements $t(17) = 2.52$, $p = .022$; and 8 identical elements $t(17) = 2.75$, $p = .014$.

An ANOVA of the error rate with cueing, the number of identical elements, and eccentricity as within-subject variables, revealed only an effect for eccentricity, $F(1, 17) = 12.79$, $p = .002$. Figure 5 shows the RT and the error rate as a function of the number of identically colored elements and cueing for the explicit cues used in Experiment 2A, separately for targets on the small and on the large presentation circle. As can be seen in Figure 5, errors are equally distributed over all conditions, except for the two conditions of eccentricity. The error rate is higher for targets on the large presentation circle than for targets on the small presentation circle. As the RT is also higher for targets on the large presentation circle than for targets on the small presentation circle, speed-accuracy trade-offs can be excluded. We will not include the effects of eccentricity in the results of Experiment 2B, because they are similar and they are not the main interest of this chapter.

***Experiment 2B.*** RTs were examined by an ANOVA with cueing and the number of identical elements as within-subject variables. The main effect of cueing, $F(1, 17) = 26.92$, $p < .001$, the main effect of the number of identical elements, $F(8, 136) = 42.47$, $p < .001$, and the Cueing × Number Of Identical Elements interaction, $F(8,136) = 3.23$, $p = .002$, were all significant.

Planned comparisons between pairs of successive conditions showed that responses were faster in the condition in which the number of identical elements was 1 versus 3, $F(1, 17) = 58.93$, $p < .001$; 3 versus 5, $F(1, 17) = 35.91$, $p < .001$; 5 versus 7, $F(1, 17) = 10.06$, $p = .006$; and 15 versus 14, $F(1, 17) = 5.71$, $p = .029$.

The mean RT for each level of the number of identical elements was compared between the color cue and the neutral cue condition. Paired samples t-tests (two-tailed) revealed faster responses in the color cue condition than in the neutral cue condition for 1 identical element, $t(17) = 4.21$, $p = .001$; 3 identical elements, $t(17) = 5.93$, $p < .001$; 5 identical elements, $t(17) = 2.65$, $p = .017$; 7 identical elements, $t(17)$

= 3.38, $p$ = .004; 8 identical elements, $t(17)$ = 2.69, $p$ = .016; 10 identical elements, $t(17)$ = 3.81, $p$ = .001; and 12 identical elements, $t(17)$ = 2.35, $p$ = .031.

Error rates were submitted to an ANOVA with cueing and the number of identical elements as within-subject variables, revealing a main effect of cueing, $F(1, 17)$ = 5.21, $p$ = .036. As can be seen in Figure 4B, the error rate is higher in the color cue condition than in the neutral cue condition. The RT is lower for targets in the color cue condition than in the neutral cue condition. Yet, it is unlikely that there is a full speed-accuracy trade-off, as the increase in error rate is mainly observable in the conditions with 1 to 5 identical elements, whereas the decrease in RT is significant in the conditions with 1 to 12 identical elements (see Figure 4B).

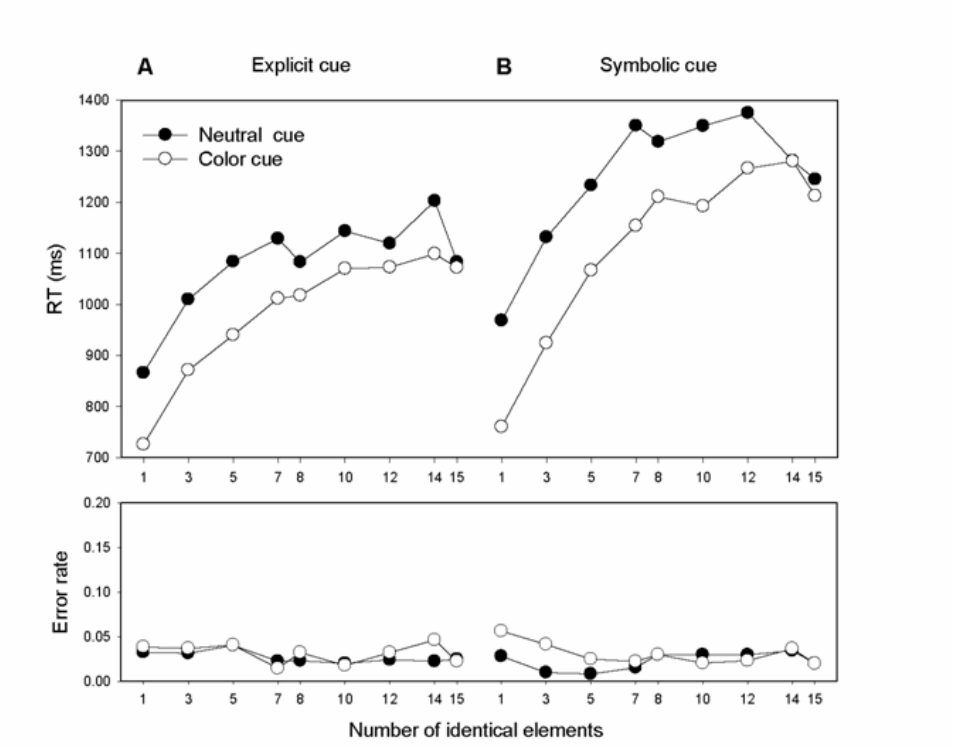


*Figure 4. (A) Response time (top) and error rate (bottom) in Experiment 2A as a function of the number of identical elements, on one of which the target was superimposed, and cueing.* (B) *Response time (top) and error rate (bottom) in Experiment 2B as a function of the number of identical elements, on one of which the target was superimposed, and cueing.*

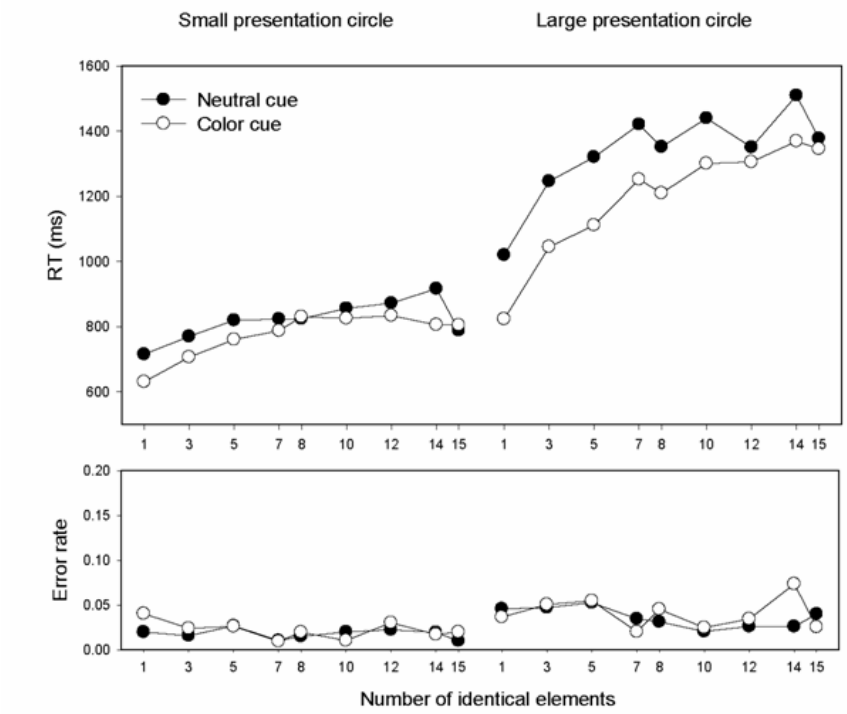


*Figure 5. (A) Response time (top) and error rate (bottom) in Experiment 2A as a function of the number of identical elements, on one of which the target was superimposed, and cueing, for targets on the small presentation circle. (B) Response time (top) and error rate (bottom) in Experiment 2A as a function of the number of identical elements, on one of which the target was superimposed, and cueing, for targets on the large presentation circle.*

***Experiments 2A and 2B.*** To examine the influence of cue type (explicit vs. symbolic), the data of Experiments 2A and 2B were analyzed together. An ANOVA of RTs with the number of identical elements and cueing as within-subject variables, and cue type as a between-subject variable, revealed that the responses tended to be faster for explicit cues than for symbolic cues, $F(1, 34) = 4.059$, $p = .052$. There was no significant interaction between cue type and the number of identical elements and between cue type and cueing.

Error rates were also submitted to a three-way ANOVA, revealing no main effect of cue type, and no interaction between cue type and the number of identical elements and between cue type and cueing.

### Discussion

The neutral cue condition in this experiment replicates the effect of gradual saliency found in Experiment 1. The addition of an oriented black line in all of the disks did not change its effect, except for perhaps making it stronger, as it made the search for the target more difficult. The most important result of Experiments 2A and 2B is that top-down visual attention speeds up the search for targets that are located on an element with the cued color, even when the target is located on a color singleton, or on an element from a minority colored set with more than one element (see Figures 4A and 4B). In fact, in Experiment 2B, top-down visual attention speeds up the search for the target more strongly with fewer elements with the cued color, on one of which the target is located. The reason probably is that a cue is more informative (i.e., selective) in conditions in which there are relatively few elements with the cued color. When all the elements have the same color, top-down visual attention does not speed up the responses. This indicates that top-down visual attention does not make participants more attentive in general. It appears to facilitate, exclusively, the selection of elements with one color among differently colored elements.

Top-down visual attention produces a stronger RT benefit for targets on the large presentation circle than for targets on the small presentation circle, as shown in Experiment 2A (see Figure 5). This finding is replicated in Experiment 2B and in the following experiments of this chapter. It will not be repeatedly discussed, because it is not the main interest of this chapter.

The findings further suggest that the explicit cues in Experiment 2A speeded up search in a similar manner as the symbolic cues in Experiment 2B, although the overall RT was somewhat slower for symbolic cues than for explicit cues (see Figures 4A and 4B). It is likely that both cue types activated top-down visual attention.

## Experiments 3A and 3B: Gradual saliency and top-down visual attention with briefly visible color displays

The previous experiments showed that elements become increasingly salient as fewer and fewer elements in the display share a color, while the benefit of shifting attention to elements from a minority colored set is restricted. Also, top-down visual attention is shown to speed up the search, when the target appears on an element from a minority colored set with more than one element, or even when it appears on a color singleton.

Yet, the colored elements were simultaneously present with the oriented lines. Accordingly, each colored element and oriented line formed a contrast. Top-down visual attention might thus have speeded up the search by *enhancing this contrast*. In Experiments 3A and 3B, we presented the colored elements only briefly before the oriented lines (i.e. the search task). If top-down visual attention (and gradual saliency) still speeds up the search, this can be attributed to *shifts of attention to locations of elements with the cued color* (and to the locations of elements that are salient). In Experiment 3A, we used colored disks as cues, whereas we used words in Experiment 3B.

## Method

### *Participants*

A total of forty-three participants from the same student population as described in Experiment 1 were tested (22 in Experiment 3A and 21 in Experiment 3B)

### *Stimuli*

Stimuli were presented on the same apparatus as in Experiment 1. The sequence of the stimuli of Experiments 2A and 2B was modified. The colored disks in the search display were now presented for 200 ms. After that, the oriented lines appeared on the preceding disks locations (see Figure 6).

The stimuli were the same as in Experiments 2A and 2B, with two differences. First, the oriented lines, which were black in Experiments 2A and 2B, were gray in order to be visible against the black background. The green and the blue disks, and the gray oriented lines were all made equiluminant. Second, the target line was horizontal or vertical. The orientation of the other lines was randomly chosen to be either 22.5° tilted to the left, 22.5° tilted to the right, 67.5° tilted to the left or 67.5° tilted to the right. Horizontal or vertical lines do not pop out between heterogeneously oriented tilted lines, meaning that participants had to rely on serial search (Theeuwes, 1992). The parallel search task of Experiments 2A and 2B was substituted by this more difficult, serial search task in order to encourage participants to make use of informative, top-down cues.

### *Procedure*

The procedure in Experiments 3A and 3B was the same as in Experiments 2A and 2B, with two exceptions. First, participants were instructed to indicate whether the orientation of the target line was horizontal or vertical, by pressing one of two

keyboard buttons. Second, participants were now instructed to direct their attention to the locations of the green disks when the cue was green, and to direct their attention to the locations of the blue disks when the cue was blue.

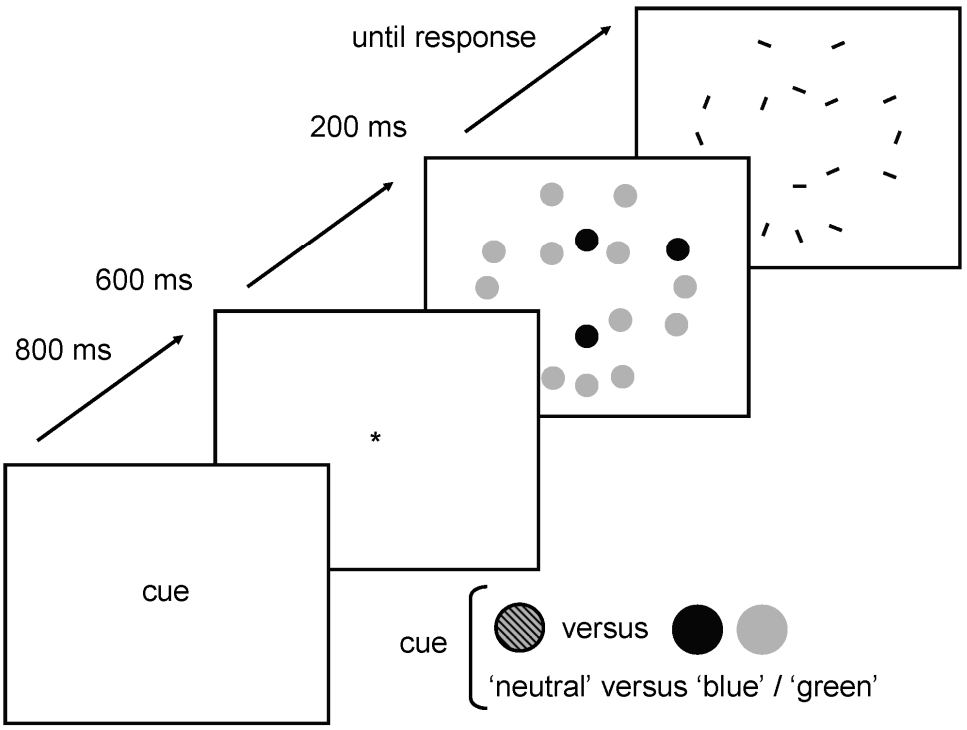


*Figure 6. Sequence of displays in Experiments 3A and 3B. Gray denotes the color blue, black denotes the color green, and the black lines denote the gray lines.*

**Results**

Data of two participants in Experiment 3A and of four participants in Experiment 3B were excluded from analysis, because they had an average error rate equal or higher than 20% over all trials. The average error rate over all other participants was 6.10% in Experiment 3A and 3.59% in Experiment 3B. RTs that were faster than 200 ms or slower than 6000 ms were excluded from the analysis. This removed 1.97% of the trials in Experiment 3A and 1.62% of the trials in Experiment 3B. Figure 7A shows the RT and the error rate as a function of the number of identically colored elements, and cueing for the explicit cues used in Experiment 3A. Figure 7B plots the RT and the error rate as a function of the

number of identically colored elements, and cueing for the symbolic cues used in Experiment 3B.

***Experiment 3A.*** RTs were examined by an ANOVA with cueing and the number of identical elements as within-subject variables. The main effect of cueing, $F(1, 19) = 49.11$, $p < .001$, the main effect of the number of identical elements, $F(8, 152) = 51.81$, $p < .001$ (Greenhouse-Geisser), and the Cueing × Number Of Identical Elements interaction, $F(8, 152) = 3.08$, $p = .014$ (Greenhouse-Geisser), were all significant.

Planned comparisons between pairs of successive conditions showed that responses were faster in the condition in which the number of identical elements was 1 versus 3, $F(1, 19) = 33.09$, $p < .001$; 3 versus 5, $F(1, 19) = 59.48$, $p < .001$; 5 versus 7, $F(1, 19) = 9.15$, $p = .007$; and 15 versus 14, $F(1, 19) = 5.51$, $p = .030$.

The mean RT for each level of the number of identical elements was compared between the color cue and the neutral cue condition. Paired samples t-tests (two-tailed) revealed faster responses in the color cue condition than in the neutral cue condition for 1 identical element, $t(19) = 6.38$, $p < .001$; 3 identical elements $t(19) = 5.11$, $p < .001$; 5 identical elements $t(19) = 4.51$, $p < .001$; 7 identical elements $t(19) = 3.83$, $p = .001$; 8 identical elements $t(19) = 2.76$, $p = .012$; and 12 identical elements $t(19) = 2.34$, $p = .030$.

Error rates were submitted to an ANOVA with cueing and the number of identical elements as within-subject variables, revealing no significant effects. Hence, speed-accuracy trade-offs cannot explain the results.

***Experiment 3B.*** RTs were examined by an ANOVA with cueing and the number of identical elements as within-subject variables. The main effect of cueing, $F(1, 16) = 18.85$, $p = .001$, the main effect of the number of identical elements, $F(8, 128) = 59.85$, $p < .001$ (Greenhouse-Geisser), and the Cueing × Number Of Identical Elements interaction, $F(8,128) = 4.75$, $p < .001$, were all significant.

Planned comparisons between pairs of successive conditions showed that responses were faster in the condition in which the number of identical elements was 1 versus 3, $F(1, 16) = 54.34$, $p < .001$; 3 versus 5, $F(1, 16) = 45.90$, $p < .001$; and 5 versus 7, $F(1, 16) = 8.22$, $p = .011$.

The mean RT for each level of the number of identical elements was compared between the color cue and the neutral cue condition. Paired samples t-tests (two-tailed) revealed faster responses in the color cue condition than in the neutral cue

condition for 1 identical element, $t(16) = 4.68$, $p < .001$; 3 identical elements, $t(16) = 4.46$, $p < .001$; 5 identical elements, $t(16) = 3.03$, $p = .008$; 7 identical elements, $t(16) = 2.91$, $p = .010$; and 8 identical elements, $t(16) = 2.28$, $p = .037$.

Error rates were submitted to an ANOVA with cueing and the number of identical elements as within-subject variables, revealing a main effect of cueing, $F(1, 16) = 5.38$, $p = .034$. The Cueing × Number Of Identical Elements interaction, $F(8, 128) = 2.32$, $p = .024$, was also significant. As can be seen in Figure 7B, the error rate is higher in the color cue condition than in the neutral cue condition, and this increase is more pronounced with fewer identical elements. Likewise, the RT is lower for targets in the color cue condition than in the neutral cue condition, and this decrease is more pronounced with fewer identical elements. Nevertheless, it is unlikely that there is a full speed-accuracy trade-off, as the increase in error rate is mainly observable in the conditions with 1 to 5 identical elements, whereas the decrease in RT is significant in the conditions with 1 to 8 identical elements (see Figure 7B).

***Experiments 3A and 3B.*** To examine the influence of cue type (explicit vs. symbolic), the data of Experiments 3A and 3B were analyzed together. An ANOVA with the number of identical elements and cueing as within-subject variables, and cue type as a between-subject variable, revealed no main effect of cue type, and no interaction between cue type and the number of identical elements and between cue type and cueing.

Error rates were also submitted to a three-way ANOVA, revealing no main effect of cue type, and no interaction between cue type and the number of identical elements. The interaction between cue type and cueing was significant, $F(1, 35) = 4.72$, $p = .037$, indicating that the error rate was higher in the color cue condition than in the neutral cue condition for symbolic cues, but not for explicit cues.

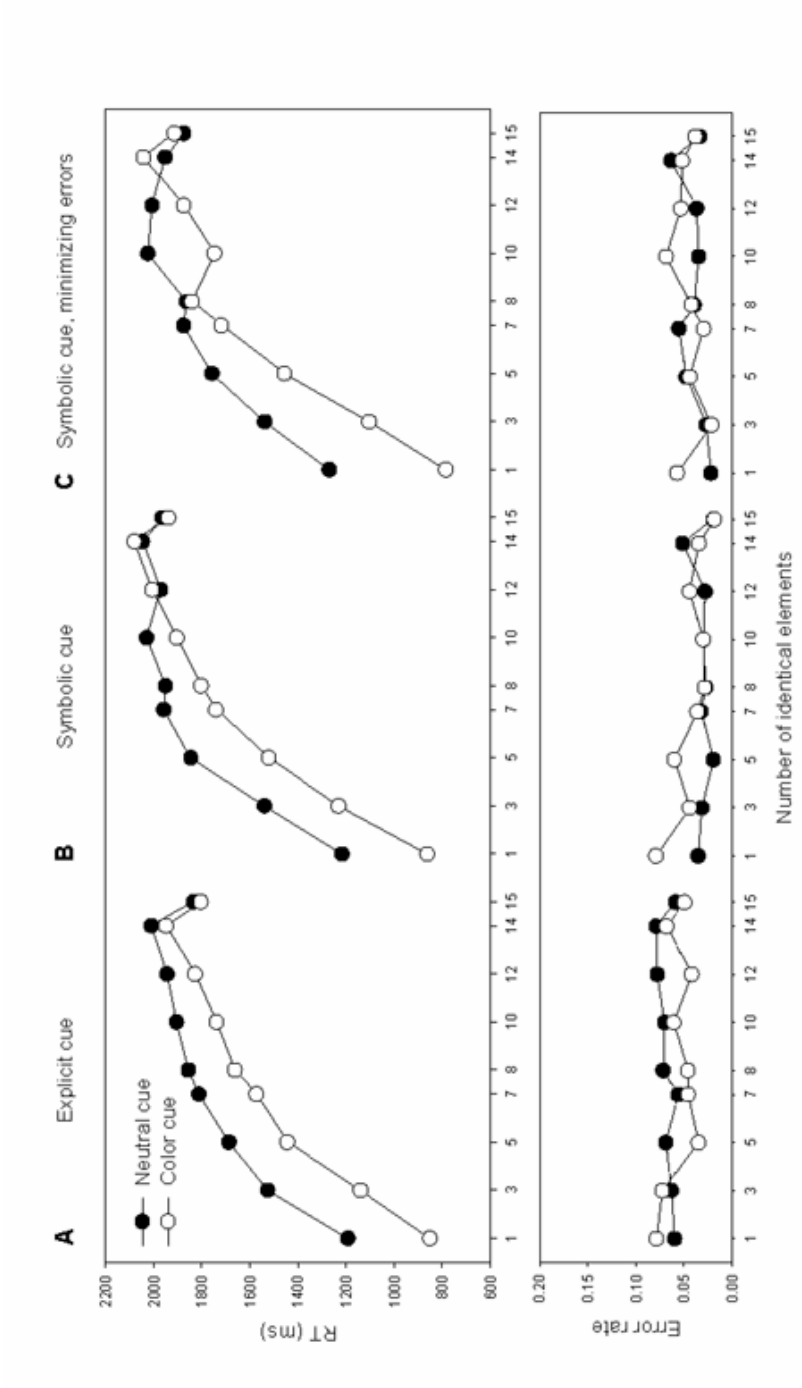


*Figure 7. (A) Response time (top) and error rate (bottom) in Experiment 3A as a function of the number of identical elements, on one of which the target was superimposed, and cueing. (B) Response time (top) and error rate (bottom) in Experiment 3B as a function of the number of identical elements, on one of which the target was superimposed, and cueing. (C) Response time (top) and error rate (bottom) in Experiment 4 as a function of the number of identical elements, on one of which the target was superimposed, and cueing.*

***Experiments 2A, 2B, 3A and 3B.*** Paired samples t-tests (two-tailed) were conducted to evaluate whether attention was shifted to distracting (color) singletons in color cue conditions. We compared the mean RT and error rate between the color cue condition with a distracting singleton (where 14 elements were in the cued color) and the color cue condition in which all the elements have the same color (where 15 elements were in the cued color), for Experiments 2A, 2B, 3A and 3B. The results are shown in Table 1. Responses are slower in the color cue condition with a distracting singleton than in the color cue condition in which all the elements have the same color, in Experiment 2B ($t(17) = 2.419$, $p = .027$) and in Experiment 3B ($t(16) = 2.488$, $p = .024$). There also is a trend of slower responses in the color cue condition with a distracting singleton than in the color cue condition in which all the elements have the same color, in Experiments 2A and 3A. Finally, in Experiments 2A, 2B, 3A and 3B there is a trend of a higher error rate in the color cue condition with a distracting singleton than in the color cue condition in which all the elements have the same color.

*Table 1*

*Response time and error rate for the color cue condition with a distracting singleton and the color cue condition in which all the elements have the same color, and the significance level of a paired samples t-test between both conditions*

| Experiment | RT | | | Error rate | | |
|---|---|---|---|---|---|---|
| | 14 | 15 | *p* | 14 | 15 | *p* |
| 2A | 1112 | 1060 | .195 | .043 | .022 | .146 |
| 2B | 1284 | 1200 | .027 | .037 | .022 | .242 |
| 3A | 2155 | 1999 | .052 | .068 | .050 | .062 |
| 3B | 2292 | 2157 | .024 | .034 | .022 | .179 |

*Note.* 14 = color cue condition with a distracting singleton; 15 = color cue condition in which all the elements have the same color.

## Discussion

The faster responses for targets on locations of color singletons and on locations of elements from minority colored sets with more than one element indicate (covert) attentional shifts toward the locations of these elements, before the presentation of the search display. Likewise, the RT benefit of top-down visual attention reflects (covert) shifts of attention toward the locations of elements with the cued color, before the presentation of the search display.

As in Experiment 2B, in Experiments 3A and 3B the RT benefit of top-down visual attention was increasingly stronger with fewer elements with the cued color. This indicates that top-down visual attention allows the selection of fewer relevant elements in those conditions (i.e., the conditions with fewer elements with the cued color). Nonetheless, this finding is more visible in Experiments 3A and 3B than in Experiments 2A and 2B. The reason might be that the increase in difficulty of the visual search task in Experiments 3A and 3B relative to Experiments 2A and 2B, resulted in a larger benefit of top-down visual attention. As in Experiments 2A and 2B, the results suggest that the explicit cues in Experiment 3A speed up the search in a similar manner as the symbolic cues in Experiment 3B. The results further show that the colored elements elicit the mechanisms that are responsible for gradual saliency, and enable selection by top-down visual attention, when they are presented for 200 ms.

Finally, in the color cue conditions of Experiments 2A, 2B, 3A and 3B, responses were generally slower and more erroneous in the condition with a distracting singleton than in the condition in which all the elements have the same color. This indicates that top-down visual attention does not confine the search to the subset of elements with the cued color. Visual attention is also shifted to distracting singletons.

Unfortunately, the error rate was significantly higher in the color cue condition than in the neutral cue condition in Experiment 3B, and this increase was stronger with fewer elements with the cued color. This issue was addressed in Experiment 4.

## Experiment 4: Prioritizing accuracy over speed

Experiment 4 is a replication of Experiment 3B, with the difference that participants were instructed to prioritize accuracy over RT. This instruction should lead to a more equal distribution of the errors over all the conditions than in Experiment 3B. Consequently, this would enable us to more clearly evaluate the effects on the response time of top-down visual attention, set by symbolic cues, and of the interaction between gradual saliency and top-down visual attention.

### Method

*Participants*

A total of fourteen participants from the same student population as described in Experiment 1 were tested.

*Stimuli*

The apparatus and the stimuli were the same as in Experiment 3B.

*Procedure*

The procedure in Experiment 4 was the same as in Experiment 3B, except for one part of the instruction. In Experiment 4, participants were explicitly requested to respond without making errors, and next, to respond as quickly as possible.

## Results

Data of three participants were excluded from analysis, because they had an average error rate equal or higher than 20% over all trials. The average error rate over all other participants was 4.25%. As in Experiment 3B, RTs that were faster than 200 ms or slower than 6000 ms were excluded from the analysis. This removed 1.19% of the trials. Figure 7C shows the RT and the error rate as a function of the number of identically colored elements and cueing for the symbolic cues.

RTs were examined by an ANOVA with cueing and the number of identical elements as within-subject variables. The main effect of cueing, $F(1, 10) = 32.32$, $p < .001$, the main effect of the number of identical elements, $F(8, 80) = 43.74$, $p < .001$ (Greenhouse-Geisser), and the Cueing × Number Of Identical Elements interaction, $F(8, 80) = 3.91$, $p = .011$ (Greenhouse-Geisser), were all significant.

Planned comparisons between pairs of successive conditions showed that responses were faster in the condition in which the number of identical elements was 1 versus 3, $F(1, 10) = 29.40$, $p < .001$; 3 versus 5, $F(1, 10) = 48.69$, $p < .001$; 5 versus 7, $F(1, 10) = 11.97$, $p = .006$; and 15 versus 14, $F(1, 10) = 5.34$, $p = .044$.

The mean RT for each level of the number of identical elements was compared between the color cue and the neutral cue condition. Paired samples t-tests (two-tailed) revealed faster responses in the color cue condition than in the neutral cue condition for 1 identical element, $t(10) = 6.73$, $p < .001$; 3 identical elements, $t(10) = 5.53$, $p < .001$; and 5 identical elements, $t(16) = 4.00$, $p = .003$.

Error rates were submitted to an ANOVA with cueing and the number of identical elements as within-subject variables, revealing no significant effects. Hence, speed-accuracy trade-offs cannot explain the results.

### Discussion

Experiment 4 replicated the RT benefits for gradual saliency and top-down visual attention that were found in Experiment 3B (see Figure 7C). Instructing participants not to make any errors resulted in an equal distribution of the errors over all the conditions. Experiment 4 thus underlines the finding of Experiment 3B that symbolic cues allow covert attentional shifts to locations of elements with the cued color, before the onset of the search display.

## Experiment 5: Gradual saliency and top-down visual attention over time

The previous experiments showed that elements from a minority colored set with more than one element are salient in an analogous manner as color singletons, albeit to a lesser extent, and that this at least partly reflects covert attentional shifts. In addition, top-down visual attention appeared to facilitate the selection of potential target locations (also at least partly by means of covert attentional shifts), even when the target location already was (gradually) salient. In this experiment we investigated the dynamics of such gradual saliency and top-down visual attention over time. Thereto, we manipulated the duration of the display consisting of the colored elements (i.e., from 50 ms to 200 ms), in addition to the number of identically colored elements, and top-down visual attention. In the neutral cue condition, the varying stimulus onset asynchrony (SOA) conditions may elucidate the dynamics of (gradual) saliency over time. In the color cue condition, the varying SOA conditions may illuminate the combined dynamics of (gradual) saliency and top-down visual attention over time.

### Method

*Participants*

A total of twenty participants from the same student population as described in Experiment 1 were tested.

*Stimuli*

Stimuli were presented on the same apparatus as in Experiment 1. The stimuli in Experiment 5 were a subset of the stimuli in Experiment 3A, but the presentation time of the colored disks was varied. The interval between the onset of the colored disks and the onset of the gray oriented lines, the SOA (equal to the presentation time of the colored disks), was 50 ms, 100 ms, 150 ms or 200 ms. In order to limit

the total number of trials, the ratio between the numbers of disks of each color was less extensively varied. Each search display was equally likely to contain 1, 3, 7, 12, or 14 disks of one color with 14, 12, 8, 3, or 1 disks of the other color. The target was equally likely to be placed in one of 1, 3, 7, 12, or 14 identically colored disks. SOA, the number of identical elements, and cueing were all randomized within each block of trials.

*Procedure*

The procedure in Experiment 5 was the same as in Experiment 4, except for the number of trials. Each participant performed three sessions, each consisting of eight blocks of 40 trials, and 24 practice trials. After the completion of both the first and the second session, participants had a mandatory, five-minute break.

## Results

Data of one participant were excluded from analysis, because it had an average error rate equal or higher than 20% over all trials. The average error rate over all other participants was 4.04%. RTs that were faster than 200 ms or slower than 6000 ms were excluded from the analysis. This removed 1.35% of the trials. Figure 8 shows the RT and the error rate as a function of the number of identically colored elements, cueing, and SOA.

RTs were submitted to an ANOVA with the number of identical elements, cueing, and SOA as within-subject variables. There were main effects of the number of identical elements, $F(4, 72) = 101.40$, $p < .001$ (Greenhouse-Geisser), and cueing, $F(1, 18) = 17.00$, $p = .001$. The Number Of Identical Elements × Cueing interaction, $F(4, 72) = 5.27$, $p = .004$ (Greenhouse-Geisser), and the Cueing × SOA interaction, $F(3, 54) = 4.52$, $p = .007$ were also significant.

The mean RT for each SOA condition was compared between the color cue and the neutral cue condition. Paired samples t-tests (two-tailed) revealed faster responses in the color cue condition than in the neutral cue condition for a SOA of 100 ms, $t(18) = 3.99$, $p = .001$; 150 ms, $t(18) = 4.00$, $p = .001$; and 200 ms, $t(18) = 3.20$, $p = .005$.

Planned comparisons between all pairs of SOA conditions showed that cueing resulted in a larger RT benefit (the difference between the neutral cue and color cue condition) in the condition in which the SOA was 100 ms versus 50 ms, $t(18) = 3.13$, $p = .006$; 150 ms versus 50 ms, $t(18) = 3.48$, $p = .003$; and 200 ms versus 50 ms, $t(18) = 2.14$, $p = .046$. Figure 9 shows the RT as a function of SOA, and cueing.

