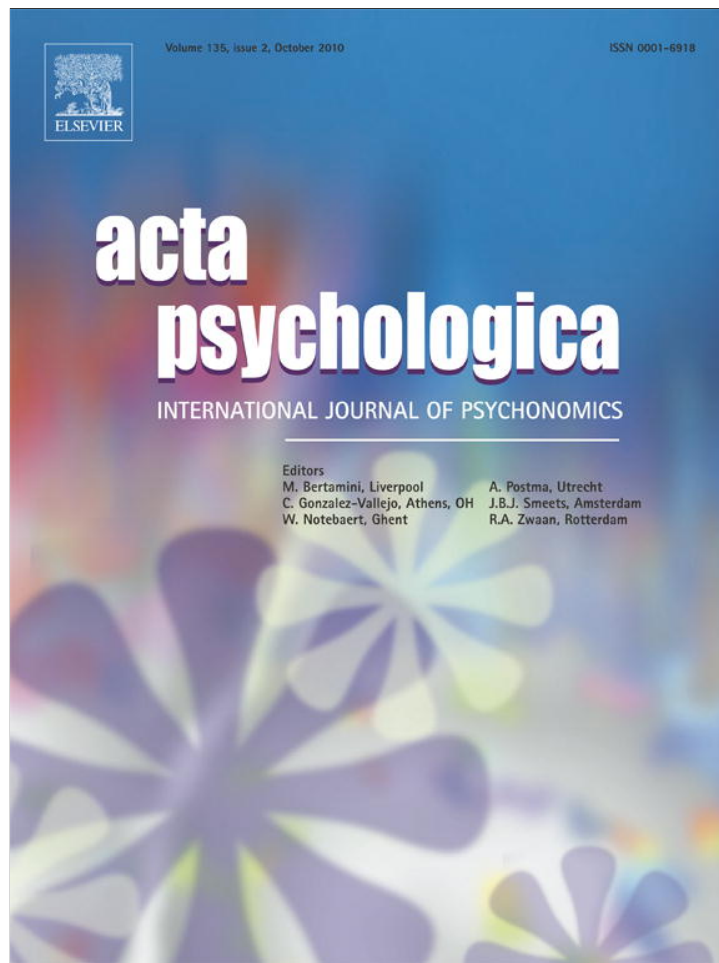


Provided for non-commercial research and education use.
Not for reproduction, distribution or commercial use.



This article appeared in a journal published by Elsevier. The attached copy is furnished to the author for internal non-commercial research and education use, including for instruction at the authors institution and sharing with colleagues.

Other uses, including reproduction and distribution, or selling or licensing copies, or posting to personal, institutional or third party websites are prohibited.

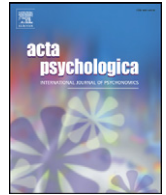
In most cases authors are permitted to post their version of the article (e.g. in Word or Tex form) to their personal website or institutional repository. Authors requiring further information regarding Elsevier's archiving and manuscript policies are encouraged to visit:

<http://www.elsevier.com/copyright>



Contents lists available at ScienceDirect

Acta Psychologica

journal homepage: www.elsevier.com/locate/actpsy

Target Article

Top-down and bottom-up control of visual selection

Jan Theeuwes*

Dept. of Cognitive Psychology, Vrije Universiteit, Van der Boechorststraat 11081 BT Amsterdam, The Netherlands

ARTICLE INFO

Article history:

Received 14 December 2009
 Received in revised form 13 February 2010
 Accepted 16 February 2010
 Available online 26 May 2010

PsycINFO classification code:
 2346

Keywords:

Attention
 Top-down control
 Bottom-up control

ABSTRACT

The present paper argues for the notion that when attention is spread across the visual field in the first sweep of information through the brain visual selection is completely stimulus-driven. Only later in time, through recurrent feedback processing, volitional control based on expectancy and goal set will bias visual selection in a top-down manner. Here we review behavioral evidence as well as evidence from ERP, fMRI, TMS and single cell recording consistent with stimulus-driven selection. Alternative viewpoints that assume a large role for top-down processing are discussed. It is argued that in most cases evidence supporting top-down control on visual selection in fact demonstrates top-down control on processes occurring later in time, following initial selection. We conclude that top-down knowledge regarding non-spatial features of the objects cannot alter the initial selection priority. Only by adjusting the size of the attentional window, the initial sweep of information through the brain may be altered in a top-down way.

© 2010 Elsevier B.V. All rights reserved.

1. Introduction

In everyday life, we constantly look around and use our visual input to guide our behavior. We may search for our favorite coffee in the supermarket; attend to curveball when the pitcher throws the baseball or search for the specific color of the sweater of our child when picking her up at the daycare. When searching for particular objects, we may sometimes experience that we attend to things in our environment for which we had no intention to look for. We may inadvertently attend to the road worker wearing a fluorescent orange safety jacket, a moving billboard along the roadside, or a waving hand in the crowd.

Over the last 25 years a considerable debate emerged regarding the extent to which selection is controlled by us in a voluntary, top-down way or by the properties of the stimulus features in the environment in a automatic, bottom-up way (e.g., see reviews Burnham, 2007; Corbetta & Shulman, 2002; Rauschenberger, 2003; Theeuwes, 2005; Theeuwes & Belopolsky, 2010; Theeuwes & Godijn, 2001; Theeuwes et al., in press). On the one hand it has been argued that the allocation of attention is initiated *voluntarily* in line with the behavioral goals (Folk, Remington, & Johnston, 1992). Alternatively, it has been claimed that events that are salient enough are selected regardless of the current top-down set (Theeuwes, 1991a, 1992, 1994b; Theeuwes, 1995c; Theeuwes, 2004). In the latter case, when objects or events receive priority of processing independent of the volitional goals of the observer one refers to this as *attentional capture*

(Theeuwes, 1992). When such an object or event not only captures attention but also triggers an exogenous saccade to the location of the event, this is referred to as *oculomotor capture* (Theeuwes et al., 1998, 1999). Before discussing top-down and bottom-up selection, it is important to define what is understood with these terms. When we speak about top-down selection we imply that selection is completely under control of the intentions of the observer. In other words, selection is completely volitional: at any time, a person can choose *at will* from the environment what to select. An experimental procedure which elegantly illustrates top-down selection is the so-called endogenous cueing procedure first described by Posner (1980). On each trial, observers receive a cue telling them the likely location of the upcoming target. Such a cue may consist of a centrally presented arrow that points to the likely target location (Posner, 1980) or a digit that tells at which location the target will appear (Theeuwes & Van der Burg, 2007). The typical finding is that observers are faster and make fewer errors when the target appears at the cued location relative to when it appears at an uncued location. The crucial point for the present discussion is each trial observers direct their attention *at will* to a particular location in space. This represents a clear example of top-down selection.

We speak of bottom-up selection when selection is determined by the feature properties present in the environment. While top-down selection is an active volitional process, bottom-up selection occurs in a passive automatic way. To demonstrate bottom-up selection, one has to ensure that the bottom-up selection is pitted against top-down selection. For example, when instructed to search for a red item among green items one cannot determine whether the selection of the red item is the result of top-down or bottom-up control. The salient feature that pops-out from the display is also the one element

* Tel.: +31 20 444 8790; fax: +31 20 444 8832.
 E-mail address: J.Theeuwes@psy.vu.nl.

people are looking for. To demonstrate true bottom-up selection, one should show that the red item is selected even though observers actively try to select something else, for example a green diamond (see for example Theeuwes, 1992).

Traditionally, bottom-up selection is associated with salience which is computed on the basis of the detection of locations whose local visual attributes significantly differ from the surrounding image attributes, along some dimension or combination of dimensions (Itti & Koch, 2001). This approach dates back to the Feature Integration Theory (Treisman & Gelade, 1980) which claimed that a number of simple visual feature dimensions are represented in the early stages of cortical visual processing such as color, edge orientation, luminance, or motion direction. Even though salience plays an important role in bottom-up selection, we also consider selection to be bottom-up when selection is driven – against the intentions of the observers – by other factors such as the emotional content of stimuli (e.g., angry faces) or previous experience. For example, visual selection that is driven by what observers experienced during the previous trial, i.e., inter-trial priming is often considered to be bottom-up (e.g., Pinto, Olivers, & Theeuwes, 2005). In these experiments, observers try to search for one object, but keep selecting the object that was the target on the previous trial. Also, selection driven by the content of our working memory while observers search for something else (e.g., Olivers et al., 2006) is bottom-up. In these experiments, observers look for a particular target object while keeping another object in memory. The results show that an object kept in memory (while being completely irrelevant for the search task) captures attention much more often than objects that were not kept in memory.

It is obvious that selection is not always bottom-up. Ultimately, we are able to select in a voluntary way those objects and events that are relevant for our behavior. However, it remains a controversy whether, initially, before light from the environment hits the retina, top-down goals can affect the way visual selection takes place. For example, when looking for my child does knowledge that she is wearing a red sweater help the initial selection process? Does pre-knowledge that I am looking for red, affects the initial selection process such that as soon as light hits the retina, red objects are prioritized over objects having other colors?

In this paper we will argue that initially, after the moment light hits the retina, visual selection is completely driven by the properties of the stimulus field. Only later, through massive recurrent feedback processing, volitional control based on expectancy and goal set will bias visual selection in a top-down manner. The idea of early bottom-up processing followed by later top-down processing is consistent with various other views (e.g., Hochstein & Ahissar, 2002; Lamme & Roelfsema, 2000; VanRullen & Koch, 2003). More specifically, Lamme & Roelfsema (2000) distinguished the feedforward sweep of initial activity from later recurrent processing. In the feedforward sweep, neural activity generated by the visual display sweeps from low-level visual areas up to temporal, parietal and frontal areas in a feedforward manner. In subsequent recurrent processing, the higher-order areas influence firing in lower visual areas through feedback connections. We argue that the initial selection is based on feedforward processing on the basis of bottom-up salience signals. By means of recurrent processing, top-down influences can bias attentional selection towards task relevant stimuli.

2. What is visual selection?

Imagine a situation in which the visual system is confronted with two different objects located at different positions in the visual field (see Fig. 1). Within the visual system, these two objects compete and question is which object wins this competition and drives neurons throughout the visual system forming an ensemble of neurons that represent this one single object. In line with the biased competition model (Desimone & Duncan, 1995) attention biases these competitive

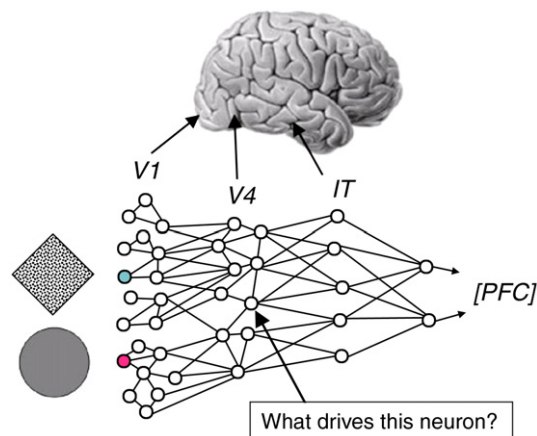


Fig. 1. A schematic drawing of the visual system. In this example, two different objects at different locations compete for attention. The question is which of the two salient objects will win the competition and drives the neuron. Bottom-up and top-down activity can resolve this competition (adapted from Serences & Yantis, 2006).

interactions such that attended stimuli receive priority over unattended stimuli. Attentional effects on resolving this competition are the result of bottom-up and top-down factors. The bottom-up signal depends on the (acquired) properties¹ of the stimulus field. Objects that are highly salient and stand out from the background may immediately receive attention priority. Indeed, it is likely that before top-down influences can have an effect, the visual system is biased towards salient stimuli that resolve the competition simply on the basis of the bottom-up input (see e.g., Hickey & Theeuwes, under review; Mathôt et al., in press; Reynolds & Chelazzi, 2004; van Zoest, Donk, & Theeuwes, 2004).

Another way to bias the competition within the visual system is through top-down volitional feedback signals that depend on the goals, intentions and expectations of the observer. For example, directing attention in a volitional way to a specific location in space increases the sensory gain for features at that location (e.g., Theeuwes & Van der Burg, 2007) and appears to alter the apparent stimulus contrast (e.g., Carrasco et al., 2004). These results imply that the directing attention to a location results in a greater neuronal sensitivity (i.e., a decreased threshold). This type of selection is endogenous and is often referred to as goal-driven selection.

As a metaphor visual attention has been compared to a spotlight that “selects” parts of the visual world around us (e.g., Posner, 1980). Visual attention allows people to select information that is relevant for their ongoing behavior. For several decades, there has been agreement that there are two functionally independent stages of visual processing (e.g., Broadbent, 1958; Neisser, 1967; Treisman & Gelade, 1980): An early visual stage, sometimes referred to as preattentive operating in parallel across the visual field and a later stage often referred to as attentive that can deal with only one (or a few items) at the same time. Even though the dichotomy between these two stages is not as strict as originally assumed, in almost all past and present theories of visual attention this basic architecture is more or less still present (e.g., Itti & Koch, 2001; Li, 2002; Treue, 2003; Wolfe, 1994). Given the two-stage framework, it is generally assumed that visual selection depends principally on the outcome of the early stage of visual processing. Processing occurring during the initial wave of stimulation through the brain, determines which element is selected and is passed on to the second stage of processing. In line with the two-stage approach, passing on an item to the second stage

¹ Bottom-up signals can also be acquired through intertrial priming, conditioning or reward contingencies. We consider these effects to be bottom-up because the effects cannot be counteracted by volitional top-down control.

of processing implies that this item has been selected for further processing (e.g., Broadbent, 1958; Treisman & Gelade, 1980).

According to this notion, from all objects that are present in the visual field (and are available at the early preattentive stage of processing), each time an object is passed on the final stage of processing, it will affect decision making and responding (see Fig. 2). This passing on from the initial stage of preattentive processing to attentive processing is what is considered to be visual selection. As is clear from Fig. 2, the selection of one object out of many objects that are available during the initial preattentive stage of parallel processing is what we consider visual selection. Note that in some conditions the preattentive, parallel stage of processing plays basically no role. In that case there is no salience calculation across the visual field and an object is selected purely on the basis of spatial information. For example, in case of endogenous cueing in which observers direct their attention to a location in space before the target is presented (cf. Posner et al., 1980) the salience of the other elements in the visual field hardly plays a role. Indeed, by endogenously directing attention to a location in space, observers even ignore irrelevant abrupt onsets which are known to be very salient (Theeuwes, 1991b; Yantis & Jonides, 1990). Also, when search is serial (or partly serial) preattentive, processing plays no or only a minor role because due to the serial nature of the task, attention is focused on a restricted spatial area thereby circumventing preattentive processing outside that area (Belopolsky et al., 2007).

Our viewpoint on what visual selection entails is derived from the classic theories on attention such as those of Broadbent (1958), Neisser (1967) and Treisman and Gelade (1980). We assume that selection ultimately takes place by space. Some theories suggest that selection does not have to be spatial. According to these theories, all features are equal; in order to select information, a target template (representing its location, its shape, or its color) has to be matched to the stimulus representation. Weights are assigned proportionally to the degree of the match (e.g., Bundesen, 1990; Bundesen, Habekost, & Kyllingsbaek, 2005; Duncan & Humphreys, 1989): the higher the weight, the higher the probability that the stimulus is selected for further processing. Even though it is undisputed that location information improves the detection of objects (e.g., Cave & Pashler, 1995; Kim & Cave, 1995; LaBerge, 1995; Lamy & Tsai, 2001; Posner, 1980; Theeuwes, 1989), some studies seem to suggest that non-spatial features may improve the entry of information into the brain without accessing location. For example, the classic study from von Wright (1970) showed efficient selection in a partial report task on the basis of simple attributes, such as color, luminance, and shape (see also, e.g., Brouwer & vanderHeijden, 1997; Bundesen, Pedersen, & Larsen, 1984). These findings have been interpreted as evidence that

non-spatial features can be used to select information. However, it should be noted that these findings do not necessarily indicate that non-spatial information is directly used to select information (as, for example, assumed by Bundesen's, 1990, theory of visual attention). The non-spatial information points to a location in the display (similar to a bar-marker indicating a location); and, ultimately, location is used as a means to select the relevant item (see also, Tsai & Lavie, 1988 but see vanderHeijden, Kurvink, deLange, deLeeuw, & vanderGeest, 1996). For example, in Moore and Egeth (1998) participants were required to detect a digit in an array consisting of letters. Targets and nontargets were either green or blue, and the participants were informed about the probability of the target being in one of these colors. Even though it appeared that selection was done purely on the basis of color, after running proper control experiments, it turned out that color was only used to guide attention to specific locations, confirming the idea that location information is used as a basis for selection. From a physiological point of view there is also every reason to assume that spatial information is special. A hallmark feature of vertebrate brain organization is that it is ordered topographically, in which sets of neuronal connections preserve the relative organization of space.

The idea that attention to a location in space is the mechanism by which selection takes place also applies to conditions of visual search that involves the detection of a simple feature. In a recent study Theeuwes, Van der Burg and Belopolsky (2008) showed that the presence of a color singleton involved a shift of spatial attention to the location of the singleton. Observers had to indicate whether a color singleton was present or not; there was no need to identify the target. The data indicated that even for such a simple response, focal attention needed to be shifted to the location of the singleton. These findings are inconsistent with claims from the main stream theories of visual attention such as feature integration theory (FIT, Treisman & Gelade, 1980) and the more modern version of FIT (Müller, Heller, & Ziegler, 1995; Müller, Reimann, & Krümmenacher, 2003; Wolfe, 1994), which assume that when observers need to detect a single feature singleton, they can check a pooled response from the relevant feature map for the presence of activity anywhere in that map. According to these theories detecting a pop-out target does not require the involvement of focal attention. Contrary to this assumption, the Theeuwes et al. (2008) study provides compelling evidence that suggests that spatial attention – the gateway of visual selection – is always necessary in order to respond to target. A response without directing spatial attention – that is without selection – is not possible.

3. Stimulus-driven selection

3.1. The theory

The basis for the theory for stimulus-driven selection was first proposed by Theeuwes (1991a, 1992) in the early 1990s. The basic premise is that when attention is divided across the visual field, preattentive processing is exclusively driven by the bottom-up properties of the stimulus field. For example, take Fig. 1 which represents a cartoon illustration of how selection may take place. After the initial preattentive analysis of a scene, one object is selected for further processing. We assume that this initial selection is completely based on (acquired) bottom-up factors such as salience. With selection – which is assumed to be spatial in nature – we imply that initially, spatial attention is shifted to the location in space containing the highest salience. This automatic shift of attention which is not the result of any top-down set on part of the observer is what is known as “attentional capture”.

As noted, the shift of spatial attention implies that the object present at the location to which attention is shifted is selected for further processing. We assume that only after the object has been

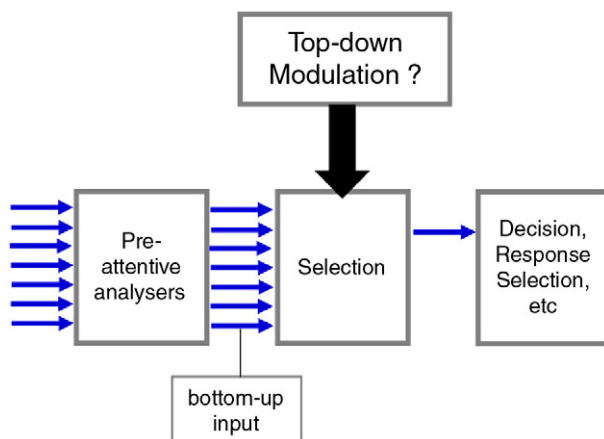


Fig. 2. A schematic drawing of the classic two-stage model of visual selection. When an object is passed on from the preattentive to the attentive stage of processing the object is thought to be selected.

selected, its identity will become available. For example, after selecting a salient singleton, it may become clear that the item is a color singleton, luminance singleton or a shape singleton. Preattentive analysis allows only the detection of locations whose local visual attributes significantly differ from the surrounding image attributes, along some dimension such as color, shape, luminance, size, etc. In other words, preattentive analyses can reveal that there are local feature differences (i.e., salient elements) but not from which feature dimensions these local differences come from. It is assumed that attention will shift in an automatic, exogenous fashion to the location having the highest local feature contrast or saliency (see also e.g., Donk & van Zoest, 2008; van Zoest et al., 2004). Only after attention has shifted to the location of the salient element, its identity becomes available. If the singleton is the target the observer is looking for, a response can be given. If it is not the target the observer is looking for, the location will be inhibited (Theeuwes & Godijn, 2004) and attention will shift in an exogenous fashion to the item that is next in line with respect to saliency.

We assume that the initial shift of attention to the most salient singleton, that is attentional capture, is the result of a bottom-up mechanism, which is triggered by the presence of feature difference signal interrupts. This notion is similar to that of Koch and Ullman (1985) who introduced the saliency map to accomplish preattentive selection (see also Theeuwes, 1994a). This map is a two dimensional map that encodes the saliency of objects in their visual environment. Neurons in this map compete among each other giving rise to a single winning location (cf. winner take all) that contains the most salient element. If this location is inhibited the next salient location will receive spatial attention. A computational version of this notion was developed by Itti and Koch (2001).

The stimulus-driven capture account recognizes several important assumptions. First, when attention is spread across the display, the calculation of a local feature difference occurs in a bottom-up fashion, which is not penetrable in a top-down, volitional way. Because the preattentive interrupt signal only can calculate differences in features, top-down knowledge (such as the fact that the observer is looking for a red singleton) cannot affect this preattentive processing stream. Once the most salient singleton is selected, its identity becomes available and then top-down knowledge (such as the fact that the observer is looking for a red target) will have an impact. If the automatically selected feature difference signal is the target singleton, a response can be given. If it is not the target singleton, top-down processing (that occurs after the item has been selected) allows a quick and fast disengagement of attention from the location having the highest saliency signal. Second, we assume that the initial feedforward sweep of information through the brain is basically stimulus-driven. The initial sweep of information provides a first preattentive analysis of the scene. There is massive recurrent top-down processing following this initial preattentive analysis. Once an item has been selected, top-down control determines whether attention resides at a particular location (engagement) or is quickly disengaged from that location. Third, spatial information which is considered to be under top-down control can affect the initial sweep of information through the brain. We assume that preattentive analysis only takes place within the attended area. By varying the attentional window from highly focused to widely distributed we are able to allow the extent to which preattentive analysis takes place (Belopolsky et al., 2007; Theeuwes, 2004; Theeuwes, Kramer, & Kingstone, 2004). Because preattentive analysis is limited to the attended area, the extent to which attention is spread also determines the occurrence of attentional capture. When attention is spread, visual search may be conducted in parallel across all items in the visual field, at the expense that any singleton relevant or irrelevant will be selected automatically. However, when the attentional window is set to a smaller size, singletons that fall outside of the attentional window will not capture attention. We propose that saliency computations are

restricted to the attentional window of the observer. Our claim is that while the size of the attentional window is under top-down control, within the attentional window top-down control cannot preclude attention from being captured by the most salient feature. It is important to realize that in situations in which search is serial or partly serial (as for example in conjunction search) the task requires that the attentional window is set to encompass a relatively small area which preclude saliency computations outside that area (Theeuwes, 2004).

It is important to note that theories that assume a very large role for top-down control over visual selection also incorporate a stage in which there seems to be basically no top-down control over selection. For example, the neural theory of visual attention (NTVA) of Bundesen et al. (2005) assumes that a perceptual cycle consists of two waves: first a wave of unselective processing followed by a wave of selective processing. According to NTVA, the outcome of the first wave is that attentional weights are computed for each object in the visual field representing their saliency. This wave of unselective processing can be considered to be the same as the initial feedforward sweep through the brain that we assume to result in attentional capture.

3.2. Evidence for stimulus-driven selection

This section discusses the most important data supporting the idea that the initial sweep of information through the brain is basically stimulus-driven. We discuss behavioral evidence as well as data from ERP, fMRI, TMS and single cell recording. All studies that are discussed in detail use variants of the additional singleton task as developed by Theeuwes (1991a, 1992).

3.2.1. Behavioral evidence

Given the theoretical framework above, experiments should reveal whether it is possible to affect the output of the preattentive stage of processing in a top-down way (see Fig. 2). As noted, it is already established that top-down set regarding location information by varying the attentional window can change the initial sweep of information through the brain (Munneke et al., 2008). However, the question is whether non-spatial top-down information (for example, top-down set for color, shape, luminance, etc.) can change the initial sweep of bottom-up activity through the brain. If one wants to answer this question one has to use a task in which top-down set for location information cannot play a role. In other words, one has to use a visual search task in which the location of the target is unknown. In addition, one has to choose a task which addresses top-down modulation on early (feedforward) vision excluding later top-down modulations arising from massive recurrent processing from extra-striate areas to primary visual areas. One such task is what is known as the feature singleton search task. In this task, the target is unique in a basic feature dimension (e.g., a red element surrounded by green elements) and therefore “pops-out” from the display. Pop-out detection tasks have been implicated to subservise the first stage of visual processing and single unit studies have implicated primary visual cortex in mediating bottom-up pop-out saliency computations (e.g., Nothdurft, Gallant, & Van Essen, 1999). Note, however, that even though the detection may occur at the preattentive stage, responding to the pop-out feature requires attention selection involving the second stage of attentive processing (Theeuwes et al., 2008).

It should be noted that visual search tasks that are relatively slow and inefficient do allow for top-down control on search. Because attention is focused on a restricted area preattentive processing is more or less circumvented. Moreover, because of the slower nature of the task there is enough time for recurrent top-down control to affect the efficiency of attentional deployment. Search in these tasks is basically serial or partially serial, not involving much, if any, preattentive processing. In this type of tasks it is not surprising that

top-down knowledge guides the deployment of attention (Wolfe, 1994). For example, Kaptein, Theeuwes and Van der Heijden (1995) showed that observers can restrict search for a color-orientation conjunction target to a color-defined subset. Thus, when searching for a red vertical line segment between red tilted and green vertical line segments, participants searched serially among the red items while they completely ignored the green line segments.

The crucial question is whether top-down knowledge can affect the initial selection of stimuli (see Fig. 2). As noted, we used a task which is known as the feature singleton search task. In this task, the target is unique in a basic feature dimension (e.g., a red element surrounded by green elements) and therefore “pops-out” from the display. The question is whether top-down knowledge modulates the salience of the “popping-out” target. In terms of neurophysiology, the question is whether neurons in early vision (e.g., primary visual cortex) can be tuned in a top-down way such that they are set for specific local features such as location, color and shape. Using a feature singleton task enables us to investigate top-down modulation on early (feedforward) vision thereby excluding later top-down modulations arising from massive recurrent processing from extrastriate areas to primary visual areas. If search is slow and effortful (as for example in a conjunction search task) there is ample time to have massive feedback from higher to lower brain areas, obscuring the bottom-up input and making it difficult to investigate initial attentional selection.

3.2.1.1. Additional singleton search task. To investigate the contribution of top-down and bottom-up control in visual selection, in the early nineties, Theeuwes (1991a, 1992) developed the so-called additional singleton task. In this visual search task, observers search for one specific clearly defined salient singleton while another irrelevant singleton is simultaneously present. Fig. 3 gives an example of a display. In the 1992 version of this task, observers always search consistently throughout the whole experiment for a green diamond singleton. In the distractor condition, one of the green circles was made red, constituting the color distractor singleton. The main finding of the additional singleton search task is that reaction time (RT) in the condition in which a unique color irrelevant distractor singleton was present (in this case the red circle) was higher than when such a distractor was not present (see Fig. 3). Importantly, an irrelevant

singleton only causes an RT increase when the distractor was more salient than the target. When the color distractor was made less salient (see Fig. 3; right panels), its presence did not affect search for the diamond target anymore. On the basis of these findings, Theeuwes (1991a, 1992, 1994a,b) postulated his notion of stimulus-driven capture, arguing that the bottom-up salience signal of the stimuli in the visual field determines the selection order. The increase in search time in conditions in which an irrelevant singleton was present was explained in terms of attentional capture. Because the irrelevant color singleton was selected exogenously (that is, captured spatial attention), it took more time before the target singleton could be selected and before a response could be emitted. Given the observation that selectivity completely depended on the relative salience of the target and the distractor singleton, it was argued early visual preattentive processing is only driven by bottom-up factors. Given Fig. 2 representing a cartoon-like drawing of visual processing stages, this implies that early on, during the first sweep of information through the brain, the competition between the two salient objects is resolved by bottom-up salience signals (Mathôt et al., in press). These signals drive the neurons at higher levels that create a perceptual coherence field, an ensemble of neurons that jointly represent the single selected object (Serences & Yantis, 2006). Top-down control over feature selection (i.e., knowing that the target is a diamond and not a red circle) cannot influence these early bottom-up salience signals.