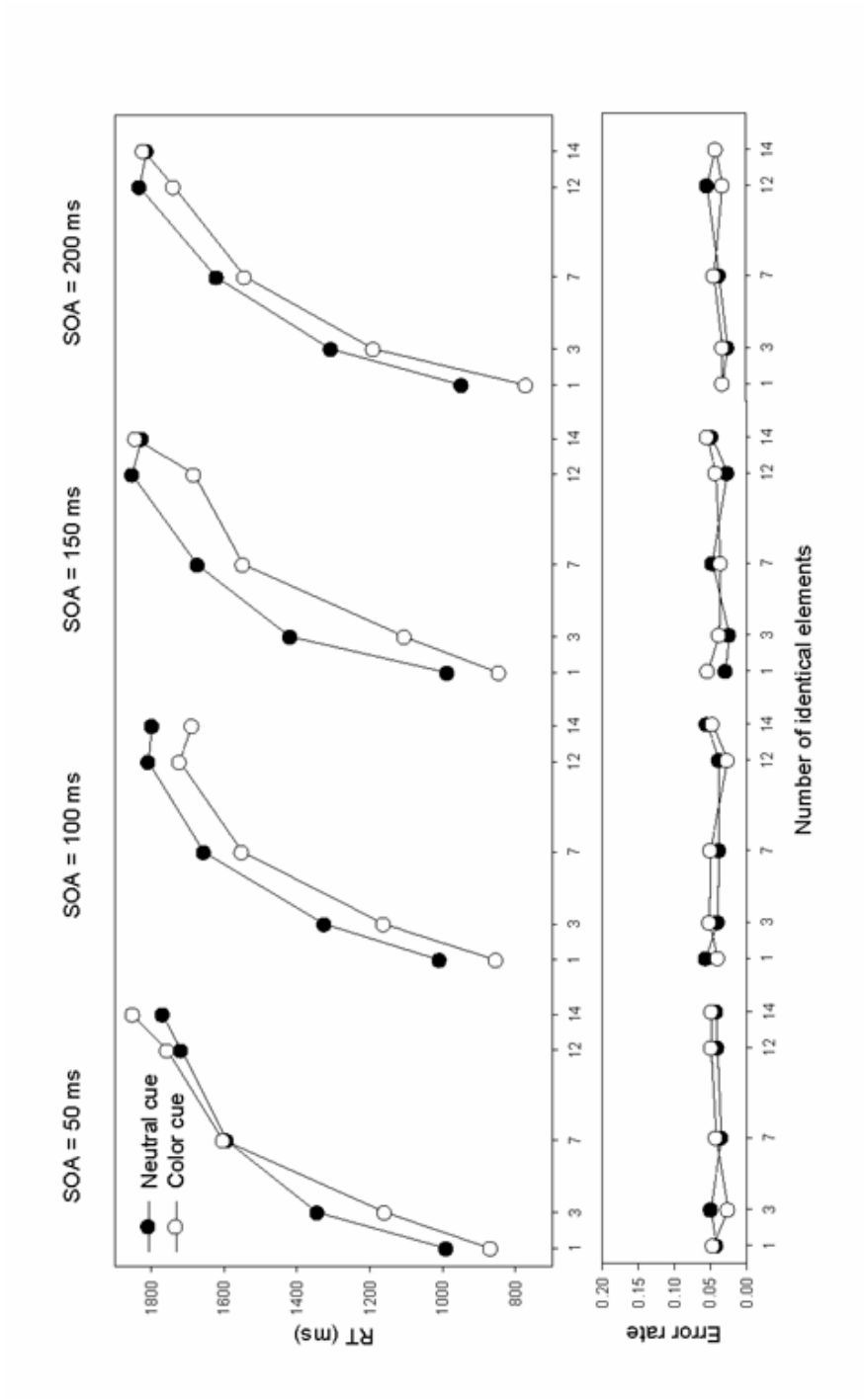


*Figure 8. Response time (top) and error rate (bottom) in Experiment 5 as a function of the number of identical elements (on one of which the target was superimposed) cueing, and SOA.*

An ANOVA of the error rate with the number of identical elements, cueing, and SOA as within-subject variables revealed no significant effects. Hence, speed-accuracy trade-offs cannot explain the results.

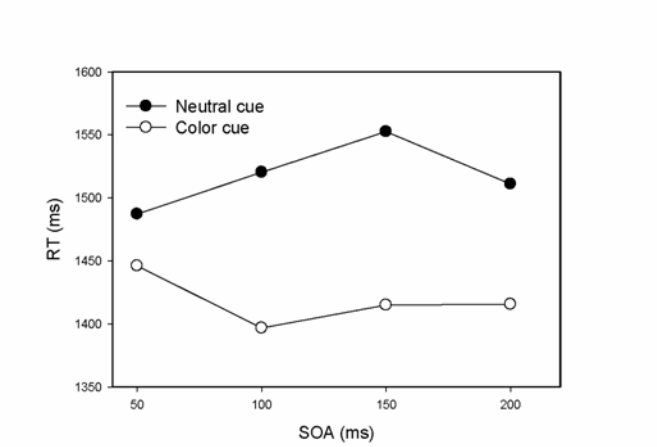


*Figure 9. Response time in Experiment 5 as a function of cueing and SOA. The data are collapsed over all conditions of the number of identical elements.*

## Discussion

The results confirm our previous findings, and show that the faster responses for targets on previously cued and salient locations are fairly stable for varying durations of the colored elements. The presentation time of the colored elements within the range of 50 ms to 200 ms did not modulate the RT for targets, whether they appeared on the location of elements from a minority colored set, or on the location of elements from a majority colored set (see Figure 8). In other words, gradual saliency for these colored elements does not develop progressively within the time range of 50 ms to 200 ms. The colored elements already trigger the mechanisms responsible for gradual saliency, when they are presented for 50 ms.

The RT benefit of top-down visual attention was smaller when the colored elements were presented for 50 ms than when the colored elements were presented for 100 ms to 200 ms (see Figure 9). In fact, responses were not reliably faster in the color cue condition than in the neutral cue condition when the colored elements were presented for 50 ms. Thus, the colored elements enable selection by top-down visual attention, when they are presented for 100 ms to 200 ms.

## General discussion

The objective of this study was to determine whether elements from a minority colored set with more than one element are salient in a similar manner as color singletons, and to investigate the interaction of this gradual saliency with top-down visual attention.

### Gradual saliency

In Experiment 1 and in the neutral cue conditions of Experiments 2-5, we found that responses are fastest for targets on color singletons, but also that responses for targets on elements from a minority colored set with more than one element are faster than responses for targets on elements from a majority colored set. This result reflects that elements from a minority colored set with more than one element are searched earlier or faster than elements from a majority colored set, and are thus prioritized in search in a similar manner as color singletons. We referred to this as gradual saliency.

Our finding of gradual saliency is consistent with earlier studies that show smaller-group search in conjunction search (Sobel & Cave, 2002; Zohary & Hochstein, 1989). In conjunction search participants have an incentive to search the smaller group first, as the target is always present among the smaller group of distracters. In our design, the target appeared with equal likelihood on one of the elements from the minority colored set or on one of the elements from the majority colored set. In principle, it is possible that participants still had an incentive to voluntarily search elements from the minority colored set before elements from the majority colored set, since each individual element from the minority colored set had a higher probability that the target was placed on it than each individual element from the majority colored set. However, an analysis of the predicted RTs according to the strategy of voluntarily searching elements from the minority colored set before elements from the majority colored set showed that this strategy does not explain our results (see Experiment 1). Hence, gradual saliency as observed here is not an artifact of strategic incentives.

Furthermore, Experiment 5 indicates that the colored elements already trigger the mechanisms responsible for gradual saliency when they are presented for 50 ms. For these colored elements, gradual saliency does not develop progressively within the time range of 50 ms to 200 ms.

## The interaction between gradual saliency and top-down visual attention

Experiments 2-5 show that top-down visual attention speeds up the search for a target, while its location is already salient. Top-down visual attention even made the search for a target faster, when it appeared on a color singleton. We observed similar results for explicit and symbolic cues. Experiments 3A and 3B excluded the possibility that top-down visual attention speeds up the search by enhancing the contrast between each element with the cued color and an oriented line. The brief presentation of the colored elements before the oriented lines (i.e. the search task) ensured that the faster responses in the color cue condition than in the neutral cue condition can be attributed to (covert) shifts of attention to locations of elements with the cued color. Furthermore, the absence of a RT benefit of top-down visual attention in the condition in which all the elements have the same color, shows that the RT benefit of top-down visual attention in the presence of (gradual) saliency cannot be explained by a generally increased level of attention. Top-down visual attention thus appears to facilitate, exclusively, the selection of elements with the cued color among all colored elements.

Finally, in Experiment 5, we found that top-down visual attention speeds up the responses for targets on locations of elements with the cued color, when the colored elements were presented from 100 ms to 200 ms. Only when the colored elements were presented for 50 ms, the RT benefit of top-down visual attention disappeared.

Our finding that top-down visual attention speeds up the search for a target, while its location is already salient, is in line with one of the findings of Sobel and Cave (2002). They found that the search for a target in a conjunction search task was mainly guided by saliency, as long as the two defining features of the target were both highly discriminable from their distracter features, and as long as the display was dense. Nevertheless, instructions to search for the target by limiting search to one type of distracters (i.e., top-down visual attention) had a small (but reliable) effect when the guiding feature was much more discriminable from its distracting feature than the other target feature. In Sobel and Cave's (2002) experiments, targets were defined by a combination of features on two feature dimensions. The attentional set thus always encompassed both defining features to some extent, independent of specific instructions to limit search to one set of distracters. Hence, specific instructions to limit the search to one set of distracters can only bias top-down visual attention slightly toward one target feature with respect to the other target feature. In our experiments, the target itself was defined by another feature

dimension than our experimental manipulation of top-down visual attention. Although the target feature was always included in the attentional set, top-down visual attention for color was either present or absent. This made the benefit from top-down visual attention (for color) very transparent in our experiments, whereas it was partly hidden in Sobel and Cave's (2002) task design.

### The attentional mechanisms underlying the interaction between gradual saliency and top-down visual attention

We found evidence for gradual saliency and further demonstrated that top-down visual attention speeds up the search for a target that is located on an element with a cued color, even when the target is located on a color singleton. Faster responses for targets on color singletons after top-down cues indicate that top-down visual attention is fast enough to interact with the mechanisms underlying saliency.

The Guided Search 2.0 (Wolfe, 1994) and FeatureGate (Cave, 1999) models of visual attention suggest that the selection of locations for attention is jointly governed by two subsystems. The bottom-up subsystem favors locations with unique features, and the top-down subsystem favors locations with features designated as target features. Each subsystem independently calculates an activation for each location, and these activations are summed to produce an overall activation for a location. Locations compete for selection on the basis of their activations. After the selection and processing of a location, the selected location is inhibited and a new competition cycle results in the selection of another location. In both Guided Search 2.0 (Wolfe, 1994) and FeatureGate (Cave, 1999) the bottom-up subsystem produces pop-out by increasing the salience of objects with features that differ from those in neighboring locations. There is no assumption that this is an all-or-none process, hence both models would predict our finding of gradual saliency. Also, their assumption that the bottom-up and the top-down subsystems determine the selection of a location in an additive manner is consistent with our finding that top-down visual attention is fast enough to interact with (gradual) saliency.

However, Guided Search 2.0 (Wolfe, 1994) and FeatureGate (Cave, 1999) would not predict that top-down visual attention speeds up the responses for targets on color singletons with the cued color. The reason is that the bottom-up subsystem already allocates a much stronger activation to the color singleton location than to other locations, and an even higher activation due to the top-down subsystem should not further speed up the selection of the color singleton location. However,

strong random noise in the (bottom-up) activations (Wolfe, 1994), could, in principle, explain the benefit of top-down visual attention in addition to saliency. Alternatively, incorporating the temporal dynamics of the competition between locations could also explain the faster responses for targets on color singletons in the presence of top-down visual attention. A color singleton location with an increased activation due to the presence of top-down visual attention may be faster to win its competition with other locations. Guided Search 2.0 (Wolfe, 1994) could also account for the benefit of top-down visual attention in addition to saliency by treating visual attention as a limited-capacity (spatially) parallel process (Wolfe, 1994), in which the rate of information processing at each location is proportional to the size of its activation (Carrasco & McElree, 2001; Wolfe, Butcher, Lee, & Hyle, 2003).

We have previously proposed a neural network model of visual object-based attention (CLAM) (Van der Velde et al., 2004), in which the identity (e.g., shape or color) of an object is used to select its location among other objects (also see, Van der Velde & De Kamps, 2001). This model consists of a feedforward network that identifies the shape and color of objects in the visual field, and a feedback network that reciprocates the connections of the feedforward network. The selectivity in the feedforward network is transferred to the feedback network using Hebbian learning (Van der Velde & De Kamps, 2001). How does this architecture allow spatial attention to shift to elements with a cued color? Suppose the feedforward network identifies elements in two colors in its visual field. The feedback network carries back information about the cued color to the lower (retinotopic) areas of the model. In these areas, interaction between the feedforward network and the feedback network (in local microcircuits) selects activation produced by elements with the cued color (Van der Velde & De Kamps, 2001). This selected activation is equivalent to directing spatial attention to the location of elements with the cued color. This neural network model can explain the RT benefit of top-down visual attention in our experiments.

In Chapter 7, we propose the Global Saliency Model (GSM). This model consists of two pathways: ventral and dorsal. The ventral pathway is based on Van der Velde and De Kamps' (2001) neural network model of visual object-based attention, and the dorsal pathway consists of a number of interacting spatial maps. The ventral and dorsal pathways interact in the model. As discussed in Chapter 7, GSM is consistent with both our finding of gradual saliency, and our finding that top-down visual attention is fast enough to interact with (gradual) saliency. In fact, in

Chapter 7 we will present a simulation which compares the model's response with the experiments of this chapter.

It remains unclear to what extent, and up to which processing stage, the mechanisms responsible for (gradual) saliency and top-down visual attention are independent. It is possible that (gradual) saliency is the result of purely bottom-up processing (e.g., Cave, 1999; Itti & Koch, 2000; Treisman & Sato, 1990; Wolfe, 1994). Alternatively, (gradual) saliency may be the result of a voluntary (Zohary & Hochstein, 1989) or an automatic process (Van der Velde, Van der Voort van der Kleij, Haazebroek, & De Kamps, in preparation), which involves top-down processing in addition to bottom-up processing. This issue will be addressed further in Chapters 6 and 7.

Does top-down visual attention always generate faster responses for targets on locations that already are salient? Here we presented one example in which it does within the color dimension. Future experiments might look at the interaction between (gradual) saliency and top-down visual attention, while the strength of (gradual) saliency is further increased (e.g., by manipulating the density or the contrast between two colors), within different dimensions (e.g. orientation, shape).

# Chapter 6 | A review of behavioral and neurophysiological studies and models of visual search

After more than two decades of visual search studies and other studies, there is still a lot of discussion about which mechanisms underlie stimulus-driven visual attention and feature-based visual attention, and how these mechanisms interact. The main questions that are addressed in this chapter are whether efficient search should be associated with processing in low cortical areas, and whether stimulus-driven visual attention is the result of bottom-up and horizontal processing, or alternatively of bottom-up, horizontal, and top-down processing. Several findings of the behavioral studies that we review suggest that efficient search cannot solely be attributed to processing in low cortical areas. The results of reviewed neurophysiological studies leave open whether stimulus-driven visual attention is the result of bottom-up and horizontal processing, or of bottom-up, horizontal, and top-down processing. Finally, an overview is presented of various models that are proposed to explain stimulus-driven and/or feature-based visual attention.

## Behavioral studies of visual search

### A qualitative distinction between parallel feature search and serial conjunction search

In a visual search task, participants generally have to indicate whether a target item is present or absent among a variable number of distracters. When the target is distinguished by a unique feature from the distracters (such as a large difference in color, orientation, or size), the response time is (relatively) independent of the number of distracters. On the other hand, when the target is distinguished by a unique conjunction of features from the distracters (such as a target defined by a conjunction of a color and an orientation among distracters that share either the target color or orientation), the response time often increases with the number of distracters.

On the basis of this observation, Treisman and Gelade (1980) made a qualitative distinction between feature and conjunction search. That is, Treisman and Gelade (1980) hypothesized that feature search reflects parallel processing of all search items across the visual field, whereas conjunction search additionally reflects serial processing of the search items. Specifically, Treisman and Gelade's (1980)

Feature Integration Theory (FIT) proposed that, in a preattentive stage, a set of simple features is registered in parallel in specialized subsystems (i.e., feature maps). When a target is distinguished by a unique feature, the target's presence or absence can be determined by monitoring whether there is (reliable) activation in the map of this feature. When a target is distinguished only by a unique conjunction of features from the distracters, however, focused attention is needed to serially scan the location of one (or at most a few) of the search items, in order to integrate and bind the features at that location. As a result of focused attention, the same location in all feature maps is selected (via a map of locations), and the features of one search item become available for recognition by higher-level areas.
Are there really two qualitatively different modes of visual search? The qualitative distinction between parallel feature search and serial conjunction search has been called into question by many behavioral studies. To remain neutral on the mechanisms underlying visual search, we will label search in which the response time is (relatively) independent of the number of distracters efficient, and search in which the response time increases with the number of distracters inefficient. Furthermore, we will refer to an item that is distinguished by one or more unique features from the other items in the search display (i.e., the distracters) as a target or singleton, and to an item that is distinguished from the other items in the search display by one or more unique features that are cued (e.g., before a session or trial) as a cued-target.

### Inefficient feature and (more) efficient conjunction searches

Some behavioral studies reported inefficient feature search, when the target differs only a little along a feature dimension from the distracters (reviewed in Duncan & Humphreys, 1989). At the same time, other behavioral studies found efficient search for color-orientation conjunctive cued-targets when the feature saliency is high enough (Wolfe, Cave, & Franzel, 1989), for cued-targets defined by a conjunction of stereoscopic disparity and color or a conjunction of stereoscopic disparity and motion (Nakayama & Silverman, 1986), and for cued-targets defined by a conjunction of motion (i.e., moving versus static) and shape (McLeod, Driver, & Crisp, 1988). Furthermore, even for inefficient conjunction searches, the response time was shown to depend not only on the number of distracters, but also on the ratio of the number of the two distracter types used (Bacon & Egeth, 1997; Egeth et al., 1984; Kaptein et al., 1995; Sobel & Cave, 2002; Zohary & Hochstein, 1989). For example, Zohary and Hochstein (1989) asked participants to

search for a red horizontal element, whereas the distracters were green horizontal elements and red vertical elements. Zohary and Hochstein (1989) varied the proportions of the two types of distracters, and showed that the search for the cued-target proceeded through the smallest group of distracters.

Thus, behavioral studies provided evidence that feature search may be inefficient, whereas conjunction search may be efficient, or more efficient than would be predicted by a strictly serial search. This is corroborated by a meta-analysis by Wolfe (1998), which indicated that the overall distribution of search slopes from 2500 experimental sessions (i.e., a single participant doing a single search task) across six categories of searches (e.g., feature searches, conjunction searches) is unimodal. Obviously, a unimodal distribution of search slopes does not provide support for a simple, data-driven distinction between parallel feature and serial conjunction search (Wolfe, 1998). The observation that feature search may be inefficient, whereas conjunction search may be (more) efficient led to two classes of models that discarded a qualitative distinction between feature and conjunction search (Mordkoff, Yantis, & Egeth, 1990).

### Early parallel processing may guide subsequent serial processing

One class of models that incorporated results of efficient conjunction search supposes that the output of early, parallel processing guides the subsequent deployment of focused attention. Wolfe, Cave, and Franzel (1989) first advanced this proposal in their Guided Search model. In this model, early, parallel processing can guide subsequent serial processing (almost) directly to the location of a conjunctive cued-target if the target features are salient enough. However, if the target features are not salient enough, early, parallel processing is noisy, and cannot guide subsequent serial processing directly to the location of the conjunctive cued-target. Treisman and Sato (1990) also maintained the distinction between early, parallel processing and subsequent serial processing of the search items. Like Wolfe et al. (1989), Treisman and Sato (1990) suggested that early, parallel processing may guide subsequent serial processing to conjunctive cued-targets if the target features are sufficiently salient. Nonetheless, the mechanism through which early, parallel processing may guide subsequent serial processing in the revised FIT (Treisman & Sato, 1990) differs from the one in Guided Search (Wolfe et al., 1989). Whereas early, parallel processing activates locations with target features for subsequent serial processing in Guided Search (Wolfe et al., 1989), early, parallel processing inhibits locations with distracter features for

subsequent serial processing in the revised FIT (Treisman & Sato, 1990). That is, Treisman and Sato (1990) extended FIT with a feature inhibition mechanism, which can simultaneously inhibit all features that are specific to the distracters.

### Parallel processing capacity may be limited for feature and conjunction search

The other class of models rejected the assumption that parallel processing is necessarily followed by strictly serial processing when searching for a conjunctive cued-target. Alternatively, this class of models supposes that a small number of search items can be processed simultaneously in both feature and conjunction search (Duncan & Humphreys, 1989; Mordkoff et al., 1990; Pashler, 1987). In other words, these models suggest a limited parallel processing capacity. As a consequence, search is inefficient when the limited parallel processing capacity is exceeded and efficient when it is not. Therefore, feature and conjunction search may either be efficient or inefficient. Several behavioral studies provided evidence for this class of models (e.g., Mordkoff et al., 1990; Pashler, 1987).

For example, Mordkoff et al. (1990) asked participants to indicate whether a cued-target defined by a conjunction of color and shape (i.e., a red X) was present, both in a redundant-target condition in which the display contained two cued-targets, and in single-target conditions in which the display contained one cued-target (with or without a distracter). Mordkoff et al.'s (1990) results not only showed that fast response times were more frequent in the redundant-target condition than in the single-target conditions, but also that the fast response times were faster in the redundant-target condition than in the single-target conditions. Even when Mordkoff et al. (1990, Experiment 3) kept the number of target features that were present in a display constant across the redundant-target and single-target conditions by using a setsize of six items, the fast response times still were faster in the redundant-target condition than in the single-target conditions (after some practice). Strictly serial processing of search items prior to response selection cannot explain these results, as this would solely yield a larger number of fast response times for displays containing two cued-targets than for displays containing one cued-target, but would not yield faster response times for displays containing two cued-targets than for displays containing one cued-target. Instead, Mordkoff et al.'s (1990) results suggest that at least two search items may simultaneously affect the decision for target presence or absence. Hence, given that search for the conjunctive cued-target was shown to be inefficient in a

separate experiment, this result is consistent with limited-capacity parallel processing models.

Deco, Pollatos, and Zihl (2002) proposed a model that abandons serial processing altogether. In their model, search items are always processed in parallel across the visual field. Yet, the model produces differences in search efficiency across conditions of feature and conjunction search due to different latencies of the model's dynamics across these conditions. We will review Deco et al.'s (2002) model and other models below.

### Associating efficient search with low cortical areas

The qualitative distinction between parallel feature searches and serial conjunction searches was accompanied by the implicit assumption that the features supporting efficient search are the same as the features of early vision (i.e., the features that are being found to excite neurons in low cortical areas, such as the primary or extrastriate visual cortex) (e.g., Treisman & Gelade, 1980). However, this assumption has been questioned in a number of ways.

First of all, the differences in orientation and color that neurons in low cortical areas are able to discriminate during visual processing with attention are finer than the differences that result in efficient search (Hochstein & Ahissar, 2002; Wolfe, 2003). In other words, the just noticeable difference is much cruder for efficient search than for early visual processing with attention. One behavioral study even suggested that efficient search can use only information about the categorical status of items (Wolfe, Friedman-Hill, Stewart, & O'Connell, 1992).

Secondly, search can be efficient over a large range of spatial scales, far exceeding the small receptive fields of neurons in the primary visual cortex (Hochstein & Ahissar, 2002; Shipp, 2004) and other low cortical areas (Hochstein & Ahissar, 2002). Hence, efficient search cannot fully be explained by the inhibitory and excitatory connections between neurons in low cortical areas. Nonetheless, the connections between neurons in low cortical areas may play a significant role in efficient search, especially at smaller spatial scales (Li, 2002). An early example of invariable search performance across a large range of spatial scales came from a study by Bergen and Julesz (1983). Bergen and Julesz (1983) tested participants' accuracy to discriminate a singleton in a search display of seven search items, which varied over a range of a factor eight in size (i.e., the stimuli subtended 2.8 - 21.8 degree of visual angle). Bergen and Julesz (1983) found that the uniform contraction or dilation of the stimulus had little effect on search performance.

Finally, efficient search is reported not only for simple features (e.g., color, orientation) that are defined by luminance contrast, but also for simple features that are defined by other properties than luminance contrast (Bravo & Blake, 1990; Wolfe, 2003) and for high-level features, which include the result of quite sophisticated processing (reviewed in Grossberg, Mingolla, & Ross, 1994; Hochstein & Ahissar, 2002). While simple features that are defined by luminance contrast are thought to be encoded at the earliest stages of cortical processing (e.g., area V1), simple features that are defined by other properties than luminance contrast and high-level features are thought to be encoded at later stages of cortical processing. Behavioral studies that report efficient search for high-level features probably provide the strongest evidence against associating efficient search with low cortical areas. Therefore, we will review a number of these studies.

### Efficient search for high-level features

Ramachandran (1988) first reported that three dimensional (3D) convex shapes ("bumps") that are conveyed by top to bottom differences in shading can be grouped together perceptually and segregated from a background of concave shapes ("cavities"). Kleffner and Ramachandran (1992) later extended this finding by demonstrating that 3D shape from top to bottom differences in shading can provide the basis for efficient search as well. Interestingly, search was not efficient for shapes that are conveyed by left to right differences in shading. The result that 3D shape from top to bottom differences in shading, but not from left to right differences in shading, can provide the basis for efficient search implies that relatively complex scene-based characteristics such as the direction of lighting influence visual search (Kleffner & Ramachandran, 1992). Further, Kleffner and Ramachandran (1992) excluded the possibility that efficient search for shapes that are conveyed by top to bottom differences in shading can simply be attributed to a difference in luminance polarity between the cued-target and distracters. Search was significantly less efficient in a control condition, in which the cued-target and distracters still differed from each other in luminance polarity, but not in 3D shape. Thus, efficient search (largely) depended on a difference in 3D shape. The cued-target and distracters in the control condition were formed by a step-change in luminance instead of a gradual change in luminance.

Similarly, Enns and Rensink (1990) let participants search for cued-targets composed of lines and polygons shaded with one of three intensities (i.e., white, gray, or black). In conditions in which the cued-target and distracters

corresponded to 3D blocks that differed in the direction of lightning, and optionally additionally in orientation, search was efficient. Search was not efficient in conditions in which the cued-target and distracters were two dimensional. Therefore, Enns and Rensink (1990) proposed that efficient search may be based on scene-based properties such as 3D orientation and the direction of lighting. Because such scene-based properties are only captured by the spatial relations among Enns and Rensink's (1990) lines and shaded polygons, they require relatively complex visual processing. Aks and Enns (1992) subsequently attempted to unravel whether possible precursors of scene-based properties, such as the type of shading gradient, the shape of the contour enclosing the gradient, and the background luminance contribute additively or interactively to the efficiency in visual search. Thereto, Aks and Enns (1992) combined these factors orthogonally in a visual search experiment. The results suggested that the type of shading gradient, the shape of the contour enclosing the gradient, and the background luminance influence the search efficiency in an additive manner. This led Aks and Enns (1992) to conclude that efficient search is not guided by specialized detectors for scene-based properties such as surface curvature and the direction of lighting, but instead by precursors to a rich 3D representation. Nonetheless, precursors to a rich 3D representation still include the result of relatively complex visual processing.

Furthermore, other behavioral studies showed that visual search follows completion processes facilitated by binocular disparity (He & Nakayama, 1992) and monocular cues (Rensink & Enns, 1998). He and Nakayama (1992) asked participants to search for an L-shape (mirrored L-shape) among mirrored L-shapes (L-shapes), while each search item was accompanied by a square. The binocular disparity was varied across conditions so that the search items either all appeared to be in a depth plane in front of the squares, or in a depth plane behind the squares. When the target and distracters appeared to be in a depth plane in front of the squares, eliminating an opportunity for perceptual completion of the target and distracters, search was efficient. Search was also efficient when the target and distracters appeared to be in a depth plane behind the squares, whilst a small gap between each L-shape and square eliminated the opportunity for perceptual completion of the target and distracters. However, when the target and distracters appeared to be in a depth plane behind the squares, and the relative position between the squares and the search items offered an opportunity for perceptual completion of the target and distracters, search was inefficient. He and

Nakayama's (1992) results indicate that binocular disparity can reduce the search efficiency when it facilitates surface completion of the target and distracters behind adjacent occluders, which makes the target and distracter perceptually more similar. As participants could not choose to apply search at a lower-level representation of feature detection, at which level search would have been easier, He and Nakayama (1992) suggested that efficient search probably has to be applied at a higher-level representation of perceived shapes or surfaces.

Finally, Wolfe, Friedman-Hill, and Bilsky (1994) had participants search for cued-targets (i.e., houses) defined by a conjunction of two colors. In conditions in which the cued-target could be characterized in hierarchical terms as a whole item of one color with a part of another color, search was (relatively) efficient. In comparison, in conditions in which the cued-target consisted of two equal parts that differed in color, search was less efficient. In two control experiments, Wolfe et al. (1994) ruled out some simple explanations in terms of the relative sizes of colored regions, for the efficient search in the part-whole condition. Wolfe et al.'s (1994) results add to the picture that efficient search incorporates quite sophisticated processing (e.g., the abstraction of part-whole relationships) beyond the mere extraction of basic features.

Other behavioral studies suggested efficient search for high-level features such as threatening faces (Ohman, Lundqvist, & Esteves, 2001) and one's own face (Tong & Nakayama, 1999). However, we will not discuss these studies, because it is highly debated whether low-level or high-level features provide the basis for efficient search in these studies (for an overview, see Wolfe & Horowitz, 2004), and because an extensive overview falls outside the scope of this chapter.

In summary, behavioral studies provided converging evidence that search for a (cued-)target that is distinguished by a high-level feature can be efficient. It also appeared that search for a target can be less efficient due to high-level completion processes, which render the target and distracters less distinguishable. It is important to remark that we do not argue which high-level features exactly were responsible for efficient search. The target-distracter pairs investigated in the studies above likely differ in multiple high-level features, and efficient search may have been based on either one of those. For example, the shading in Ramachandran (1988), Kleffner and Ramachandran (1992), and Aks and Enns' (1992) study may have created high-level features such as surface orientation, direction of lighting, and precursors to a rich 3D representation. Irrespective of which high-level feature led to efficient search, it is evident that efficient search

may, and sometimes even has to, be based on the results of later stages of cortical processing. Hence, efficient search is not confined to (cued-)targets distinguished by features that are encoded at the earliest stages of cortical processing.

## Neurophysiological evidence

Besides behavioral data, results from neurophysiological studies also constrain the development of neurally plausible models of visual search. Behavioral studies allow only inferential conclusions about the stages of processing that make up response times. In contrast, neurophysiological measures (i.e., neuronal activity) can provide markers that distinguish between the end of one stage of processing and the beginning of another (Schall & Thompson, 1999). The neural mechanisms of the selection of a target among distracters can be studied with the highest spatial and temporal resolution by recording the activity of single neurons in monkeys (Schall & Thompson, 1999).

A number of neurophysiological studies have investigated the selection of a target among distracters in (efficient) feature search (Bichot et al., 2005; Bichot & Schall, 2002; Constantinidis & Steinmetz, 2001; McPeek & Keller, 2002; Schall, Hanes, Thompson, & King, 1995; Thompson, Hanes, Bichot, & Schall, 1996) and (inefficient) conjunction search for cued-targets defined by a unique combination of shape and color (Bichot et al., 2005; Bichot & Schall, 1999; Gottlieb, Kusunoki, & Goldberg, 1998).[4] As noted in the previous section, in feature search a target is distinguished from distracters by a unique feature, although the exact value of this feature typically changes over trials (e.g., a white target among black distracters or a black target among white distracters). In contrast, in conjunction search a target is distinguished from distracters by a unique combination of features that usually remains the same over trials. Therefore, in feature search in which the target features are unknown, attentional mechanisms have to select the target by virtue of its physical saliency, whereas in conjunction search in which the target features are known, attentional mechanisms may also use top-down knowledge about the target features to select the cued-target. For this reason, feature search is associated with stimulus-driven visual attention, and conjunction search with a combination of stimulus-driven visual attention and top-down visual attention for one or more features (i.e., feature-based visual attention).

Nevertheless, top-down knowledge about the target features may also play a role in feature search when the target features are unknown. For example, (implicit) expectations about the target and distracters features, which are raised by

repetitions of the target and distracters features in previous trials, decrease the response time in feature search (i.e., priming of pop-out) (Maljkovic & Nakayama, 1994). Furthermore, even when the target features are unknown, the behavioral task of detecting a target in feature search may employ specific attentional mechanisms. In order to examine whether attention automatically selects an item that is distinguished from other items by a unique feature or combination of features, some neurophysiological studies presented search displays to monkeys when they were irrelevant to the behavioral task (i.e., when monkeys only maintained fixation) (Constantinidis & Steinmetz, 2005; Hegdé & Felleman, 2003; Thompson, Bichot, & Schall, 1997).