The additional singleton task has particular features that make this task particularly suited for investigating top-down and bottom-up control of attention. First, the color distractor singleton is never the target so there is no reason for observers to attend to the distractor. If – against the intentions of the observer – this distractor singleton is still selected first, it provides strong evidence for stimulus-driven, bottom-up selection. Second, the top-down goal of the observers is very clear. Throughout the experiment observers consistently search for the same target (in the Theeuwes, 1992 version for a green diamond). In other words, there is no uncertainty for what observers should search for. Third, observers search for the green diamond target but respond to the orientation of the line segment inside the diamond shape. This is a so-called compound search task (Duncan, 1985) which makes it possible to disentangle factors affecting the selection of the target from those affecting response selection. In other

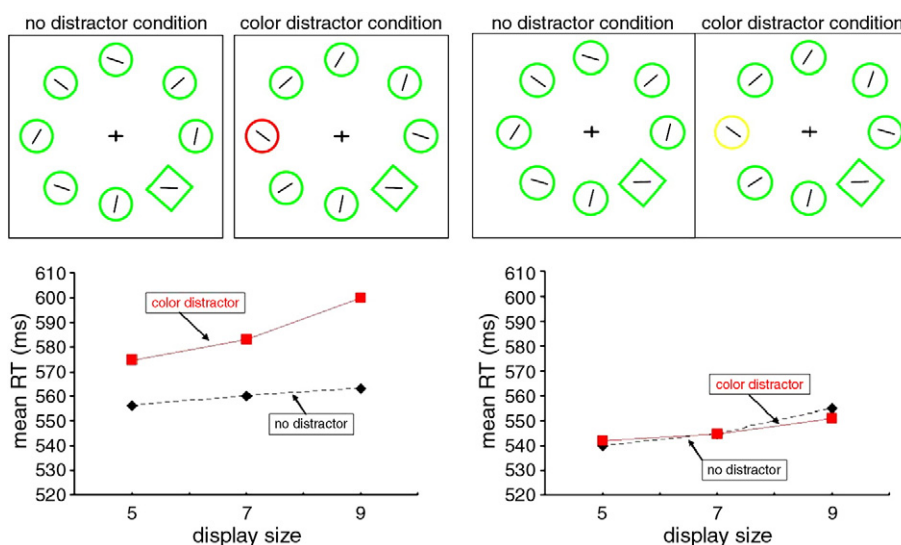


Fig. 3. Stimuli and data from Theeuwes (1992). Observers search throughout the whole experiment for a shape singleton, a green diamond presented among a variable number of circles. Observers respond to the orientation (horizontal or vertical) of the line segment presented within the target diamond shape. On the left side: the color distractor singleton captures attention and causes a Reaction Time (RT) increase because the color distractor is more salient than the target singleton (the green diamond). On the right side: Finding the shape singleton is not affected by the presence of the color singleton because the color singleton is in this condition less salient than the target singleton (the green diamond). These results indicate that even though observers always search for a diamond singleton, this top-down set cannot prevent the selection of the color singleton. Selection appears to be completely controlled by the salience of the stimuli in the visual field. This result is taken as evidence for bottom-up attentional capture.

words, observers search for one aspect (the diamond shape) but respond to another aspect (the line orientation inside the target). Because we employ a compound search task, the response requirements remain the same over the various conditions ensuring that the RT effects caused by the presence of the color distractor are due to perceptual interference and not to response interference. Fourth, the target and distractor singleton could appear at any location within the visual field ensuring that observers cannot focus spatial attention before display onset to a circumvented area. As noted, focusing attention prevents the capture of attention outside the attended area (Theeuwes, 1991b). Fifth, by varying the number of items in the display we can check whether search is indeed conducted in parallel across the visual display. As argued, pop-out tasks have been implicated to subservise the first stage of visual processing and single unit studies have shown the involvement of primary visual cortex in mediating bottom-up pop-out saliency computations. Sixth, the target and distractor singleton are simultaneously present. This is important because only when items are simultaneously present, there is competition (Desimone & Duncan, 1995) which may then reveal bottom-up and top-down attentional effects on resolving this competition. Mathôt et al. (in press) have shown that there is hardly any biased competition between a target and onset distractor singleton when they are presented sequentially. Also, Theeuwes, Atchley, and Kramer et al. (2000) have shown that the interference effect of a color distractor singleton (as in the Theeuwes, 1992 task) is no longer present when the color distractor is presented 150 ms before the target singleton. In other words, to reveal competition between top-down and bottom-up control target and distractor singletons have to be presented simultaneously.

Note that in the original additional singleton paradigm (Fig. 3) and in other versions of this paradigm (the singleton cueing paradigm, see

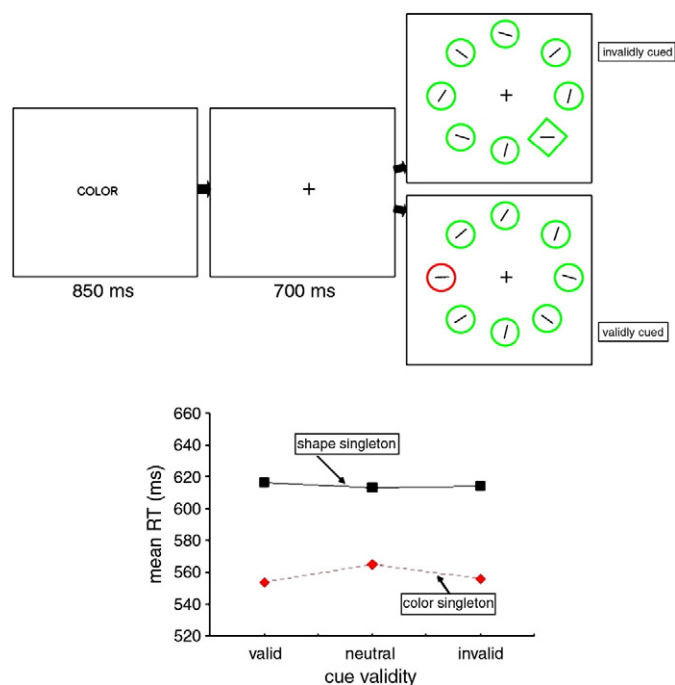


Fig. 4. Stimuli and data from Theeuwes et al. (2006). Observers responded to the orientation of the line segment inside the singleton. The singleton containing the target line segment was either a diamond (shape singleton) or a red circle (color singleton). The verbal cue indicated with 83% validity the target singleton dimension for the upcoming trial (the word cue “shape”, “color” or “neutral”). The RT data show that observers were faster to respond to the color singleton than to the shape singleton. Importantly however the validity of the cue had no effect on responding. This data indicates that non-spatial top-down knowledge about the upcoming target singleton has no effect of visual search for a singleton.

Fig. 4) the line segment inside the target singleton that observers responded to was either horizontal or vertical while the line segments in the nontarget elements were slightly tilted. It has been argued that this setup makes it possible for observers to directly search for the vertical or horizontal line segment while ignoring the circle and diamond shapes. Control experiments have shown that such a strategy is not possible. If observers would use such a strategy search becomes very slow and serial with search slopes of about 88 ms/item (see Experiments 1 and 2 Theeuwes, 1991a). Since the data (Fig. 3) indicate that in the additional singleton paradigm search is efficient and conducted in parallel (involving preattentive processing), it is clear that observers cannot use the strategy.

It is important to realize that in the Theeuwes (1992) experiment, the no distractor and distractor conditions were varied between blocks of trials. In other words, in the distractor condition, observers knew that on every trial a red distractor singleton would be present. Still even when observers knew that the red singleton was never the target they could not apply sufficient top-down control to prevent its selection. Also in Theeuwes (1992) we examined whether training would reduce the interference effect. Even though the overall RTs became significantly faster the interference effect caused by the irrelevant color distractor remained present even after 1800 trials of training. Thus even practice does not change the pattern of results, and cannot induce sufficient top-down control to overcome the interference caused by a salient distractor.

In the Theeuwes (1992) study, the target was always a green diamond among a variable number of green circles. In the distractor condition one of the green circles was red (see Fig. 3). In an earlier version of this task (Theeuwes, 1991a), we let observers also search for a unique shape but randomized among trials, the target shape singleton could be a diamond among circles or a circle among diamonds. So even though observers knew they had to look for a unique shape they did not know which shape it was. The same was true for the distractor conditions: the distractor was either a red circle among green circles or a green circle among red circles which was varied randomly within a block. The results of this task in which both the target and distractor singleton constantly changed roles were basically the same except that the interference effects were much larger. Indeed, in the Theeuwes (1992) task, interference effects were up to about 25 ms (see Fig. 3) while in the 1991 study, these effects were about 120 to 150 ms.

A study by Pinto et al. (2005) addressed the difference in interference between these two studies. Consistent with the earlier studies of Theeuwes (1991a, 1992) Pinto et al. (2005) showed that when observers searched for a shape, a color distractor singleton had a larger slowing effect in a mixed block, in which both the target shape could change from trial to trial, than in a pure block, in which the target shape remained the same. Importantly, however, this increased distractor singleton effect was entirely traced back to inter-trial priming, since the increased costs occurred only on trials in which the target and the distractor singleton swapped identity (Experiment 1) or on trials in which the target alone changed identity while the distractor singleton remained constant (Experiment 2). It was concluded that target uncertainty itself does not lead to strategic changes in the attentional selection of singletons. Instead, selection is driven by relatively automatic bottom-up priming mechanisms. Note that regardless of the distractor effects, RTs in the mixed condition were overall slower than those in the pure condition. Because this slowing was not affected by the presence of a distractor singleton, it is unlikely that it was caused by a general widening of the attentional selection process. Instead, the overall character of the slowing suggests that additional processing after selection had taken place, is responsible (but see Lamy, Carmel, Egeth, & Leber, 2006). One such process may be the comparison of the selected item with a target template in visual short term memory (VSTM, Duncan & Humphreys, 1989). When a target template is extended to include more target types, matching of the items in VSTM to the target template

takes more time (cf. Sternberg, 1969). In other words, in conditions of target uncertainty, attentional dwell times on the target and distractor singleton are increased relative to when the target and distractor remain fixed. When there is more target uncertainty, it simply takes more time to decide whether the item selected is either a target or a distractor. The target uncertainty has no effect on the attentional selection.

Since its introduction in 1991, the basic findings of additional singleton task has been replicated by many researchers. For example, Bacon and Egeth (1994), replicated Theeuwes 1992 and showed that it did not matter whether the distractor conditions were varied within or between blocks. Kumada (1999) examined between dimension (e.g., color and orientation) and within dimension (e.g., orientation) interference in a simple and compound version of the additional singleton task. The results showed interference in the within dimension condition both in compound and simple search. In the between dimension condition, there was only interference in the compound search condition. Leber and Egeth (2006) used an extensive training scheme and replicated Theeuwes' basic findings with heterogeneous displays. Mounts (2000) used similar display to those of Theeuwes (1992) and showed that the identification of a letter was slowed by the presence of an irrelevant color singleton. Importantly, Mounts (2000) showed that the identification of the letter was slowest when it was located next to the irrelevant color singleton (see also for a similar result Hickey & Theeuwes, under review; Mathôt et al., in press). Kim and Cave (1999) employed the additional singleton search task in combination with a probe detection task and showed that at the 60 ms SOA the probe RT at the location of the distractor singleton was about 20 ms faster than at the target singleton location. At the 150 ms SOA, however, this pattern was reversed: the probe RT at the target location was about 15 ms faster than at the distractor location. It was concluded that early on – at 60 ms after display onset – attention was first captured by uniquely colored distractor, while soon thereafter (at 150 ms) the probe at the target singleton received more attentional activation. Geyer et al. (2008) showed that capture depended on the frequency of occurrence of the distractor singleton. Lu and Han (2009) showed that when the search task becomes more difficult (more serial) attentional capture is reduced or even absent (see also Proulx & Egeth, 2006). Schubö (2009) showed large RT interference effects when observers searched for a shape singleton while a color singleton was presented, but not the reversed. Dalton and Lavie (2007) reported an equivalent effect of attentional capture in the auditory domain: Irrelevant high intensity singletons interfered with an auditory search task when the target itself was also a feature singleton. Others have demonstrated similar effects using measures related to d-prime (Theeuwes & Chen, 2005; Theeuwes et al., 2004), saccadic eye movements (Godijn & Theeuwes, 2002; Ludwig & Gilchrist, 2002; Mulckhuysen, Van der Stigchel, & Theeuwes, 2009; Theeuwes et al., 1998; Theeuwes et al., 1999); and hand movements (Hunt, von Muhlenen, & Kingstone, 2007).

3.2.1.2. Cueing paradigm. In addition to the additional singleton paradigm, in recent years, cueing procedures were used to examine the boundary conditions of top-down and bottom-up control of attention (see Theeuwes et al., 2006; Theeuwes & Van der Burg, 2007, 2008). For reasons discussed above, displays used in a salient singleton were present. If top-down information can be used to select the target singleton as assumed by almost all theories on visual search (e.g., Müller et al., 2003; Wolfe, 1994) then providing information about the upcoming target should improve visual search. In other words, if we tell observers that the upcoming target is a red color singleton then observers should be faster when the target is indeed a red singleton then when it is a green singleton. For example, in Theeuwes et al. (2006) we cued observers with 80% validity about the likely feature property of the upcoming target singleton (see Fig. 4). We gave observers about 1.5 s to get optimally prepared for the feature defining the target singleton which would be presented on

the upcoming trial. In the example (Fig. 4), observers received as a cue the word “color” (or the word “red”) and knew with 80% validity that the line segment they were looking for would be presented within the red colored circle. In 20% invalid trial, the target line segment would appear in the shape singleton (the green diamond). In the neutral condition we did not provide information about the property of the target singleton. As is clear from the data, providing this information had no effect on the efficiency of target selection. Whether one knows whether the target is red or whether it is a diamond did not improve performance.

It is important to realize that in these studies, we again used a compound search task (looking for a singleton and responding to the line inside the singleton) ensuring that the RT costs and benefits obtained reflect an improvement in the perceptual representation and not reflect effects on later (response) decision processes (e.g., Mortier, Theeuwes, & Starreveld, 2005; Prinzmetal, McCool, & Park, 2005; Starreveld, Theeuwes, & Mortier, 2004). Previous studies that have shown top-down RT effects on visual search (e.g., Müller et al., 2003; Treisman, 1988; Wolfe, Butcher, Lee, & Hyle, 2003) may have shown effects that are related to the speed of responding and not so much to the speed of visual selection (see for a discussion Mortier et al., 2005).

In subsequent studies we extended these claims showing that non-spatial cueing has no effect on perceptual sensitivity (Theeuwes & Van der Burg, 2007). In this study we compared the efficiency of spatial and non-spatial cues. As in our previous study (Theeuwes et al., 2006) the cues were presented 1.5 s before the presentation of the display, giving ample opportunity to prepare for the upcoming display (similar to the procedure shown in Fig. 4). These experiments showed that top-down set for spatial information has a strong effect on perceptual sensitivity (in terms of A-prime). However, a top-down set for non-spatial information does not change perceptual sensitivity. So knowing where the target singleton will appear affects perceptual selectivity; knowing what it is, does not help selectivity. It is important to note that measures of sensitivity (A-prime or d-prime) are assumed to reflect early perceptual processing such as the sensory encoding of a stimulus (see e.g., Handy, Jha, & Mangun, 1999; Müller & Findlay, 1988).

These cueing studies showed that cueing with the word “red” or “color” to prepare for the upcoming target display did not affect RT or perceptual sensitivity. However, in both studies, showing the actual target (i.e., showing a red circle as a cue) did have an effect of the speed of selection and on perceptual sensitivity. In other words, non-spatial cueing can result in a more optimal selection of the target singleton. Importantly however, the effect occurred regardless of cue validity. For example, in Theeuwes and Van der Burg (2007) showing the target as a cue had the same effect on A-prime regardless whether the cue was highly valid (83% validity) or whether the validity was at chance level (50%). In Theeuwes et al. (2006) we pushed this effect even further and made the cue counter predictive. For example, when a red circle was shown as a cue, there was a high chance (83%) that the target would be a green diamond. In other words, a red circle as a cue indicated that observers should prepare for a green diamond because in the majority of trials, a green diamond was the target. The results indicated that when a red circle served as cue even when it was counter predictive it still had an effect on RT such that observers were faster when the ‘unlikely’ red target singleton was presented. The same was true for the reverse (diamond cue, red circle target). In fact, the RT benefits due the cue were the same regardless of whether cue was counter predictive (17%) or highly predictive (80%). These findings strongly suggest that the effect of the cue on visual selection is not the result of top-down volitional control. Obviously visual search benefitted regardless of the cue validity.

On the basis of these studies we suggested that by showing the actual target as a cue one can obtain non-spatial cueing effects that can only be attributed to improved visual selection (and response selection). However, because these effects are independent of cue

validity, we claim that these effects are automatic and bottom-up. In fact, these effects are likely to be the same as inter-trial priming effects that have been reported in studies investigating “priming of pop-out”. For example in [Maljkovic and Nakayama \(1994\)](#) observers searched for a feature singleton that was defined in either two different ways: a red target among green distractors or a green target among red distractors. They showed that even when repetition of the feature value was at chance level (i.e., repetition was no more likely than alternation), repeating a target (but not the response) improved performance. They concluded that this effect is due to automatic (visual) priming which is basically not affected by top-down set (see also [Kristjánsson, Wang, & Nakayama, 2002](#)). These conclusions are fully consistent with our study of inter-trial priming ([Pinto et al., 2005](#)) that we discussed earlier to account for the RT differences between the [Theeuwes \(1991\)](#) and [Theeuwes \(1992\)](#) version of the additional singleton task.

Note that the idea that only spatial (and not non-spatial information) can have an effect on the earlier stages of processing is consistent with several electrophysiological studies. Event-Related Potential studies (ERP) have shown early sensory-specific ERP components (P1 and N1) elicited over posterior visual cortical areas in response to visual stimuli at attended compared to unattended locations (e.g., [Eimer, 1995](#); [Hillyard & Munte, 1984](#); [Mangun, 1995](#)). Amplitude modulations in response to these spatial cues start as early as 80 ms after stimulus onset. In contrast, when attending to non-spatial features, such as color or shape, there is an enhanced sustained negativity that starts much later, at about 150 ms post-stimulus (e.g., [Eimer, 1995](#); [Heslenfeld, Kenemans, Kok, & Molenaar, 1997](#)). With respect to spatial attention ERP studies have shown that the cortical mechanism involved consists of interconnected dorsolateral prefrontal (DLPFC) and posterior parietal (PPC) cortical areas. These areas can modulate incoming visual information via projections to both dorsal and ventral extrastriate visual pathways ([Martinez et al., 1999](#)). ERP studies show that selection based on non-spatial information has completely different ERP pattern than those on the basis of spatial information ([Hillyard & Munte, 1984](#)), indicating that cortical mechanism involved in non-spatial selection is qualitatively different from that involved in the selection on the basis of space.

3.2.2. ERP evidence

The study by [Hickey et al. \(2006\)](#) was among the first to examine attentional capture by means of ERPs. [Hickey et al. \(2006\)](#) used the 1991 version of the additional singleton paradigm of [Theeuwes](#). In this version, observers search for a unique singleton in the shape dimension (either a diamond among circles or a circle among diamond, randomized within blocks). On some trials, a color distractor was present (either a red circle among greens or a green circle among reds). In each display there were 10 stimuli. To investigate whether the salient but for the task irrelevant singleton captured attention, [Hickey et al. \(2006\)](#) focused on a specific component of the visual ERP that has been linked to the deployment of attention in visual space ([Luck, Woodman, & Vogel, 2000](#); [Woodman & Luck, 1999](#)). This component, known as the N2pc, is a negative-going deflection that is observable in the ERPs recorded over the posterior scalp roughly 175–300 ms post-stimulus. Specifically, the N2pc is defined as a larger negative voltage at electrodes contralateral to an attended stimulus than at electrodes ipsilateral to that stimulus, and is named for its polarity, latency, and topography (posterior contralateral negativity in the latency of the N2 component). Even though there is no complete consensus of what exactly the N2pc reflects, it is established that it is related to the attentional selection either by the suppression of surrounding items ([Luck & Hillyard, 1994](#)) or by processing of target related features ([Eimer, 1995](#)). Even though some have suggested that the occurrence of N2pc reflects the actual capture of attention to a location in space (e.g., [Eimer & Kiss, 2007](#)) more recent studies clearly refute this idea and

suggest that the N2pc reflects attentional processing at a location in space ([Kiss, Van Velzen, & Eimer, 2008](#)). Before discussing the various results of ERP studies, it is important to realize that the N2pc does not reflect the shift of attention itself but the enduring attentional processing at the location to where attention shifted.

[Fig. 5](#) presents the main results of [Hickey et al. \(2006\)](#). Panel A shows the no distractor condition revealing a clear N2pc to the target. Panel B shows a larger N2pc when target and distractor singleton are located within the same hemifield. The most important result is displayed in panel C when the target and distractor were presented in opposite visual hemifields. As is clear from this figure the contralateral distractor elicited an ERP negativity in the latency of the N2pc that was first contralateral to the distractor singleton followed only later to a ERP negativity contralateral to the target. In other words, the data thus suggest that attention was initially oriented to the distractor stimulus before being reoriented to the target singleton. This pattern of results provides compelling evidence that after attention is captured by the distractor singleton (the most salient element in the display), there is first attentional processing at the location of the distractor, followed by attentional processing at the location of the target. Because [Hickey et al. \(2006\)](#) employed the [Theeuwes, 1991a](#) version of additional singleton task in which target and distractors switch from trial to trial, substantial processing is necessary both at the location of the target and distractor singleton (see also [Pinto et al., 2005](#)).

The obvious question that needs to be answered is what the ERPs will look like when the target and distractor remain fixed throughout a block of trials (as in [Theeuwes 1992](#)). Preliminary results suggest that with version of the paradigm there is no N2pc towards the color distractor singleton ([Hickey & Theeuwes, 2008](#)). Even though the absence of an N2pc to the distractor singleton may suggest that the salient singleton did not capture attention, one should be cautious drawing such a conclusion. Indeed, we know that the interference in the [Theeuwes \(1992\)](#) additional singleton task when target and distractors remain fixed is typically only about 25 ms. When target and distractor change from trial to trial as in [Theeuwes \(1991a\)](#) this effect is about 100 to 150 ms. As [Pinto et al. \(2005\)](#) showed this large increase in RT interference is due to additional processing after the item has been selected. Indeed, when there is target uncertainty more time is needed to decide whether the selected item is a target or a distractor. Because the N2pc does not reflect the attentional shift per se but the processing of features at a location ([Kiss et al., 2008](#)) it may not be surprising that when the distractor and target remain fixed there is not much, if any, processing at the location of the distractor. Because target and distractor remain fixed, once the distractor has been selected, not much time is needed to decide that the item is a distractor and not the target. Therefore the absence of an N2pc in these experiments does not say anything about whether attention was captured by the distractor singleton or not; it only tells that there may not have been enough attentional processing going on at the location of the distractor to reveal an N2pc. The crucial point to make is that the presence of an N2pc reveals that attention has shifted to a location in space in order to process information at that location. Finding no N2pc does not necessarily imply that attention never shifted to that location; it may never have shifted but it may also imply that attention has shifted but that there was not enough processing at that location to reveal an N2pc.

Similar to [Hickey and Theeuwes \(2008\)](#), [Schubö \(2009\)](#) also used the [Theeuwes \(1992\)](#) additional singleton paradigm in which target and distractors do not change over trials. Also in this study, there was no N2pc to the distractor singleton. However, consistent with our unpublished data, the amplitude of the N2pc towards the target singleton was reduced when a contralateral distractor singleton was present. [Schubö \(2009\)](#) argued that the reduced N2pc to the target was due to a reduced focusing of attention to the target singleton in case a contralateral distractor was present. If some attention was allocated to the location of the distractor singleton (at least on subset

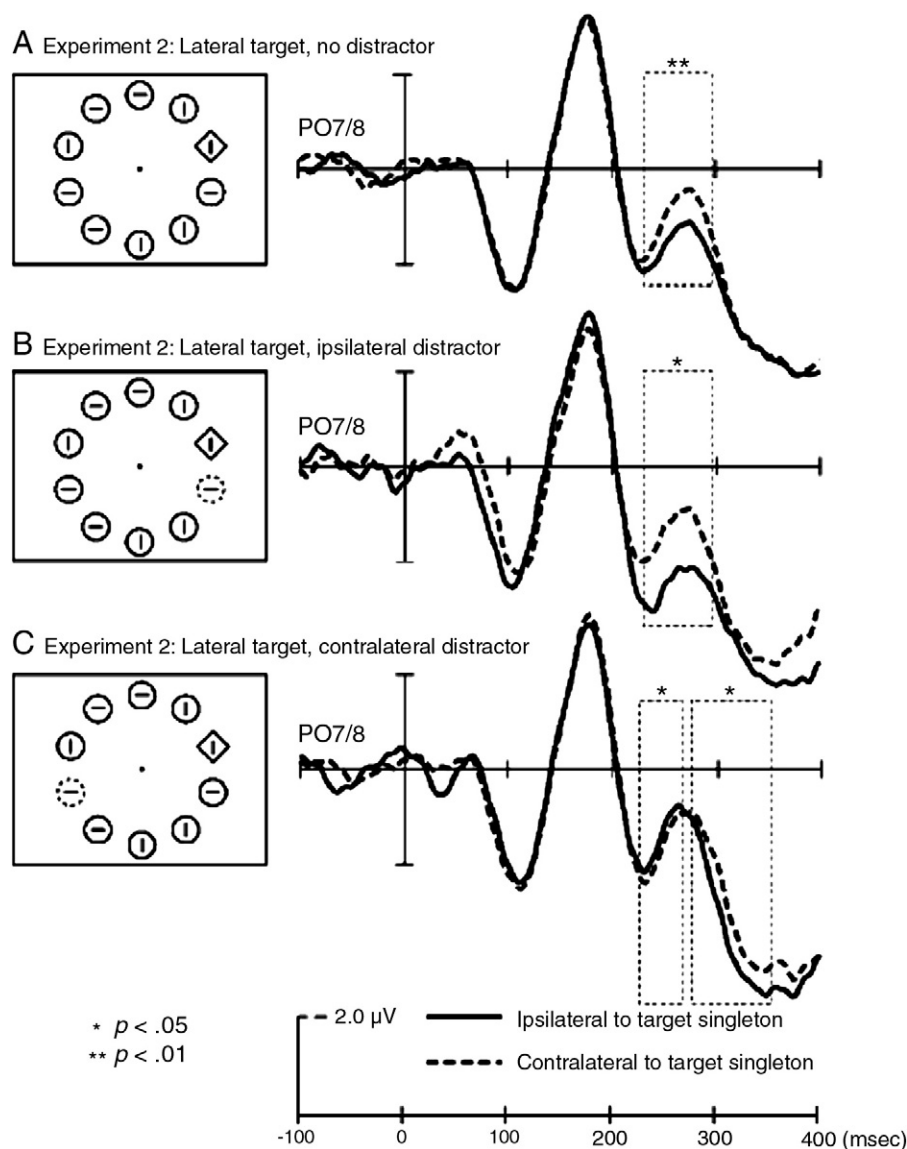


Fig. 5. Stimuli and data from Hickey et al. (2006). Grand-averaged ERPs for observers searching for a shape singleton (in this case a diamond) in the additional singleton paradigm. Search arrays that contained a lateralized target singleton elicited a posterior ERP negativity in the latency of the N2pc at electrodes contralateral to the target, both when the distractor was absent (panel A), and when the distractor was on the same side of the vertical meridian (panel B). Search arrays that contained a lateralized target and a contralateral distractor elicited a posterior ERP negativity in the latency of the N2pc that was initially contralateral to the distractor and only later contralateral to the target (panel C). This latter panel shows that there is first attentional processing at the location of the distractor singleton before attention is shifted to the target singleton to allow the processing of the target.

of trials), this will result in a reduced attention allocation at the target location.

In another version of the additional singleton paradigm Wykowska and Schubö (in press) showed a clear N2pc to the target singleton while there was no N2pc to the distractor singleton. Importantly, however they showed that the onset of the N2pc to the target was delayed when the irrelevant singleton was presented in the opposite hemifield relative to the same hemifield. On the basis of the absence of the N2pc they argued that attention was never allocated to the distractor singleton even though the irrelevant singleton produced enough interference to delay the allocation of attention to the target. The delayed allocation of attention to the target in the presence of a distractor in the opposite hemifield is consistent with our interpretation that the irrelevant singleton does capture attention but only for a brief time. However, Wykowska and Schubö (in press) discard this interpretation on the basis of an additional experiment in which they locked ERPs to probe onsets presented at the target and distractor singleton locations. There was a probe-locked P1 sensory gain effect at the location of the target singleton but not at that of the distractor

singleton which led to the conclusion that the presence of the distractor singleton has a non-specific non-spatial effect on the allocation of attention to the target singleton. This interpretation is viable, yet it is possible that the probe onsets were presented not early enough at the distractor location to reveal its sensory gain effect. Indeed, it is known that attentional disengagement may be fast, possibly even faster than the 80 to 120 ms interval that was used in the Wykowska and Schubö (in press) study.

In addition to ERP studies using versions of the additional singleton tasks, in recent years several studies have investigated attentional capture using variants of the Folk, Remington and Johnston (1992) contingent cueing paradigm (see Fig. 7 for a detailed explanation). For example, in Eimer and Kiss (2008) when observers had to report the orientation of a uniquely colored target bar among nontarget bars (color task), an N2pc was obtained towards the color cue which was presented 150 ms before the target display. However, that very same cue display did not generate an N2pc when observers were required to report the orientation of the target bar presented without distractors (the onset task). These results were interpreted as

evidence for contingent capture, since the occurrence of an N2pc depended on the task set. Notably, Eimer and Kiss (2008) argued that the “early cue-induced N2pc [is] indicative of rapid attentional capture by color singleton cues” (p. 1432). However, this interpretation is disputed because another of their studies clearly demonstrated that “the N2pc triggered in response to pop-out visual search targets does not reflect processes involved in covert shifts of spatial attention, but is instead linked to spatially”. In other words, the occurrence of the N2pc does not say anything about attentional capture but about the post-selection processing occurring at a particular location. Given this latter interpretation, it is not surprising that when observers are looking for a color target singleton, attention will strongly engage at the color singleton when just before the target display the same color is presented as a cue. Obviously, observers have to decide whether the color singleton is a target or a cue. However, when looking for an onset, it will not take much processing, if any, to disengage attention from the color cue because it will be immediately clear that it is the cue and not the target.

Another ERP study also investigated whether salient distractors capture attention in an exogenous way (Leblanc, Prime, & Jolicoeur, 2008). In this study observers were looking for a target which was color-defined item embedded in an RSVP stream. Target items were preceded by distractor displays consisting of two items presented to the left or right of the centrally presented stream. The color of one of these distracting stimuli matched that of the color observers were looking for in the central stream. The results showed that target-colored distractors elicited significant N2pc waves, whereas distractors that had a nontarget color did not result in an N2pc. Leblanc et al. (2008) interpreted these results as evidence for contingent attentional capture: only when the color in the periphery matched that of target there was an N2pc. Even though this study undoubtedly shows that there was more attentional processing at the location of the distractor when it looked like the target (similar to Eimer & Kiss, 2008), the conclusion that the “significant N2pc waves [indicates] that visuospatial attention had been drawn to their location” (p. 669) may not necessarily be correct. As we outlined before, the N2pc may have nothing to do with attentional capture per se but with the processing of information at the attended location. Another aspect of this study makes it even more questionable whether we are dealing with attentional capture of stimuli presented in the periphery (as is argued in this paper). It seems that the distractors were in fact not actually presented in the periphery. The method section indicates that distractor stimuli were presented only one degree of visual angle (contour to contour) away from the central stream. Therefore, it is unlikely that there was an actual shift of attention because the distractor stimuli were presented very close to the focus of attention.