What have neurophysiological studies revealed about the neural process of discriminating an item from other items, while monkeys passively view search displays, perform a feature search task, or perform a conjunction search task? That is, which neural correlates at the single cell level (and other levels) have been found for stimulus-driven visual attention (both in a passive fixation task and feature search task) and the combination of stimulus-driven and feature-based visual attention respectively?

## Neural correlate of stimulus-driven visual attention

*V1 neurons are not specifically selective for feature discontinuities leading to efficient search*

Previous neurophysiological studies have shown that already in the primary visual cortex (area V1) many neurons respond more strongly to pop-out center-surround stimuli, in which a single item in the classical receptive field (CRF) is surrounded by items that differ in a feature, than to homogeneous center-surround stimuli, in which the item centered on the CRF is identical to the items in the surround (e.g., Knierim & Van Essen, 1992). This result, which is also found in anesthetized animals (Kastner, Nothdurft, & Pigarev, 1999; Nothdurft, Gallant, & Van Essen, 1999), has been interpreted as evidence that pop-out results from selection at the earliest stages of cortical processing, (largely) independent of top-down processing of visual information (e.g., Kastner et al., 1999; Knierim & Van Essen, 1992; Li, 2002).

However, a study by Hedgé and Felleman (2003) recently challenged this interpretation. Hedgé and Felleman (2003) presented a set of 36 different stimuli, consisting of a single bar of a preferred or non preferred color and orientation in the CRF and none or 58-109 bars in the surround, to monkeys that had to

maintain fixation. The set of stimuli contained center-alone stimuli, and homogenous, pop-out and conjunction center-surround stimuli. Hedgé and Felleman's (2003) results indicated that, according to many different response measures, neurons in area V1 typically respond similarly to pop-out and conjunction center-surround stimuli. Hence, neurons in area V1 appear to be selective for feature discontinuities in general, and not specifically for the kind of feature discontinuities that lead to perceptual pop-out (i.e., efficient search).

*The time course of neural target discrimination in the PP, the FEF and the SC*

Neurophysiological studies have provided converging evidence that neurons in the posterior parietal cortex (PP), the frontal eye field (FEF), and the superior colliculus (SC) distinguish an item that is defined by a unique feature from other items (i.e., a singleton), regardless of whether monkeys passively view stimuli (Constantinidis & Steinmetz, 2005; Thompson et al., 1997) or search for the singleton (Constantinidis & Steinmetz, 2001; McPeek & Keller, 2002; Thompson et al., 1996). The time course of the neuronal target discrimination is also investigated in the PP, the FEF, and the SC. Interestingly, several neurophysiological studies indicated that the first feedforward sweep of visual information through the brain does not discriminate a target from distracters in these areas, even when the target is distinguished by a unique feature from the distracters (Constantinidis & Steinmetz, 2001, 2005; McPeek & Keller, 2002; Thompson et al., 1997; Thompson et al., 1996). Instead, the neuronal discrimination of a singleton from distracters in the PP (Constantinidis & Steinmetz, 2001, 2005), the FEF (Thompson et al., 1997; Thompson et al., 1996), and the SC (McPeek & Keller, 2002) appears to occur in the following epoch, which involves both horizontal and feedback processing (J. H. Fecteau, personal communication, January 24, 2006).

Constantinidis and Steinmetz (2005) trained monkeys to maintain fixation while they presented single items (i.e., green or red squares), arrays of nine items of which one item differed in color (i.e., a green square among red squares or a red square among green squares), or arrays of nine identical items (i.e., green squares or red squares). They analyzed the responses of neurons in area 7a of the PP that displayed significant selectivity for the spatial location of a single item. These neurons responded most strongly to a single item in the center of their receptive field, and stronger to a singleton in the center of their receptive field than to one of the 'distracters' in the center of their receptive field. The responses to one of the

homogenous items in the center of the receptive field were weaker than the responses to a singleton in the center of their receptive field, but stronger than one of the 'distracters' in the center of their receptive field.

Furthermore, Constantinidis and Steinmetz's (2005) results showed that responses to a singleton in the center of the receptive field and to one of the 'distracters' in the center of the receptive field initially were remarkably similar. After a burst of activity from 50 ms to 150 ms after stimulus onset, responses to these stimuli slightly decreased. Only after 180 ms, responses to a singleton in the center of the receptive field became reliably stronger than responses to one of the 'distracters' in the center of the receptive field. Not only do Constantinidis and Steinmetz's (2005) results suggest that attentional mechanisms automatically select an item that is distinguished from other items by a unique feature such as color, but also that this selection occurs only after about 180 ms.

In Thompson et al.'s (1996) study, monkeys were trained to shift gaze to a target that was distinguished by either color or form from the distracters in either of two complementary feature search displays (e.g., a red target among green distracters, or a green target among red distracters). Initially, the neural activity of visually responsive neurons in the FEF did not discriminate between the presence of the target or a distracter in their receptive field. After about 100 ms, a selection process occurred that resulted in a higher level of neural activity when the target versus a distracter was present in the receptive field of visually responsive neurons in the FEF. Visually responsive neurons in the FEF thus ultimately indicated the location of the target. A later study showed that this neuronal process of discriminating the target from the distracters was not dependent of the planning of a saccade (i.e., a fast eye movement) (Thompson et al., 1997). Even when monkeys maintained fixation at the center of the search display, the neurons in the FEF still discriminated the singleton from a 'distracter'.

McPeek and Keller (2002) investigated the time course of the neuronal target discrimination in the SC. They trained monkeys to make a saccade to a target that was distinguished by a unique color from three distracters (i.e., a red target among green distracters, or a green target among red distracters). McPeek and Keller (2002) found that a subset of visuo-movement (VM) neurons (i.e., neurons showing significant visual and saccade-related activity) discriminated the target from a distracter at a time that was nearly independent of saccade latency. Thus, this subset of VM neurons may be primarily involved in the selection of the target, as opposed to eye movement commands.

Initially, the activity of the subset of VM neurons that were selective for the target in the SC did not discriminate the target from a distracter, as was the case for neurons in area 7a of the PP (Constantinidis & Steinmetz, 2001, 2005) and the visually responsive neurons in the FEF (Thompson et al., 1997; Thompson et al., 1996). In fact, the discrimination time for the subset of VM neurons that were selective for the target was typically about 100-130 ms for VM burst neurons and about 140-150 ms for VM prelude neurons. For VM burst neurons, the discrimination time coincided with their second burst of activity. The timing of the neuronal target discrimination for these VM neurons in the SC is quite similar to the timing of the neuronal target discrimination for the visually responsive neurons in the FEF (Thompson et al., 1997; Thompson et al., 1996).

### Neural correlate of stimulus-driven visual attention and top-down knowledge about the target and distracter features in previous trials

As we briefly discussed above, expectations about the target and distracters features that are raised by repetitions of the target and distracters features in previous trials may be considered as top-down knowledge, whether it is generated in areas of the ventral stream (e.g., within local circuits of the inferotemporal cortex) or somewhere else in the brain (e.g., in the prefrontal cortex) (reviewed in Bichot & Schall, 2002). How does such top-down knowledge influence the neuronal discrimination of a target from distracters in feature search? A study by Bichot and Schall (2002) suggests that in the FEF top-down information about the target and distracters features in previous trials modifies the same neural correlate as the discrimination of a singleton from distracters.

Bichot and Schall (2002) recorded neurons in the FEF, while monkeys performed a feature search task in which either both the target and distracter features or only the distracter features switched across trials with a certain probability, or in blocks of 10 trials. As in the studies that we discussed above, neurons in the FEF initially did not respond selectively to the target. Only in the following epoch, the activity of neurons in the FEF discriminated the target from a distracter. Moreover, the neuronal target discrimination occurred increasingly earlier in time with an increasing number of trials in which the distracter and or target features remained constant. For instance, in trials immediately after a change of the target-distracter relationship the discrimination time was about 200 ms, whereas in trials following 4-9 repetitions of the same target-distracter relationship the discrimination time was about 120 ms. These changes in the time of neuronal

target discrimination caused by change of the target-distracter relationship predicted changes in behavioral performance such as median saccade latency.
In addition, Bichot and Schall (2002) found that as the accuracy of monkeys increased, the target activity increased and the distracter activity decreased. Hence, this suggests that the neuronal target discrimination is mediated by both target enhancement and distracter suppression. Bichot and Schall's (2002) results were similar, regardless of whether both the target and distracter features changed over trials, or only the distracter features.

### Neural correlate of the combination of stimulus-driven and feature-based visual attention

So far we have looked at the neuronal process of discriminating a singleton from distracters when the target features change over trials, i.e., stimulus-driven visual attention. When the target (and distracter) features are known, top-down knowledge about these features may assist the search for the cued-target, in addition to stimulus-driven visual attention. Two recent neurophysiological studies have investigated the neuronal target discrimination in conjunction search in which the target features were known, in the FEF (Bichot & Schall, 1999) and area V4 (Bichot et al., 2005).
Bichot and Schall's (1999) study suggests that the time course of the neuronal discrimination of a (known) conjunctive cued-target from distracters is rather similar to the time course of neuronal discrimination of a (unknown) singleton from distracters in the FEF, even though top-down knowledge about the target features is available to select a cued-target. The monkeys in their study had to execute a saccade to a cued-target defined by a unique combination of color and shape (differently across sessions) among three or five distracters in trials. Bichot and Schall (1999) reported the activity of neurons in the FEF in trials in which the first saccade was directed to the cued-target. As in Thompson et al.'s (1996) study, neurons in the FEF initially responded the same to each search item that appeared in their receptive field. Only some time after stimulus presentation (about 100-130 ms for two of the FEF neurons), neurons in the FEF responded more strongly to the cued-target than to a distracter in their receptive field.
Moreover, Bichot and Schall's (1999) results indicated that FEF neurons not solely discriminated the cued-target from distracters in their receptive field, but also distracters that shared one feature with the cued-target from distracters that shared no feature with the cued-target. FEF neurons additionally discriminated

distracters that had been the cued-target during the previous session from other distracters in the receptive field. These two modulations of FEF activity correlated with the monkeys' tendency to make erroneous saccades to distracters that either shared a cued-target feature or had been the cued-target during the previous session. Bichot and Schall's (1999) finding that FEF neurons discriminate distracters that share one feature with a conjunctive cued-target from distracters that share no feature with the conjunctive cued-target is consistent with models of visual search that suppose that the search efficiency in conjunction search, and in search in general, depends on the similarity between the (cued-)target and distracters (Duncan & Humphreys, 1989).

As neurons in the FEF are not typically selective for visual features (Mohler, Goldberg, & Wurtz, 1973), it is likely that the neuronal discrimination of the cued-target from distracters and of distracters that share one feature with a conjunctive cued-target from distracters that share no feature with the conjunctive cued-target originates from areas of the ventral stream. Indeed, a recent study by Bichot et al. (2005) is consistent with this idea. Bichot et al. (2005) let monkeys freely scan complex search displays in both feature search (e.g., color, shape) and conjunction search. Bichot et al. (2005) recorded neurons in area V4 whose receptive field contained a search item that was not selected for the next saccade. The activity of these V4 neurons was greatest and most strongly synchronized when a preferred stimulus in their receptive field was the cued-target. Moreover, the activity of V4 neurons was greater and more strongly synchronized when a preferred stimulus in their receptive field was a distracter that shared one feature with the conjunctive cued-target, or resembled the feature cued-target, than when a preferred stimulus in their receptive field was a distracter that shared no feature with the conjunctive cued-target, or did not resemble the feature cued-target.

It is important to remark that Bichot et al.'s (2005) study demonstrates that feature-based visual attention enhances the activity of neurons that represent target features in parallel throughout the visual field, at least within area V4. This was suggested by previous neurophysiological and imaging studies, which found that attention for features or objects results in modulated neural activity within visual areas that represent the attended features (Chawla et al., 1999; Martinez-Trujillo & Treue, 2004; Motter, 1994a, 1994b; Saenz et al., 2002) or objects (Chelazzi et al., 1993; O'Craven et al., 1999). Many researchers hypothesized that the neural activity within visual areas that represent the attended features is modulated by feedback signals from the prefrontal cortex, which maintains a

representation of the relevant features (e.g., Deco et al., 2002; Hamker, 2004; Van der Velde & De Kamps, 2001; Van der Velde et al., 2004). Although it is evident how such a mechanism of feature-based visual attention may help to select a cued-target (in combination with spatial visual attention[5]), it is not yet clear how a target is selected when the target features are unknown.

## Implications of behavioral and neurophysiological studies

The results of the behavioral and neurophysiological studies that we reviewed have important implications for models of visual search. First, behavioral studies have reported several findings that suggest that efficient search cannot solely be attributed to processing in low cortical areas: differences in orientation and color that neurons in low cortical areas are able to discriminate during visual processing with attention are finer than the differences that result in efficient search (Hochstein & Ahissar, 2002; Wolfe, 2003); search can be efficient over a large range of spatial scales, far exceeding the small receptive fields of neurons in the primary visual cortex and other low cortical areas (Hochstein & Ahissar, 2002; Shipp, 2004); and efficient search may (Enns & Rensink, 1990; Kleffner & Ramachandran, 1992; Wolfe et al., 1994) and sometimes even has to (He & Nakayama, 1992; Rensink & Enns, 1998) be based on the results of later stages of cortical processing.
Second, several neurophysiological studies have found that stimulus-driven visual attention does not modulate the first feedforward sweep of visual information through the brain (Constantinidis & Steinmetz, 2001, 2005; McPeek and Keller, 2002; Thompson et al., 1996; Thompson et al., 1997). Only in the epoch following the first feedforward sweep of visual information through the brain, responses of neurons in the SC (McPeek and Keller, 2002), FEF (Thompson et al., 1996; Thompson et al., 1997) and PP (Constantinidis & Steinmetz, 2001, 2005) discriminate a singleton from a distracter in the receptive field. This epoch involves both horizontal and top-down processing, and it is not yet clear whether the neuronal target discrimination depends on horizontal and/or top-down processing. Hence, neurophysiological studies leave open whether stimulus-driven visual attention is the result of bottom-up and horizontal processing, or of bottom-up, horizontal, and top-down processing.[6]
Third, in the FEF the neuronal discrimination of a singleton from distracters is faster after repetitions of the distracter and/or target features over consecutive trials than after a change of the distracter and/or target features (Bichot & Schall, 2002). This result is in line with behavioral studies that found faster response

times or saccade latencies (and a higher accuracy) with the repetition of target and/or distracter features over consecutive trials in feature search (Maljkovic & Nakayama, 1994; McPeek, Maljkovic, & Nakayama, 1999).

Finally, even though neurons in the FEF of monkeys are primarily selective for location information, they are also shown to encode the features of a cued-target such as color and shape in a conjunction search task (Bichot & Schall, 1999). This selectivity for target features other than location of the FEF neurons probably reflects the result of feature-based visual attention originating in areas of the ventral stream, such as in area V4. Indeed, in area V4, feature-based visual attention is found to enhance the activity of neurons that represent target features in parallel throughout the visual field (Bichot et al., 2005).

## Models of visual search

In this section we will review various models of visual search. The models incorporate mechanisms of stimulus-driven visual attention (Itti & Koch, 2000; Koch & Ullman, 1985; Li, 2002) feature-based visual attention (Deco et al., 2002; Hamker, 2004; Van der Velde & De Kamps, 2001; Van der Velde et al., 2004) or stimulus-driven and feature-based visual attention (Cave, 1999; Tsotsos et al., 1995; Wolfe, 1994; Wolfe et al., 1989). Most models that incorporate mechanisms of stimulus-driven visual attention assume that stimulus-driven visual attention results from bottom-up and horizontal processing (Cave, 1999; Itti & Koch, 2000; Koch & Ullman, 1985; Li, 2002; Wolfe, 1994). Only one model that incorporates mechanisms of stimulus-driven visual attention suggests that stimulus-driven visual attention results from bottom-up, horizontal, and top-down processing (Tsotsos et al., 1995). Naturally, models that incorporate mechanisms of feature-based visual attention have to rely on top-down processing, because they employ top-down knowledge about relevant features. We have organized our review of models on the basis of the types of visual attention that the models explain (i.e., stimulus-driven visual attention, stimulus-driven and feature-based visual attention, or feature-based visual attention) and the types of processing that are proposed to explain stimulus-driven visual attention (i.e., bottom-up and horizontal processing, or bottom-up, horizontal and top-down processing). Table 1 shows a classification of models based on these two characteristics.

We define bottom-up processing as the processing of a stimulus from lower-level areas to higher-level areas in the visual processing hierarchy, and horizontal processing as the processing of a stimulus within an area in the visual processing

hierarchy. Top-down processing is defined as the processing of a stimulus from higher-level areas to lower-level areas in the visual processing hierarchy.

*Table 1*

*Classification of models based on the types of visual attention that the models explain (i.e., stimulus-driven visual attention, stimulus-driven and feature-based visual attention, or feature-based visual attention) and the types of processing that are proposed to explain stimulus-driven visual attention (i.e., bottom-up and horizontal processing, or bottom-up, horizontal and top-down processing) and feature-based visual attention (i.e., bottom-up, horizontal and top-down processing)*

| Types of VA | Types of processing | Models |
|---|---|---|
| Stimulus-driven VA | Bottom-up and horizontal processing | Koch and Ullman (1985)<br>Itti and Koch (2000)<br>Li (2002) |
| | Bottom-up, horizontal, and top-down processing | |
| Stimulus-driven VA<br>Feature-based VA | Bottom-up and horizontal processing | Wolfe (1994)<br>Cave (1999) |
| | Bottom-up, horizontal, and top-down processing | Tsotsos et al. (1995) |
| Feature-based VA | Bottom-up, horizontal, and top-down processing | Humphreys and Müller (1993)<br>Van der Velde and De Kamps (2001)<br>Deco et al. (2002)<br>Hamker (2004) |

*Note.* VA = visual attention.

## Models of stimulus-driven visual attention

### *Stimulus-driven visual attention results from bottom-up and horizontal processing*

*Koch and Ullman (1985)* were the first to suggest a neurally plausible circuitry of stimulus-driven visual attention. Koch and Ullman's (1985) model (in principle) implements a separate, retinotopic map for each feature that enables pop-out search. The activation in each feature map is linearly summed for each retinotopic location into a saliency map. In the saliency map, representations at different locations, which can be representations of different features (i.e., color, intensity, orientation), compete with each other. The location that is most highly activated in the saliency map wins the competition and attention is directed to that location.

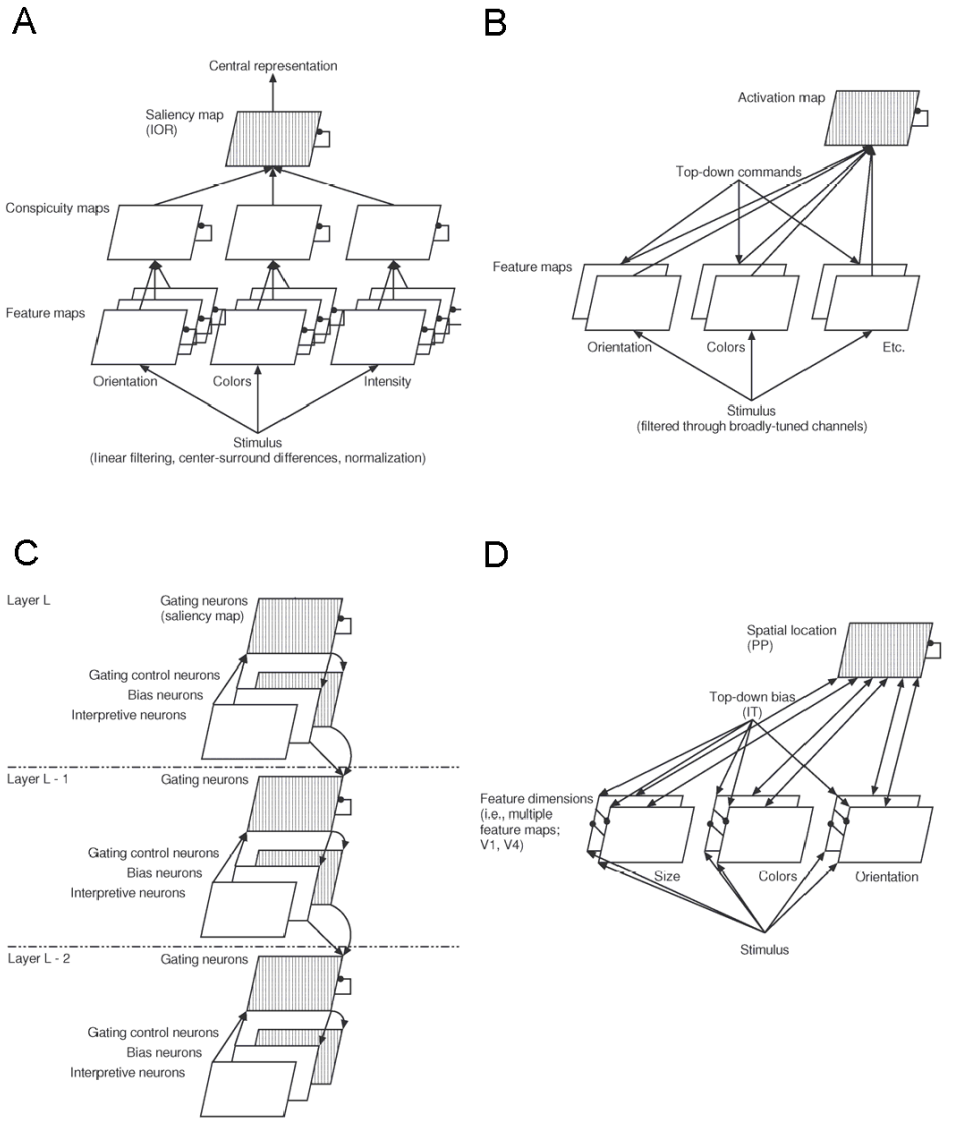


*Figure 1. Models of visual search. Each model has been redrafted to preserve its unique architecture but using a standard pictography to show equivalent elements across models. Lines with an arrowhead at the end denote excitatory connections, while lines with a filled dot at the end denote inhibitory connections. (A) Itti and Koch's (2000) model. (B) Wolfe's (1994) model. (C) Tsotsos et al.'s (1995) model. (D) Deco et al.'s (2002) model.*

Then, the selected location and its neighbors become inhibited in the saliency map and attention switches to the next-most salient location. Following Koch and Ullman (1985), many models of stimulus-driven visual attention have incorporated an implicit or explicit saliency map, i.e. a two-dimensional map that encodes for saliency at every location within the visual field. Yet, these models differ in the mechanisms that process a stimulus to compute saliency.

*Itti and Koch (2000)* presented a new approach to combine information from a variety of feature maps into a saliency map.[7] In Koch and Ullman's (1985) model, representations at different locations do not compete with each other within feature maps, so that the activation in each feature map is directly summed for each retinotopic location into the saliency map. Instead, Itti and Koch (2000) implemented competition between representations at different locations within each feature map (see Figure 1A). As a consequence, the most highly activated representation in each feature map wins the competition. Moreover, after competition within each feature map, the activation in feature maps is summed (across multiple spatial scales) into three separate conspicuity maps (i.e., color, intensity, orientation). For example, the activation in two feature maps encoding for color (i.e., red/green, blue-yellow) is summed into the conspicuity map for color. Within each conspicuity map, representations at different locations again compete with each other. The competition within feature and conspicuity maps allows for the selection of the most highly activated representation in these maps, and thus diminished the likelihood that many comparably activated representations cancel each other out in the saliency map.

In conclusion, the models of Koch and Ullman (1985) and Itti and Koch (2000) suppose that stimulus-driven visual attention results from bottom-up and horizontal processing. Although Koch and Ullman (1985) and Itti and Koch (2000) did not specifically relate the feature (and conspicuity) maps to one or more cortical areas, the processing of low-level features in these maps implies that the feature (and conspicuity) maps are associated with low cortical areas.

*Li (2002)* hypothesized that area V1 provides a saliency map, in which the activation of each neuron increases monotonically with the saliency of the visual input (given a visual scene) in its classical receptive field. Accordingly, two neurons in area V1 are thought to be equally active when the visual input in their classical receptive field is equally salient, even though the two neurons are selective and responding to different features (e.g., one neuron is color selective and the other neuron is motion selective). Li (2002) proposed that bottom-up processing and horizontal

processing within area V1 computes saliency, but his model does not implement explicit feature maps. In line with known excitatory and inhibitory contextual influences observed in area V1 physiology, Li's (2002) model implements iso-orientation (iso-feature) suppression and contour enhancement.

Iso-orientation suppression refers to the suppression of the activation of a neuron with a bar of a certain orientation within its classical receptive field, when the bar is surrounded by other bars of the same orientation. Iso-orientation (iso-feature) suppression results from disynaptically inhibitory connections between pyramidal neurons that code for similar orientations (features). Contour enhancement refers to the enhancement of the activation of a neuron with a bar of a certain orientation within its classical receptive field, when the bar is surrounded by other oriented bars that together form a smooth (isolated) contour. Contour enhancement results from monosynaptically excitory connections between pyramidal neurons.

Because Li's (2002) model attributes efficient search to the activity in area V1, Li's (2002) model does not explain efficient search for high-level features (Enns & Rensink, 1990; Kleffner & Ramachandran, 1992; Wolfe et al., 1994). Li's (2002) model also does not account for the finding that search can be efficient over a large range of spatial scales, far exceeding the small receptive fields of neurons in the primary visual cortex and other low cortical areas (Hochstein & Ahissar, 2002; Shipp, 2004). Nevertheless, Li (2000) gave an interesting explanation for the finding that the differences in orientation and color that neurons in low cortical areas are able to discriminate during visual processing with attention are finer than the differences that result in efficient search (Hochstein & Ahissar, 2002; Wolfe, 2003). That is, Li (2002) suggested that in addition to the response tuning of neurons in area V1, the specificity of their horizontal connections determines the saliency of a target, and consequently efficient search.

## Models of stimulus-driven and feature-based visual attention

### *Stimulus-driven visual attention results from bottom-up and horizontal processing*

*Wolfe (1994)* and *Cave (1999)* too proposed that stimulus-driven visual attention results from bottom-up and horizontal processing in low cortical areas (e.g., extrastriate cortex). In addition to mechanisms of stimulus-driven visual attention, Wolfe's (1994) Guided Search 2.0 and Cave's (1999) FeatureGate incorporate mechanisms of feature-based visual attention. In these models, a bottom-up subsystem favors locations with unique features (i.e., stimulus-driven

visual attention), and a top-down subsystem favors locations with features designated as target features (i.e., feature-based visual attention). Together the bottom-up and top-down subsystem determine the selection of locations for attention. More specific, each subsystem independently calculates an activation for each location, and these activations are summed in the activation map to produce an overall activation for a location (see Figure 1B). Locations compete for selection on the basis of their activations in the activation map. After the selection and processing of the most highly activated location in the activation map, the selected location is inhibited and a new competition cycle results in the selection of the next-most highly activated location in the activation map.

In Guided Search 2.0 (Wolfe, 1994) and FeatureGate (Cave, 1999), the bottom-up subsystem, which is responsible for stimulus-driven visual attention, compares features of each location to those in neighboring locations. Specifically, it increases the activation of locations with features that differ from those in neighboring locations. This is done separately for each feature dimension (i.e., color and orientation). In Guided Search 2.0, this activation is calculated at only one spatial level (Wolfe, 1994). Instead, FeatureGate reduces the number of long range connections that are necessary to compare features of each location to those in other locations by implementing a hierarchy of spatial levels (Cave, 1999). At each spatial level, features of each location are compared to only those in nearby locations. As the size of the receptive fields increases while climbing up the hierarchy, features of each location are compared to those in an increasingly larger area of the visual field.

*Stimulus-driven visual attention results from bottom-up, horizontal, and top-down processing*

Another model of stimulus-driven and feature-based visual attention was presented by *Tsotsos et al. (1995)*. Tsotsos et al.'s (1995) selective tuning model implements stimulus-driven visual attention by bottom-up, horizontal and top-down processing. This model processes visual information in a hierarchy of layers in two cycles. First, visual information is processed by interpretive neurons in a bottom-up manner (see Figure 1C). Each interpretive neuron is linked to a gating neuron. The gating neuron receives input from this interpretive neuron and from a bias neuron. Bias neurons enable the inhibition of specific features or locations that are not task-relevant (i.e., feature-based visual attention).

After the bottom-up cycle, gating neurons at the highest level of the visual processing hierarchy compete with each other; a winner-takes-all (WTA) process. There may be multiple winning gating neurons. Each winning gating neuron activates one more WTA process across the inputs of its associated interpretative neuron at the preceding layer (via a gating control neuron). At the same time, each gating neuron that does not win the competition shuts down the WTA across the inputs of its associated interpretative unit at the preceding layer. As a result, at the preceding level of the visual processing hierarchy, some gating neurons compete with each other, while other gating neurons become inactive. Again, each winning gating neuron selectively activates one more WTA process across the inputs of its associated interpretative neuron at the now preceding layer. Hence, the second cycle of processing consists of a top-down cascade of WTA processes.

As a result, an increasing number of gating neurons becomes inactive at each level of top-down processing. This changes subsequent bottom-up processing, as bottom-up processing by interpretive neurons is restricted to those with an active gating neuron. Eventually, the top level of the visual processing hierarchy represents only the most salient location(s), either or not biased by feature-based visual attention.

## Models of feature-based visual attention

Duncan and Humphreys (1989) presented a theory of visual search when the target features are known (i.e., feature-based visual attention with or without stimulus-driven visual attention). According to this theory, the similarity between search items determines the search efficiency. Specifically, Duncan and Humphreys (1989) hypothesized that search efficiency decreases with increasing target-distracter (T-D) similarity and with decreasing distracter-distracter (D-D) similarity. In addition, T-D similarity and D-D similarity are thought to interact. When the target and the distracters are highly dissimilar, decreasing the D-D similarity does not make search less efficient. When all distracters are highly similar, decreasing the T-D similarity makes search only slightly less efficient. However, search is very inefficient when the T-D similarity is high and the D-D similarity is low. This is the case in which the distracters have many in common with the target, but rather less in common with one other.

In particular, Duncan and Humphreys (1989) suggested that structural units, such as the target and distracters, compete for selection on the basis of selection weights. The selection weight for each structural unit increases in proportion to

the match of the structural unit to the target template. Accordingly, Duncan and Humphreys (1989) distinguished two processes within visual search that are influenced by T-D similarity and D-D similarity. First, the similarity between all possible targets and distracters determines to what extent each search item matches the target template, and thus influences the selection weights. Second, the similarity between all targets and distracters in the visual display additionally determines perceptual grouping between structural units. Perceptual grouping influences the selection weights, because any change in selection weight for one structural unit is distributed to other structural units in proportion to the strength of perceptual grouping between structural units (i.e., weight linkage).

*Humphreys and Müller (1993)* implemented a neural network model of visual search that is in part based on Duncan and Humphreys' (1989) theory. Humphreys and Müller's (1993) Search via Recursive Rejection model (SERR) is a parallel processing model that generates efficient search for a shape conjunction (i.e., an inverted T) among homogenous shape conjunctions (i.e., upright T's) and inefficient search for a shape conjunction (i.e., an inverted T) among heterogeneous shape conjunctions (i.e., upright T's, left-oriented T's, or right-oriented T's). In SERR, objects that are identical group together. The objects that group most strongly are selected and then rejected from further search. Search proceeds until either the cued-target is selected (i.e., target present response) or all objects are rejected (i.e., target absent response). When the distracters form a single group that can be rejected, search is efficient. When there are multiple distracter groups, the probability increases that one of the rejected distracter groups accidentally includes the cued-target. For that reason a time consuming check process is required to reduce the miss rate, which makes search inefficient.

Grouping in SERR is implemented in match maps. There is a separate match map for each target and distracter (i.e., one for an inverted T, upright T, left-oriented T, and right-oriented T), which accumulates evidence for the presence of its object (via the activation of a corresponding template neuron). Connections between neurons within each match map are excitatory, whereas connections between neurons between match maps are inhibitory. The excitory connections between neurons within each match map result in grouping of identical objects. When a group of objects within a match map activates its corresponding template neuron, which codes for the cued-target, search ends. However, when a group of objects within a match map activates its corresponding template neuron, which codes for a distracter, this template neuron inhibits all the neurons in its corresponding

match map and inhibits all locations in which there is evidence only for this object. As a consequence, search proceeds without these distracters.

*Deco et al. (2002)* implemented a neural network model of feature-based visual attention in which visual attention arises as a consequence of continuous competitive interactions within and between modules. Hence, the model does not include an explicit saliency map. Attention for specific features or locations bias this competition, such that the competition is resolved in favor of attended features or locations. Although Deco et al.'s (2002) model processes visual information in parallel across the visual field, it produces differences in search efficiency across conditions of feature and conjunction search. This is due to different latencies of the model's dynamics across these conditions.

Deco et al.'s (2002) model includes a ventral and a dorsal pathway of modules. The ventral pathway consists of the modules V1, V4 and IT, and the dorsal pathway of the modules V1, V4 and PP (see Figure 1D). Modules in the ventral pathway (i.e., V1 and V4) process different feature dimensions (e.g., color, size) of a visual item. Each feature dimension consists of multiple feature maps (e.g., big, small), which extract the values of the features for an item at each position. The PP module represents the location of a visual item. The PP module is bidirectionally connected with the different feature maps, and can bind the different feature dimensions for an item location. Importantly, there is independent competition within each feature dimension. That is, a neuron (i.e., a population of neurons) inhibits all other neurons within a feature dimension, except for neurons that belong to the same feature map. Neurons in the PP module compete at all locations with each other.

When a visual search display is presented, neurons coding for a feature at a location that is present in the display receive excitatory sensory input. Feature-based visual attention is implemented by adding an extra excitatory input to the neurons in the feature maps that correspond to the attended feature(s) at each location. Likewise, spatial visual attention can be implemented by adding an extra excitatory input to the neurons in PP that correspond to the attended location(s). The top-down bias for specific features, which is hypothesized to come from the IT module, biases the competition within each feature dimension so that only the neurons corresponding to the attended feature(s) are able to win the competition. The reason is that these neurons receive both excitatory sensory input and excitatory top-down input. Interaction between the feature maps and the PP

module (intermodular attentional biasing) subsequently results in the selection of the spatial location of the attended feature(s) in PP.

*Hamker (2004)* suggested a similar neural network model of feature-based visual attention. As in Deco et al.'s (2002) model, the location of objects with attended features is selected within a continuous dynamic process in Hamker's (2004) model. Also, Hamker's (2004) model does not include an explicit saliency map. In Hamker's (2004) model, feature-based visual attention biases processing in ventral modules such as IT and V4 via top-down connections from prefrontal areas. First, inputs into IT that match the top-down bias for specific features get enhanced. Next, information about relevant features is transferred to module V4 via top-down processing between the modules IT and V4, and inputs into module V4 that match top-down cues get enhanced.

Unlike Deco et al. (2002), Hamker (2004) does not implement strong competition within the ventral modules IT and V4. Instead, Hamker (2004) hypothesized that spatial competition is embedded in the visuo-motor system by competition in areas that serve for action selection, such as the FEF. Hence, in Hamker's (2004) model, processing within the ventral pathway (i.e., module V4) provides the source for spatial selection, but the spatial competition takes place in the premotor map of the FEF[8] (or PP). The premotor map affects subsequent processing in module V4 via spatially organized connections between the premotor map and module V4. This slow spatial reentry leads to facilitated processing of certain items, but does not fully suppress the activity of not spatially attended items.

*Van der Velde and De Kamps (2001)* proposed a neural model of object-based visual attention. This model and the related CLAM have already been discussed in preceding chapters and will not be described here.

# Chapter 7 | The Global Saliency Model

Most models incorporating mechanisms of global saliency assume that global saliency is the result of bottom-up and horizontal processing in the ventral pathway, i.e. of within-feature competition (Cave, 1999; Itti & Koch, 2000; Li, 2002; Wolfe, 1994). However, consistent with neurophysiological evidence (e.g., Constantinidis & Steinmetz, 2001, 2005; Hegdé & Felleman, 2003), we present the Global Saliency Model (GSM), in which global saliency results from interaction between bottom-up, horizontal and top-down processing in the ventral pathway and bottom-up and horizontal processing in the dorsal pathway. The ventral pathway is based on Van der Velde and De Kamps' (2001) model of object-based visual attention, while the dorsal pathway consists of a number of interacting spatial maps. This architecture solves some problems with within-feature competition models, e.g., an explosion of the number of necessary (inhibitory) horizontal connections, and an early reduction of information. The model presented here can explain several findings in visual search, including the selection of a singleton among distracters, the effects of target-distracter and distracter-distracter similarity, and the findings of the behavioral experiments in Chapter 5.

## Introduction

An object stands out, or *pops out*, among a number of distracters when it is distinguished from the distracters by a large difference along a feature dimension (e.g., color, orientation, size). An example is a red ball among a number of blue balls. The object, the *singleton*, pops out in the sense that the number of distracters does not affect the time it takes to correctly identify its absence or presence in visual search (Wolfe & Horowitz, 2004). The selection of a singleton is automatic. That is, a singleton is selected among distracters even when the singleton and the distracters are irrelevant to the behavioral task (i.e., when the task is only to maintain fixation) (e.g., Constantinidis & Steinmetz, 2005; Thompson et al., 1997).

An object against a uniform background also stands out (e.g., a red ball on a green lawn). This can be explained by the response of neurons in early stages of cortical processing. Neurons in the early stages of cortical processing respond vigorously to a local discontinuity as given by a contour, or a change in color or shading

(Coren, Ward, & Enns, 2003). Accordingly, these neurons respond vigorously to the discontinuity between an object and its adjacent (local) background, making the (location of the) object salient. We propose to call this form of saliency *local saliency*.

Local saliency cannot account for the selection of a singleton among distracters (e.g., a red ball among blue balls on a green lawn). Both the singleton (i.e., the red ball) and the distracters (i.e., the blue balls) are locally salient with respect to the background, since they all form a discontinuity with the adjacent background, resulting in vigorous activity of neurons in the early stages of cortical processing. Hence, the selection of a singleton among distracters is the result of another (additional) process. We refer to this process as *global saliency*, because the singleton and the distracters may be distributed over a large region in the visual field.

Neurophysiological studies found that already in area V1 many neurons respond more strongly to pop-out center-surround stimuli, in which a single item in the classical receptive field (CRF) is surrounded by items that differ in a feature, than to homogeneous center-surround stimuli, in which the item centered on the CRF is identical to the items in the surround (e.g., Knierim & Van Essen, 1992). This result, which is also found in anesthetized animals (Kastner et al., 1999; Nothdurft et al., 1999), has been interpreted as evidence that pop-out results from selection at the earliest stages of cortical processing (in the ventral pathway), (largely) independent of top-down processing of visual information (e.g., Kastner et al., 1999; Knierim & Van Essen, 1992; Li, 2002).

Therefore, most models that incorporate mechanisms of global saliency assume that the selection of a singleton among distracters results from bottom-up and horizontal processing (Cave, 1999; Itti & Koch, 2000; Li, 2002; Wolfe, 1994) (see Chapter 6). More specific, these models implement competition between neurons that represent the same features.[9] This *within-feature competition* is organized either in a separate map for each feature, a feature map, (Itti & Koch, 2000; Wolfe, 1994) or without separate feature maps (Li, 2002). We will refer to this class of models as within-feature competition models. Of these models, Itti and Koch's (2000) model has probably been most influential to explain the selection of a singleton among distracters.

How is a singleton selected among distracters in Itti and Koch's (2000) model? Suppose that one red ball and a number of blue balls are present in the visual field of the model. The red ball activates neurons in the feature map "red" (i.e., the red/green feature map) at its corresponding location, and the blue balls activate

neurons in the feature map "blue" (i.e., the blue-yellow feature map) at their corresponding location. In each feature map, representations at different locations compete with each other (i.e., there is a WTA process). As a result, the representations of the blue balls in feature map blue diminish each other. In contrast, the representation of the red ball in the feature map red is unaffected by the WTA process, since only one red ball is represented. Next, the activation in each feature map is combined into an overall saliency map (via processing in conspicuity maps (see Chapter 6)). Within the saliency map, representations at different locations again compete with each other. As a consequence, the most highly activated location, the location of the red ball, wins the competition in the saliency map. Thus, within-feature competition models can account for the selection of any singleton among distracters, given that there is competition within a feature that is absent in the singleton, but present in the distracters.

Behavioral studies reported a number of findings that suggest that global saliency cannot solely be attributed to processing in low cortical areas (see Chapter 6). First, search can be efficient over a large range of spatial scales, far exceeding the small receptive fields of neurons in the primary visual cortex (Hochstein & Ahissar, 2002; Shipp, 2004) and other low cortical areas (Hochstein & Ahissar, 2002). Second, efficient search is reported not only for simple features (e.g., color, orientation) that are defined by luminance contrast, but also for simple features that are defined by other properties than luminance contrast (Bravo & Blake, 1990; Wolfe, 2003) and for high-level features, which include the result of quite sophisticated processing. Thus, efficient search may (Enns & Rensink, 1990; Kleffner & Ramachandran, 1992; Wolfe et al., 1994) and sometimes even has to (He & Nakayama, 1992; Rensink & Enns, 1998) be based on the results of later stages of cortical processing.

These behavioral findings pose problems for models that attribute within-feature competition exclusively to low cortical areas, such as area V1 (Li, 2002). Although other models relate within-feature competition to relatively high cortical areas (e.g., extrastriate areas) (Wolfe, 1994), low and high cortical areas (Cave, 1999), or do not relate within-feature competition to one or more cortical areas (e.g., Itti & Koch, 2000), these models can still be questioned in a number of ways.

First, the assumption that there is competition within each (simple and high-level) feature that can provide the basis for efficient search, entails an explosion of the number of horizontal, inhibitory connections (in feature maps) across different stages of cortical processing. Second, within-feature competition models suggest a

clear dichotomy between features (present in the distracters, but not in the target) that enable global saliency and those that do not, depending on whether there is or is not competition within a feature. In turn, this should result in a dichotomy between search slopes in visual search experiments, which has not been found experimentally (Wolfe, 1998). Third, within-feature competition models are based on automatic competition within each feature. This is a form of reducing information that could be needed in later stages of visual processing (cf., Wolfe & Horowitz, 2004). For example, similar features on different locations could belong to the same object. Competition among these features reduces the effectiveness of recognizing the object.

In addition, several neurophysiological studies call the assumption that global saliency results from bottom-up and horizontal processing into question. Recently, Hedgé and Felleman (2003) challenged the interpretation that many neurons in area V1 are already selective for pop-out center-surround stimuli (e.g., Knierim & Van Essen, 1992). Hedgé and Felleman (2003) presented a set of 36 different stimuli, consisting of a single bar of a preferred or non preferred color and orientation in the CRF and none or 58-109 bars in the surround, to monkeys that had to maintain fixation. The set of stimuli contained center-alone stimuli, and homogenous, pop-out and conjunction center-surround stimuli. Hedgé and Felleman's (2003) results indicated that, according to many different response measures, neurons in area V1 typically respond similarly to pop-out and conjunction center-surround stimuli. Hence, neurons in area V1 appear to be selective for feature discontinuities in general, and not specifically for the kind of feature discontinuities that lead to efficient search.

Other neurophysiological studies indicated that the first feedforward sweep of visual information through the brain does not discriminate a target from distracters in these areas, even when the target is distinguished by a unique feature from the distracters (Constantinidis & Steinmetz, 2001, 2005; McPeek & Keller, 2002; Thompson et al., 1997; Thompson et al., 1996). Instead, the neuronal discrimination of a singleton from distracters in the PP (Constantinidis & Steinmetz, 2001, 2005), the FEF (Thompson et al., 1997; Thompson et al., 1996), and the SC (McPeek & Keller, 2002) appears to occur in the following epoch, which involves both horizontal and feedback processing (J. H. Fecteau, personal communication, January 24, 2006) (see Chapter 6). Taken together, several neurophysiological studies (Constantinidis & Steinmetz, 2001, 2005; Hegdé & Felleman, 2003; McPeek & Keller, 2002; Thompson et al., 1997; Thompson et al.,

1996) indicate that global saliency may result just as well from a combination of bottom-up, horizontal and top-down processing, as from solely bottom-up and horizontal processing .

In this chapter we present a model of global saliency that is not based on within-feature competition across different stages of cortical processing and only bottom-up and horizontal processing. Instead, in the Global Saliency Model (GSM), there is competition within features only at the latest stage of cortical processing in the ventral pathway, and global saliency results from interaction between bottom-up, horizontal, and top-down processing in the ventral pathway and bottom-up and horizontal processing in the dorsal pathway. We propose that the mechanisms of global saliency and object-based visual attention partly overlap. Therefore, bottom-up, horizontal, and top-down processing in the ventral pathway is related to Van der Velde and De Kamps' (2001) model of object-based visual attention.

After introducing GSM, we present several simulations. First, the selection of a singleton among distracters is simulated. We then simulate the behavioral experiments in Chapter 5 that investigate whether global saliency is an all-or-none or a gradual phenomenon, and discuss how GSM can explain the finding of the behavioral experiments in Chapter 5 that top-down visual attention speeds up the response to a target, even when the location of the target is already (globally) salient. In addition, the effects of target-distracter and distracter-distracter similarity (Duncan & Humphreys, 1989) are simulated. Finally, we explore how illuminance may influence the saliency of objects in GSM.

## The model

### Architecture

Figure 1 illustrates the model for the selection of a singleton (e.g., cross) among distracters (e.g., triangles). The model consists of two pathways: ventral and dorsal. The ventral pathway processes object identification. When the identity of an object is selected in ventral area AIT, it generates feedback activity which interacts with stimulus activity in the ventral retinotopic areas (as in Van der Velde & De Kamps, 2001). The result is the selection of activity related to the object's location in these retinotopic areas. This selection (activation) is transmitted to the dorsal pathway. In the case of object-based attention (Van der Velde & De Kamps, 2001), the identity of the target is selected due to the memorization of the target. In the case of singleton selection discussed here, either the identity of the singleton (Figure 1A) or the identity of the distracter (Figure 1B) is selected. This

selection is due to a competition process in AIT (Chelazzi et al., 1993). In the case of object-based attention (Van der Velde & De Kamps, 2001) this process is influenced by the memorization of the target (Chelazzi et al., 1993). In the case of singleton selection, the competition is assumed to be random. However, the model in Figure 1 selects the location of the singleton, both when the singleton or when the distracter is selected in AIT.

In the dorsal pathway, the objects generate activation in an "input" retinotopic map. Activation is location related, not identity related. Each object is locally salient, so there is no difference in activation between the objects in the input map. The input map activates a "contrast" retinotopic map in a point-to-point manner (i.e., retinotopically). In the contrast map, WTA competition occurs between different spatial representations.

The ventral pathway activates a "ventral" retinotopic map, in a point-to-point manner. The ventral map inhibits the representations in the contrast map in a point-to-point manner. The input and ventral map interact in the contrast map, so that the activation ("location") that is not selected (enhanced) in the ventral map is selected. The ventral map also activates a "top-down" retinotopic map (point-to-point). In the top-down map, WTA competition occurs between different spatial representations. Finally, the contrast and top-down map activate a "saliency" retinotopic map (point-to-point). In the saliency map, WTA competition occurs between different spatial representations (as in Cave, 1999; Itti & Koch, 2000; Koch & Ullman, 1985; Wolfe, 1994).

Figure 1A shows what happens when the singleton (cross) is selected in AIT. Its location is selected in the ventral map, and thus in the top-down map. The contrast map represents the locations of the distracters (triangles), because the location of the cross is inhibited by the ventral map. Due to WTA, distracter representations are (more) reduced in the contrast map. As a result, the singleton's location is most strongly activated in the saliency map. The singleton wins the WTA competition, and its location is selected.

Figure 1B shows what happens when the distracter (triangle) is selected in AIT. The locations of the triangles are selected in the ventral map, and thus in the top-down map. But due to WTA, distracter representations are (more) reduced in the top-down map. The contrast map represents the location of the singleton (cross), because the distracter locations are inhibited by the ventral map. As a result, the location of the singleton (cross) is most strongly activated in the saliency map. The singleton wins the WTA competition, and its location is selected.

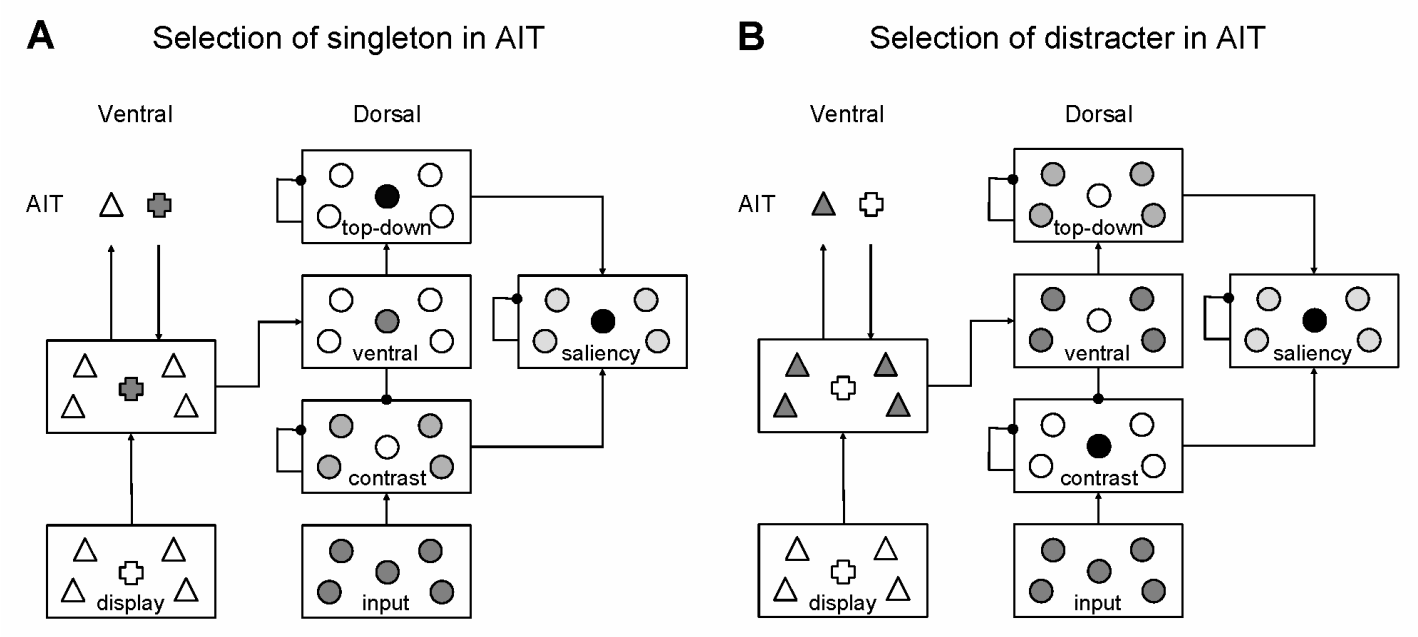


*Figure 1. The Global Saliency Model. (A) The identity of the singleton is selected in AIT (in the ventral pathway). The competition in the spatial maps of the model (in the dorsal pathway) results in the selection of the location of the singleton. (B) The identity of the distracter is selected in AIT (in the ventral pathway). The competition in the spatial maps of the model (in the dorsal pathway) again results in the selection of the location of the singleton.*

## Implementation

The model is implemented in terms of neuron populations in the spatial maps of the dorsal pathway. Each spatial map in the dorsal pathway consists of a retinotopic layer of $31 \times 31$ excitatory neuron populations. The activation that each object generates in the input map is simulated by injecting external input into the input map. Specifically, excitatory neuron populations in the input map that represent the location of an object receive excitatory external input. Excitatory neuron populations in the input map that do not represent the location of an object receive inhibitory external input.[10]

The ventral pathway is based on Van der Velde and De Kamps' (2001) model. In the simulations below, processing in the ventral pathway is not explicitly implemented. Instead, the selection achieved in the ventral pathway (Van der Velde & De Kamps, 2001) is simulated by injecting external input into the ventral map. More specific, excitatory neuron populations in the ventral map that represent the location of an object of which the identity is selected in the ventral pathway receive excitatory external input, while excitatory neuron populations in the ventral map that represent the location of an object of which the identity is not selected in the ventral pathway receive inhibitory external input. Excitatory neuron populations in the ventral map that do not represent the location of an object also receive inhibitory external input. We assume that the selection of an object identity in AIT is random, unless the identity of the target is cued (i.e., a cued-target). We therefore present the results for each object identity that may be selected in AIT.

The objects in our simulation are disks. The location of a disk is represented by 12 excitatory neuron populations in a spatial map. Excitatory neuron populations in the input and ventral map receive external input from the onset of a simulation (i.e., time = 0) until the activation of the excitatory neuron populations in the saliency map converges to a stable state (i.e., time = 50 ms).

WTA competition in the top-down, contrast and saliency map is implemented through an inhibitory neuron population (Deco et al., 2002; Usher & Niebur, 1996). The inhibitory neuron population receives input from all excitatory neuron populations in a spatial map via excitatory connections, and inhibits all excitatory neuron populations in the spatial map via inhibitory connections. Thus, the inhibitory neuron population inhibits each excitatory neuron population in the spatial map in proportion to the sum of activation over all excitatory neuron populations in the spatial map. As a result, excitatory neuron populations that

receive the highest net input win the competition from excitatory neuron populations that receive lower net input in the spatial map. At the same time, the overall level of activation in the spatial map is regulated.

The activation (average neuron activity) of an excitatory neuron population is given by:

$$\tau_E \frac{dI(t)}{dt} = -I(t) + \sum_m W_m F(I_m(t)) + I_{extern} + I_{bg}.$$

The current $I$ determines the average firing rate in the excitatory neuron population. The input from other populations is received through $W_m$, with $W_m > 0$ for excitatory input and $W_m < 0$ for inhibitory input. $I_{extern}$ is external input (only for the input map and ventral map), and $I_{bg}$ is background noise ($I_{bg}$ = 0.025). The average firing rate is given by:

$$F(I) = \frac{k}{\left(1 + e^{-\beta(I-\theta)}\right)},$$

with $k = 80$ Hz, $\theta = 4.0$ and $\beta = 1.0$.

All simulations are performed with the same set of parameters, unless the value of a parameter was systematically varied for the purpose of a simulation. The parameters and their default values are described in detail in the Appendix at the end of this chapter. We will specify the value(s) of a parameter in the simulations below, when it deviates from the default value.

## Simulations

### Selecting a singleton among distracters

Figure 2 shows the response of the model when a singleton is presented among four distracters, both when the singleton (Figure 1A) or the distracter (Figure 1B) is selected in AIT. In both cases, the location of the singleton is selected in the saliency map (s-SM). Distracter activity in the saliency map is low in both cases (d-SM). With singleton selection in AIT (Figure 2A), singleton activity in the top-down map (s-TDM) is stronger than distracter activity in contrast map (d-CM). With distracter selection in AIT (Figure 2B), singleton activity in the contrast map (s-CM) is stronger than distracter activity in top-down map (d-TDM). These differences in activation in favor of the singleton determine the selection of the singleton in the saliency map, in the manner as illustrated in Figure 1.

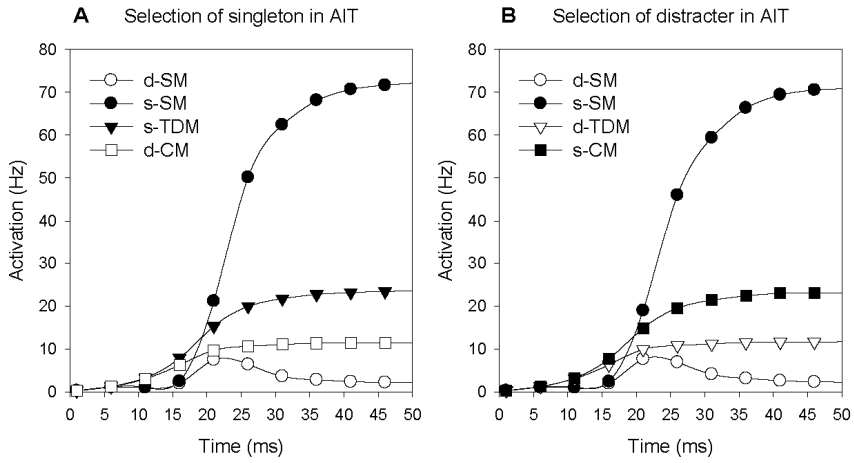


*Figure 2. (A) The activation in the model over time when the singleton is selected in AIT. (B) The activation in the model over time when the distracter is selected in AIT. (s-SM = singleton in saliency map, s-TDM = singleton in top-down map, s-CM = singleton in contrast map, d-SM = distracter in saliency map, d-TDM = distracter in top-down map, d-CM = distracter in contrast map.)*

Figure 2 also shows that the distracter activity in the saliency map is lower than the distracter activity in the contrast map (Figure 2A), or lower than the distracter activity in the top-down map (Figure 2B). This is due to the fact that in the saliency map the WTA process is dominated by the singleton, which is not the case for the contrast map or the top-down map. (The role of the contrast map versus the top-down map in this case results from selecting either the identity of the singleton or the identity of the distracter in AIT, as illustrated in Figure 1.)

The difference in distracter activity between the salience map and the contrast map or the top-down map suggests that distracter activity in the saliency map would be higher when the singleton is not present. This suggestion is corroborated by a simulation of the model: without the presence of a singleton, distracter activity in the saliency map is similar to distracter activity in the contrast map or the top-down map. This is also true when the singleton is replaced by a distracter, so that there are more distracters in that case (which would result in more competition). The result of this simulation is in line with an observation of distracter activity in posterior parietal area 7a (Constantinidis & Steinmetz, 2005). In this experiment, distracter activity in this area was higher when the singleton was absent, compared to distracter activity when the singleton replaced one of the

distracters. The activity of the singleton, however, was the highest in all cases, as in Figure 2.

The key characteristic of search for a singleton among distracters is that the time it takes to correctly determine the presence of the singleton is unaffected by the number of distracters (Wolfe & Horowitz, 2004). Figure 3 shows a simulation of the model in which a singleton is presented among 1 to 7 identical distracters. The activation of the singleton and the distracter in the saliency map (s-SM and d-SM) is shown at 50 ms after the onset of the simulation, when the activity of the singleton and the distracter has converged to a stable state, the activation at convergence. The activation at convergence of an object that is presented alone is plotted as a reference (single object-SM). The results of the simulation are analogous, whether the singleton (Figure 3A) or the distracter (Figure 3B) is selected in AIT. As expected, the singleton activity and the distracter activity in the saliency map are nearly equally strong when the singleton is presented with one distracter. This is logical because either one of the two presented objects may be considered as the singleton or the distracter. However, the singleton activity is much stronger than the distracter activity in the saliency map when the singleton is presented with two or more distracters. Hence, the location of the singleton is selected in the saliency map when the distracters outnumber the singleton.

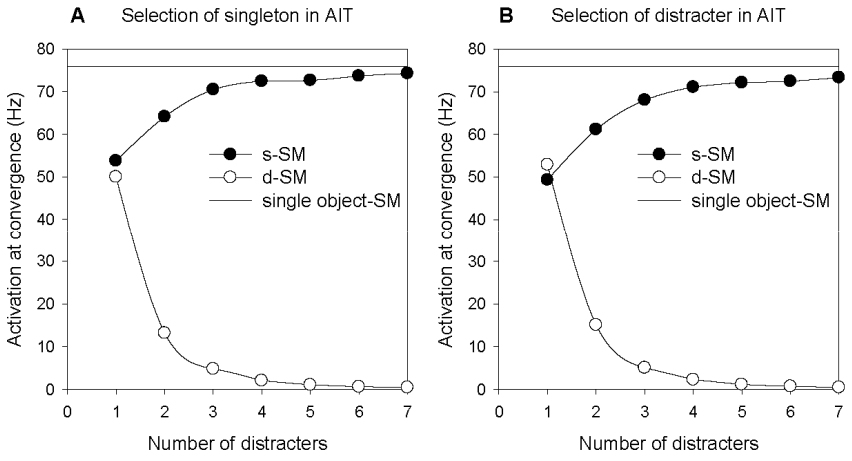


*Figure 3. The activation at convergence in the saliency map of the model as a function of the number of distracters, when the singleton (A) or the distracter (B) is selected in AIT. The activation at convergence of an object that is presented alone (single object) is plotted as a reference. (s-SM = singleton in saliency map, d-SM = distracter in saliency map, single object-SM = single object in saliency map.)*

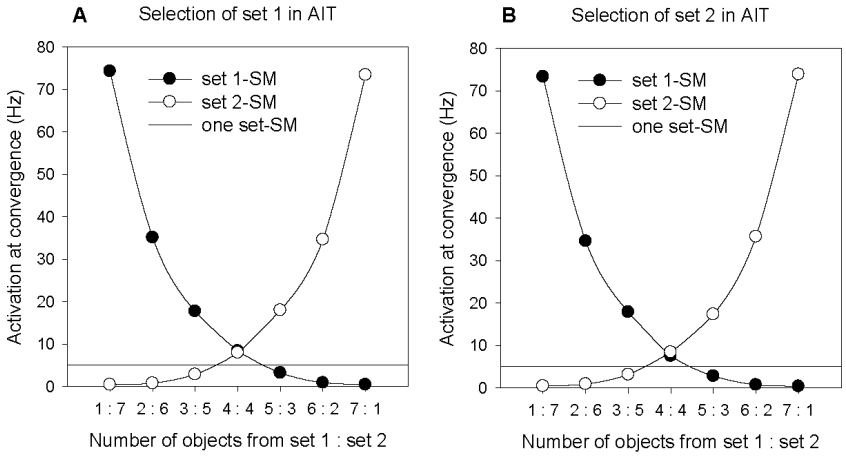


*Figure 4. The activation at convergence in the saliency map of the model as a function of different proportions of set 1 and set 2, when set 1 (A) or set 2 (B) is selected in AIT. The activation at convergence of an object when eight objects from one set are presented is plotted as a reference (one set). (set 1-SM = set 1 in saliency map, set 2-SM = set 2 in saliency map, one set-SM = one set in saliency map).*

## Gradual global saliency

As described in Chapter 5, we investigated whether elements from a minority colored set with more than one element are salient in a similar manner as color singletons. In our experiments, participants had to search for a target that was superimposed on one of fifteen colored disks. Each search display was equally likely to contain 0, 1, 3, 5, 7, 8, 10, 12, 14, or 15 disks of one color with 15, 14, 12, 10, 8, 7, 5, 3, 1, or 0 disks of the other color. The target was equally likely to be placed in one of 1, 3, 5, 7, 8, 10, 12, 14, or 15 identically colored disks. We found that responses are fastest for targets on color singletons, but also that responses for targets on elements from a minority colored set with more than one element are faster than responses for targets on elements from a majority colored set. This result reflects that elements from a minority colored set with more than one element are searched earlier or faster than elements from a majority colored set, and are thus prioritized in search in a similar manner as color singletons. We referred to this as gradual saliency.

We tested whether our model also produces gradual saliency. Therefore, we presented 8 objects to the model, which were divided in two sets: set 1 and set 2. The proportion of both sets was varied. The model was presented 1, 2, 3, 4, 5, 6, or 7 objects from set 1 and 7, 6, 5, 4, 3, 2, or 1 objects from set 2. In case of set 1

selection in AIT, the external input injected into the excitatory neuron populations in the ventral map that represent the location of objects from set 1 was 2, and the external input injected into the excitatory neuron populations in the ventral map that represent the location of objects from set 2 was -2. Likewise, in case of set 2 selection in AIT, the external input injected into the excitatory neuron populations in the ventral map that represent the location of objects from set 2 was 2, and the external input injected into the excitatory neuron populations in the ventral map that represent the location of objects from set 1 was -2.

Figure 4 shows the activation at convergence of objects from set 1 and 2 in the saliency map (set 1-SM and set 2-SM) as a function of different proportions of both sets, both when set 1 (Figure 4A) or set 2 (Figure 4B) is selected in AIT. The activation at convergence of an object when eight objects from one set are presented is plotted as a reference (one set-SM). Both when set 1 or set 2 is selected, the activity of objects from a set in the saliency map is highest when only one object from that set and seven objects from the other set are presented. This situation is identical to the presentation of a singleton among seven distracters (see Figure 3). Interestingly, the activity of objects from a set in the saliency map gradually decreases as more and more objects from that set are presented. Naturally, the activity of objects from set 1 and 2 in the saliency map is equally strong when the same number of objects is presented from each set (i.e., 4 objects from set 1 and 2). Thus, our model most strongly selects the location of a singleton, but the model also to some extent selects the locations of objects from a minority set in the saliency map.