3.2.3. fMRI evidence

De Fockert, Rees, Frith and Lavie (2004) (see also Lavie & de Fockert, 2006) investigated the neural substrates of attentional capture using a variant of the Theeuwes' additional singleton paradigm. Observers searched for a unique shape target (a circle among diamonds) while neural activity was measured using functional magnetic resonance imaging (fMRI). The original task was slightly altered such that the appropriate comparisons could be made. In the color singleton present condition, there was always one color singleton which either could be the target singleton (observers search for a circle which happened to be the only red element in the display) or could be the distractor singleton (observers search for a green circle and a red distractor singleton was present). In the color singleton absent condition all display elements were green and observers searched for a green circle. To have a full factorial design, there was also a distractor condition in which one of the distractors was reduced in size, thus creating a distractor condition that did not contain a color singleton. The classic comparison, color distractor present-absent showed a large interference effect of 94 ms (809 ms

versus 713 ms). When the target happened to be the color singleton this only had a small non-significant benefit of 9 ms relative to when the shape singleton did not have a unique color, possibly due to a floor effect.

The fMRI data comparing color distractor singleton presence versus absence showed the neural activity associated with attentional capture. The presence of a color distractor singleton resulted in bilateral activation within the superior parietal cortex (Brodmann area 7). Importantly, this activity has previously been associated with spatial shift of attention (Corbetta & Shulman, 2002), providing converging evidence that the presence of the color distractor singleton caused in a shift of spatial attention, a conclusion which is consistent with the ERP data reported by (Hickey et al., 2006). In addition to the parietal activity, the presence of the color singleton was associated with activity within the left lateral precentral gyrus (BA 6) of the frontal cortex. This frontal activity is assumed to reflect the resolution of subsequent competition between the salient distractor and target. Interestingly, de Fockert et al. (2004) found that this frontal activity was negatively correlated with the interference effect in magnitude RT caused by the color distractor singleton. In other words, it appears that the more frontal activity the less behavioral capture is observed, suggesting a role for top-down control in the modulation of attentional capture. However, it should be noted that in contrast to the strong negative correlation between the activity in the frontal cortex and the behavioral capture effect, there was no correlation between activity in the superior parietal cortex and the amount of capture. This suggests that the signal in the superior parietal cortex reflects bottom-up capture of attention towards the location of the salient singleton, with little or no variation in the strength of the signal. Subsequently, the activity in the frontal cortex reflects the processes resolving attentional capture, occurring after attention has been shifted to the salient singleton. These findings fit nicely with our notion that top-down control operates on the speed with which attention can be disengaged from the location of the salient singleton. If the disengagement is quick and efficient it will result in a small distractor effect reflecting strong top-down control in overcoming attentional capture. If attention sticks relatively long at the location of the color singleton, it will result in a large RT effect reflecting less top-down control. Assuming that the activity in the frontal cortex reflects top-down control, the data of de Fockert et al. (2004) elegantly show the strength of top-down control may help to disengage attention from the distractor location. Top-down control has no effect on the initial shift of attention to the color distractor singleton since the signal strength in the superior parietal areas is not related to behavioral interference.

In another study, Lavie and de Fockert (2005) examined capture under conditions of high and low memory load. Observers performed a version of the additional singleton task, and had to keep in memory five digits that were either presented in a sequential order (0, 1, 2, 3, 4, and 5) constituting the low memory load or in a random order (0, 3, 2, 4, and 1) making up the high memory load condition. The results indicated that the RT interference effect caused by the color singleton was larger in the high than in the low memory load condition. These findings point in the same direction as the fMRI data: when frontal cognitive control functions are occupied by an additional memory task, observers have more trouble disengaging attention from the distractor location causing larger RT effects than when frontal areas are not occupied by an additional task. Note however, that even when there is no memory load at all, the presence of a distractor always caused some interference confirming the observations that capture always occurs in this type of tasks (see Lavie & de Fockert, 2006 for a similar reasoning).

3.2.4. TMS evidence

In a TMS study (Hodsoll, Mevorach, & Humphreys, 2009) observers performed the additional singleton task while the parietal cortex was stimulated by means of sustained transcranial magnetic

stimulation (rTMS) over the Posterior Parietal Cortex (PPC). In Experiment 1 (Hodsoll et al., 2009) observers always searched for a circle among five diamonds. Similar to the de Fockert et al. (2004)'s design, in case a color singleton was present, it was the target (the target circle was red) on half of the trials and the distractor on the other half (a red diamond as a distractor while a green circle served as a target). Similar to Pinto et al. (2005) this design made it possible to look at inter-trial effects. For example, a distractor singleton effect is expected to be larger when on the previous trial the singleton was the target relative to a condition when in the previous trial it was a distractor singleton. Overall Hodsoll et al. found a large interference effect: when a distractor was present observers were 137 ms slower than when it was absent. Importantly, the size of the distractor effect was modulated by rTMS such that it was reduced to 96 ms when rTMS was applied to the right parietal cortex but not when applied to the left parietal cortex or when no rTMS was applied (152 ms versus 162 ms). Hodsoll et al. concluded that right parietal stimulation reduced the interference effect of the distractor singleton. Consistent with Pinto et al. (2005) Hodsoll et al. found strong inter-trial priming such that interference of the color distractor was much larger when in the previous trial the color singleton was the target relative to when on the previous trial it was the distractor. Importantly, however, this inter-trial effect was also strongly modulated by right PPC stimulation such that there was basically no inter-trial priming when there was right PPC stimulation while priming was fully intact following left PPC stimulation. In Hodsoll et al.'s second experiment, inter-trial priming was removed by presenting all trials in separate blocks. Thus, in a block of distractor singleton trials, the color singleton was always the distractor so there was no reason for observers to look for a color singleton (as for example in Theeuwes, 1992). Not surprising in this design the overall interference effect was much smaller (the distractor condition was about 45 ms slower than in the no distractor condition). Again the interference effect interacted with PPC stimulation: in case of left PPC stimulation the distractor effect remained basically intact (about 36 ms) but importantly with right PPC stimulation, the inference effect completely disappeared. In fact, there was no difference between the distractor and no distractor condition. Note that the effect was specific for right parietal cortex and was equivalent for both the contralateral and ipsilateral visual field.

The results of Hodsoll et al. (2009) indicate an important role for the right PPC in the occurrence of attentional capture. The result extends the fMRI findings of de Fockert et al. (2004) who also showed that the parietal areas are important except that this study showed bilateral activation. The crucial point is that in Hodsoll et al. (2009) Experiment 2, the color singleton was never the target implying that capture was truly exogenous (as in Theeuwes 1992). Even when capture was automatic and bottom-up, stimulation of the right PPC abolished the capture effect. The data suggest that the right PPC is important in biasing attention to salient stimuli in the visual field. The right PPC may be the brain area of automatic bottom-up capture.

Another important finding of the study of Hodsoll et al. (2009) is that TMS effects were only restricted to the distractor singleton condition. In other words, TMS had no effect when the color singleton was the target or when no singleton was present. This was also found in Experiment 2 in which all these conditions were mixed within blocks. These findings are consistent with results of Ashbridge, Walsh and Cowey (1997) who showed no effect on feature search following right parietal TMS. It seems that competition between a salient target and distractor singleton is necessary for TMS to have an effect. When the target and distractor are in competition, TMS to the right parietal area appears to block the activity of the most salient distractor in the visual field, i.e., the color distractor and performance is as if no color distractor singleton was present. If only one singleton is present (and there is no competition) then TMS does not have an effect. It seems that biased competition with the PPC is necessary for TMS to disrupt the response to salience.

3.2.5. Single cell evidence

In a paper Ogawa and Komatsu (2004) provided elegant data regarding the neural process underlying stimulus-driven bottom-up selection (see also Connor, Egeth, & Yantis, 2004 for a discussion of this paper). In a version of Theeuwes additional singleton task, monkeys were trained to adopt one of two top-down strategies in different experimental sessions: either search for the color singleton or search for the shape singleton. Monkeys responded by making a saccade to the location of the target singleton. To ensure that the neural modulation would not reflect the execution of saccades the recording in V4 neurons only included the firing of neurons before the onset of a saccade. Fig. 6 presents data from a single neuron in area V4 during search for the shape singleton (on the left) or search for the color singleton (on the right) a typical monkey.

The monkey was required to remain fixated in the center until the search array came on and then had to execute a saccade to the target singleton. In the figure the red frame indicates the condition in which the shape singleton fell in the receptive field (RF) of the recorded neuron (indicated by the gray circle). Similarly the green frame indicates the condition when the color singleton was in the RF. The figure shows for the first 175 ms after display onset an identical firing rate pattern regardless of whether the animal was searching for a shape singleton (on the left) or a color singleton (on the right). In other words, for the first 175 ms there is no evidence that top-down set played a role. Or stated differently: for the first 175 ms post-stimulus neurons that matched the search goal showed exactly the same firing rate as neurons that did not match the search goal. Whether a singleton was task relevant or irrelevant had no effect on the initial firing rate. This study elegantly demonstrates that there is no top-down modulation on the initial sweep of information through the brain. Initial selection is completely bottom-up, representing the neural correlates of bottom-up capture as advocated by Theeuwes (1991a, 1992). It is important to note that after 175 ms top-down modulation starts having an effect. When searching for a color singleton, when a color singleton is in the RF the firing rate stays up at 50 spikes/s while the firing rate goes down to about 20 spikes/s when this color singleton is task irrelevant. In other words, top-down

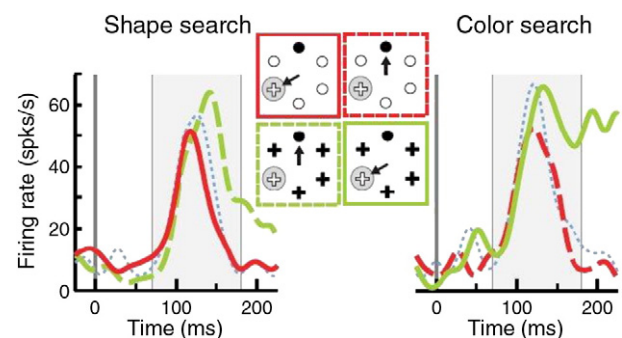


Fig. 6. Data of Ogawa and Komatsu (2004). Response of a single V4 neuron when a typical monkey searches for a shape singleton (left) or a color singleton (right). The monkey is required to make a saccade to the target singleton (indicated by the arrow). The red frame indicates the shape singleton in the receptive field (indicated by the gray circle); the green frame indicates the color singleton in the RF. It is important to note that the firing rate when the shape singleton is in the RF (red frames) is basically the same whether the animal is looking for the shape singleton (solid red line) or is looking for a color singleton (dashed red line). Similarly, when a color singleton is in the RF (green frames) the initial firing rate is the same whether the animal is looking for a color singleton (solid green line) or is looking for a shape singleton (dashed green line). Only about 175 ms after search array onset, top-down modulation has an effect: when the color singleton in the RF is relevant (the animal is looking for a color singleton, right panel) the firing rate stays up near 50 spikes/s (solid green line) while only a firing rate of 20 spikes/s is observed for the same color singleton when it is irrelevant (dashed green line, left panel). The blue line represents the response functions to nontarget elements.

control certainly plays a role, but only later because top–down effects require recurrent processing from top to bottom.

Another aspect of this data set is important. As can be seen in Fig. 6, the color singleton response function (green lines) starts to separate from the shape singleton (red) and distractor (blue) responses at about 125 ms. This is consistent with the idea that the bottom–up activity generated by a color singleton is larger than that generated by a shape singleton. At the neural population level, the bottom–up difference between the shape and color singleton first became significant at about 120 ms, while top–down effects first became significant at about 195 ms. The study of Ogawa and Komatsu (2004) shows that the neural singleton effect is basically bottom–up for the first 150 to 200 ms with more bottom–up neural activity for the color singleton than for the shape singleton completely consistent with psychophysical evidence of Theeuwes (1991a, b, 1992). Moreover, the neural activity for the first 150 to 200 ms is not modulated by task set; only after 175 ms the top–down modulation starts having an effect. This is also consistent with the stimulus-driven capture account that claims that early preattentive processing (<150 ms) is completely data-driven and only top–down processing kicks in at a later point in time.

The study of Ogawa and Komatsu (2004) shows also other important parallels to the data of additional singleton paradigm. Similar to findings using the additional singleton paradigm in eye movement studies (Godijn & Theeuwes, 2002; Theeuwes et al., 1998; Theeuwes et al., 1999) the monkeys in Owaga and Kumatsu study often made erroneous saccades to the irrelevant singleton, much more often than to a non-salient non-singleton. In other words, just like in the additional singleton paradigm with human subjects, monkeys (who typically have much more training than humans) show attentional and oculomotor capture by irrelevant salient stimuli.

Other single cell studies also provide evidence that initially – about 150 to 200 after stimulus onset – neurons respond unselective. For example (Chelazzi, Duncan, Miller, and Desimone, 1998; Chelazzi, Miller, Duncan, & Desimone, 2001) measured the response of neurons in anterior IT cortex and V4 during visual search. While the monkey maintained fixation, a cue stimulus was presented at the center, followed by a blank delay period after which a search array came on containing 2 to 5 stimuli. The monkey had to detect the stimulus that matched the cue. When a target was present in the array, strong effects of attention were found. Importantly however during the first 150–200 ms after stimulus onset, the IT neuron responded in the same manner regardless of whether the good (effective in driving the cell) or the poor stimulus (ineffective in driving the cell) was the target. These studies confirm the notion that initially processing is bottom–up not modulated by top–down (feedback) signals.

Other single cell studies are consistent with these findings. For example, Constantinidis and Steinmetz (2005) showed that neurons in PPC increased their firing rate when a salient item fell within their receptive field. Importantly however, the firing rate was not dependent on the specific feature value but only whether the element was salient or not. The authors conclude that the PPC automatically detects and encodes the location of salient stimuli regardless whether they are relevant or not. In a recent study Buschman and Miller (2007) recorded neural activity simultaneously from frontal and parietal areas while the monkeys performed a serial search and a pop-out task. The results showed that fast bottom–up target selection is seen first neurons in LIP. Top down selection is slower and occurs first in frontal areas. Again, this data suggests that parietal neurons represent the salience map responsible for bottom up selection (see also Bisley & Goldberg, 2003).

3.3. The specifics of the stimulus-driven account

After reviewing the most important data providing evidence for stimulus-driven capture, in this section we will discuss some specific

features of the stimulus-driven capture account and the evidence associated with these characteristics.

3.3.1. The irrelevant singleton causes a shift of spatial attention

The basic finding of the additional singleton task is that the presence of the irrelevant singleton causes an increase in RT relative to the condition in which the singleton is absent. According to the stimulus-driven account the increase in RT is the result of attention being first shifted to the location of the irrelevant singleton before attention can be directed to the target singleton. This shift of attention constitutes stimulus-driven capture. Folk and Remington (1998) offered an alternative explanation for the increase in RT in conditions in which a distractor was present. Instead of assuming that the irrelevant singleton captured attention, they argued that the increase in search time caused by the irrelevant singleton is due to “filtering costs” a notion first introduced by (Kahneman, Treisman, & Burkell, 1983). According to this idea, the presence of an irrelevant singleton in the additional singleton task slows the deployment of attention to the target item by requiring an effortful and time-consuming filtering operation (see also Becker, 2007). This line of reasoning holds that attention goes directly to the target singleton, yet, because of the presence of another singleton, directing attention to the target may take more time than when no such irrelevant singleton is present. Note that this view does not assume that attention is ever shifted to the location of the irrelevant singleton. The filtering cost account explains the results of the additional singleton paradigm without assuming a bottom–up shift of spatial attention to the irrelevant singleton. Instead it assumes that top–down control selectively guides spatial attention to the target singleton.