In order to relate the response time of our experiments (see Chapter 5) in a qualitative manner to the activation in the saliency map of our model, we normalized both measures. Thereto, we defined the condition, in which the target is located on a singleton, as the baseline condition (Experiments, 1:14; Simulation, 1:7), and the condition, in which a singleton is present, but the target is located on another object, as the reference condition (Experiments, 14:1; Simulation, 7:1). Next, the increase in response time or activation with respect to the baseline condition was computed for each condition (Experiments, 1:14, 3:12, 5:10, 7:8, 8:7, 10:5, 12:3, 14:1, 15:0; Simulation, 1:7, 2:6, 3:5, 4:4, 5:3, 6:2, 7:1, 8:0). Then, we normalized the increase in response time or activation with respect to the baseline condition for each condition by taking the increase of the reference condition as norm. Thus, the normalized increase is given by:

(Condition - Baseline condition) / (Reference condition - Baseline condition)

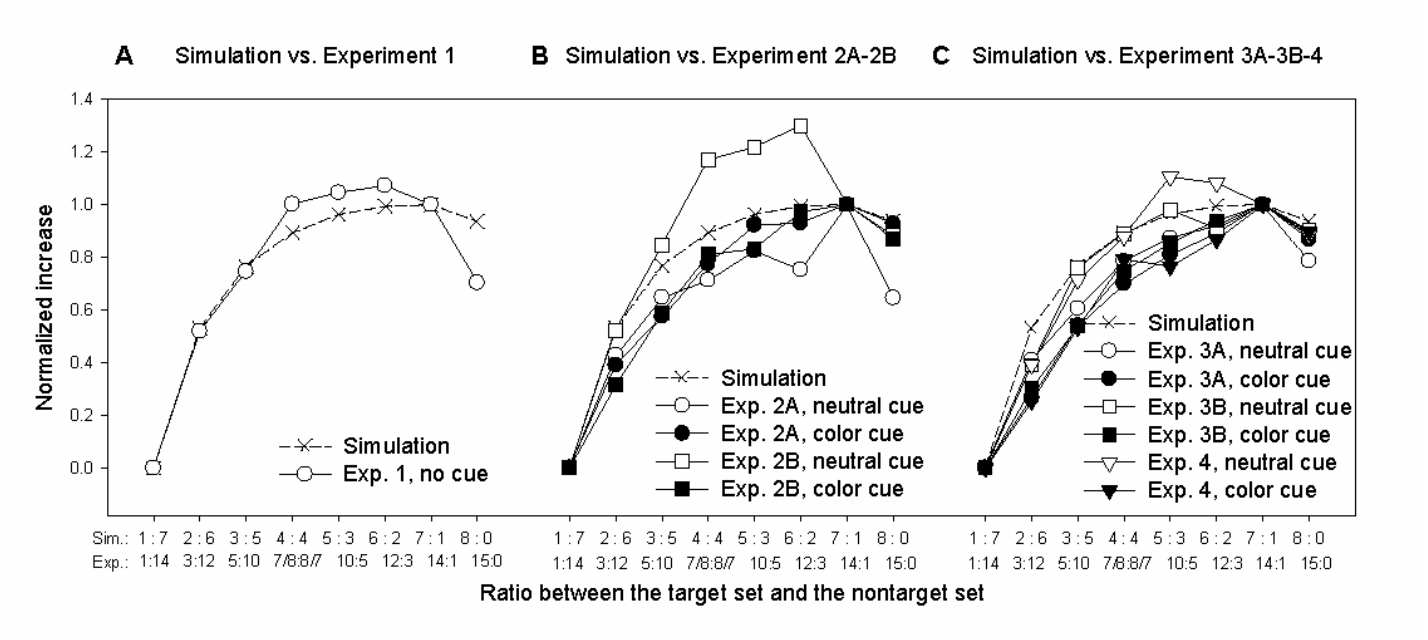


*Figure 5. The normalized increase as a function of the ratio between the set on which the target is located and the set on which the target is not located (see the text for explanation). (A) Data for the simulation and Experiment 1. (B) Data for the simulation and Experiments 2A and 2B. (C) Data for the simulation and Experiments 3A, 3B, and 4.*

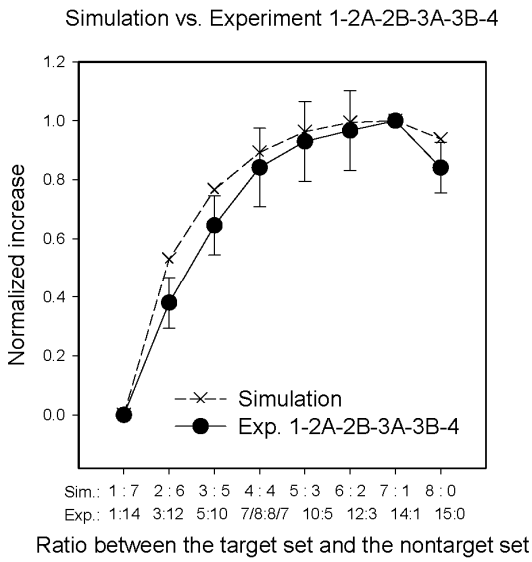


*Figure 6. The normalized increase as a function of the ratio between the set on which the target is located and the set on which the target is not located, for the simulation and averaged over the eleven conditions of Experiments 1, 2A, 2B, 3A, 3B, and 4 (see the text for explanation). The error bars indicate the root mean squared error of the normalized increase that is averaged over the eleven conditions of Experiments 1, 2A, 2B, 3A, 3B, and 4.*

In other words, the normalized increase indicates the increase in response time or decrease in activation with respect to the baseline condition, while taking the increase or decrease of the reference condition as the unit of measurement. Consequently, the normalized increase is 0 in the baseline condition and 1 in the reference condition.

The conditions in the simulation were mapped in a qualitative manner onto the conditions in the experiments. Besides mapping the conditions in the simulation in which a singleton is present (1:14 and 14:1) onto the corresponding conditions in the experiments (1:7 and 7:1), the condition in the simulation in which only objects from one set are present (8:0) was mapped onto the corresponding condition in the experiments (15:0). Furthermore, we merged the conditions 7:8 and 8:7 of the experiments, and mapped it onto the 4:4 condition in the simulation. Finally, the intermediate conditions in the simulation, in which the ratio between the set on which the target is located and the set on which the target

is not located was 2:6, 3:5, 5:3, or 6:2, were mapped onto the conditions in the experiments, in which the ratio was respectively 3:12, 5:10, 10:5, or 12:3.

Figure 5 shows the normalized increase as a function of the ratio between the set on which the target is located and the set on which the target is not located, for the simulation, and Experiment 1 (Figure 5A), Experiments 2A and 2B (Figure 5B), and Experiments 3A, 3B, and 4 (Figure 5C). The normalized values clearly indicate that the decrease in activation in the saliency map of our model is qualitatively similar to the increase in response time in the experiments as the ratio between the set on which the target is located and the set on which the target is not located increases. This fit is comparable for the no cue condition (Experiment 1), the neutral cue conditions (Experiments 2A, 2B, 3A, 3B, and 4), and the color cue conditions (Experiments 2A, 2B, 3A, 3B, and 4).

Figure 6 shows the normalized increase as a function of the ratio between the set on which the target is located and the set on which the target is not located, for the simulation, and averaged over Experiments 1, 2A, 2B, 3A, 3B, and 4. The error bars indicate the root mean squared error (RMSE) of the normalized increase that is averaged over the eleven conditions of Experiments 1, 2A, 2B, 3A, 3B, and 4. Again, the normalized values clearly indicate that the decrease in activation in the saliency map of our model is qualitatively similar to the increase in response time in the experiments as the ratio between the set on which the target is located and the set on which the target is not located increases. As can be seen in Figure 6, the model somewhat underestimates the saliency when the ratio between the set on which the target is located and the set on which the target is not located in the simulation is 2:6, 3:5, and 8:0.

Although the data do not allow a quantitative comparison between the simulation and the experiments, since the conditions of the simulations are only qualitatively mapped onto the conditions of the experiments, it appears that our model is consistent with the finding of gradual saliency in our experiments (Chapter 5). In fact, the model's decrease in saliency as more and more objects share a characteristic is qualitatively similar to the increase in response time that we observed in the experiments.

We also investigated the interaction of gradual saliency with top-down visual attention (see Chapter 5). In our experiments, top-down visual attention was either set by a color cue at the beginning of each trial, or was absent due to a neutral cue. We found that top-down visual attention speeds up the search for a target, while the location of the target is already salient. Top-down visual

attention even made the search for a target faster, when it appeared on a color singleton. This finding is predicted by the architecture of our model. In our model, top-down visual attention for a color biases the competition process in AIT. By biasing the competition process in AIT, top-down visual attention for a color speeds up the selection of an object identity (i.e., the attended color) in AIT. As a result, the spatial maps of our model are able to compute global saliency earlier in time.

### Target-distracter (T-D) similarity and distracter-distracter (D-D) similarity

Several visual search studies varied the difference between the singleton and the distracters along a feature dimension (e.g., varying the distracter orientation, while fixing the target orientation) (for overviews, see Duncan & Humphreys, 1989; Wolfe & Horowitz, 2004). In line with numerous visual search studies, the findings indicated that as long as the singleton and distracters differ largely along a feature dimension (e.g., color, orientation or size), the number of distracters does not affect the time it takes to correctly determine the presence of the singleton. However, the findings also indicated that the time it takes to correctly determine the presence of the singleton increases with an increasing number of distracters as the singleton differs less and less from the distracters along a feature dimension.

Furthermore, a visual search study by Duncan and Humphreys (1989) showed that the time it takes to correctly determine the presence of a cued-target was largely unaffected by the number of distracters when the distracters were homogeneous, but increased with the number of distracters when the distracters were heterogeneous.

Based on these findings and other findings, Duncan and Humphreys (1989) proposed that the similarity between search items determines the search efficiency, i.e. the degree to which the time it takes to correctly determine the presence of a target is unaffected by the number of distracters. Specifically, Duncan and Humphreys (1989) hypothesized that search efficiency decreases with increasing target-distracter (T-D) similarity and with decreasing distracter-distracter (D-D) similarity.

We tested how our model responds to varying T-D and D-D similarity. Thereto, we increased T-D similarity in one simulation, and decreased D-D similarity in another simulation of the model. In both simulations, a singleton and four distracters were presented to the model. We used a fixed number of distracters, as

we are interested in demonstrating the mechanisms of the model that determine global saliency in this chapter.

*T-D similarity*

In our model, the T-D similarity affects the selection achieved in the ventral pathway. As the singleton and distracters become more and more similar, selection of the singleton (distracter) in AIT results not only in the selection of activity related to the singleton's location (distracters' location), but also to some extent in the selection of activity related to the distracters' location (singleton's location) in the ventral retinotopic areas.

We simulated the effect of increasing T-D similarity for the selection achieved in the ventral pathway by making the external input of excitatory neuron populations in the ventral map that represent the location of an object of which the identity is selected in AIT (the selected object current in the ventral map) increasingly similar to the external input of excitatory neuron populations in the ventral map that represent the location of an object of which the identity is not selected in AIT (the unselected object current in the ventral map). The selected object current in the ventral map was fixed at 2. However, the unselected object current in the ventral map was varied from -2 to 2. Accordingly, the T-D similarity is lowest when the unselected object current in the ventral map is -2, and the T-D similarity is highest when the unselected object current in the ventral map is 2 (i.e., in that case, the singleton and the distracters are identical).

Figure 7 shows the activation at convergence of the singleton and the distracter in the saliency map as a function of the T-D similarity, when the singleton is selected in AIT. The results are analogous when the distracter is selected in AIT. The singleton activity (s-SM) is much higher than the distracter activity in the saliency map (d-SM) as long as the T-D similarity is low (e.g., the unselected object current in the ventral map is $< 1.8$). Only when the T-D similarity becomes very high, the distracter activity starts approximating the singleton activity in the saliency map. As a result, the location of the singleton can no longer (uniquely) be selected in the saliency map. In that case, the time it takes to correctly determine the presence of the singleton will increase with an increasing number of distracters. Our model is thus consistent with Duncan and Humphreys' (1989) theory and visual search studies that varied the difference between the singleton and the distracter along a feature dimension (for overviews, see Duncan & Humphreys, 1989; Wolfe & Horowitz, 2004).

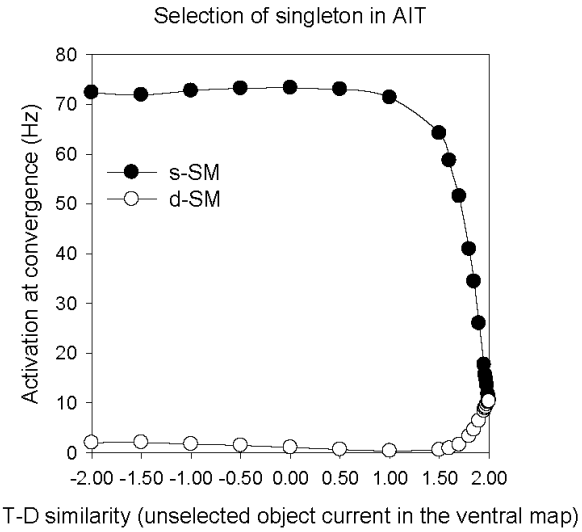


*Figure 7. The activation at convergence in the saliency map of the model as a function of the T-D similarity, when the singleton is selected in AIT. T-D similarity increases with an increasing value of the unselected object current in the ventral map (see the text for explanation). (s-SM = singleton in saliency map, d-SM = distracter in saliency map.)*

### *D-D similarity*

The D-D similarity also affects the selection achieved in the ventral pathway in our model. Suppose that a singleton is presented among two distracters types, which are highly dissimilar: distracter type 1 and distracter type 2. Then, the selection of distracter type 1 in AIT results only in the selection of activity related to distracter type 1's location, but not in the selection of activity related to distracter type 2 and the singleton's location in the ventral retinotopic areas. Likewise, the selection of distracter type 2 in AIT results only in the selection of activity related to distracter type 2's location, but not in the selection of activity related to distracter type 1 and the singleton's location in the ventral retinotopic areas. The selection of the singleton in AIT results only in the selection of activity related to the singleton's location, but not in the selection of activity related to distracter type 1 and 2's location in the ventral retinotopic areas (given low T-D similarity).

We simulated the effects of decreasing D-D similarity for the selection achieved in the ventral pathway as follows. In the case of singleton selection in AIT, the external input injected into the excitatory neuron populations in the ventral map

that represent the singleton's location was 2, and the external input injected into the excitatory neuron populations in the ventral map that represent the location of either distracter type was -2. In the case of distracter type 1 selection in AIT, the external input injected into the excitatory neuron populations in the ventral map that represent the location of distracter type 1 was 2, the external input injected into the excitatory neuron populations in the ventral map that represent the singleton's location was -2, and the external input injected into the excitatory neuron populations in the ventral map that represent the location of distracter type 2 was varied from -2 to 2 (variable *a*). Likewise, in the case of distracter type 2 selection in AIT, the external input injected into the excitatory neuron populations in the ventral map that represent the location of distracter type 2 was 2, the external input injected into the excitatory neuron populations in the ventral map that represent the singleton's location was -2, and the external input injected into the excitatory neuron populations in the ventral map that represent the location of distracter type 1 was varied from -2 to 2 (variable *a*). Accordingly, the D-D similarity is highest when variable *a* has value 2 (i.e., in that case, both distracter types are identical), and the D-D similarity is lowest when variable *a* has value -2.

Figure 8 shows the activation at convergence of the singleton, distracter type 1 and distracter type 2 in the saliency map as a function of the D-D similarity. When the singleton is selected in AIT (Figure 8A), the location of the singleton is selected in the saliency map (singleton-SM), independently of the D-D similarity. However, when distracter type 1 is selected in AIT (Figure 8B), the location of the singleton is only selected in the saliency map as long as the D-D similarity is high enough ($a > 0.25$). As the D-D similarity decreases ($a < 0.25$), the locations of distracter type 1 are instead selected in the saliency map (distracter type 1-SM). Similarly, when distracter type 2 is selected in AIT (Figure 8C), the location of the singleton is only selected in the saliency map as long as the D-D similarity is high enough ($a > 0.25$). As the D-D similarity decreases ($a < 0.25$), the locations of distracter type 2 are instead selected in the saliency map (distracter type 2-SM). Given our assumption that the selection of an object identity in AIT is random, the results suggest that in addition to the singleton, the distracters are frequently selected when the D-D similarity is low. In that case, the time it takes to correctly determine the presence of the singleton will increase with an increasing number of distracters. Our model is thus consistent with Duncan and Humphreys' (1989) theory and their visual search experiments, which investigated the effect of distracter homogeneity and heterogeneity.

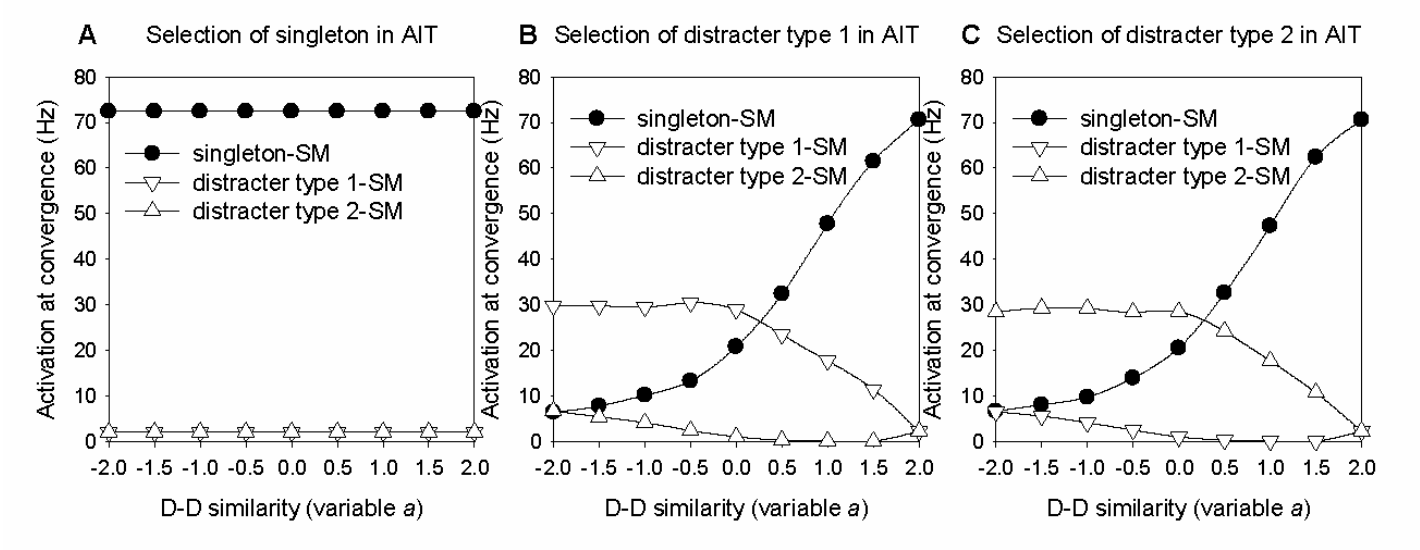


*Figure 8. The activation at convergence in the saliency map of the model as a function of the D-D similarity, when the singleton (A), distracter type 1 (B) or distracter type 2 (C) is selected in AIT. D-D similarity increases with an increasing value of variable a (see the text for explanation). (singleton-SM = singleton in saliency map, distracter type 1-SM = distracter type 1 in saliency map, distracter type 2-SM = distracter type 2 in saliency map.)*

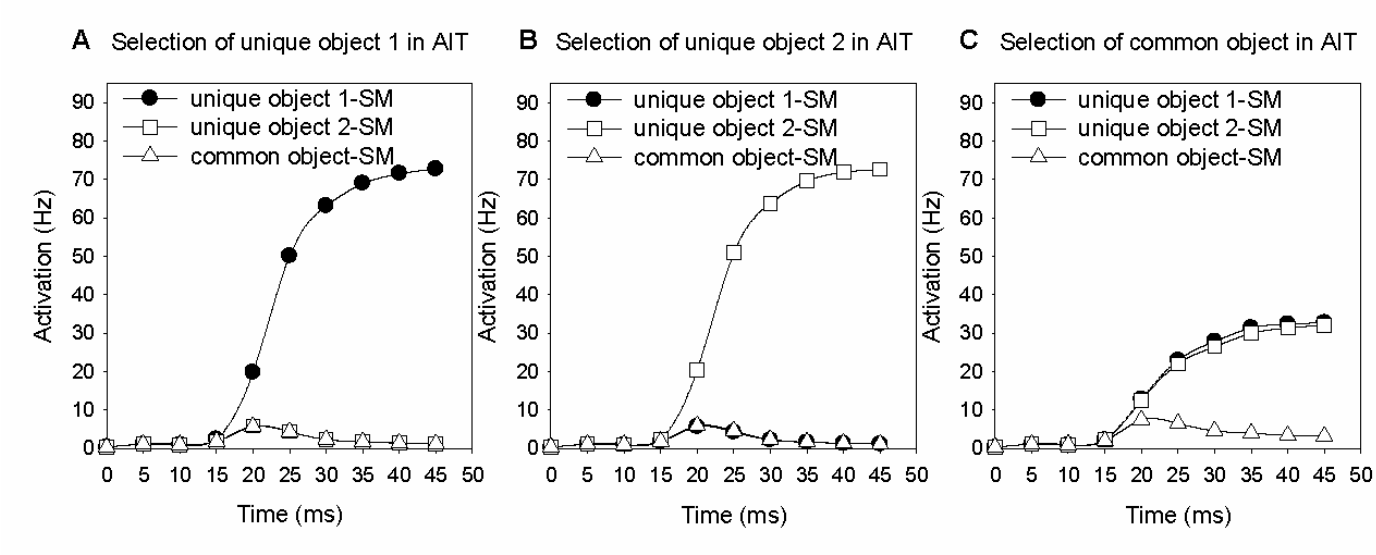


*Figure 9. The activation in the saliency map of the model over time when six objects are presented to the model, of which two objects are unique compared to the other objects. (A) Unique object 1 is selected in AIT. (B) Unique object 2 is selected in AIT. (C) The common object is selected in AIT. (unique object 1-SM = unique object 1 in saliency map, unique object 2-SM = unique object 2 in saliency map, common object-SM = common object in saliency map.)*

### Two unique objects among other objects

Figure 9 shows the response of the model when six objects are presented to the model, of which two objects are unique compared to the other objects (e.g., a red and green disk among four gray disks). Suppose that the identity of either object (e.g., the color red, green or gray) can independently be selected in the ventral stream. The activity of unique object 1, unique object 2 and the common objects in the saliency map (unique object 1-SM, unique object 2-SM, and common object-SM) then depends on whether unique object 1 (Figure 9A), unique object 2 (Figure 9B) or the common objects are selected in AIT (Figure 9C). When unique object 1 is selected in AIT, the location of unique object 1 is selected in the saliency map. Likewise, when unique object 2 is selected in AIT, the location of unique object 2 is selected in the saliency map. When the common objects are selected in AIT, however, the locations of unique object 1 and 2 are selected in the saliency map.

Yet, the location of a unique object is selected less strongly in the saliency map when the common objects are selected in AIT than when the unique object itself is selected in AIT. This is due to the fact that the selection of the common objects in AIT results in the selection of both unique objects in a spatial map with WTA competition (i.e., the contrast map), while the selection of unique object 1 (unique object 2) in AIT results in the selection of only unique object 1 (unique object 2) in a spatial map with WTA competition (i.e., top-down map). Given our assumption that the selection of an object identity in AIT is random, the results indicate that the locations of unique object 1 and 2 are as frequently selected in the saliency map.

As has already been mentioned, we assume that object-based visual attention biases the competition process in AIT due to the memorization of the target, so that the identity of the attended object is (more frequently) selected. Thus, when unique object 1 (unique object 2) is attended, unique object 1 (unique object 2) is selected in AIT. As a consequence, the location of unique object 1 (unique object 2) is selected in the saliency map (Figures 9A and 9B). When the common objects are attended, the common objects are selected in AIT. Consequently, the locations of unique object 1 and 2 are selected in the saliency map (Figure 9C).

In conclusion, our model predicts that when no objects are cued (i.e., in the absence of object-based visual attention) both unique objects are as frequently selected in the saliency map. Conversely, when one of the unique objects is cued, this object can be selected. This prediction is consistent with Bacon and Egeth's (1994) proposal that spatial attention is automatically shifted to any singleton

when participants are searching for a singleton (i.e., singleton detection mode), but not when participants are able to direct top-down visual attention exclusively to the relevant feature of a target (i.e., feature search mode).

### Saliency by illuminance

Highly illuminant stimuli (or stimuli with an abrupt onset) activate the photoreceptors in the retina more strongly than lowly illuminant stimuli (or stimuli with a gradual onset). This neural activity is projected to both the dorsal and the ventral pathway (Coren et al., 2003). In the dorsal pathway, we simulate a highly illuminant object by increasing the external input to excitatory neuron populations in the input map that represent the location of a highly illuminant object from 2 tot 2.15. In the ventral pathway, we simulate a highly illuminant object by increasing the external input to excitatory neuron populations in the ventral map that represent the location of a highly illuminant object from 2 tot 2.15, provided that the highly illuminant object is selected in AIT. [11]

Figure 10A shows the response of the model when five objects are presented, which are identical except that one of the objects is highly illuminant (e.g., four gray disks and one highly illuminant gray disk). All the objects are selected in AIT, as they have the same object identity (e.g., they are all gray). The location of the highly illuminant object is selected in the saliency map (highly illuminant object-SM). Activity of the other objects in the saliency map is low (other object-SM). The activation in the contrast map (highly illuminant object-CM and other object-CM) is very low, since the input and ventral map receive the same external input. This is due to the fact that all the objects are selected in AIT. The activity of the highly illuminant object in the top-down map (highly illuminant object-TDM) is higher than the activity of the other objects in the top-down map (other object-TDM). This difference in activation in favor of the highly illuminant object determines the selection of the highly illuminant object in the saliency map.

Figure 10B shows the activation at convergence of the highly illuminant object and the other objects in the saliency map (highly illuminant object-SM and other object-SM) as function of the difference in illuminance. The highly illuminant object is increasingly strongly selected in the saliency map as its difference in illuminance with the other objects increases.

In this simulation (and in the other simulations), the input and the ventral map are activated at the same time. That is, both the excitatory neuron populations in the input map and the excitatory neuron populations in the ventral map receive

external input from the onset of a simulation (i.e., time = 0). Nonetheless, it is reasonable to assume that the activation that each object generates in the input map evolves earlier in time than the activation in the ventral map that results from the selection achieved in the ventral pathway. If this is indeed the case, the activation in the contrast map, and consequently the activation in the saliency map, initially would be influenced primarily by the activation in the input map. As the activity of the highly illuminant object is higher than the activity of the other objects in the input map, the location of the highly illuminant object would still be selected in the saliency map.

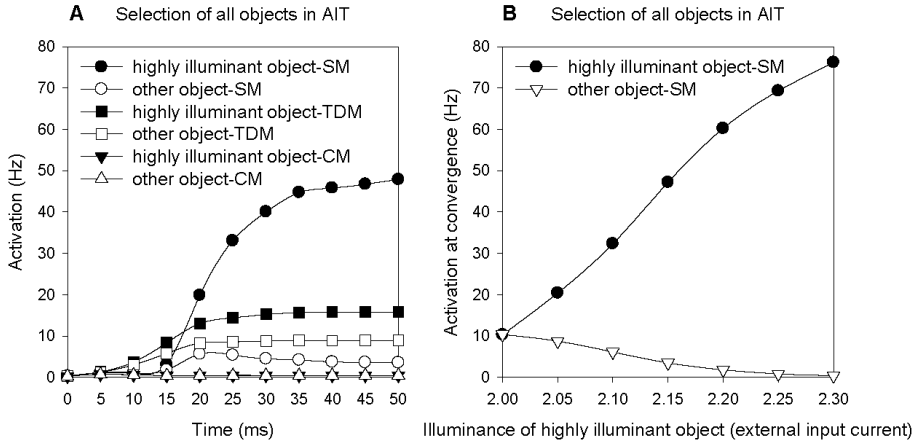


*Figure 10. The activation in the model when five objects are presented, which are identical except that one of the objects is highly illuminant. All objects are selected in AIT. (A) The activation in the model over time. (B) The activation at convergence in the saliency map of the model as a function of the illuminance of the highly illuminant object. (highly illuminant object-SM = highly illuminant object in saliency map, highly illuminant object-TDM = highly illuminant object in top-down map, highly illuminant object-CM = highly illuminant object in contrast map, other object-SM = other object in saliency map, other object-TDM = other object in top-down map, other object-CM = other object in contrast map.)*

After the selection in the ventral pathway has taken place, the activation in the ventral map neutralizes the initial domination of the highly illuminant object in the contrast map. As noted above, the reason is that the activation in the input and in the ventral map then become equivalent, and cancel each other out in the contrast map. At the same time, the activity of the highly illuminant object in the top-down map becomes higher than the activity of the other objects in the top-

down map. This difference in activation in favor of the highly illuminant object subsequently determines the selection of the highly illuminant object in the saliency map (see Figure 10A).

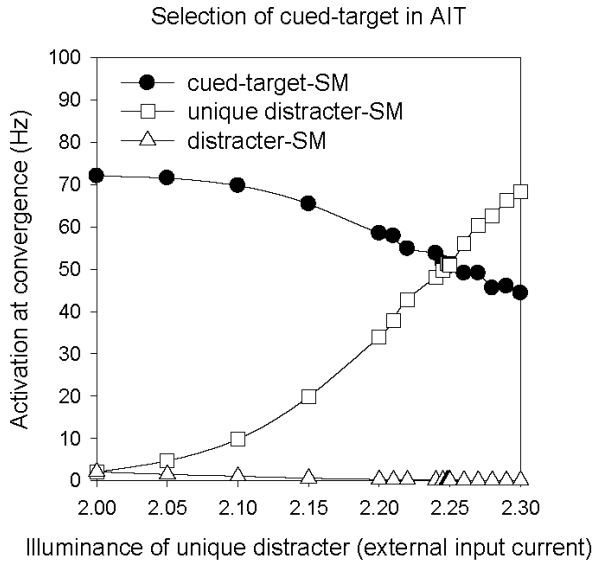


*Figure 11. A unique object that is attended (cued-target) is presented among one distracter that is unique and highly illuminant (unique distracter) and four distracters. Object-based visual attention results in the selection of the cued-target in AIT. The graph shows the activation at convergence in the saliency map of the model as a function of the illuminance of the unique distracter. The illuminance of the unique distracter increases with an increasing value of the corresponding external input current (see the text for explanation) (cued-target-SM = cued-target in saliency map, unique distracter-SM = unique distracter in saliency map, distracter-SM = distracter in saliency map.)*

### A unique, attended object among distracters, of which one is unique and highly illuminant

We are interested in the response of our model when a unique object, which is attended (cued-target), is presented among one distracter that is unique and highly illuminant (unique distracter) and four other distracters (e.g., a red target among a highly illuminant green distracter and four gray distracters). Figure 11 shows the activation at convergence of the cued-target, the unique distracter, and the other distracters in the saliency map (cued-target-SM, unique distracter-SM, and distracter-SM) as a function of the illuminance of the unique distracter.

Object-based visual attention results in the selection of the cued-target in AIT. The cued-target activity is higher than the unique distracter activity in the saliency map as long as the difference in illuminance between the cued-target and the unique distracter is low (e.g., external input current of the unique distracter is < 2.25). Only when the illuminance of the unique distracter becomes much higher than the illuminance of the cued-target (e.g., external input current of the unique distracter is >> 2.25), the unique distracter activity surpasses the cued-target activity in the saliency map. As a result, the location of the cued-target can no longer (uniquely) be selected in the saliency map. Hence, our model predicts that even when participants are searching for a unique, attended object, spatial attention may automatically be shifted to a unique distracter, given a high enough illuminance of the unique distracter (or a distracter with an abrupt onset).

## Conclusion

The Global Saliency Model can explain several findings in visual search. In simulations, we showed that a singleton is selected in GSM, as long as the distracters outnumber the singleton. That is, the location representation of the singleton in the dorsal pathway wins the competition in the saliency map. This location representation can subsequently influence the ventral pathway to select the identity (e.g., shape, color) of the singleton as well, in particular when the distracter was initially selected in AIT (Figure 1B). In this way, the interaction between the dorsal and ventral pathway in the model binds location information with identity information (Van der Velde & De Kamps, 2001, 2006).

Other simulations demonstrated that GSM is consistent with the findings of the behavioral experiments in Chapter 5. Global saliency appears to be gradual in GSM. The model's decrease in global saliency as more and more objects share a characteristic is even qualitatively similar to the increase in response time that we observed in the experiments. GSM can also account for the finding that top-down visual attention speeds up the search for a target, when the target location is already globally salient. In the architecture of GSM, top-down visual attention (i.e., object-based visual attention) speeds up the competition process in AIT, by biasing the competition process toward the attended object identity. Therefore, the spatial maps of the model are able to compute global saliency earlier in time.

Furthermore, GSM is able to simulate the effects of T-D and D-D similarity (Duncan & Humphreys, 1989). The T-D and D-D similarity affect the selection in the ventral pathway in GSM. When the T-D similarity is high, or the D-D

similarity is low, distracters are frequently selected in the saliency map of our model. Accordingly, the time it takes to correctly determine the presence of the singleton will increase with an increasing number of distracters, in line with Duncan and Humphreys' (1989) theory and visual search experiments (for overviews, see Duncan & Humphreys, 1989; Wolfe & Horowitz, 2004).

Finally, we explored how illuminance may influence the saliency of objects in GSM. In GSM, an object can also be selected among identical objects (i.e., all the objects have the same object identity) in the saliency map, when it is more illuminant than the other objects. In fact, when a unique object that is attended (i.e., a cued-target) is presented among one distracter that is unique and highly illuminant and a number of other distracters (e.g., a red target among a highly illuminant green distracter and a number of gray distracters), GSM predicts that the highly illuminant object is increasingly strongly selected in the saliency map as its difference in illuminance with the other objects increases.

Although other models incorporating mechanisms of global saliency may also account for these findings in visual search (e.g., Cave, 1999; Wolfe, 1994), GSM does so without implementing within-feature competition across different stages of cortical processing. In GSM, there is no competition within features until the latest stage of cortical processing in the ventral pathway, in AIT. This avoids the drawback of within-feature competition models (Cave, 1999; Itti & Koch, 2000; Li, 2002; Wolfe, 1994) that information is reduced, which could be needed in later stages of visual processing (cf., Wolfe & Horowitz, 2004). In AIT, the identity of an object is selected due to competition. When the identity of an object is selected in AIT, it generates feedback activity which interacts with stimulus activity in the ventral retinotopic areas. The result is the selection of activity related to the object's location in these retinotopic areas. This selection (activation) in the ventral pathway, related to Van der Velde and De Kamps' (2001) model of object-based visual attention, is transmitted to the dorsal pathway. In the dorsal pathway, there are several maps in which (neurons coding for) different locations compete with each other. In the top-down map, the locations that are selected in the ventral pathway compete. In the contrast map, all other locations, which are not selected in the ventral pathway, compete. The activation in the top-down map and the contrast map is combined into the saliency map. Hence, the model has two distinctive features.

First, it is based on a combination of bottom-up, horizontal, and top-down processing. This differs from most other models that incorporate mechanisms of

global saliency, which assume that global saliency is the result of only bottom-up and horizontal processing (Cave, 1999; Itti & Koch, 2000; Li, 2002; Wolfe, 1994), but is consistent with neurophysiological evidence (Constantinidis & Steinmetz, 2001, 2005; Hegdé & Felleman, 2003; McPeek & Keller, 2002; Thompson et al., 1997; Thompson et al., 1996) (see Chapter 6).

Second, we propose a strong overlap between the mechanisms of global saliency and object-based visual attention. GSM supposes that global saliency and object-based visual attention mainly differ in the nature of object selection in AIT. When the identity of the target is unknown (i.e., when the target is defined as a singleton), the competition process in AIT is assumed to be random. When the identity of a target is known (in the presence of object-based visual attention), however, the competition process in AIT is assumed to be biased (and speeded up) toward the attended object (Van der Velde & De Kamps, 2001), due to memorization of the target (Chelazzi et al., 1993). From the selection of an object in AIT onwards, global saliency and object-based visual attention (Van der Velde & De Kamps, 2001) operate in the same way.

GSM can in principle explain efficient search for any (simple and high-level) feature or conjunction of features as long as the feature can be identified in an area such as AIT (i.e., it is represented in an area such as AIT) (cf., Ahissar & Hochstein, 2004; Wolfe, 2003), and its representation in AIT generates feedback activation to the retinotopic areas of the visual cortex, which enables the selection of activity specifically related to the feature's location in these retinotopic areas. Hence, GSM predicts a range of search slopes, depending on the effectiveness of the feedback activation to distinguish between activity related to the target and activity related to the distracters (we hypothesize that these feedback connections can be trained to some extent, see Chapter 4). This prediction is in line with the observation that the overall distribution of search slopes is unimodal (Wolfe, 1998). Instead, within-feature competition models predict a bimodal distribution of search slopes.

Evidently, GSM assumes specific feedback connections for (simple and high-level) features that can lead to efficient search. Nonetheless, feedback connections are anyhow needed to select visual information on the basis of top-down information (e.g., knowledge, expectations, goals), such as in the case of feature-based visual attention (Chawla et al., 1999; Martinez-Trujillo & Treue, 2004; Motter, 1994a, 1994b; Saenz et al., 2002) and object-based visual attention (Chelazzi et al., 1993; O'Craven et al., 1999). In contrast to within-feature competition models, GSM

does not additionally assume an explosion of the number of horizontal, inhibitory connections (in feature maps) across different stages of cortical processing.

## Appendix

### The spatial maps in the dorsal pathway of GSM

The dorsal pathway of GSM consists of five spatial maps: the input map (IM), contrast map (CM), ventral map (VM), top-down map (TDM) and saliency map (SM). Each spatial map is made up of a retinotopic layer of $31 \times 31$ excitatory neuron populations. The excitatory neuron populations in a spatial map are connected in a point-to-point manner to the excitatory neuron populations in other maps: the IM is connected to the CM, the VM to the CM, the VM to the TDM, the CM to the SM, and the TDM to the SM (see Figure 1).

### Excitatory neuron populations in the spatial maps

The excitatory neuron populations are modeled in terms of average neuron activity, which represents the overall activity of a neuron population. The average neuron activity is given by equations that regulate the input currents to a neuron population, and a response function that transforms these input currents into the discharge rate.

The equations that regulate the input currents to the excitatory neuron populations in the IM, CM, VM, TDM and SM are:

$$\tau_E \frac{dI_{i,j}^{IM}}{dt} = -I_{i,j}^{IM} + I_{i,j}^{dorsal} + I_{bg},$$

$$\tau_E \frac{dI_{i,j}^{CM}}{dt} = -I_{i,j}^{CM} + W_{IM\ to\ CM} F\left(I_{i,j}^{IM}\right) + W_{VM\ to\ CM} F\left(I_{i,j}^{VM}\right) + W_{to\ CM} F\left(L_{i,j}^{CM}\right) + I_{bg},$$

$$\tau_E \frac{dI_{i,j}^{VM}}{dt} = -I_{i,j}^{VM} + I_{i,j}^{ventral} + I_{bg},$$

$$\tau_E \frac{dI_{i,j}^{TDM}}{dt} = -I_{i,j}^{TDM} + W_{VM\ to\ TDM} F\left(I_{i,j}^{VM}\right) + W_{to\ TDM} F\left(L_{i,j}^{TDM}\right) + I_{bg},$$

$$\tau_E \frac{dI_{i,j}^{SM}}{dt} = -I_{i,j}^{SM} + W_{CM\ to\ SM} F\left(I_{i,j}^{CM}\right) + W_{TDM\ to\ SM} F\left(I_{i,j}^{TDM}\right) + W_{to\ SM} F\left(L_{i,j}^{SM}\right) + I_{bg}.$$

In these equations, $I_{i,j}^{k}$ is the current in the excitatory neuron population in spatial map $k$ at retinotopic location *(i, j)*. Furthermore, $\tau_E$ is the time membrane constant for excitatory neuron populations and $-I_{i,j}^{k}$ is the decay (leakage) of the excitatory

neuron population in spatial map *k* at retinotopic location *(i, j)*. The parameter $W_{k1\ to\ k2}$ represents the synaptic weight of a connection from spatial map *k1* to spatial map *k1*.

Moreover, $L^k_{i,j}$ is the lateral input current to the excitatory neuron population in spatial map *k* at retinotopic location *(i, j)*. The parameter $W_{to\ k}$ represents the synaptic weight of the connection from the inhibitory neuron population that provides the lateral input current (as described below) to the excitatory neuron population.

The excitatory neuron populations in the IM receive an external input current $I^{dorsal}_{i,j}$, and the excitatory neuron populations in the VM receive an external input current $I^{ventral}_{i,j}$. All neuron populations receive an input current reflecting background noise, $I_{bg}$, which is randomly selected from a Gaussian with mean $MI_{bg}$ and standard deviation $sdI_{bg}$.

The function *F(I)* represents the response function that transforms the input currents into the discharge rate *A:*

$$A = F(I) = \frac{k}{\left(1 + e^{-\beta(I-\theta)}\right)}.$$

### Inhibitory neuron populations in the CM, TDM, and SM

WTA competition in the CM, TDM, and SM is implemented through an inhibitory neuron population (Deco et al., 2002; Usher & Niebur, 1996). The TDM, CM and SM are linked to an inhibitory neuron population that receives input from all excitatory neuron populations in a spatial map via excitatory connections and inhibits all excitatory neuron populations in that spatial map via inhibitory connections. Consequently, each excitatory neuron population within a spatial map receives the same amount of inhibition $\forall ij : L^k_{i,j} = L_k$. The excitatory neuron populations are inhibited by choosing a negative weight for $W_{to\ k}$. The input currents to the inhibitory neuron population of spatial map *k*, $L_k$, are regulated by the following equation:

$$\tau_I \frac{dL_k}{dt} = -L_k + \sum_{i,j} W_{from\ k} F\left(I^k_{i,j}\right) + I_{bg}.$$

In the above equation, $\tau_I$ is the time membrane constant for inhibitory neuron populations. The parameter $W_{from\ k}$ represents the synaptic weight of the

connections from the excitatory neuron populations to the inhibitory neuron population in spatial map $k$.

### Parameter settings

In our simulations, we use $\tau_E$ = 5 ms, $\tau_I$ = 5 ms, $k$ = 80 Hz, $\theta$ = 4.0, $\beta$ = 1.0, $MI_{bg}$ = 0.025, and $sdI_{bg}$ = 0.03. The synaptic weights are set at the following values:

$$W_{IM\ to\ CM} = W_{VM\ to\ TDM} = W_{CM\ to\ SM} = W_{TDM\ to\ SM} = 0.5,$$

$$W_{VM\ to\ CT} = -0.5,$$

$$W_{to\ CM} = W_{to\ TDM} = W_{to\ SM} = -0.1,$$

$$W_{from\ CM} = W_{from\ TDM} = W_{from\ SM} = 0.005.$$

The values of the external input currents $I_{i,j}^{dorsal}$ and $I_{i,j}^{ventral}$ are shown in Table 1.

As can be seen in Table 1, $I_{i,j}^{dorsal}$ = 2 for excitatory neuron populations in the IM that represent the location of an object, and $I_{i,j}^{dorsal}$ = -2 for excitatory neuron populations in the IM that do not represent the location of an object. The value of the external input current $I_{i,j}^{ventral}$ is 2 for excitatory neuron populations in the VM that represent the location of an object of which the identity is selected in the ventral pathway, and -2 for excitatory neuron populations in the ventral map that represent the location an object of which the identity is not selected in the ventral pathway and that do not represent the location of an object. These values of $I_{i,j}^{dorsal}$ and $I_{i,j}^{ventral}$ are used in simulations, in which the illuminance of the presented objects is hypothesized to be 'standard' and the selection in the ventral pathway to be 'effective' (i.e., the T-D similarity is low). When the values of these parameters differ from the values in Table 1, their values in a simulation are given in the text.

*Table 1*
*Default external input currents*

| | Ventral | Dorsal |
|---|---|---|
| No object | -2.0 | -2.0 |
| Object selected in the ventral pathway | 2.0 | 2.0 |
| Object not selected in the ventral pathway | -2.0 | 2.0 |

# Chapter 8 | The inhibitory annulus of attention: Is it pre-attentive inhibition?

It has been proposed that the surrounds of the focus of spatial attention are inhibited. Such inhibitory surrounds have been inferred from longer search times for targets near attention-grabbing distracters, relative to targets far from such distracters. Here, we investigate the existence of such an inhibitory surround in two psychophysical experiments. In Experiment 1, evidence for an inhibitory surround accompanying attention was only found for inconspicuous targets. In Experiment 2, near targets benefited from spatial attention when spatial attention was manipulated through cueing, instead of through salient distracters. An alternative explanation for findings of an inhibitory surround may be that salient distracters inhibit surrounding elements not after grabbing attention, but pre-attentively through lateral inhibition.

## Introduction

It has long been known that visual stimuli in the focus of attention are detected more easily than those outside it. This has also been found at the neural level: attended stimuli elicit larger responses, and elicit responses at lower levels of contrast than unattended stimuli (Reynolds & Chelazzi, 2004). In the brain, attention also has other effects. Attention to one object within the receptive field of extrastriate neurons also results in smaller responses to other stimuli within the receptive field of the same cell (Moran & Desimone, 1985; Reynolds & Chelazzi, 2004). Responses to unattended objects close to objects in the focus of attention thus seem to be suppressed.

Evidence for inhibition of unattended objects close to the focus of attention has also come from psychophysiological research. Caputo and Guerra (1998) asked participants to detect an increase in the length of a line segment presented within a target form. They also presented a distracter with a unique color, which, in such circumstances, can capture attention (Theeuwes, 1991; Theeuwes, 1992). In conditions, in which the target had a unique but changing shape (i.e., its shape and those of the nontargets switched trial by trial), the line length threshold increased as the distance from the target to the distracter with a unique color decreased. Caputo and Guerra surmised that in their experiments attention was

first grabbed by the distracter. This then caused surrounding elements to be inhibited, making discrimination of surrounding objects more difficult (including, in some trials, the target). Similarly, Mounts (2000) let participants search for a target letter in displays that also contained attention-grabbing distracters. When the target letter was close to the attention-grabbing distracter, it was detected more slowly than when it was at some distance of the distracter. This was not the case if the distracter did not grab attention. Although the effect was small, it was also reported by Theeuwes and Godijn (2001) with different stimuli. Similar conclusions were reached with other experimental setups (Bahcall & Kowler, 1999; Cave & Zimmerman, 1997; Cutzu & Tsotsos, 2003; Müller, Mollenhauer, & Rösler, 2005), which we will review in the general discussion.

Most theories of visual search can accommodate these results. Many models include lateral inhibition between stimuli in each other's vicinity (Itti & Koch, 2000; Wolfe, 1994). Although such models usually simulate events only up to the moment that attention is focused upon one location, one can easily imagine that lateral inhibition of distracters becomes stronger when attention boosts the signal associated with an attended target stimulus (see Spivey & Spirn, 2000). One model, the Selective Tuning model of attention (Tsotsos et al., 1995), includes an explicit inhibitory surround around the focus of attention: within a map of features, attention enhances the signal at the attended location, but dampens signals in the vicinity of that location.

Originally, we set out to test two accounts of inhibition around the focus of attention, namely strengthened lateral inhibition and an explicit inhibitory annulus around the focus of attention. We did this by manipulating the similarity between target and distracter. As lateral inhibition is usually assumed to be strongest within feature maps (Itti & Koch, 2000; Wolfe, 1994), we reasoned that a lateral inhibition account would predict that the inhibitory surround of a distracter would be stronger when a distracter shared features with the target than when it did not. The inhibitory annulus account would predict no such interaction. To preview the results, we indeed found no interaction. In fact, we found no evidence at all of an inhibitory surround when the target was also somewhat salient. In a second experiment, we then found that an attention-grabbing cue did not produce an inhibitory surround.

## Experiment 1: Modulation of inhibitory effects between two feature singletons by lateral inhibition within color

In the experiments of Caputo and Guerra (1998) and Mounts (2000), the distance between a target and an attended location (i.e., the location of a feature singleton) was varied, and participants' latency to identify or match the target decreased with a larger distance between the attended location and the target. Our first experiment was designed to test whether a salient distracter that captures attention inhibits a close target stimulus stronger when it shares its defining feature with the target, than when it does not, as would be predicted by a lateral inhibition account (see above).

The target was identifiable by its unique shape, within which participants had to identify the orientation of a line. The distracter was defined by color. We manipulated the distance between the target and the distracter. Moreover, there were five conditions in the experiment: in the first and second condition, the target and the distracter were both colored, and either had the same color or a different color (Same and Different condition). In a third condition, similar to experiments of Mounts (2000), a gray target was accompanied by a colored distracter. In a fourth condition, the target itself was the only color singleton. The final one was a control condition, in which there was no distracter and all elements including the target were gray.

### Methods

*Participants*

Participants were eleven students at the Vrije Universiteit Amsterdam, who were paid for their participation.

*Stimuli*

Stimuli were presented on 21" SVGA color (Philips Brillance 201 P) monitors, with a resolution of 1024 to 768 pixels and a refresh rate of 120 Hz. Each trial started with the presentation of a fixation symbol for 700 ms. The fixation symbol was a gray "+" (lines 0.92 degrees of visual angle) located at the center of a black background. The fixation symbol remained visible in the search display.

Search displays were adapted from Theeuwes (1992) (see Figure 1A). They consisted of twelve elements randomly placed against a black background on fixed locations on an imaginary circle; one diamond, the target, and eleven nontarget disks (of which one could be a colored distracter). The diameter of the imaginary

circle was about 16.33 degrees of visual angle, while the diamond and disks measured 4.01 and 3.21 respectively in degrees of visual angle.

Each element contained an oriented, gray line with a length of approximately 1.1 degree of visual angle. The line in the target was horizontal or vertical. The orientation of lines in the distracter and nontargets was randomly chosen to be 22.5° or 67.5° tilted to the left, or 22.5° or 67.5° tilted to the right. Since horizontal or vertical lines do not pop out between heterogeneously oriented tilted lines, the target line in such displays is usually found through the unique shape surrounding it (Theeuwes, 1992).

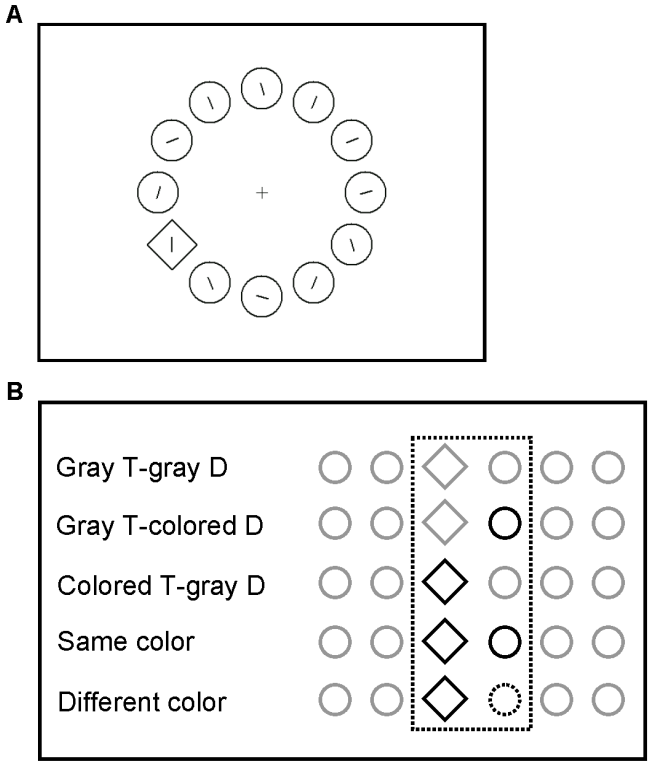


*Figure 1. Experiment 1. (A) A screenshot of the search display in the gray target and the gray distracter condition. (B) A schematic drawing of the five conditions, leaving out the oriented lines and the configuration of the stimuli. Gray denotes gray elements. Black denotes the color green, and the dashed, black line denotes the other color; red.*

*Table 1*
*Combinations of target and distracter color, and how they map onto the five conditions in Experiment 1*

| Target color | Distracter color | | |
|---|---|---|---|
| | Gray | Green | Red |
| Gray | Gray T-gray D | Gray T-colored D | Gray T-colored D |
| Green | Colored T-gray D | Same color | Different color |
| Red | Colored T-gray D | Different color | Same color |

*Note.* T = target; D = distracter.

The target was equally likely to be gray, red or green, which were all made equiluminant. In every trial, at least ten of the eleven nontarget disks were gray. The last disk, the distracter, was equally likely to be gray, red or green (the gray 'distracter' was equivalent to an 11th nontarget disk). This resulted in nine combinations of target and distracter color, which can be grouped into the five conditions listed above (see Table 1 and Figure 1B).
The locations of the target and the distracter -if present- were independently varied. This implied that the distance between the diamond and the colored disk was equally likely to be 1, 2, 3, 4, 5, or 6 locations.

*Procedure*

Participants were seated in a darkened room at approximately 70 cm of the screen. Each trial started with the presentation of a fixation symbol for 700 ms, after which the search display appeared. Participants were instructed to indicate whether the orientation of the line in the diamond was horizontal or vertical, by pressing one of two keyboard buttons. They were requested to respond as quickly as possible without making mistakes, and received visual feedback for 400 ms following errors. The response was followed by an interval of 200 ms until the onset of the fixation symbol for the following trial.
The experiment consisted of ten blocks of 54 trials, preceded by 24 practice trials. After each block, participants received feedback about their average response time and their accuracy in the last block, and a comparison to the previous block. Feedback also functioned as a self-paced break.

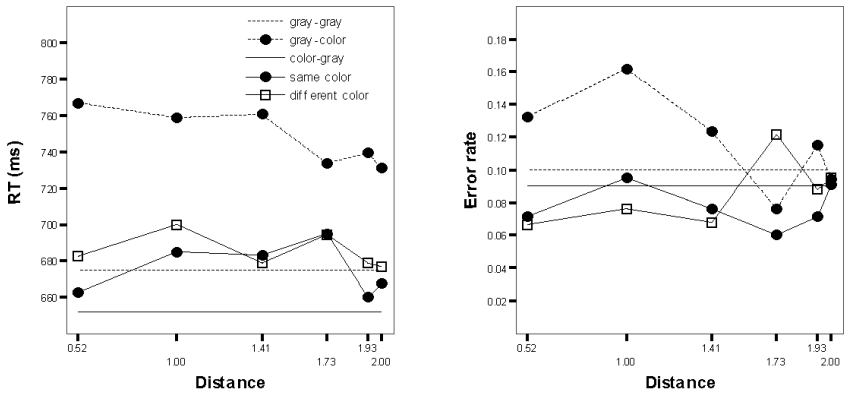


*Figure 2. Response time (left) and error rate (right) as a function of the distance between the target and the distracter for each condition. Note that there is no unique distance between the target and the distracter in conditions in which the distracter is gray. The values at the x-axis indicate the distance between the target and the distracter, which are computed with sinusoidal functions from the number of locations (i.e., 1, 2, 3, 4, 5, or 6) that the distracter is distant from the target.*

## Results

### *Response times*

RTs that were slower than 1200 ms were excluded from the analysis. This removed 4.88% of the trials. The average error rate over the remaining trials was 9.19% (one participant had a high error rate). Subsequent analyses were carried out over accurate trials only. Figure 2 shows average RTs for correct trials and error rates as a function of condition and of the distance between the target and the distracter.

An analysis of variance (ANOVA) performed on the RTs showed that there was an effect of condition, $F(4, 40) = 18.67$, $p < .001$. Planned comparisons between pairs of conditions revealed that in conditions with gray targets responses were slowed by the presence of a distracter (RT of 758.12 ms with distracter, vs. 673.47 ms without distracter), $t(10) = -8.47$, $p < .001$. The same was true for conditions with a colored target. RTs were slowed by the presence of an identically colored distracter (654.18 ms vs. 685.95 ms), $t(10) = -4.66$, $p = .001$, and also by the presence of a differently colored distracter (654.18 ms vs. 697.69 ms), $t(10) = -5.74$, $p = .000$. This indicates that the presence of a colored distracter slows down the identification of the target, independent of whether the target was itself colored or not. The colored distracter thus reliably captured attention.

A colored target made responses faster, both in conditions without a distracter, $t(10) = 5.78$, $p < .000$, and in conditions with a differently colored distracter, $t(10) = 3.67$, $p < .004$.

Mean RT was not different in the condition with a colored target and identically colored distracter, and that with a colored target and a differently colored distracter. This result is inconsistent with the hypothesis that lateral inhibition within one feature map (i.e., color map) increases competition for attentional selection between a colored target and a colored distracter.

*Linear regressions of response times over distance*

To evaluate the modulation of inhibitory effects by distance, we plotted RT as a function of the target-distracter distance in each condition in which there was a colored distracter, for each individual participant. We then fitted linear regression lines, and performed our statistical analyses on the slope parameters found for each participant in each condition. T-tests (two-tailed) were conducted to evaluate whether regression coefficients in each of the three conditions differed significantly from zero (see Figure 3). The regression coefficients indicated a negative slope only in the condition with a gray target and a colored distracter, $t(9) = -2.98$, $p < .015$. There was no significant slope in the other two conditions with a colored distracter.

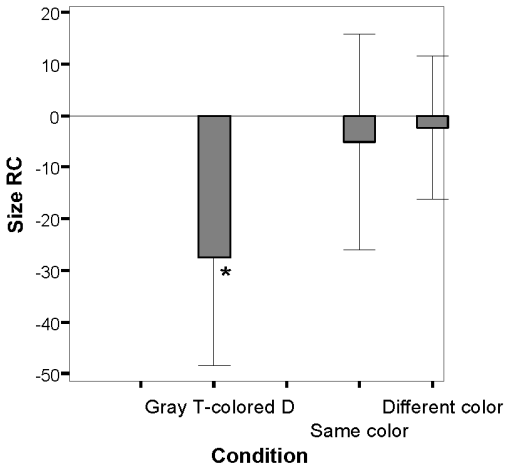


*Figure 3. The size of regression coefficients averaged over all participants, for each condition in Experiment 1. An asterisk indicates a slope significantly different from zero, p < .05, as tested in a two-tailed, one-sample t-test.*

Previous studies have noted that there may be costs when attention crosses the midline (Hughes & Zimba, 1985; Zimba & Hughes, 1987). As distance between the distracter and the target grows, the likelihood increases that the two are in different hemifields, and that attention will have to cross the midline when it is redeployed from the distracter to the target. To investigate whether this factor camouflages part or all of our results, we divided gray-target colored-distracter trials[12] post hoc into three categories: trials in which the target and distracter were in the same hemifield, where they were in different hemifields, and where one of the two was placed on the midline. While there was a trend for overall RTs to be longer in the condition in which either the target or the distracter appeared on the midline, this difference was not significant, $F(2, 18) = 2.82$, $p = .086$ (RT same hemifield: 767 ms; in different hemifields: 751 ms; target or distracter on midline: 772 ms). Gradients were noisy due to few trials per cell, but they too did not seem to be influenced by whether target and distracter were in the same hemifield or not (see Figure 4). There was a trend for RTs to be higher on trials in which target and distracter were near, than on trials in which they were further apart, $F(1, 9) = 3.56$, $p = .09$, and there was no effect of hemifield in which target and distracter appeared, $F < 1$.

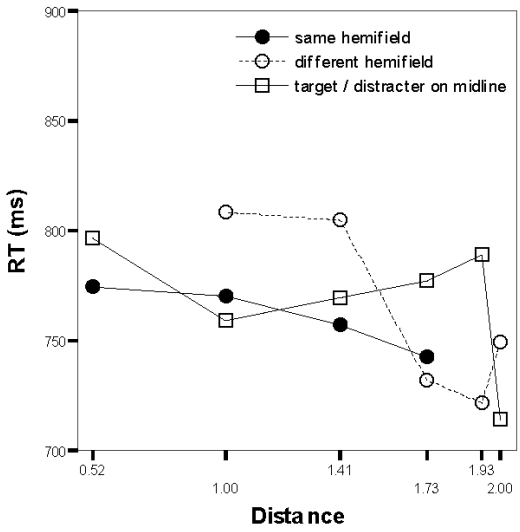


*Figure 4. Response time as a function of the distance between a gray target and a colored distracter, for trials in which the target and distracter were in the same hemifield, where they were in different hemifields, and where one of the two was placed on the midline.*

*Error rates*

An analysis of variance of the error rate with distance and condition as within-subject variables revealed no significant effects. As can be seen in Figure 2, errors are equally distributed over all conditions, except for the conditions in which a gray target and a colored distracter are relatively close to each other. In these conditions, the error rate is relatively high. As RTs were also high in these conditions, speed-accuracy tradeoffs can be excluded.

### Discussion

In summary, we found evidence for an inhibitory surround around an attention-capturing distracter when the target was gray, replicating Mounts (2000). When the target was colored itself, no such surround was apparent, although we found evidence for attentional capture by colored distracters also in those conditions. This is inconsistent with the hypothesis that a modulation of lateral inhibition causes the inhibitory surround: a colored distracter does not seem to inhibit similarly colored targets more than differently colored targets.

However, our results are also not entirely consistent with theories positing an inhibitory annulus around the focus of attention. If such an annulus were to exist, it is difficult to explain why such an annulus would affect gray, but not colored targets, as we found. One can speculate that the colored targets were more salient, and therefore more or less immune to the inhibitory surround, or that the colored distracters therefore attracted attention less often or less totally than when the target was gray (although this is unlikely in the face of evidence for attentional capture by the colored distracter even when the target was colored as well).

## Experiment 2: Modulation of inhibitory effects of a feature singleton by spatial attention

In a second experiment we further investigated the hypothesis of an explicit inhibitory annulus around the focus of attention. We now manipulated spatial attention more directly, by means of a spatial cue, in addition to attentional capture by color singletons. The cue, which preceded the search display, either directed attention to the location of the target or to the location of the distracter. In the *cued target location* condition the cue appeared on the location of the target, whereas in the *cued distracter location* condition the cue appeared on the location of the distracter. Given that the cue will capture attention, the inhibitory annulus hypothesis should predict an inhibitory surround around the location of the cue.

If the cue appeared at the location of the distracter, we should thus find longer RTs when the target appeared close to the distracter, than when the target appeared far from the distracter. We should thus find a gradient in the RT as a function of target-distracter distance in this condition.

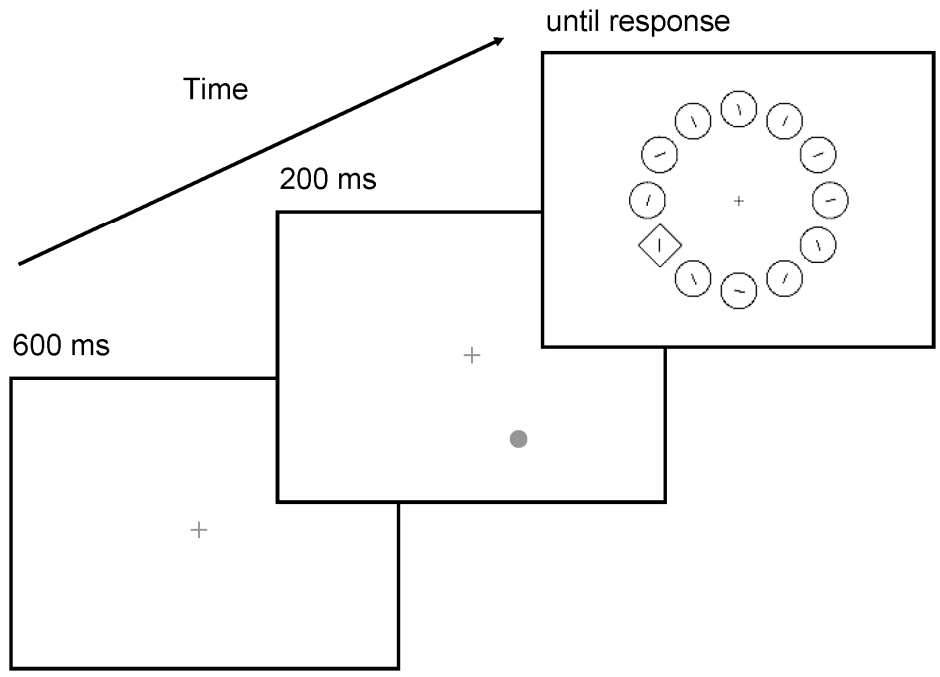


*Figure 5. The sequence of displays within a trial in Experiment 2. Gray denotes gray elements, and black denotes the color green.*

## Methods

### *Participants*

Participants were ten students at the Vrije Universiteit Amsterdam, who were paid for their participation.

### *Stimuli*

Stimuli were presented on the same apparatus as in Experiment 1. Search displays were equal to those in Experiment 1, with the difference that targets and distracters were never red, but only gray or green (ratio 1:1). Moreover, search

displays were now preceded by a small, circular gray dot that was visible at one of the 12 element locations for 200 ms (see Figure 5). It subtended 0.34 degree of visual angle. The spatial cue indicated the location of the upcoming target in 50% of the trials. In the remaining trials, the cued appeared at the location of an upcoming distracter. This was the colored distracter in the condition where there was one, and one of the gray disks, a nontarget, in conditions without a colored distracter.

*Procedure*

The procedure and instructions in Experiment 2 were the same as in Experiment 1.

## Results

*Response times*

RTs that were slower than 1200 ms were excluded from the analysis. This removed 7.10% of the trials. The average error rate over the remaining trials was 6.28%. Subsequent analyses were carried out over accurate trials only. For the cued distracter location and the cued target location condition, RTs and error rates are plotted as a function of condition and the distance between the target and distracter in Figure 6.

An analysis of variance (ANOVA) was performed on the RTs, treating the cued location and condition (i.e., the combination of target and distracter color) as within-subject variables. There were main effects of cued location, $F(1, 9) = 32.01$; $p < .001$, and condition, $F(3, 27) = 20.46$, $p < .001$ (Greenhouse-Geisser). The main effect of cueing confirms that participants paid attention to the location cue. Participants were faster when the target's location was cued (649.03 ms) than when the distracter's location was cued (706.20 ms).

The mean RT in the condition with a gray target and no distracter (672.25 ms) was compared to the mean RT in the condition with a gray target and a colored distracter (751.59 ms). Again, mean RT was higher when a colored distracter was present in the display, $t(9) = -4.74$, $p = .001$. The same was true when the target was colored: the presence of a colored distracter slowed the responses (671.63 ms vs. 646.22 ms), $t(9) = -3.97$, $p = .003$. As in Experiment 1, attention is captured by the colored distracter both with gray and colored targets. As in Experiment 1, a colored target was easier to find than a gray target when a distracter was present, $t(9) = 4.88$, $p < .001$.