Even though the exact mechanism underlying the filtering costs account is not well defined (for a discussion see Schreij, Theeuwes, & Olivers, in press), the idea that an increase in RT in the distractor condition is not necessarily due to a shift of spatial attention is quite feasible. However, three different lines of research have specifically addressed this issue using different approaches demonstrating that the results of the additional singleton paradigm can only be explained by assuming a shift of spatial attention to the irrelevant singleton rather than by a non-spatial filtering operation.

In one study Theeuwes (1995b) used what has been called the “identity intrusion technique” (see also Theeuwes & Burger, 1998) to determine whether spatial attention had shifted to the location of the salient singleton. The basic idea underlying this identify intrusion technique is related to the Eriksen and Eriksen (1974) congruency manipulation which typically has been used for focused attention tasks. In this paradigm the singleton element that observers had to ignore is either identical or different from the target element observers are searching for. For example, in Theeuwes and Burger (1998) participants searched for the target letter E or R among a variable number of nontarget letters. In each display there was one letter that had a unique color, constituting the singleton that had to be ignored. The singleton to be ignored was either identical to the target letter (‘congruent’ condition: i.e., both letters were Es or Rs) or different from the target letter (‘incongruent’ condition: the singleton was an ‘E’ and the target letter was an ‘R’; vice versa). If participants could filter out color singleton successfully, then it is expected that the identity of the singleton would have no effect on search for the target element. Alternatively, when participants would not be capable of completely ignoring the color singleton it is expected that the processing of a response-incongruent singleton will produce performance costs relative to a response-compatible singleton. This latter effect Theeuwes and Burger (1998) labeled the identity intrusion effect (see also Gibson & Bryant, 2008).

Theeuwes (1995b) used this technique to determine whether attention was shifted to the location of the irrelevant singleton. Observers performed a typical additional singleton search as Theeuwes (1992). Instead of having a neutral element at the location

of the irrelevant singleton, the congruency of the character at the location of the irrelevant distractor was manipulated. In half of the trials the character at the distractor location was associated with the same response as was required by the target, while on the other half of it was the opposite of what was required by the target. Theeuwes (1995b) reasoned that the identity of the character at the location of the irrelevant singleton could only have an effect on responding when at some point spatial attention would have been employed at the location of the distractor. If however attention would never go the location of the irrelevant singleton as is assumed by the filtering account then there should be no congruency effect of a character presented at the irrelevant singleton location. In other words, if attention never goes to the location of the irrelevant singleton, it is impossible that the identity of a character can have any effect on responding. However, Theeuwes (1995b) did find clear congruency effect which provided strong evidence that before a response was given, spatial attention was at the location of the irrelevant singleton. This finding was completely in line with the notion that spatial attention was captured by the irrelevant singleton (but see Folk, Remington, & Wu, 2009 for a different interpretation).

The second line of evidence for the involvement of spatial attention in stimulus-driven capture comes from studies using inhibition of return (IOR) as a marker for spatial attention. The general consensus is that IOR is an attentional effect (Posner & Cohen, 1984) resulting from an inhibition of returning attention to a location that previously was attended. It is assumed that when attention is drawn to a location in space, and attention is subsequently disengaged from that location, an inhibitory mechanism starts to operate, inhibiting the return of attention to the location that was previously attended (see Klein, 2000). An important hallmark of IOR is that it only follows for a location to which attention has shifted reflexively. It does not follow endogenous shifts of attention (Posner & Cohen, 1984; Pratt, Sekuler, & McAuliffe, 2001). Thus if IOR occurs, it can only be the result of a shift of spatial exogenous attention.

Theeuwes and Chen (2005) used a variant of the additional singleton paradigm. Instead of measuring RT, this study employed measures of signal detection (d' -prime) to determine the perceptual sensitivity at the location of the target singleton. Unlike in previous experiments, in this study the color singleton could appear at chance level at the location of the shape target singleton. Note that there was no reason for observers to specifically search for the color singleton because on most trials it would signal a nontarget location. In this experiment the irrelevant color singleton preceded the search display with either a short (60 ms) or a long (500 ms) delay. The results indicated at the location of the irrelevant color singleton, there was first an increased sensitivity, which was followed by a reduced sensitivity, relative to a condition in which there is no distractor present. The increased sensitivity implies that spatial attention was directed at the location of the distractor singleton whereas the reduced sensitivity suggests that there was inhibition at the location of the distractor. This pattern of results showing initial facilitation followed by inhibition is the classic biphasic pattern signifying the occurrence of IOR. Note that IOR is hallmark of spatial attentional orienting typically occurring in studies using abrupt onsets to draw attention in a bottom-up way (Klein, 2000). The fact that IOR occurs at the location of the irrelevant singleton strongly suggests that spatial attention was drawn to this location (see also for similar findings Folk & Remington, 2006; Lamy & Egeth, 2003; Theeuwes & Godijn, 2002).

The final line of evidence comes from the ERP study of Hickey et al. (2006) discussed earlier. As outlined, if one observes an N2pc in the ERP signal it is ensured that there is attentional processing at a location in space. As is clear from Panel C, there is first processing of information at the location of the distractor singleton followed by attentional processing at the location of the target singleton. This pattern of results provides compelling evidence that there is first spatially specific processing at the location of the distractor singleton

before attention shifts over to the target location at which there is subsequent spatial processing.

3.3.2. *The shift of spatial attention to the irrelevant singleton is short-lived*

We assume that the initial sweep of information through the brain is bottom-up, selecting the location of the element having the highest salience. Only after attention has been directed to the location having the highest salience, the identity of the salient element becomes available. In other words, only after attention has shifted to the high salience location, the visual system can discern whether the high salience at the location is the result of for example, a color, luminance or shape difference (see also Donk & van Zoest, 2008; van Zoest et al., 2004). Once attention is shifted to the location of the highest salience, massive recurrent top-down processing makes it possible to have a fast and swift disengagement of attention from that location. The consequence is that when observers consistently search for one particular target (e.g., a green diamond as in Theeuwes, 1992) they also consistently ignore the very same distractor singleton (e.g., a red circle). Because they ignore the same distractor on each and every trial, the interference effect caused by the distractor is relatively small. Because all conditions are blocked, disengagement is fast and efficient. Even though the interference effect is relatively small (about 20 to 25 ms) it should be noted that with extensive training the interference effect does not go away. In Theeuwes (1992) observers received up to 1800 trials. Even though the overall RT because faster over blocks of trials, the interference effect remained present.

It is important to realize that the capture effects (i.e., the size of the RT effect due to the presence of a distractor) can be very much exacerbated when the disengagement of attention from the distractor singleton is more difficult. Finding a larger effect on RT does not mean that there is more capture; it means that it simply may take longer to disengage attention from the distractor location thereby increasing the effect on RT. For example, as we have shown in Theeuwes (1991a), when targets and distractors change roles from trial to trial the disengagement of attention is slow causing RT interference effect of about 100 to 150 ms (see also Pinto et al., 2005). When elements switch roles from trial to trial, once attention is captured by the singleton, substantial top-down processing is necessary to determine whether what is selected is the target or not. Consequently, in these types of experiments, interference effects are much larger. The same holds for the fMRI experiment conducted by de Fockert et al. (2004) and the TMS experiment of Hodsoll et al. (2009; Experiment 1). In these experiments, interference effects were also much larger, up to 150 ms. We assume that capture effects are usually small unless uncertainty about the target and distractor requires a relatively long dwell time at the location of the distractor.

As discussed before Theeuwes et al. (2000) specifically addressed the question how long it would take to overcome capture by an irrelevant singleton. In this study, with different SOAs, the irrelevant color singleton was presented before the target singleton. The reasoning was that if the irrelevant color singleton would have been presented enough ahead of time before the target singleton, there should be sign of capture anymore. If the color singleton is presented well before the target singleton, performance should have been as if no distractor was presented. The results showed that the color distractor singleton only had an effect when it was presented in very close temporal proximity to the target (less than 100 ms). When the distractor singleton was presented a considerable time (SOAs 150 to 300 ms) before the presentation of the target singleton, it was possible to exert sufficient top-down control such that attention was already disengaged from the location of the distractor when the target singleton was presented. At these SOAs there was no sign of attentional capture.

Kim and Cave (1999) came to a similar conclusion in a study using the additional singleton search task in combination with a probe detection task. In this study, probes were presented either 60 or

150 ms following the presentation of the search display at the location of the target and of the distractor singleton. It was hypothesized that if early processing is solely driven by bottom-up salience as the stimulus-driven capture account suggests, then the location of the salient distractor singleton should be attended first. Therefore, the probe RT at the distractor location should be faster than at any of the other locations in the short SOA condition regardless of whether the unique feature is relevant or not. Because attention will get disengaged from the distractor location and reshifted to the target location, it is expected that in the late SOA condition, attention will no longer be at the distractor location but instead will be at the location of the target singleton. For conditions in which target and distractor singleton were locally unique (and therefore salient enough) Kim and Cave (1999) did indeed exactly find these results. At the 60 ms SOA the probe RT at the location of the distractor singleton was about 20 ms faster than at the target singleton location. At the 150 ms SOA however this pattern was reversed: the probe RT at the target location was about 15 ms faster than at the distractor location. The crucial finding of Kim and Cave is that after 150 ms, attention was no longer at the location of the distractor but instead at the location of the target.

Lamy, Tsai and Egeth (2003) replicated the results of Kim and Cave (1999) using another variant of the additional singleton task. Observers searched for shape singleton while a color singleton was present. As in Kim and Cave (1999) with different SOAs, a small probe dot could appear at one of the square or circle locations, and observers had to respond as quickly as possible to the probe (and at the end of the trial gave an unspeeded response as to whether the target circle was present). Lamy et al. (2003) basically replicated Kim and Cave (1999) showing that early in time, observers were faster in responding to the probe presented at the distractor location indicating that it had captured attention. At a later point in time observers were faster to the probe at the shape target singleton suggesting that top-down control had taken over.

It is important to note that these findings showing early bottom-up capture followed by top-down control that kicks in around 150 ms are completely consistent with the earlier discussed study of Ogawa and Komatsu (2004) who recorded the electrical spike activity of individual neurons in area V4 (see Fig. 6). This study showed that up until about 175 ms after display onset, neurons show similar response enhancement for a color and shape singleton regardless of the task of the monkey. Only after about 175 ms one can see that top-down modulation takes an effect: after 175 ms, when the monkey is looking for a color singleton, the firing rate of a neuron for the color dimension stays up near 50 spikes/s while for that very same neuron the firing rate of only 20 spikes/s when the monkey is looking for a shape singleton. This latter difference that occurs around 175 ms represents how top-down control takes over initial bottom-up capture.

3.3.3. The shift of spatial attention to the irrelevant singleton is stimulus-driven

In the past there have been several arguments challenging the stimulus-driven capture account. One claim that has been quite prominent is the suggestion that capture in the additional singleton paradigm is a top-down choice on part of the observer. According to this view which was first suggested by Bacon and Egeth (1994) observers choose to be in what has been called the “singleton detection mode”. When observers decide to be in this mode they choose to direct attention to the location having the largest feature contrast. When engaged in this mode, the most salient singleton will capture attention regardless of whether it is the target or not. Theeuwes (1991a, 1992) suggested that these conditions represent a clear case of bottom-up capture while Bacon and Egeth (1994) argued that this is a top-down strategy (i.e., the strategy is to engage in a singleton detection mode). There is an alternative mode of search which is labeled “feature search”. When observers choose to be in this

mode they choose to direct their attention to a particular feature (e.g., a green circle) and when choosing this mode there will be no attentional capture. By definition the act of choosing a specific search mode implies the involvement of top-down control and therefore suggests that capture is not necessarily stimulus-driven.

Bacon and Egeth (1994) came to these conclusions on the basis of experiments which were variants of the additional singleton paradigm. Bacon and Egeth first replicated Theeuwes' (1992) additional singleton data showing that the irrelevant color singleton interfered with search for the target shape singleton. In the subsequent experiments, they added additional shapes (i.e., squares and triangles) to the display so that the shape target singleton was no longer unique. In this condition the irrelevant color singleton did not interfere anymore. Bacon and Egeth (1994) suggested that under these conditions observers could not simply use “uniqueness” to find the target. They argued that by adding additional shapes observers could no longer rely on a difference signal detection (referred to as “singleton detection mode”) and had to switch strategies and rely on a “feature search mode”. In a feature search mode, observers are able to exclusively direct their attention to the relevant feature and irrelevant singletons no longer interfere. These results suggest that when observers ‘choose’ a feature search mode, attention capture by irrelevant singletons can be eliminated. The notion that choosing a search strategy allows attentional control suggests that capture is very much under top-down control.

The interpretation of Bacon and Egeth (1994) has been widely accepted and is used as an argument against pure bottom-up capture. On the face of it, the interpretation of Bacon and Egeth (1994) seems reasonable. However, if one adheres this idea it is not immediately clear why in the original additional singleton experiments observers would ‘choose’ a singleton detection mode when this mode will result in erroneous capture towards the color singleton. If choosing these search modes is really under top-down control why do people not choose the more efficient strategy of looking for the specific feature? For example in Theeuwes (1992) observers always search for a green diamond (see Fig. 3). Observers know that the distractor is always a red circle. Why would participants in this condition not switch to a feature detection mode (e.g., pick up the green diamond) when this would prevent attentional capture by the red circle? Note that observers cannot even do this after having received up to 1800 trials (Theeuwes, 1992; Experiment 2).

Also on theoretical grounds the two search modes are questionable. Each time one observes attentional capture, the claim is that people ‘choose’ to have this capture even when this capture is harmful for task. In other words, according to this line of reasoning capture can never be bottom-up because each time one observes capture, the claim is that people choose to be captured. Because people cannot set themselves at will to choose one or the other search mode (at least there is no empirical evidence for this) the concept of a search mode does not explain much, if anything. Moreover, in a recent study Kawahara (unpublished data) showed that even when people report they are searching for a specific feature (in a feature search mode looking for example for a green diamond), the data revealed that they operated in a singleton detection mode. Given the fact that observers are not able to set themselves volitionally for one or the other search mode and even if they claim they do, the data show otherwise the concept of search modes is not very useful.