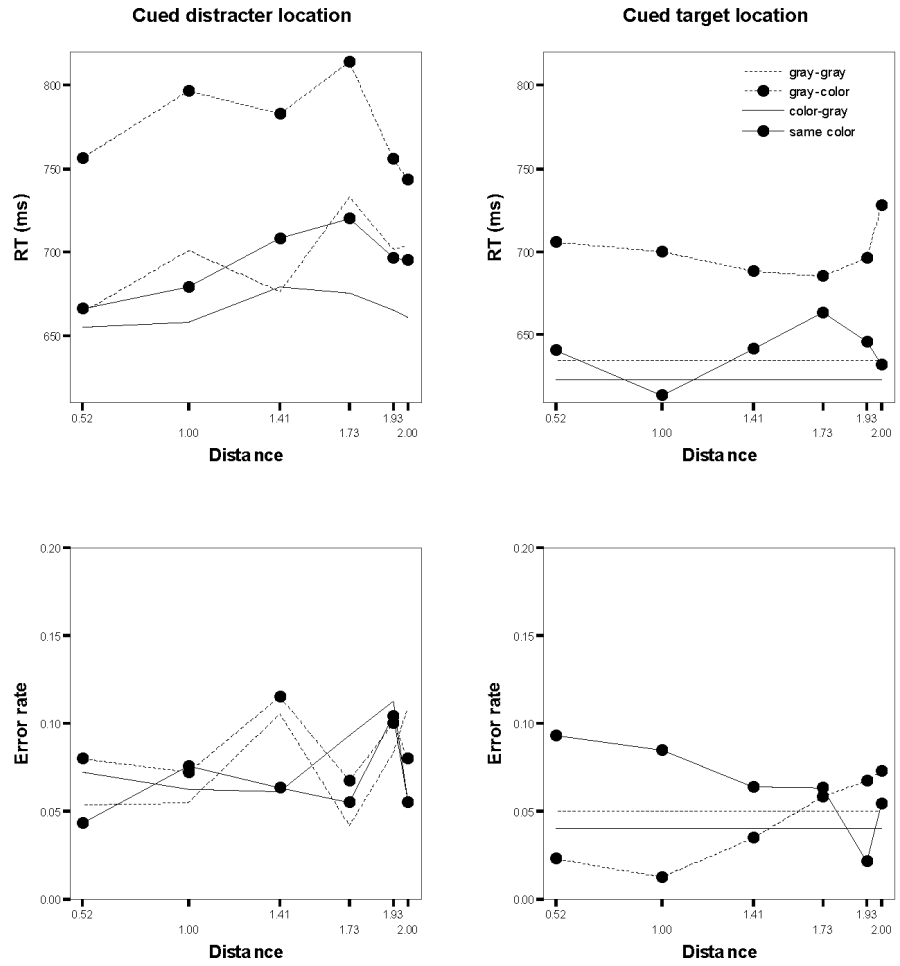


*Figure 6. Response time (top) and error rate (bottom) as a function of the distance between the target and the distracter for the cued distracter location (left) and the cued target location (right) separately, for each condition.*

### *Linear regressions of response time over distances*

To evaluate the hypothesis of an explicit inhibitory annulus around the focus of attention, we plotted RT as a function of the target-distracter distance in each condition in which there was a colored distracter or in which the cue indicated the location of the upcoming distracter, for each individual participant. We then fitted linear regression lines, and performed our statistical analyses on the slope parameters found for each participant in each condition.

One sample t-tests (two-tailed) were conducted to evaluate whether regression coefficients in each of the six conditions (cueing × combination of target and distracter color) differed significantly from zero (see Figure 7). The regression coefficients indicated a positive slope in the condition with a gray target and no distracter, in which the location of a gray nontarget was cued, $t(9) = 2.44$, $p < .037$. There was no significant slope in the other conditions in which there was a colored distracter or in which the cue indicated the location of the upcoming distracter. There was a trend towards a positive slope, however, when the location of the colored distracter was cued, $t(9) = 1.83$, $p < .010$.

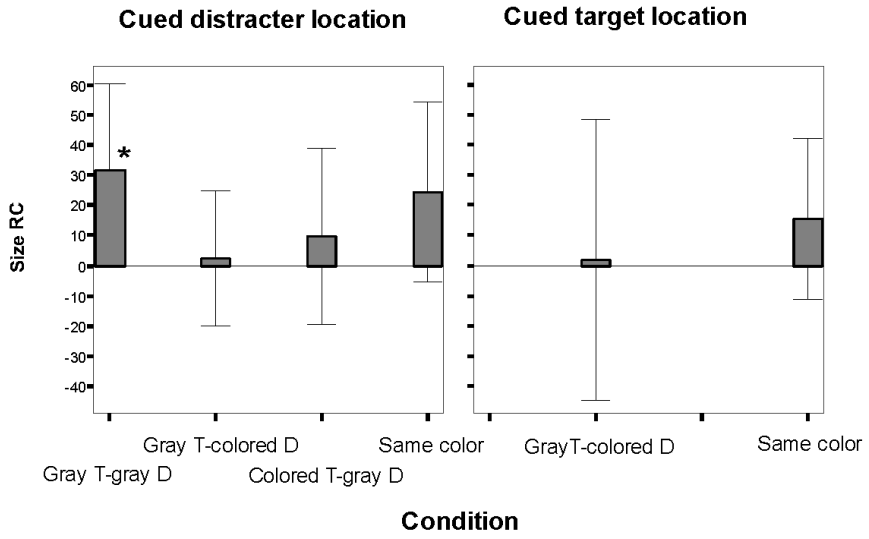


*Figure 7. The size of regression coefficients averaged over all participants, for the conditions in Experiment 2 in which the location of the distracter was cued (left) and for the conditions in which the location of the target was cued (right). An asterisk indicates a slope significantly different from zero, p < .05, as tested in a two-tailed, one-sample t-test.*

### *Error rates*

An analysis of variance of the error rate with the cued location, distance, and condition as within-subject variables, revealed only a main effect of cued location, $F(1, 9) = 11.24$; $p < .008$. Participants made more errors (M = 0.073) when the location of the distracter (or of a nontarget) was cued than when the location of the target was cued (M = 0.049). As this increase in errors rate goes hand in hand with an increase in RT, speed-accuracy tradeoffs cannot explain our results.

### Discussion

We found evidence for attentional capture both by the colored distracter, and by the spatial cue. Whether or not the cue captured attention in a purely exogenous way cannot be determined, because its 50% validity may have given participants an incentive to heed the cue. Whether or not participants did this, is not of importance for the results.

The most important result of Experiment 2 was that neither the cue nor the distracter gave rise to an inhibitory surround. The inhibitory effect that was present in the gray target / colored distracter condition in Experiment 1 disappeared when the location of the distracter was cued. Following the hypothesis of an inhibitory annulus around the focus of attention, the addition of a cue that indicates the location of the upcoming distracter should result in a steeper gradient in the condition with a gray target and a colored distracter than the gradient that is found in Experiment 1. This is because the cue should enhance the potential of the distracter to capture attention, and therefore generate a stronger inhibitory annulus. This was not found.

In the condition in which no distracter was present and the target was gray, facilitation was even found when the target appeared close to a cued nontarget location. Evidently, spatial attention is not automatically accompanied by inhibition of the immediate surroundings. Instead, attending a location seems to be accompanied by a facilitatory surround.

## General discussion

### No inhibitory annulus around the focus of attention?

Both Caputo and Guerra (1998) and Mounts (2000) found that when a target appears near an attention-grabbing distracter, it is found more slowly or less reliably than when it is at some distance from the distracter. We replicated these findings in Experiment 1, where we found that the RT increased for gray targets near a colored distracter that captured attention. However, the negative slope disappeared when the target was also colored, although the colored distracter still captured attention in this condition. This result goes against the hypothesis of an inhibitory annulus around the focus of attention.

More evidence against an inhibitory annulus was found in Experiment 2. When attention was manipulated by presenting a spatial cue, no inhibitory surround around the cued location was evident. Instead, targets close to the cued location were found faster than those further away (i.e., in the condition with a gray target

and no distracter, in which the location of a gray nontarget was cued), suggesting that attention has facilitatory effects, not inhibitory, around its focus. This corroborates older findings, in which endogenous cues were used to predict the location of upcoming targets. Benefits were found for either the whole hemisphere around the cue (Hughes & Zimba, 1985; Zimba & Hughes, 1987), or with a gradient around the cued position (Downing & Pinker, 1985; Rizzolatti, Riggio, Dascola, & Umilta, 1987; Zimba & Hughes, 1987).

It is entirely possible that an inhibitory annulus accompanies attention in some situations, but not in others. More research should then clarify when it does, and when it does not occur. Alternatively, the focus of attention may in fact not be accompanied by an inhibitory annulus. A new explanation will then have to be found for the findings that suggest such an annulus (Caputo & Guerra, 1998; Mounts, 2000, our Experiment 1). We will offer such an alternative explanation and then discuss other evidence for an inhibitory surround (Bahcall & Kowler, 1999; Cave & Zimmerman, 1997; Cutzu & Tsotsos, 2003; Müller et al., 2005).

### A pre-attentive inhibitory annulus around each salient location

Our alternative explanation proceeds from the relatively uncontroversial mechanism of pre-attentive lateral inhibition. Such a mechanism, in which nearby stimuli reduce one another's signal, is assumed in many models of visual search (Itti & Koch, 2000; Wolfe, 1994), and has also been found in the brain (Reynolds & Chelazzi, 2004). More salient stimuli have been shown to inhibit responses to other stimuli more strongly than less salient stimuli (Reynolds & Chelazzi, 2004). Many models of visual search have implemented lateral inhibition within feature maps (separate maps for each color, each orientation, etcetera) at lower levels in the visual hierarchy (Itti & Koch, 2000; Wolfe, 1994). Although this is possible, lateral inhibition within one or more spatial maps, independent of any specific feature value, can also explain our results.

In conditions in which the distracter is much more salient than the target, the result of such a mutual inhibition would be an overshadowing of the target by the salient distracter. This can explain the results of Caputo and Guerra (1998), Mounts (2000) and from the gray target-colored distracter condition in Experiment 1, which show an increasing latency to identify targets near a salient feature singleton. Pre-attentive lateral inhibition between all salient stimuli would also clarify the absence of evidence for an inhibitory surround in conditions of Experiment 1, in which both the target and the distracter are colored. In those

conditions, both the target and the distracter are relatively salient. Both stimuli will inhibit one another, and therefore their relative saliency will remain the same, independent of the distance between the two stimuli. For example, at small distance the colored target and the colored distracter may inhibit one another strongly. This will make both less salient, but the colored target and the colored distracter will remain more salient than the gray nontargets and therefore likely to be chosen for attentional selection. The same is true when the target and distracter are far from one another, and thus inhibit one another less strongly.

Mounts (2000) found that when color singletons fail to capture attention, no inhibitory surround is present. Although Mounts interpreted this as pointing to a role of attention in producing the inhibition, it can also be explained as a result of diminished saliency. The ability of distracters to capture attention is tightly linked to their saliency (Theeuwes & Godijn, 2001). In conditions in which distracters do not capture attention, they are not salient and therefore may not exert much pre-attentive lateral inhibition on target stimuli. In colored-target conditions in Experiment 1, we found evidence for attentional capture but not for an inhibitory surround, showing that saliency and not attentional capture may be the better predictor of an inhibitory surround.

Experiment 2 and similar experiments by others (Downing & Pinker, 1985; Rizzolatti et al., 1987; Zimba & Hughes, 1987) suggest that when attention is directed to a location by a cue, stimuli at surround locations are facilitated. The interplay of this facilitation and pre-attentive lateral inhibition could explain why the negative slope, found in the gray target and colored distracter condition in the Experiment 1, disappeared when the location of the colored distracter was cued in Experiment 2.

### Evidence for inhibition not relying on capture

Four experiments have yielded evidence for an inhibitory surround of attention without relying on attentional capture. Cave and Zimmerman (1997) used a probe technique to investigate the spread of attention after a search task, while both Bahcall and Kowler (1999) and Cutzu and Tsotsos (2003) let observers compare two locations. Müller and coauthors (Müller et al., 2005) found evidence of an inhibitory surround in a flanker task.

Cave and Zimmerman (1997) had observers search for a target letter within a briefly presented eight-letter array. In a portion of the trials they presented a probe dot on one of the eight positions following the array. Response times were

faster for probes on target locations than for probes at distracter locations. As participants received more practice, a second effect appeared. Response times became slower for probes at distracter locations near the target than for probes at distracter locations more distant from the target. This inhibitory surround was stronger when distracter letters near the target shared features with the target letter (and thus interfered more), suggesting to Cave and Zimmerman that the strength of spatial attention, and consequently the strength of its inhibitory surround, is flexibly adjusted according to the amount of interference between the target and distracter shapes. Cave and Zimmerman suggested that attention was allocated to inhibit distracter locations, and therefore to diminish interference, and that its strength (and its precision) increased with practice.

How might pre-attentive lateral inhibition between all salient stimuli explain this result? Although all the letters were equally salient with respect to color, extensive practice in search for the target letter could have made the target letter more salient through the development of a more elaborated representation. Increasing neural response to a target with extended practice is known from the animal literature (Bichot, Schall, & Thompson, 1996), and training was surely extensive in the experiment (for many participants, increases in RT for locations near the target only became significant after 19200 trials).

Two studies used two targets, and investigated the effect of the distance between the two on performance. Bahcall and Kowler (1999) measured the accuracy with which two target letters could be identified amidst a circular array of 24 characters. In different conditions, either the targets were cued by their unique color, or the locations of the target letters were cued by uniquely colored letters or by specific characters (numbers between letters) in a prior display that also consisted of 24 characters. In all cue conditions, the identification of the two target letters improved with a larger distance between the two target letters. Cutzu and Tsotsos (2003) had participants match the shape of two targets amidst distracters, after one of the target's locations was cued. The accuracy increased with a larger distance between the two targets.

Both Bahcall and Kowler (1999) and Cutzu and Tsotsos (2003) propose an explanation for their results in which cueing of a location results in an inhibitory surround around the focus of attention. The results of our second experiment and previous studies (Downing & Pinker, 1985; Rizzolatti et al., 1987; Zimba & Hughes, 1987) make us propose that cueing of a location has facilitatory effects at surrounding locations. How can we explain these inconsistent results? Pre-

attentive lateral inhibition can explain some results of Bahcall and Kowler (1999) and Cutzu and Tsotsos (2003). In Cutzu and Tsotsos' (2003) experiments the two targets were always colored, whereas distracters were black. Both target locations were consequently salient, and may therefore have pre-attentively inhibited one another when they were close. This does not explain results in Bahcall and Kowler's (1999) conditions in which cues were numbers between letters, and all characters were black. An alternative explanation for these results was provided by Bahcall and Kowler (1999) themselves. They proposed that targeting of attention could become less precise with a smaller separation between two targets. This explains why the two target letters are harder to identify with smaller target separation. That there is some location insecurity in the visual system has long been argued by several researchers (Ashby, Prinzmetal, Ivry, & Maddox, 1996; Bundesen, 1990, 1998; Intriligator & Cavanagh, 2001; Treisman & Schmidt, 1982).

Our finding of a facilitatory surround around the focus of spatial attention is in line with results from previous flanker experiments, showing that interference between incompatible and task-irrelevant flankers and a target decreases monotonically with an increasing target-irrelevant flanker distance (e.g., Eriksen & Hoffman, 1972; Eriksen & St James, 1986). However, Müller et al. (2005) recently reported evidence for what they called a Mexican hat-shaped distribution of attention in an adapted flanker paradigm. In their study a target letter always appeared on the same location. Flankers immediately adjacent to the target interfered most with target identification, but a flanker at the second position from the target interfered less than a flanker at the third position. This implies an inhibitory region (the brim of the hat at the second position) around a cone of facilitation (first position). Our range of the target-cued distracter distances included the visual degrees at which Müller et al. found evidence for inhibition, and cannot explain the inconsistent findings. It is possible that task differences may have induced different distributions of attention in the two studies. In Müller et al.'s experiments targets were presented at one fixed location, while in our second experiment the spatial cue was only fifty percent valid. Participants may thus have adopted a wider attentional window in our experiments than in those of Müller et al. (2005). As a result, the inhibitory surround may have been attenuated or made too distant in our Experiment 2. In fact, conditions with colored distracters, in which the location of the distracter (or a nontarget) is cued, show a nonsignificant decrease in RT for the two positions in which target-distracter

distance is largest. This would be in line with an inhibitory surround. Alternatively, it is possible that the Mexican hat distribution of attention in Müller et al.'s study is an artifact. The pairwise comparison between the second and third position in Müller et al.'s study was barely significant, and not corrected for the number of tested comparisons.

## Conclusion

Neurophysiological studies have suggested that although attention facilitates stimuli in its focus, it inhibits responses to stimuli that are at some distance from this focus (Moran & Desimone, 1985; Reynolds & Chelazzi, 2004). Psychophysical studies have also found support for this pattern. Here, we found that the support is less robust than it seemed, and that attention may instead facilitate the processing of stimuli near its focus. We propose that attention-capturing distracters may slow search for near-targets through pre-attentive lateral inhibition instead of through an inhibitory annulus accompanying the capture of attention. Although this can explain our findings and some older ones, other findings remain difficult to explain without assuming an inhibitory annulus. More research is needed to resolve these inconsistencies.

# Chapter 9 | Conclusions

This thesis set out to investigate the mechanisms of global saliency, the mechanisms of top-down visual attention, and the interaction between these mechanisms, in visual search. Following the outline of CLAM (Van der Velde et al., 2004), simulations in the preceding chapters explored mechanisms of visual working memory in the prefrontal cortex and of object recognition in the ventral pathway, and specified mechanisms of spatial selection in the dorsal pathway. Behavioral experiments additionally addressed several questions regarding global saliency and top-down visual attention in visual search, and their interaction. The findings of the simulations and behavioral experiments have implications for CLAM in particular, and for the mechanisms of global saliency and top-down visual attention in general. An overview of the main findings of the simulations and behavioral experiments in this thesis is presented below, together with conclusions that may be drawn from these findings.

## Visual working memory in the prefrontal cortex

Behavioral research has shown that the number of objects that can be maintained in visual working memory (VWM) without interference (i.e., loss of information) is limited (to about four), but the number of object features (e.g., shape, color, location, motion, etc.) is unlimited for each of these objects (Vogel et al., 2001). The simulations in Chapter 2 indicate that the architecture of visual working memory that was proposed in Van der Velde and De Kamps (2003) and in CLAM has a qualitatively similar capacity limit. Naturally, the fact that this blackboard architecture of visual working memory shows a capacity limit that is also shown by its human counterpart does not allow the inference that the visual working memory in humans is based on the architecture in CLAM. Other models of visual working memory can account for this finding as well (e.g., for an account based on neural synchronization within object representations and inhibition between object representations, see Raffone & Wolters, 2001). Nonetheless, it is possible that the capacity limit of the human visual working memory arises from an architecture, in which objects are represented in a blackboard (Van der Velde & De Kamps, 2006; Van der Velde et al., 2004). The representation of an object in the VWM-blackboard is used to bind the features of the object, which are either located in the ventral and dorsal stream or in PFC itself (or both). As the number of

objects increases, the representation of an object with a specific feature (i.e., shape or location) cannot reliably be selected among the other object representations in the VWM-blackboard in PFC, due to interference between the object representations. As a result, it becomes impossible to bind the features of an object that is represented in visual working memory.

## Object recognition in the ventral pathway

Chapter 3 suggested a process which might contribute to location invariant object recognition in the ventral pathway. Central to this proposal is a learning scheme in which learning in the feedforward network of the ventral pathway is built up. The feedforward network first learns to identify simple features (e.g., oriented lines, edges) at all locations and therefore becomes selective for location invariant features. Subsequently, the feedforward network of the ventral pathway learns to identify objects partly by learning new conjunctions of these location invariant features. Simulations showed that such a learning scheme enabled the feedforward network of the ventral pathway to identify an object at a new location (to some extent).

Efficient search for a cued-target among distracters not only requires that the feedforward network of the ventral pathway is able to identify the target at any (trained and new) location (i.e., location invariant object recognition), but also that the feedback network of the ventral pathway carries information about the cued-target to the retinotopic areas (Van der Velde & De Kamps, 2001; Van der Velde et al., 2004). In fact, we argued in Chapter 7 that top-down processing is even involved in search for a singleton, i.e., in the absence of object-based visual attention. Hence, learning in the feedforward network of the ventral pathway needs to be transferred to the feedback network of the ventral pathway (Van der Velde & De Kamps, 2001). Simulations indicated that transferring the selectivity in the feedforward network to the feedback network (using Hebbian learning), in the building up learning scheme as used in Chapter 3, is not sufficient to reliably find a cued-target at new locations among distracters. Accordingly, we hypothesized that, under this learning scheme, additional, location dependent features are needed to reliably find a cued-target among distracters, and that this can be achieved by supervised learning once the feedforward network is able to identify an object at a new location. This hypothesis predicts that the generalization to new locations by the visual system is more restricted when we

have to find an object among other objects (Van der Velde & De Kamps, 2001), than when we have to recognize a single object.

This prediction was not supported by findings of the behavioral experiments in Chapter 4. It was found that search for a digital 2 (digital 5) among digital 5's (digital 2's), which was highly inefficient before training (also see, Wang et al., 1994) became more (but not fully) efficient through training. However, this increase in search efficiency (and the general decrease in response time) was hardly specific for trained locations, but generalized substantially from trained to untrained locations. Evidently, building up learning is only one approach to obtain some degree of location invariance for object recognition (and localization) in the ventral pathway. There are interesting other suggestions about how location invariant representations may be achieved in the ventral pathway, and these may together or instead underlie the impressive human (and primate) performance (e.g., Riesenhuber & Poggio, 2000; Wallis & Rolls, 1997).

The digital 2 and the digital 5 were chosen as the search items in the behavioral experiments in Chapter 4, because inefficient search for this target-distracter pair (Wang et al., 1994) seems inconsistent with recent findings about the effect of familiarity in visual search (Malinowski & Hübner, 2001; Shen & Reingold, 2001). Malinowski and Hübner (2001) and Shen and Reingold (2001) found that the familiarity of the distracters largely determines the search efficiency, i.e., search is (more) efficient when the distracters are familiar and (more) inefficient when the distracters are unfamiliar. Since the digital 2 and the digital 5 are both hypothesized to be familiar, search for the digital 2 (digital 5) among digital 5's (digital 2's) would not be expected to be as inefficient as reported by Wang et al. (Wang et al., 1994).

The finding of the first behavioral experiment in Chapter 4 that the familiarity of the digital 2 and digital 5 can be improved significantly through training indicates that the digital 2 and digital 5 were not as familiar as assumed by Wang et al. (1994). This undermines Wang et al.'s (1994) conclusion that search is not efficient when both the target and the distracters are familiar.

Although search for the digital 2 (digital 5) among digital 5's (digital 2's) became substantially more efficient through training, it did not become fully efficient. Apparently, the intensive training (more than 5760 search trials) in Experiment 1 did not allow the representations of the digital 2 and digital 5 in the visual cortex to become as independent as required for highly accurate, parallel search. This

suggests that, after intensive training, objects are still (partially) recognized on the basis of relatively simple features, which are similar for the digital 2 and digital 5.
Moreover, the improvement in search performance was largely specific to the trained target-distracter pair (i.e., the digital 2 among digital 5's or the digital 5 among digital 2's). This difference in search performance between the trained and the untrained target-distracter pair cannot be explained by a difference in familiarity between the target stimulus and the distracter stimulus. The digital 2 and the digital 5, defined as the target and the distracter stimulus in terms of their use in the trained search task, were equally familiar after training (this was tested by measuring RTs in an identification task). Thus, the results of the behavioral experiments in Chapter 4 suggest that search efficiency does not only depend on the familiarity of the distracters (Malinowski & Hübner, 2001; Shen & Reingold, 2001; Mruczek & Sheinberg, 2005) or on the difference in familiarity between the target and the distracters (Wang et al., 1994).
Apparently, learning the distracters as a group also affects the search efficiency, even though it does not result in an increased familiarity of the distracter stimulus as compared to the familiarity of the target stimulus. Since the increase in search efficiency (and the general decrease in response time) for the trained search task generalized substantially from trained to untrained locations, we propose that in the first experiment in Chapter 4 a grouping of distracters was learned with a representation at a high level of the visual hierarchy, in which neurons have large receptive fields. Finally, it was found in Chapter 4 that the effect of learning was quite robust over time, i.e., it was still (partly) present two months after training, and was largely specific to the actual stimuli used.

## Interaction between object recognition in the ventral pathway and spatial selection in the dorsal pathway

The behavioral experiments in Chapter 5 provided evidence that global saliency is a gradual phenomenon. Elements from a minority colored set with more than one element were searched earlier or faster than elements from a majority colored set, and are thus prioritized in search in a similar manner as color singletons. In contrast to conjunctive search studies that explained findings of smaller-group search by gradual global saliency (Sobel & Cave, 2002; Zohary & Hochstein, 1989), the benefit of shifting attention to elements from a minority colored set was restricted in our experiments. Moreover, it was shown that our findings could not be explained by the strategy of voluntarily searching elements from the minority

colored set before elements from the majority colored set. The findings of the behavioral experiments in Chapter 5 further demonstrated that top-down visual attention speeds up the response to a target, even when the location of the target is already globally salient. Regarding the dynamics of the mechanisms of global saliency and top-down visual attention over time, evidence was obtained that colored elements already activate the mechanisms responsible for global saliency when they are presented for 50 ms, whereas they enable the selection by top-down visual attention when they are presented for 100 ms.

Chapter 6 presented an extensive review of behavioral and neurophysiological studies and models of visual search. Based on the findings of the reviewed behavioral studies, it was concluded that global saliency cannot solely be attributed to processing in low cortical areas. From the reviewed neurophysiological studies (Constantinidis & Steinmetz, 2001, 2005; Hegdé & Felleman, 2003; McPeek & Keller, 2002; Thompson et al., 1997; Thompson et al., 1996), the conclusion was drawn that global saliency does not necessarily have to be the result of solely bottom-up and horizontal processing. Instead, the findings of the reviewed neurophysiological studies are also consistent with the hypothesis that global saliency is the result of a combination of bottom-up, horizontal, and top-down processing.

Chapter 7 presented a model of global saliency, GSM, which can account for several important findings in visual search. The model differs from other models that incorporate mechanisms of global saliency (Cave, 1999; Itti & Koch, 2000; Li, 2002; Wolfe, 1994). These models implement competition between neurons that represent the same features. In order to explain the behavioral findings that efficient search may (Enns & Rensink, 1990; Kleffner & Ramachandran, 1992; Wolfe et al., 1994) and sometimes even has to (He & Nakayama, 1992; Rensink & Enns, 1998) be based on the results of later stages of cortical processing, which we reviewed in Chapter 6, within-feature competition models need to assume that there is within-feature competition across different stages of cortical processing. This assumption entails an explosion of the number of horizontal, inhibitory connections across different stages of cortical processing.

In GSM there is also competition within features, but only at the latest stage of cortical processing in the ventral pathway, for example in AIT. In addition to reducing the number of horizontal, inhibitory connections, this avoids the drawback of within-feature competition models (Cave, 1999; Itti & Koch, 2000; Li, 2002; Wolfe, 1994) that information is reduced, which could be needed in later

stages of visual processing (cf., Wolfe & Horowitz, 2004). In GSM, all the visual information remains present in the feedforward network of the ventral pathway (Van der Velde & De Kamps, 2001) and in the input map of the dorsal pathway.

We suggested in Chapter 7 that after the selection of an object identity in AIT (following the processing of visual information in the feedforward network of the ventral pathway), the selected object identity generates activity in the feedback network of the ventral pathway, which interacts with the activity in the retinotopic areas of the feedforward network in the ventral pathway. The result is the selection of activity related to the object's location in these retinotopic areas. This selection (activation) in the ventral pathway, related to Van der Velde and De Kamps' (2001) model of object-based visual attention, is transmitted to the dorsal pathway. In the dorsal pathway, spatial selection, which was not yet specified in CLAM, is hypothesized to take place in several spatial maps in which (neurons coding for) different locations compete with each other. Together, the interaction between object recognition in the ventral pathway and spatial selection in the dorsal pathway results in global saliency.

Hence, it is proposed in this thesis that global saliency results from top-down processing (in the ventral pathway), in addition to bottom-up and horizontal processing (in the ventral and dorsal pathway). This differs from within-feature competition models, which assume that global saliency is the result of only bottom-up and horizontal processing (Cave, 1999; Itti & Koch, 2000; Li, 2002; Wolfe, 1994), but is consistent with neurophysiological evidence. This conclusion was drawn from the neurophysiological studies that we reviewed in Chapter 6 (Constantinidis & Steinmetz, 2001, 2005; Hegdé & Felleman, 2003; McPeek & Keller, 2002; Thompson, Bichot, & Schall, 1997; Thompson, Hanes, Bichot, & Schall, 1996).

Thus, we hypothesize that the mechanisms of global saliency and object-based visual attention largely overlap, and that they primarily differ in the nature of the selection of an object identity in AIT in the ventral pathway. In the case of object-based visual attention, the competition between object identities in AIT is biased toward the attended object identity due to memorization of the attended object in visual working memory (Van der Velde & De Kamps, 2001; Van der Velde et al., 2004), while in the case of global saliency the competition between object identities in AIT is random. Therefore, the selection of an object identity in AIT is speeded up in the presence of object-based visual attention as compared to in the absence of object-based visual attention. As a result, the spatial maps in the dorsal

pathway of our model are able to compute global saliency earlier in time. This hypothesized interplay between the mechanisms of global saliency and the mechanisms of object-based visual attention is consistent with the behavioral finding in Chapter 5 that top-down visual attention speeds up the response to a target, when the location of the target is already globally salient.

Behavioral studies found that the response time to identify or match a target decreases with a larger distance between the target and an attended location (i.e., the location of a feature singleton) (e.g., Caputo & Guerra, 1998; Mounts, 2000). These results and other results have been interpreted as evidence that there is an inhibitory annulus around the focus of attention. Chapter 8 tested whether inhibition around the focus of attention might result from pre-attentive lateral inhibition between objects that is stronger when objects share features with another than when they do not, as assumed by within feature-competition models (Cave, 1999; Itti & Koch, 2000; Li, 2002; Wolfe, 1994). The first behavioral experiment tested this prediction by manipulating the similarity between a target and distracter. No interaction was found. In fact, we found no evidence of an inhibitory surround if the target was also salient, even when a salient distracter grabbed attention. Moreover, in a second behavioral experiment it was found that a spatial cue, which grabbed attention, produces a facilitatory surround. Hence, the findings of the behavioral experiments in Chapter 8 suggest that the support for an inhibitory annulus around the focus of attention is less robust than it seemed, and that attention may instead facilitate the processing of stimuli near its focus. In line with GSM, we propose that salient objects inhibit surrounding objects pre-attentively through lateral inhibition and not after grabbing attention, but irrespective of whether they share features or not.

The fact that the response time to identify or match a target may depend on the distance between a target and a distracter, as in the condition in which the distracter but not the target was salient in Experiment 1 in Chapter 8, and in other studies (e.g., Caputo & Guerra, 1998; Mounts, 2000), suggests that the (spatial) competition between salient objects is not completely homogeneous across the visual field. Instead, the strength of the competition between salient objects seems to depend (partly) on the distance between objects, i.e. it is *gradual* (or has a gradual component). In the future, the spatial competition in the saliency map of GSM, which is yet completely homogenous across the visual field, may therefore be adapted to reflect these findings.

## References


Ahissar, M., & Hochstein, S. (1997). Task difficulty and the specificity of perceptual learning. *Nature, 387,* 401-406.

Ahissar, M., & Hochstein, S. (2004). The reverse hierarchy theory of visual perceptual learning. *Trends in Cognitive Sciences, 8,* 457-464.

Aks, D. J., & Enns, J. T. (1992). Visual search for direction of shading is influenced by apparent depth. *Perception and Psychophysics, 52,* 63-74.

Amit, Y., & Mascaro, M. (2003). An integrated network for invariant visual detection and recognition. *Vision Research, 43,* 2073-2088.

Ashby, F. G., Prinzmetal, W., Ivry, R. B., & Maddox, W. T. (1996). A formal theory of feature binding in object perception. *Psychological Review, 103,* 165-192.

Bacon, W. F., & Egeth, H. E. (1994). Overriding stimulus-driven attentional capture. *Perception and Psychophysics, 55,* 485-496.

Bacon, W. F., & Egeth, H. E. (1997). Goal-directed guidance of attention: Evidence from conjunctive visual search. *Journal of Experimental Psychology: Human Perception and Performance, 23,* 948-961.

Bahcall, D. O., & Kowler, E. (1999). Attentional interference at small spatial separations. *Vision Research, 39,* 71-86.

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

Bichot, N. P., Rossi, A. F., & Desimone, R. (2005). Parallel and serial neural mechanisms for visual search in macaque area V4. *Science, 308,* 529-534.

Bichot, N. P., & Schall, J. D. (1999). Effects of similarity and history on neural mechanisms of visual selection. *Nature Neuroscience, 2,* 549-554.

Bichot, N. P., & Schall, J. D. (2002). Priming in macaque frontal cortex during popout visual search: Feature-based facilitation and location-based inhibition of return. *Journal of Neuroscience, 22,* 4675-4685.

Bichot, N. P., Schall, J. D., & Thompson, K. G. (1996). Visual feature selectivity in frontal eye fields induced by experience in mature macaques. *Nature, 381,* 697-699.

Bravo, M., & Blake, R. (1990). Preattentive vision and perceptual groups. *Perception, 19,* 515-522.

Bundesen, C. (1990). A Theory of Visual Attention. *Psychological Review, 97,* 523-547.

Bundesen, C. (1998). A computational theory of visual attention. In G. W. Humphreys, J. Duncan & A. Treisman (Eds.), *Attention, space, and action: Studies in cognitive neuroscience* (pp. 54-71). Oxford: Oxford University Press.

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

Carrasco, M., & McElree, B. (2001). Covert attention accelerates the rate of visual information processing. *Proceedings of the National Academy of Sciences, 98,* 5363-5367.

Cave, K. R. (1999). The FeatureGate model of visual selection. *Psychological Research, 62,* 182-194.

Cave, K. R., & Wolfe, J. M. (1990). Modeling the role of parallel processing in visual search. *Cognitive Psychology, 22,* 225-271.

Cave, K. R., & Zimmerman, J. M. (1997). Flexibility in spatial attention before and after practice. *Psychological Science, 8,* 399-403.

Chawla, D., Rees, G., & Friston, K. J. (1999). The physiological basis of attentional modulation in extrastriate visual areas. *Nature Neuroscience, 2,* 671-676.

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

Constantinidis, C., & Steinmetz, M. A. (2001). Neuronal responses in area 7a to multiple stimulus displays? Neurons encode the location of the salient stimulus. *Cerebral Cortex, 11,* 581-591.

Constantinidis, C., & Steinmetz, M. A. (2005). Posterior parietal cortex automatically encodes the location of salient stimuli. *Journal of Neuroscience, 25,* 233-238.

Corbetta, M., & Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nature Neuroscience, 3,* 201- 215.

Coren, S., Ward, L. M., & Enns, J. T. (2003). *Sensation and perception* (5th ed.). New York: John Wiley & Sons.

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

De Kamps, M., & Van der Velde, F. (2001). From artificial neural networks to spiking neuron populations and back again. *Neural Networks, 14,* 941-953.

Deco, G., Pollatos, O., & Zihl, J. (2002). The time course of selective visual attention: Theory and experiments. *Vision Research, 42,* 2925-2945.

Downing, C. J., & Pinker, S. (1985). The spatial structure of visual attention. In M. I. Posner & O. S. M. Marin (Eds.), *Attention and Performance XI: Mechanisms of Attention* (pp. 171-188). Hillsdale, NJ: Erlbaum.

Duncan, J., & Humphreys, G. W. (1989). Visual search and stimulus similarity. *Psychological Review, 96,* 433-458.

Egeth, H. E., Virzi, R. A., & Garbart, H. (1984). Search for conjunctively defined targets. *Journal of Experimental Psychology: Human Perception and Performance, 10,* 32-39.

Enns, J. T., & Rensink, R. A. (1990). Influence of scene-based properties on visual search. *Science, 247,* 721-723.

Eriksen, C. W., & Hoffman, J. E. (1972). Temporal and spatial characteristics of selective encoding from visual displays. *Perception and Psychophysics, 12,* 201-204.

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

Fukushima, K. (2004). Neocognitron capable of incremental learning. *Neural Networks, 17,* 37-46.

Gottlieb, J. P., Kusunoki, M., & Goldberg, M. E. (1998). The representation of visual salience in monkey parietal cortex. *Nature, 391,* 481-484.

Grossberg, S., Mingolla, E., & Ross, W. D. (1994). A neural theory of attentive visual search: Interactions of boundary, surface, spatial, and object representations. *Psychological Review, 101,* 470-489.

Hamker, F. (2004). A dynamic model of how feature cues guide spatial attention. *Vision Research, 44,* 501-521.

He, Z. J., & Nakayama, K. (1992). Surfaces versus features in visual search. *Nature, 359,* 231-233.

Hegdé, J., & Felleman, D. J. (2003). How selective are V1 cells for pop-out stimuli? *Journal of Neuroscience, 23,* 9968-9980.

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

Hughes, H. C., & Zimba, L. D. (1985). Spatial maps of directed visual attention. *Journal of Experimental Psychology: Human Perception and Performance, 11,* 409-430.

Humphreys, G. W., & Müller, H. J. (1993). SEarch via Recursive Rejection (SERR): A connectionist model of visual search, *Cognitive Psychology, 25,* 43-110.

Humphreys, G. W., Quinlan, P. T., & Riddoch, M. J. (1989). Grouping processes in visual search: Effects with single- and combined-feature targets. *Journal of Experimental Psychology: General, 118,* 258-279.

Intriligator, J., & Cavanagh, P. (2001). The spatial resolution of visual attention. *Cognitive Psychology, 43,* 171-216.

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

Kaptein, N. A., Theeuwes, J., & Van der Heijden, A. H. C. (1995). Search for a conjunctively defined target can be selectively limited to a color-defined subset of elements. *Journal of Experimental Psychology: Human Perception and Performance, 21,* 1053-1069.

Karni, A., & Sagi, D. (1991). Where practice makes perfect in texture segmentation: Evidence for primary visual cortex plasticity. *Proceedings of the National Academy of Sciences, 88,* 4966-4970.

Kastner, S., Nothdurft, H. C., & Pigarev, I. (1999). Neuronal responses to motion and orientation contrast in cat striate cortex. *Visual Neuroscience, 16,* 587-600.

Kleffner, D. A., & Ramachandran, V. S. (1992). On the perception of shape from shading. *Perception and Psychophysics, 52,* 18-36.

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

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

Leonards, U., Rettenbach, R., Nase, G., & Sireteanu, R. (2002). Perceptual learning of highly demanding visual search tasks. *Vision Research, 42,* 2193-2204.

Li, Z. (2002). A saliency map in primary visual cortex. *Trends in Cognitive Sciences, 6,* 9-16.

Malinowski, P., & Hübner, R. (2001). The effect of familiarity on visual-search performance: Evidence for learned basic features. *Perception and Psychophysics, 63,* 458-463.

Maljkovic, V., & Nakayama, K. (1994). Priming of pop-out: I. Role of features. *Memory & Cognition, 22,* 657-672.

Martinez-Trujillo, J. C., & Treue, S. (2004). Feature-based attention increases the selectivity of population responses in primate visual cortex. *Current Biology, 14,* 744-751.

McLeod, P., Driver, J., & Crisp, J. (1988). Visual search for a conjunction of movement and form is parallel. *Nature, 332,* 154-155.

McPeek, R. M., & Keller, E. M. (2002). Saccade target selection in the superior colliculus during a visual search task. *Journal of Neurophysiology, 88,* 2019-2034.

McPeek, R. M., Maljkovic, V., & Nakayama, K. (1999). Saccades require focal attention and are facilitated by a short-term memory system. *Vision Research, 39,* 1555-1566.

Mohler, C. W., Goldberg, M. E., & Wurtz, R. H. (1973). Visual receptive fields of frontal eye field neurons. *Brain Research, 61,* 385-389.

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

Mordkoff, J. T., Yantis, S., & Egeth, H. E. (1990). Detecting conjunctions of color and form in parallel. *Perception and Psychophysics, 48,* 157-168.

Motter, B. C. (1994a). Neural correlates of attentive selection for color or luminance in extrastriate area V4. *Journal of Neuroscience, 14,* 2178-2189.

Motter, B. C. (1994b). Neural correlates of feature selective memory and pop-out in extrastriate area V4. *Journal of Neuroscience, 14,* 2190-2199.

Mounts, J. R. W. (2000). Attentional capture by abrupt onsets and feature singletons produces inhibitory surrounds. *Perception and Psychophysics, 62,* 1485-1493.

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

Müller, N. G., Mollenhauer, M., & Rösler, A. (2005). The attentional field has a Mexican hat distribution. *Vision Research, 45,* 1129-1137.

Nakayama, K., & Silverman, G. (1986). Serial and parallel processing of visual feature conjunctions. *Nature, 320,* 264-265.

Nothdurft, H. C., Gallant, J. L., & Van Essen, D. C. (1999). Response modulation by texture surround in primate area V1: Correlates of "popout" under anesthesia. *Visual Neuroscience, 16,* 15-34.

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

Ohman, A., Lundqvist, D., & Esteves, F. (2001). The face in the crowd revisited: A threat advantage with schematic stimuli. *Journal of Personality and Social Psychology, 80,* 381-396.

Parkhurst, D., Law, K., & Niebur, E. (2002). Modeling the role of salience in the allocation of overt visual attention. *Vision Research, 42,* 107-123.

Pashler, H. (1987). Detecting conjunctions of color and form: Reassessing the serial search hypothesis. *Perception and Psychophysics, 41,* 191-201.

Raffone, A., & Wolters, G. (2001). A cortical mechanism for binding in visual working memory. *Journal of Cognitive Neuroscience, 13,* 766-785.

Ramachandran, V. S. (1988). Perception of shape from shading. *Nature, 331,* 163-166.

Rensink, R. A., & Enns, J. T. (1998). Early completion of occluded objects. *Vision Research, 38,* 2489-2505.

Reynolds, J. H., & Chelazzi, L. (2004). Attentional modulation of visual processing. *Annual Review of Neuroscience, 27,* 611-647.

Riesenhuber, M., & Poggio, T. (2000). Models of object recognition. *Nature Neuroscience, 3,* 1199-1204.

Rizzolatti, G., Riggio, L., Dascola, I., & Umilta, C. (1987). Reorienting attention across the horizontal and vertical meridians: Evidence in favor of a premotor theory of attention. *Neuropsychologia, 25,* 31- 40.

Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology, 8*, 382-439.

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

Schall, J. D., Hanes, D. P., Thompson, K., & King, D. J. (1995). Saccade target selection in frontal eye field of macaque I: Visual and premovement activation. *Journal of Neuroscience, 15,* 6905-6918.

Schall, J. D., & Thompson, K. G. (1999). Neural selection and control of visually guided eye movements. *Annual Review of Neuroscience, 22,* 241-259.

Shen, J., & Reingold, E. M. (2001). Visual search asymmetry: The influence of stimulus familiarity and low-level features. *Perception and Psychophysics, 63,* 464-475.

Shipp, S. (2004). The brain circuitry of attention. *Trends in Cognitive Sciences, 8,* 223-230.

Sigman, M., & Gilbert, C. D. (2000). Learning to find a shape. *Nature Neuroscience, 3,* 264-269.

Sobel, K. V., & Cave, K. R. (2002). Roles of saliency in conjunction search. *Journal of Experimental Psychology: Human Perception and Performance, 28,* 1055-1070.

Spivey, M. J., & Spirn, M. J. (2000). Selective visual attention modulates the direct tilt effect. *Perception and Psychophysics, 62,* 1525-1533.

Tanaka, K. (1996). Representation of visual features of objects in the inferotemporal cortex. *Neural Networks, 9,* 1459-1475.

Theeuwes, J. (1991). Cross-dimensional perceptual selectivity. *Perception and Psychophysics, 50,* 184-193.

Theeuwes, J. (1992). Perceptual selectivity for color and form. *Perception and Psychophysics, 51,* 599-606.

Theeuwes, J., & Godijn, R. (2001). Attentional and oculomotor capture. In C. L. Folk & B. S. Gibson (Eds.), *Attraction, distraction, and action: Multiple perspectives on attentional capture* (pp. 121-150). Amsterdam: Elsevier.

Theeuwes, J., Reimann, B., & Mortier, K. (2006). Visual search for featural singletons: No top-down modulation, only bottom-up priming. *Visual Cognition, 14,* 466-489.

Thompson, K. G., Bichot, N. P., & Schall, J. D. (1997). Dissociation of target selection from saccade planning in macaque frontal eye field. *Journal of Neurophysiology, 77,* 1046-1050.

Thompson, K. G., Hanes, D. P., Bichot, N. P., & Schall, J. D. (1996). Perceptual and motor processing stages identified in the activity of macaque frontal eye field neurons during visual search. *Journal of Neurophysiology, 76,* 4040-4055.

Tong, F., & Nakayama, K. (1999). Robust representations for faces: Evidence from visual search. *Journal of Experimental Psychology: Human Perception and Performance, 25,* 1016-1035.

Treisman, A. (1982). Perceptual grouping and attention in visual search for features and for objects. *Journal of Experimental Psychology: Human Perception and Performance, 8,* 194-214.

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

Treisman, A., & Sato, S. (1990). Conjunction search revisited. *Journal of Experimental Psychology: Human Perception and Performance, 16,* 459-478.

Treisman, A., Vieira, A., & Hayes, A. (1992). Automaticity and preattentive processing. *American Journal of Psychology, 105,* 341-362.

Treisman, A. M., & Schmidt, H. (1982). Illusory conjunctions in the perception of objects. *Cognitive Psychology, 14,* 107-141.

Treue, S. (2001). Neural correlates of attention in primate visual attention. *Trends in Cognitive Sciences, 24,* 295-300.

Tsotsos, J. K., Culhane, S. M., Wai, W., Lai, Y., Davis, N., & Nuflo, F. (1995). Modeling visual attention via selective tuning. *Artificial Intelligence, 78,* 507-545.

Turatto, M., Galfano, G., Gardini, S., & Mascetti, G. G. (2004). Stimulus-driven attentional capture: An empirical comparison of display-size and distance methods. *The Quarterly Journal of Experimental Psychology, 57A,* 297-324.

Usher, E., & Niebur, M. (1996). Modeling the temporal dynamics of IT neurons in visual search: A mechanism for top-down selective attention. *Journal of Cognitive Neuroscience, 8,* 311-327.

Van der Velde, F., & De Kamps, M. (2001). From knowing what to knowing where: Modeling object-based attention with feedback disinhibition of activation. *Journal of Cognitive Neuroscience, 13,* 479-491.

Van der Velde, F., & De Kamps, M. (2003). A model of visual working memory in PFC. *Neurocomputing, 52-54,* 419-424.

Van der Velde, F., & De Kamps, M. (2006). Neural blackboard architectures of combinatorial structures in cognition. *Behavioral and Brain Sciences, 29,* 37-108.

Van der Velde, F., De Kamps, M., & Van der Voort van der Kleij, G. T. (2004). CLAM: Closed-loop attention model for visual search. *Neurocomputing, 58-60,* 607-612.

Van der Velde, F., Van der Voort van der Kleij, G. T., Haazebroek, P., & De Kamps, M. (in preparation). The Global Saliency Model.

Van der Voort van der Kleij, G. T., De Kamps, M., & Van der Velde, F. (2003). A neural model of binding and capacity in visual working memory. *Lecture Notes in Computer Science, 2714,* 771-778.

Vogel, E. K., Woodman, G. F., & Luck, S. J. (2001). Storage of features, conjunctions, and objects in visual working memory. *Journal of Experimental Psychology: Human Perception and Performance, 27,* 92-114.

Wallis, G., & Rolls, E. T. (1997). A model of invariant object recognition in the visual system. *Progress in Neurobiology, 51,* 167-194.

Wang, Q., Cavanagh, P., & Green, M. (1994). Familiarity and pop-out in visual search. *Perception and Psychophysics, 56,* 495-500.

Wilson, F. A. W., Scalaidhe, S. P. O., & Goldman Rakic, P. S. (1993). Dissociation of object and spatial processing domains in primate prefrontal cortex. *Science, 260,* 1955-1957.

Wolfe, J. M. (1994). Guided Search 2.0: A revised model of visual search. *Psychonomic Bulletin and Review, 1,* 202-238.

Wolfe, J. M. (1998). What can 1 million trials tell us about visual search? *Psychological Science, 9,* 33-39.

Wolfe, J. M. (2003). Moving towards solutions to some enduring controversies in visual search. *Trends in Cognitive Sciences, 7,* 70-76.

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,* 483-502.

Wolfe, J. M., Cave, K. R., & Franzel, S. L. (1989). Guided search: An alternative to the feature integration model of visual search. *Journal of Experimental Psychology: Human Perception and Performance, 15,* 419-433.

Wolfe, J. M., Friedman-Hill, S. R., & Bilsky, A. B. (1994). Parallel processing of part-whole information in visual search tasks. *Perception and Psychophysics,* 55, 537-550.

Wolfe, J. M., Friedman-Hill, S. R., Stewart, M. I., & O'Connell, K. M. (1992). The role of categorization in visual search for orientation. *Journal of Experimental Psychology: Human Perception and Performance, 18,* 34-49.

Wolfe, J. M., & Horowitz, T. S. (2004). What attributes guide the deployment of visual attention and how do they do it? *Nature Reviews Neuroscience, 5,* 1-7.

Yantis, S., & Serences, J. T. (2003). Cortical mechanisms of space-based and object-based attentional control. *Current Opinion in Neurobiology, 13,* 187-193.

Zimba, L. D., & Hughes, H. C. (1987). Distractor-target interactions during directed visual attention. *Spatial Vision, 2,* 117-149.

Zohary, E., & Hochstein, S. (1989). How serial is serial processing in vision? *Perception, 18,* 191-200.

## Endnotes

[1] For all experiments, analyzing RTs with setsize as an additional within-subject variable yielded similar results, with RT × setsize interactions mirroring the effects of search slopes. For clarity and conciseness, we only report the interactions between setsize and other variables in the analyses of search slopes.

[2] An analysis that compared the performance directly after training with the performance two months after training revealed that both for the trained and the untrained search task (averaged over trained and untrained locations), the RTs and the search slopes were respectively slower and steeper two months after training than directly after training [RTs: trained search task, $F(1, 7) = 8.60$, $p = .022$; untrained search task, $F(1, 7) = 6.06$, $p = .043$; Search slopes: trained search task, $F(1, 7) = 7.41$, $p = .030$; untrained search task, $F(1, 7) = 7.25$, $p = .031$]. However, both for the trained and the untrained search task (averaged over trained and untrained locations), the error rate was lower two months after training than directly after training [Error rates: trained search task, $F(1, 7) = 20.34$, $p = .003$; untrained search task, $F(1, 7) = 19.17$, $p = .003$], obscuring a clear interpretation of these data.

[3] Although both for the trained and the untrained search task search became more efficient through training, it is not possible to define the exact scope of learning due to an accompanying increase in error rate.

[4] In neurophysiological studies investigating visual search, monkeys are often required to make a fast eye movement to the target. To distinguish between the neural activity related to the eye movement command and the neural activity related to the selection of the target, monkeys in some studies were trained to withhold the saccade to the target until a cue is presented (i.e., throughout a delay period). In other studies, monkeys were taught to make the saccade to the target as fast as possible. In those studies, a neuron's activity is taken to be related to either the eye movement command or the selection of the target, depending on whether the time at which the neuron discriminates the target is correlated or unrelated respectively to the saccade latency (Thompson et al., 1996).

[5] Following parallel, feature-based visual attention, spatial visual attention may select the enhanced representation of one or more search items in a serial manner for further processing (e.g., Bichot et al., 2005; Hamker, 2004).

[6] Nonetheless, if stimulus-driven visual attention results from bottom-up processing in combination with horizontal processing, we would expect that the neuronal target discrimination occurs faster than the timing of more than 100 ms after stimulus onset that is observed.

[7] The features in Itti and Koch's (2002) model are extracted from high-resolution photographs.

[8] In Hamker's (2004) model, the FEF is separated into a perceptual and a premotor map. The perceptual map receives input across all feature dimensions (e.g., color and orientation) from V4. In turn, the perceptual map gives excitatory input to the premotor map. As cells in the premotor map inhibit each other (i.e., there is surround inhibition), cells that receive more perceptual input than inhibitory input become more highly activated, and cells that receive less perceptual input than inhibitory input become more highly activated. Fixation cells further regulate the level of activation in the premotor map. The activation of the premotor cells decreases with an increasing activation of the fixation cells.

[9] The model of Wolfe (1994) is not a neural network model. Nonetheless, it is based on rules that implement a mechanism similar to within-feature competition.

[10] The results of our simulations do not depend on the use of inhibitory and excitatory external input per se, but only on their difference. Using varying levels of excitatory external input instead of inhibitory and excitatory external input yields qualitatively similar results. We start with inhibitory and excitatory external input, to explore a range of possible combinations (see Figure 7).

[11] The stimulus activity of a highly illuminant object is higher than the stimulus activity of a lowly illuminant object in the ventral pathway. As stimulus activity in the ventral retinotopic areas interacts with the feedback activity that is generated in AIT, we suppose that the selected activity in the ventral retinotopic areas is higher when a highly illuminant object is selected in AIT than when a lowly

illuminant object is selected in AIT. However, we suppose that the illuminance of objects that are not selected in AIT does not influence the level of activity of unselected objects in the ventral retinotopic areas.

[12] We also investigated gradients on other conditions in which distracters were presented. In none of these conditions there was any gradient when trials were separated into same-hemifield, different-hemifield or midline trials.

# Samenvatting

Het menselijke visuele systeem is gelimiteerd in de hoeveelheid visuele informatie die het op een bepaald moment kan verwerken. Onze omgeving projecteert een overdosis aan visuele informatie op onze ogen. Om hiermee om te kunnen gaan, selecteert ons visuele systeem telkens slechts een gedeelte van de beschikbare visuele informatie voor uitgebreide verwerking en verwerkt de rest van de informatie minder uitgebreid. Dit proces wordt *selectieve visuele aandacht* genoemd.
Bovengenoemde selectie van visuele informatie gebeurt niet alleen op basis van kennis, verwachtingen en doelen, maar ook onafhankelijk hiervan. Het eerste wordt *top-down visuele aandacht* genoemd, het tweede *stimulus-driven visuele aandacht*. Een vorm van stimulus-driven visuele aandacht is de automatische selectie van een object dat zich door een uniek kenmerk van andere objecten onderscheidt. Dit wordt in dit proefschrift *global saliency* genoemd.
Selectieve visuele aandacht wordt vaak bestudeerd in visuele zoektaken, waarin proefpersonen moeten zoeken naar een doelobject (*target*) tussen een aantal afleidende objecten (*distracters*). In deze zoektaken wordt het aantal afleidende objecten gevarieerd en wordt meestal de tijd gemeten, die nodig is om te bepalen of het doelobject wel of niet aanwezig is. Er wordt onderscheid gemaakt tussen *efficient zoeken* en *inefficient zoeken*. Bij efficient zoeken heeft het aantal afleidende objecten geen of nauwelijks invloed op de reactietijd, bij inefficient zoeken neemt de reactietijd toe met een toenemend aantal afleidende objecten.
Dit proefschrift onderzoekt de mechanismen van stimulus-driven en top-down visuele aandacht, aan de hand van zowel gedragsstudies (visuele zoektaken) als simulaties van hersengebieden betrokken bij de verwerking van visuele informatie. Hierbij wordt uitgegaan van een bestaand model voor top-down visuele aandacht, het *Closed-Loop Attention Model* (CLAM). In CLAM is top-down visuele aandacht in visuele zoektaken het resultaat van de interactie tussen het visuele werkgeheugen in de prefrontale cortex, objectherkenning in de ventrale route en spatiële selectie in de dorsale route. Dit model wordt in hoofdstuk 1 besproken.
Hoofdstuk 2 laat zien dat de architectuur van het visuele werkgeheugen in CLAM kan verklaren waarom het aantal objecten dat een mens kan onthouden in zijn werkgeheugen gelimiteerd is, terwijl er geen limiet is voor het aantal kenmerken van elk onthouden object. In CLAM wordt elk object dat moet worden onthouden,

abstract gerepresenteerd op een zogenaamd schoolbord (*blackboard*). Zo'n abstracte objectrepresentatie wordt gebruikt om alle kenmerken van het object te selecteren. Ruimtelijk bestaat er overlap tussen de objectrepresentaties op het schoolbord. De simulaties in dit hoofdstuk tonen aan dat de overlap toeneemt naarmate er meer objecten onthouden moeten worden, waardoor het moeilijker wordt om een bepaalde objectrepresentatie en de bijbehorende kenmerken te selecteren.

Hoofdstuk 3 onderzoekt met behulp van simulaties een proces waarmee locatie-invariante objectherkenning tot stand kan komen, zonder dat het object op alle mogelijke locaties geleerd hoeft te worden. In het voorgestelde proces wordt het leren van objectkenmerken als volgt opgebouwd. Eerst worden simpele kenmerken geleerd op alle mogelijke locaties. Hierdoor wordt herkenning van deze kenmerken locatie-invariant. Vervolgens wordt geleerd om objecten te herkennen, gedeeltelijk door het leren van nieuwe conjuncties van deze locatie-invariante kenmerken. Hoewel objecten hierdoor op nieuwe locaties inderdaad herkend werden, bleek de kennis onvoldoende om een object op een nieuwe locatie tussen andere objecten te selecteren. Wij concluderen dat hiervoor ook locatie-afhankelijke kenmerken geleerd moeten worden.

Hoofdstuk 4 onderzoekt de relatie tussen de bekendheid van objecten en de zoekefficientie. De bekendheid van de objecten werd getraind in een identificatietaak en de zoekefficientie in een zoektaak. De digitale 2 en digitale 5 werden als objecten gebruikt. De resultaten tonen aan dat de zoekefficientie toeneemt als zowel doelobject als afleidend object individueel bekender worden (en niet berust op een verschil in bekendheid tussen doelobject en afleidend object), maar dat het leren van de afleidende objecten als een groep in een visuele zoektaak de zoekefficientie nog verder verhoogt, zonder dat de bekendheid van het afleidende object hierdoor toeneemt ten opzichte van de bekendheid van het doelobject. De toename in zoekefficientie beperkte zich niet tot getrainde locaties, maar generaliseerde aanzienlijk naar nieuwe locaties. Bovendien was het effect van training ook twee maanden later nog zichtbaar in de resultaten.

Hoofdstuk 5 onderzoekt of de global saliency van objecten gradueel toeneemt naarmate een kleiner aantal van de objecten op een display dezelfde kleur heeft. Uit de resultaten blijkt dit inderdaad zo te zijn. Daarnaast wordt de interactie onderzocht van deze graduele global saliency met top-down visuele aandacht voor kleur. De resultaten laten zien dat top-down visuele aandacht helpt bij het

selecteren van de locatie van het doelobject, zelfs wanneer deze locatie al global salient is.

Hoofdstuk 6 bespreekt bevindingen uit verscheidene gedragsstudies en neurofysiologische studies en modellen van visueel zoeken. Op basis van de bevindingen van de onderzochte gedragsstudies concluderen we dat global saliency niet alleen kan worden toegewezen aan de verwerking in lagere corticale gebieden. Uit de onderzochte neurofysiologische studies is de conclusie getrokken dat global saliency niet noodzakerlijkerwijs het resultaat is van enkel bottom-up en horizontale verwerking, maar ook het resultaat kan zijn van een combinatie van bottom-up, horizontale en top-down verwerking.

Hoofdstuk 7 presenteert het *Global Saliency Model* (GSM). Dit model stelt een mechanisme van global saliency voor en specificeert de interactie van dit mechanisme met de mechanismen van top-down visuele aandacht. Het model gaat uit van de hypothese dat global saliency het resultaat is van een interactie tusssen objectherkenning in de ventrale route en spatiële selectie in de dorsale route. Spatiële selectie in de dorsale route vindt plaats in een aantal interacterende spatiële kaarten (*maps*). In overeenstemming met de conclusies uit Hoofdstuk 6 is global saliency in GSM het resultaat van top-down verwerking in de ventrale route, naast bottom-up en horizontale verwerking (in de ventrale en dorsale routes). Simulaties tonen aan dat het model een aantal belangrijke bevindingen van visueel zoeken kan verklaren, zoals het efficient zoeken van een uniek object tussen afleidende objecten en de invloed op visueel zoeken van de gelijkenis tussen een doelobject en afleidende objecten en van afleidende objecten onderling. Bovendien kan GSM de resultaten van de gedragsstudies in Hoofdstuk 5 verklaren.

Hoofdstuk 8 onderzoekt of een *inhiberende annulus* om een object waarop de aandacht gericht is (de *focus of attention*) het gevolg is van inhibitie tussen objecten nog voordat er een object geselecteerd wordt, die sterker is wanneer objecten bepaalde kenmerken met elkaar delen dan wanneer zij dit niet doen. In een visuele zoektaak werd de gelijkenis tussen één doelobject en één afleidend object gemanipuleerd temidden van andere objecten. Het bleek niet uit te maken of het doelobject en het afleidend object al dan niet dezelfde kleur hadden. Er was überhaupt geen evidentie voor een inhiberende annulus om het afleidende object als het doelobject ook salient was, hoewel het saliente afleidende object wel aandacht trok. In een ander experiment werd verder gevonden dat een spatiële aanwijzing die de aandacht trok juist de verwerking van objecten faciliteerde,

evenredig met hun afstand tot de spatiële aanwijzing. In lijn met GSM suggereren we dat saliente objecten elkaar inhiberen door middel van laterale inhibitie voordat er een object geselecteerd wordt, onafhankelijk van het feit of ze wel of niet kenmerken delen.

# Dankwoord

Het is al vijf jaar geleden dat ik vol goede moed aan mijn promotie-onderzoek begon. Zonder koffie was dit boekje waarschijnlijk nooit tot stand gekomen, maar nog belangrijker waren de inspiratie, de afleiding en de steun van mensen om mij heen. Hiervoor wil ik al mijn collega's van de sectie Cognitieve Psychologie bedanken. Een paar mensen wil ik expliciet noemen. Jan-Rouke, Michiel en andere aio's, bedankt voor de vele gezellige lunches, koffie- en Swirlpauzes en andere time-outs. Jiska, ons plan om een bar te maken in onze kamer is weliswaar nooit gerealiseerd, maar het was vaak gezellig op 2A49. Albertien, fantastisch zoals je altijd, maar dan ook écht altijd voor me klaar stond.

Gracias, twee geweldige studenten die hun scriptie-onderzoek bij mij hebben gedaan: Roel ter Winsen en Pascal Haazebroek. Ik heb het erg leuk gevonden om met jullie samen te werken. Pascal, de uren schoten voorbij als we enthousiast aan het praten waren, maar toch had ik er geen uur van willen missen. Dit laatste geldt ook voor jou, Joost. Ik vond het heel interessant om ideeën uit te wisselen vanuit onze verschillende onderzoeksgebieden, naast alle andere onderwerpen die we aansneden:)

Natuurlijk wil ik iedereen bedanken die bij het overkoepelende NWO-project betrokken was. De meetings die we hadden waren niet alleen wetenschappelijk interessant, maar ook zeer gezellig. Martijn, bedankt voor de coördinatie van het NWO-project, de goede samenwerking en de waardevolle feedback en adviezen die je me gegeven hebt. Judith, ik ben ontzettend blij dat jij ook een aio was binnen het NWO-project. Onze afspraken en de lange emails waarin we elkaar alle onderzoekservaringen vertelden, positief én negatief, waren voor mij een super steun, evenals het witbier op terrasjes en andere non-work ontmoetingen.

Further I would like to thank all organizers and teachers of "The European Diploma in Cognitive and Brain Sciences". The four "summerschools" in fall and spring were not only inspiring for research, but were also ideal breaks of daily PhD-life thanks to my wonderful co-students. Despite the fact that I was physically exhausted after each 10-day work- and partymarathon, mentally I always felt uplifted! Laura, Laura and Jane, our weekends in Dublin and Tuscany were grand. Laura McAvinue, you're fabulous.

Alan, you are probably the only one who filled my inbox faster than the very interesting emails I received about upcoming colloquia, conferences and other

research-related stuff. Thank you for sharing your inspiring, disturbing, funny and interesting perspective, and of course your music.
In de laatste jaren van mijn promotietijd nam mijn behoefte aan koffie- en theepauzes significant toe;) Het was dan ook een mooi toeval dat ik halverwege mijn promotie een paar aio's van de sectie Sociale Psychologie heb leren kennen. Sezgin en Krispijn, ik waardeer onze koffie-breaks en wandelingen zeer, ondanks de terugkerende dreiging dat jij, Sezgin, me het brugje af zou duwen:) Sezgin, behalve het krijgen van frisse lucht, heb ik talloze keren mijn hart bij je kunnen luchten. Ik ben je daar zeer dankbaar voor en voor de vriendschap die we tijdens onze promotie hebben opgebouwd.
Rosien, Nadir, Mar, Bas, Judie, Tijmen (merci voor je opmaak design), Arnvid en andere vrienden, zoals de film "It's a wonderful life" duidelijk maakt, is elk mens geslaagd zolang hij/zij vrienden heeft. Lieve Friends, bedankt voor alle mooie momenten in de afgelopen vijf jaar die mijn leven zo veel rijker maakten dan onderzoek alleen.
Lieve John en Lia, de aanlokkelijke locaties van mijn conferenties zorgden er vaak voor dat we niet naar Italie kwamen:) maar grazie dat jullie altijd geïnteresseerd bleven in de voortgang en mij altijd met open armen ontvingen.
Lieve Emer, Mark, Ardan en Petra, dank voor de uren die jullie mij over mijn proefschrift hebben willen aanhoren. Ardan en Emer, ik ben ontzettend blij met jullie als broer en zus! Dimf en John, heel erg bedankt voor jullie liefde en oprechte interesse! John, ik waardeer het dat je soms zo betrokken was dat je zelfs gerelateerde artikelen naar me doorstuurde, ook al werd mijn stapel "te lezen papers" hierdoor nog hoger:)
Olwen en Thierry, onze lange avonden en nachten bij De Mexicaan en thuis waren het meest effectief om mijn onderzoek te relativeren en werkten als een levenselixer. Olwen, je hebt me zo vaak moed in gepraat in de afgelopen jaren. Je bent écht een supersis. Thierry, jouw living for the moment spirit werkte bij mij zeer aanstekelijk. I love u2. Jullie zijn vraiment superbe.
Vincent, ik weet eigenlijk niet waar ik moet beginnen met jou te bedanken. De best te verdedigen stelling luidt zonder twijfel "Zonder jou, Vincent, zou dit proefschrift er nooit gekomen zijn". Je hebt me niet alleen tegengehouden toen ik op het punt stond om mijn gehele folder "Proefschrift" te deleten:) maar je hebt ook mijn papers en mijn gehele proefschrift (twee keer) gelezen en van constructief commentaar voorzien. Ik kan me dan ook geen betere paranimf voorstellen. Lieve lieve Vins, de afgelopen vijf jaar (eigenlijk tien jaar) ben je er elke

dag voor mij geweest. ♫ You're the one, you're the one, you're the one for me ♫ Je t'aime, infinitely.

## Curriculum Vitae

Gwendid van der Voort van der Kleij was born in Leiderdorp, the Netherlands, on 20 June 1977. She attended Gymnasium at the Bonaventura College in Leiden from 1989 to 1995. After completing her secondary education, she chose to study Psychology at Leiden University. As part of her undergraduate studies in Cognitive Psychology, she did a research project in the Dynamics of Adaptive Behavior Research Group at the Case Western Reserve University in Cleveland, U.S.A. In 2001 she graduated *cum laude*. She was first a PhD student at the Université Libre de Bruxelles in Brussels, Belgium, and then a researcher at the Nyfer Forum for Economic Research in Breukelen, before she became a PhD student in the Cognitive Psychology Unit of Leiden University in 2002.