Recently, Theeuwes (2004) challenged the notion of search modes directly and gave a much simpler explanation for the absence of capture in the displays used by Bacon and Egeth (1994). Theeuwes (2004) argued that their results may have been caused by the heterogeneity of the search displays (i.e. the stimulus) rather than the assumed search modes chosen by the observers. More complex stimuli may force subjects to search in a more focused fashion, which may partly preclude attention capture. By adding different shapes and rendering the display less homogenous, not only may the distractor

singleton become less salient, but also the target singleton may become less salient. As a consequence, one cannot detect the target singleton by parallel (preattentive) search. Observers must engage in serial search to find the target singleton, and as pointed out, it is known that the distraction effect of a singleton is strongly attenuated during partly serial search (see Gibson & Peterson, 2001; Theeuwes, 1991a, 1992; Theeuwes & Burger, 1998). Importantly, even though in Bacon and Egeth (1994) the search slopes were relatively flat (up to 11.5 ms/item), they always differed significantly from zero, suggesting that search may have been partly serial. To make sure that the target and distractor singleton were salient enough to be picked up by parallel search Theeuwes (2004) also used different shape singletons (forcing observers to engage in a 'feature search mode'), but at the same time by adding more nontarget elements increased the salience of target and distractor singleton. Because the target singleton was salient enough, search became completely parallel again (less than 1 ms/item) and the interference by the color singleton re-emerged. So even in conditions in which observers could only use the assumed 'feature search mode' the irrelevant singleton captured attention. The crucial point is that for capture to occur search has to be conducted by the preattentive parallel stage of processing. As soon as search becomes serial or partly serial, the capture effect (which is relatively small to begin with) ceases to exist.

Leber and Egeth (2006) challenged the claims of Theeuwes (2004) in a study which also used heterogeneous displays forcing a feature search mode. After an extensive training phase with the instruction to search for a specific feature (a circle), participants were able to avoid capture by the salient singleton even though search was conducted in parallel. On the face of it this finding seems problematic for Theeuwes' account because this is a case of parallel search without capture by the salient distractor singleton. In that same study, Leber and Egeth (2006) also had a group of observers trained in singleton search. These data show the classic interference by the salient distractor. More importantly however observers in this group responded faster in the distractor absent condition than observers in that very same condition that had learned to search for a feature. In footnote 3, Leber and Egeth (2006) speculate that the distractor interference effect in the feature search group may have been concealed by slower responding of observers that have learned to use the feature search mode. Consistent with our claims, by slowing down search (possibly by setting a smaller attentional window) the effect of the distractor should disappear. It seems that the only thing that observers learned in feature search is to narrow their attentional window and to slow down search. Recent data of Leber (2009) looking at whole brain pretrial activity to predict capture effects on the upcoming trial are consistent with this idea. Leber (2009) showed that on trials in which pretrial signal was low, capture was substantial; when pretrial signal was high, capture was negligible. However, the behavioral data showed that in trials in which capture was negligible observers responded relatively slow while when capture was present observers responded relatively fast. It seems that the only thing that pretrial activity predicts is whether observers are going to be slow or fast on the next trial which in turn will determine whether there is capture or not.

On the basis of these findings, there is no need to assume the hypothetical search modes. The extent to which attention is spread across the display seems to determine whether a singleton, being a target or a distractor, will capture attention. If observers search in parallel for the target singleton, attention needs to be divided across the display (a wide attentional window), any salient element within that window will capture attention in a stimulus-driven way.

3.3.4. The role of the size of the attentional window

One of the premises of the stimulus-driven capture approach is that salience computations are restricted to the attentional window of the observer. Within the attentional window, top-down control

cannot preclude attention from being captured by the most salient feature singleton. Even though capture itself within the attended window is completely stimulus-driven, the extent to which people divide their attention across the display (the size of the window) is under top-down control. As outlined in the previous paragraph, if search is not conducted in parallel across the display but more focused as seen in the non-flat search function of Bacon and Egeth (1994), salient singletons do not capture attention. In effect, we claim that the only top-down control over capture is through varying the attentional window from highly focused to widely distributed. Because preattentive analysis is limited to the attended area, the extent to which attention is spread determines the occurrence of attentional capture. When attention is spread, visual search may be conducted in parallel across all items in the visual field, at the expense that any singleton relevant or irrelevant, will be selected. However when the attentional window is set to a smaller size, singletons that fall outside of the attentional window will not capture attention. When observers are in a focused mode, selection of the next item is random (as for example in Jonides & Yantis, 1988). We propose that salience computations are restricted to the attentional window of the observer. Our claim is that while the size of the attentional window is under top-down control, within the attentional window top-down control cannot preclude attention from being captured by the most salient feature (Belopolsky et al., 2007; Theeuwes, 2004).

The notion that focusing attention prevents attentional capture dates back to studies conducted in the early nineties (Theeuwes, 1991b; Yantis & Johnston, 1990). For example, Theeuwes (1991a) showed that focusing attention on a location in space by means of an endogenous central cue prevents attention capture by abrupt onset presented elsewhere in the display (see also Yantis & Johnston, 1990 for a similar finding). Similarly, focusing attention before display onset prevents capture of the eyes by abrupt onsets (Theeuwes et al., 1998). In a recent study by Belopolsky et al. (2007) the size of the attentional window was explicitly manipulated. To ensure that observers spread their attention across the display, before they could start searching, they had to make a judgment regarding the global spatial lay-out of the elements in the display. Specifically, before searching, observers had to decide whether the elements in the display formed an upward or downward pointing triangle. Only when it was pointing upward they could start their search. In order to make this judgment about the lay-out of the display, observers had to spread their attention. In another condition, Belopolsky et al. ensured that observers focused their attention before they could start searching. In this case only when the fixation point was a circle were observers allowed to start searching. The results showed that when attention was initially focused in the center (focused attention condition), the salient color singleton was basically ignored confirming the findings of Jonides and Yantis (1988). However, when attention was initially diffused over the global stimulus arrangement (diffuse attention condition), the irrelevant color singleton was often selected first.

Recently, Belopolsky and Theeuwes (in press) tested the idea of the attentional window using the classic additional singleton task of Theeuwes (1992). They found that when observers were in a diffuse attentional state, the classic attentional capture interference effect was found: the presence of an irrelevant color singleton slowed search for a shape singleton. However, this very same capture effect was abolished when just before the presentation of the display attention was in a focused state. If attention was not spread over the display, but focused in the center, the presence of an irrelevant singleton no longer captured attention. Belopolsky and Theeuwes (in press) concluded that the attentional window is a determining factor in the occurrence of attentional capture. The findings of Belopolsky and Theeuwes are consistent with results reported by Proulx and Egeth (2006). They showed that the distraction caused by an irrelevant feature (a bright singleton) was

modulated by target–nontarget similarity. With increasing target–nontarget similarity search became more difficult, and with increasing search difficulty, the effect of the presence of the irrelevant bright singleton was reduced. These findings are fully consistent with the idea that when search becomes more difficult, the attentional window needs to be smaller, causing a reduced effect of the irrelevant distractor on search. As outlined before, the same reasoning holds for the findings reported by Bacon and Egeth (1994).

3.4. Conclusions

We argue that the first sweep of information through the brain is basically stimulus-driven and automatic. We assume that the attentional capture – the initial shift of attention to the most salient singleton – is the result of a bottom–up mechanism, which is triggered by the presence of feature difference signal interrupt. Independent of top–down control, attention will first shift to the most salient object in the visual field. Only by varying the spatial window before the display comes on, the first sweep of information through the brain can be modulated. It is assumed that salient information outside the attended area does not capture attention anymore. We argue that attentional capture is spatial such that attention is shifted to the location having the highest salience. Only after attention has shifted to that location (after the object has been selected), the identity of the object at that location becomes available. If the selected object does not look like the target at all, disengagement from that location is fast and swift. If the selected object resembles the target disengagement may be slow and effortful. Disengagement of attention after attention has been captured is very much under top–down control. The size of the capture effect in ms RT depends on the speed with which disengagement is possible. Capture has been associated with activity in the parietal areas (possibly only the right side); disengagement of attention is associated to the frontal areas.

4. Competing viewpoints

This section discusses the major competing viewpoint regarding attentional capture. We will first explain the basic premises of each of these viewpoints and then discuss alternative explanations for these findings.

4.1. Contingent capture

4.1.1. Evidence

Completely opposite to the idea of stimulus-driven attentional capture is the hypothesis that capture is always contingent on top–down control settings. This view, known as the contingent capture hypothesis (e.g., Folk, Remington, & Johnston, 1992; Folk, Remington, & Wright, 1994) argues that selection always depends on the explicit or implicit goals held by the observer at any given time. The idea that selection is completely under volitional top–down control comes from a paradigm known as the spatial precueing paradigm developed by Folk and Remington (1992). In this paradigm the search display is preceded by a cue display. In their classic 1992 study, the target display consisted of either a color or an onset singleton and observers were required to identify the unique element. In the color display, the target was red while the other three elements were white. In the onset display, only one element was presented, and so the target was characterized as being the only element presented with an abrupt onset. Immediately preceding the target display at an 150 ms SOA, a cue display was presented: this cue display either consisted of a color cue (in which one location was surrounded by red dots and the other three locations were surrounded by white dots) or an onset cue (in which one location was surrounded by an abrupt onset of white dots and the remaining locations remained empty). All conditions were

factorially combined and the target type remained constant within block of trials (see Fig. 7).

The critical finding of Folk et al.'s studies was that only when the search display was preceded by a to-be-ignored featural singleton (the “cue”) that matched the singleton for which observers were searching, did the cue capture attention. Thus, when searching for a red target singleton, attention automatically shifted to the location of the irrelevant red cue that preceded the search display, while the irrelevant onset had no effect on performance. The result suggests that the top–down attentional set determines the selection priority: when observers are set for a particular feature singleton, only elements that match this top–down attentional set will capture attention. Feature singletons that do not match top–down attentional sets will simply be ignored: “With a control setting established, events exhibiting the critical properties will involuntarily summon attention, whether or not the event is actually relevant to task performance. Stimuli not exhibiting these properties will not involuntarily summon attention” (p. 1041, Folk, Remington & Johnston, 1992). The underlying theoretical notion of the contingent capture hypothesis proposed by Folk and colleagues is that capture is fully contingent on the top–down set adopted by the observer and salient objects with irrelevant properties are simply filtered out (Folk & Remington, 1998). The contingent capture hypothesis appears to be the generally accepted way to account for top–down control in visual search. The classic findings of Folk have been replicated many times using various modifications of the classic paradigm (Folk & Remington, 1998; Folk et al., 1994; Gibson & Amelio, 2000; Gibson & Kelsey, 1998; Pratt et al., 2001; Remington, Folk, & McLean, 2001 but see Yeh & Liao, 2008).

4.1.2. Alternative interpretation of contingent capture

The contingent capture hypothesis represents a view that is completely opposite of the stimulus-driven capture account. To explain these differences it is important to analyze the differences between the paradigms on which the empirical evidence is based. The contingent capture hypothesis is based on a cueing paradigm in which observers have to ignore a “cue” that is presented just before the

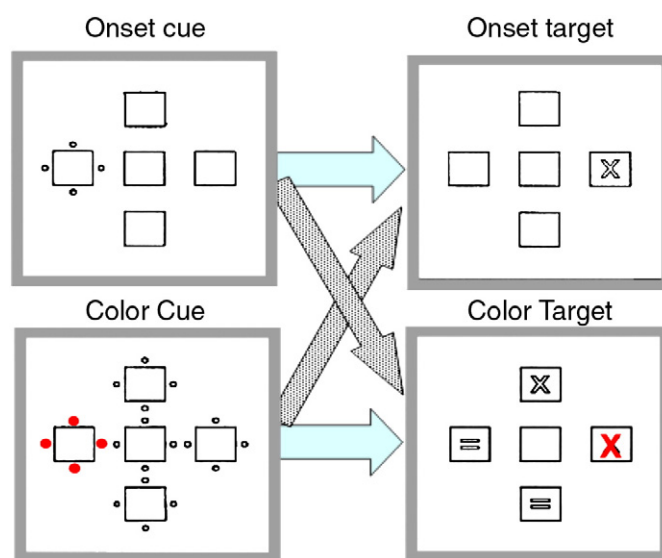


Fig. 7. The contingent capture paradigm of Folk et al. (1992). Observers had to respond to a target singleton (either an “X” or an “=”). The target was defined as singleton which had a unique color (“color target” condition, bottom-right) or was the only element presented as an onset (“onset target” condition, top-right). Each type of target display was preceded by a cue display. The cue display consisted of either an onset cue (top left) or a color cue (bottom-left). All conditions were factorially combined. The important finding was that each cue type (onset versus color cue) only captured attention when observers were set to look for it. In this example, the cue is invalid because the cue appears at a location that does not coincide with the target.

presentation of the search display. When the search display is preceded by a to-be-ignored featural singleton (the “cue”) that matches the singleton for which observers are searching, the cue captures attention, as evidenced by a prolonged reaction time to identify the target (i.e., when the cue and target appear in different spatial locations). On the other hand, if the to-be-ignored featural singleton “cue” does not match the singleton for which they are searching, its appearance did not have an effect on responding, i.e., the cue did not capture attention. The critical finding in these studies is that a cue that does not match the top-down search goal (i.e., the defining property of the target) does not affect RT (i.e., a null effect), while a cue that matches the search goal has an effect on RT.

The fact that in this paradigm the target and the distractor (the cue) are never presented simultaneously, and that top-down control is inferred on the basis of a null effect, allows an alternative interpretation. As discussed in Section 3.3.2, we have provided evidence that when the distractor and target are dissimilar, capture is short-lived. In other words, if attention is captured by a distractor that does not look like the target, observers are able to immediately withdraw attention from that location. In our study (Theeuwes et al., 2000) we demonstrated that in the additional singleton paradigm capture is no longer present when the distractor precedes the target with an SOA of 150 ms. This implies that capture may only last 100 ms and possibly even less when the distractor does not look at all like the target. Therefore, it is quite feasible that also in Folk et al.'s spatial cueing paradigm the irrelevant cue did capture attention. Because there is typically a delay of 150 ms between the presentation of the cue and the search display, observers may have been able to overcome the attentional capture by the time the search display was presented (but see Chen & Mordkoff, 2007). More specifically, in the spatial cueing paradigm of Folk, Remington and Johnston (1992) disengagement of attention from the cue may have been relatively fast when the cue and target do not share the same defining properties (e.g., the cue is red and the target is an onset), while disengagement from the cue may be relatively slow in the case where the cue and target share the same defining properties (e.g., both were red). Such a mechanism could explain why there are RT costs when the cue and target have the same defining characteristics and no costs when cue and target are different. In this view, the contingent capture hypothesis can explain why it may be easier to disengage attention from a particular location when an element presented at that location is not in line with the top-down control settings. However, this does not imply that there is no capture of attention by the irrelevant cue singleton; it simply indicates that after a certain time participants are able to exert top-down control over the erroneous capture of attention by the irrelevant singleton.

In a recent study, Anderson and Folk (2010) provided evidence for the idea that spatial cueing says something about the disengagement of attention after attention has been captured rather than about capture itself. In this study Anderson and Folk (2010) used cues that contained a feature that varied in the way it resembled the target feature observers were looking for. For example, observers were set to look for red because the target was red. The cue consisted of the color red which was identical to the target (as in the original experiment of Folk et al) or could be an orange color similar to red, or an orange color similar to green, etc. The data showed that the capture effect (valid RTs minus invalid RTs) depended on the resemblance of the color of the target with that of the cue: More capture when the cue looked like the target than when it did not look like the target. Importantly, however, almost all effects of the target distractor similarity were obtained in invalid trials: the less the cue looked like the target, the less time it took to disengage attention from the invalid cue location. Importantly, the RTs in valid trials were about the same independent of whether the color of the cue looked like the target or not. In other words, attentional capture was not affected by cue-target resemblance but the speed of disengaging attention from the invalidly cued location was. These data provide compelling evidence that the cueing effect in the cueing paradigm of Folk et al. is at least strongly related to the speed

of disengaging attention from the invalid cue location. Note however that Anderson and Folk (2010) interpret these data in quite a different way. They argue that the similarity between the target and distractor determines the extent to which attention is captured by the distractor. In other words, they claim that when the cue looks more like the target the RTs on valid trials should be faster than when the cue does not look much like the target. However, the data seem to indicate differently. The variation in the capture effect due to target-distractor similarity (valid RTs minus invalid RTs) is for the larger part determined by the RTs on invalid trials. The RTs on valid trials are not systematically affected by the similarity between the target and the distractor.

In a recent study Belopolsky, Schreij and Theeuwes (2010) further challenged the contingent capture hypothesis. They used exactly the same spatial cueing paradigm as Folk, Remington and Johnston (1992). Rather than keeping the target fixed over a whole block of trials (as was originally done with contingent capture experiments), participants were encouraged to adopt a top-down set before the start of each trial. In other words, observers were cued at the beginning of each trial to either look for a unique color or the unique onset. If, as claimed by the contingent capture hypothesis, top-down attentional set determines which property captures attention, then one would expect that only properties that match the top-down set would capture attention. Belopolsky et al. showed that even though participants knew what the target would be on the upcoming trial, both relevant and irrelevant properties captured attention. In other words, there was no sign of contingent capture, instead both the relevant cue that matched the target as well as the irrelevant cue, captured attention.

In subsequent experiments Belopolsky et al. (2010) forced observers to adopt the strongest possible top-down set. On each trial observers had to choose themselves which attentional set they wanted to establish for that particular trial (e.g., either search for color or for an onset). Also, observers had to repeat out loud the chosen target feature until the search display appeared. Observers were also free to choose when to start the trial and to decide how much time they needed to establish the top-down set. Asking observers to rehearse attentional set vocally minimized reliance on working memory and prevent potential lapses of attentional set associated with lapses in working memory maintenance. With these measures it was ensured that on each trial observers established and maintained their top-down control for a specific feature. In these experiments, Belopolsky et al. found a clear suppression of the cue that did not match the search feature. For example, when searching for an onset, observers were extremely slow when the target was presented at the location of the color cue. Unlike in Folk et al.'s experiments, observers did not simply ignore the irrelevant cue; they actively suppressed it. Obviously, such suppression represents a form of top-down control. Note, however, that it is a very different type of control than the one suggested by the contingent capture hypothesis. According to the contingent capture hypothesis, the cues that do not match the top-down set simply do not cause involuntary attentional shifts (Folk, Remington & Johnston, 1992). Such cues are assumed to be filtered out, which sometimes results in a non-spatial filtering cost (Folk & Remington, 1998). Belopolsky et al. reported spatially-selective suppression at the location that contained a feature that did not match the top-down set. This was the case for both color cue as well as onset cues. Belopolsky et al. claimed that such suppression can only take place after attention has been at that location. Other studies have reported a similar type of suppression at the location of the irrelevant singletons but mostly only for color cues (Kim & Cave, 1999; Lamy & Egeth, 2003; Lamy, Leber, & Egeth, 2004).

Belopolsky et al. concluded that even when observers adopt a top-down attentional set on a trial-by-trial basis, they cannot prevent attentional capture by the irrelevant cue. Forcing observers to adopt a top-down set on every trial sped the disengagement of attention, consistent with the disengagement hypothesis discussed earlier. If the

top-down set is very strong, disengagement can lead to suppression of the location of the irrelevant cue. It is assumed that this spatially-selective suppression occurs after attention has been captured by the irrelevant cue. This view is consistent with the idea that the primary role of the top-down set is to control the disengagement of attention from the features that do not match it. The speed of disengagement is determined by the strength of the top-down set: if the top-down set is strong — attention will be quickly disengaged from the salient item that does not match it; if the top-down set is weak — attention will linger at that location.

Belopolsky et al. made another important observation. An inter-trial analysis of their data showed that previous contingent capture findings may to a large extent be explained by inter-trial priming. Indeed, in the typical Folk et al.'s paradigm observers search a whole block for one particular target (a color singleton or an onset), which makes it possible that inter-trial priming drives the contingent capture effect rather than the assumed top-down set. Their results suggest that the concept of “attentional set” as currently proposed by the contingent capture hypothesis on the basis of the cueing paradigm is not purely top-down. In fact, when the search target is fixed, voluntary top-down control does not have to be present in a continuous fashion on each trial during the search task. Indeed such a strategy would make our behavior very slow and would require a lot of resources. Once the target parameters defined by instruction have led to a correct response, attentional selection can carry on based on the bottom-up inter-trial priming. This means that voluntary control over selection is not necessary and attentional set is established after the first few experiences with the target. The function of top-down control involves only the monitoring — verification and checking of whether the selected information matches the goal, as well as disengagement, but top-down control itself does not control the selection process.

In another study, Schreij, Owens and Theeuwes (2008) using the spatial cueing paradigm of Folk et al., had observers establish a top-down set for color because they searched for a whole block for a uniquely colored target. Consistent with the contingent capture hypothesis, Schreij et al. reported cueing effects of the relevant color cue. However, even though the data showed this clear attentional set for color, the presentation of an irrelevant onset caused RT interference during search. Schreij et al. (2008) attributed this increase in RT to attentional capture by the irrelevant onset. However, Folk et al. (2009) attributed this effect to non-spatial filtering instead of attentional capture. Schreij et al. (2010) provided compelling evidence that the interference effect of the onset was not caused by non-spatial filtering. Moreover, in more recent experiments Schreij, Theeuwes and Olivers (in press) showed that there was IOR at the location of the onset providing evidence that attention was captured by the onset, since it is assumed that IOR is only observed when attention is exogenously captured to a location in space (Posner & Cohen, 1984). Interestingly, there was no sign of IOR at the location of the color cue (see also Gibson & Amelio, 2000 for a similar result; Pratt et al., 2001), which led Schreij et al. (in press) to conclude that the capture that is assumed to be contingent on top-down setting may not be capture at all. As noted, if it would have been capture, which is by definition is exogenous in nature, one would have expected to observe an IOR effect at the location of the color cue. Since no IOR was found, it was concluded that the shift of attention to the matching cue may have been endogenous in nature (for a similar argument see Yantis, 1993).

In conclusion, the contingent capture hypothesis claims that attentional capture is dependent on top-down settings. On the basis of recent findings (Belopolsky et al., 2010; Schreij et al., 2008, 2010, in press) we believe that capture that is contingent on what people are looking for is not exogenous in nature. When attention moves to the cue that matches the feature observers are looking for, we claim that this shift of attention is endogenous in nature. This is not surprising because people are looking for example for red and just before the target display is presented, another display is presented that contains

a feature that matches that of the target (see Fig. 7). The shift of attention is endogenous in nature because it fits the goal of the observer, and therefore will not generate IOR. When a cue is presented that does not fit the top-down goal (one is looking for color and an onset is used as a cue) then there is true exogenous capture by the non-matching cue but this capture is very brief and will not show up in the typical spatial cueing task as used by Folk and colleagues because of the time interval between the cue and the search display. Capture will show up however when the onset is presented during the search display (as in Schreij et al., 2008, 2010, in press). This fast disengagement of attention can result in suppression at the location of the irrelevant cue as shown by Belopolsky et al. (2010). Suppression which is endogenous in nature occurs after spatial attention has exogenously been captured by the location containing the irrelevant cue.

4.2. Only onsets capture attention

4.2.1. Evidence

On the one hand the contingent capture hypothesis claims that all capture is dependent on top-down settings while on the other hand the stimulus-driven capture account argues that initially capture is purely bottom-up in nature. A view that is somewhere in the middle is the notion that only abrupt onsets have the ability to capture attention in a truly bottom-up way, with other properties, such as color, being dependent on top-down settings. In the late 1980s, Yantis and colleagues (Jonides & Yantis, 1988; Yantis & Egeth, 1999; Yantis & Jonides, 1984) conducted several studies investigating whether feature singleton receive attentional priority (see also Theeuwes, 1990). Yantis and colleagues adopted a visual search task, such that the target of search was a non-singleton letter. This type of search is not efficient because search times increased linearly with the number of elements present in the display. In each display, there was always one salient element and the question addressed was whether search would automatically start at the salient element. With N as the number of elements in the display, the salient element was the target on $1/N$ of the trials, indicating that the chance that the salient element was the target was the same as for any other letter. Since the salient element was the target at chance level, there was no incentive to deliberately start searching at the salient singleton (see Fig. 8). Jonides and Yantis (1988) showed that observers did not start searching at the salient element in the display. When the unique element happened to be the target (e.g., an element with a unique color or unique luminance), the search slopes were basically the same as in the condition in which a non-unique element was the target (see panels B and C; compare the “unique” and “common” search slopes). It was concluded that salient static singletons are treated in the same way as other non-salient elements in the visual field. Uniqueness in color or luminance is not sufficient to capture attention when it is irrelevant to the top-down goal. More importantly, however, Jonides and Yantis (1988) showed that elements appearing with an abrupt onset have a special status in capturing attention irrespectively of the top-down settings (see panel A).

Overall, Jonides and Yantis (1988; see also Yantis & Egeth, 1999) claimed that a feature singleton (such as an element having a unique color or brightness) is not automatically selected. Only when the element is presented with abrupt onset, it receives attentional priority. Yantis & Egeth (1999) claimed that selection is under top-down control except for elements that are presented with abrupt onset when constituting a new object (e.g., Yantis & Hillstrom, 1994, but see Franconeri, Hollingworth, & Simons, 2005). Others have claimed that motion captures attention in a bottom-up way (e.g., Abrams & Christ, 2003; Theeuwes, 1990). The previously discussed study of Schreij et al. (2008) confirmed the special role for onsets showing attentional capture by onsets while observers had a top-down set for color.

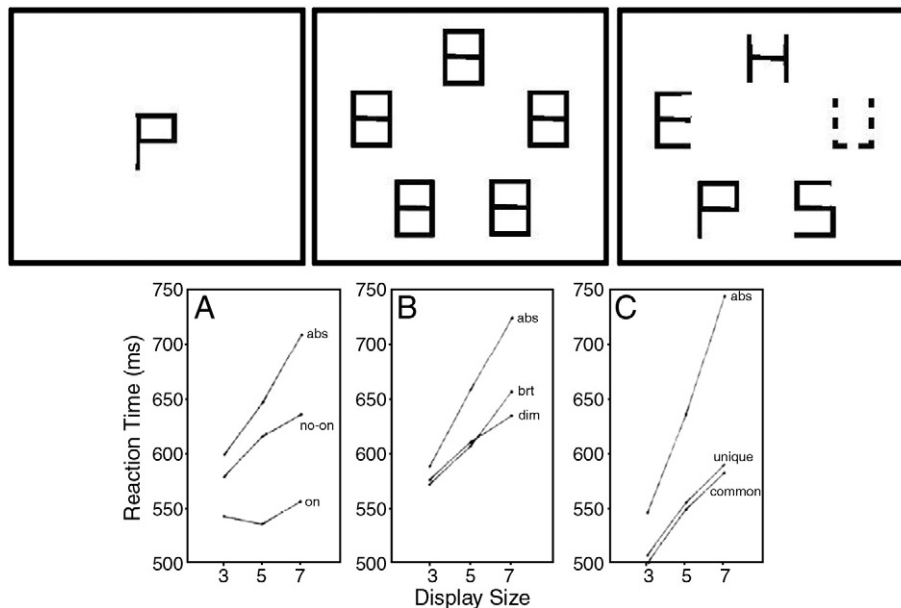


Fig. 8. Paradigm and data from Jonides and Yantis (1988). In the first display a target letter was displayed for 1000 ms (in this case the letter P) followed by a premask display for 1000 ms. In the search display one letter had a unique color (dotted lines). At chance level this letter could be the target. The results show that observers do not start searching at the unique feature (panel B: unique brightness; panel C: unique color). Note that when the unique feature is an abrupt onset (panel A), observer does start searching at the unique feature (i.e., the abrupt onset), on = present-onset; no-on = present-no-onset; abs = absent; brt = present-bright; dim = present-dim; common = present-common color; unique = present-unique color.

4.2.2. Alternative interpretation

Even though it is clear that due to their salience onsets (and motion) should be able to capture attention in a bottom-up way, the observation that other salient singletons such as color or luminance singleton do not capture attention in the Yantis type search task appears to be inconsistent with the stimulus-driven capture account. Indeed, in the additional singleton paradigm Theeuwes (1991a, 1992, 1994) showed that static singletons of color, shape and luminance do capture attention. One way to reconcile why static singleton in the Yantis type of search task do not capture attention while these very same singletons do capture attention in the additional singleton task, is the idea that the attentional window adopted by the observers may have played a determining role in the occurrence of capture. In the classic Yantis task (see Fig. 8) observers are looking for a target letter among a multiple number of other nontarget letters, a search task which requires effortful serial search as is clear from the large search slopes. This implies that observers may have adopted a small attentional window to allow focused serial search. As discussed before (Section 3.3.4) when a small window is used, no static singletons will capture attention, explaining why the static singleton do not capture attention in the serial search task of Yantis and colleagues. Along similar lines, in the additional singleton task, observers spread attention across the display allowing preattentive parallel search and in that case any singleton within the attended area grabs attention.

As discussed earlier, the Belopolsky et al. (2007) tested this idea directly by manipulating the size of the attentional window in a task that was very similar to that of Jonides and Yantis (1988). As noted, the results revealed that observers which were in a focused attention state, the salient color singleton was examined just as frequently as the other elements in the display, a result similar to the “classic” finding of Jonides and Yantis (1988). However, when attention was initially diffused over the global stimulus arrangement (diffuse attention condition), attention more frequently went to the location of the color distractor, which was evidenced by faster responses and a significantly reduced search slope when the colored element happened to be the target. This finding suggest that contrary to the belief that only onsets can capture attention, static singletons do so as well as long it is ensured that observers spread their attention across the display.

Even though the Belopolsky et al. study clearly indicates that the attentional window plays a role in attentional capture of static

singletons, it should be noted that abrupt onsets may be special in the sense that they can capture attention even when people are not dividing attention across the display (as for example in Jonides and Yantis, 1988). The most obvious reason is that onsets are more salient than static singleton and therefore they may capture attention even when people do not divide their attention across the display. There are many studies showing that onsets have a special status in attentional prioritization (see e.g., Belopolsky, Theeuwes, & Kramer, 2005; Donk & Theeuwes, 2001, 2003; Theeuwes, 1994b, 1995a).

4.3. Top-down weighting of stimulus features or stimulus dimensions

4.3.1. Evidence

There is ample empirical data that suggests that top-down settings can affect the selection of a feature singleton. For example, Treisman (1988) showed that knowing the dimension of the target singleton (whether it would be a color or shape singleton) speeded search with about 100 ms. Treisman (1988) suggested that there is no top-down selectivity within dimensions; yet, across dimensions knowing in what dimension the target will be presented speeds up search significantly. Along similar lines, Müller et al. (1995) had observers search for three possible targets which all were defined within one dimension (e.g., orientation) or defined across dimensions (e.g., orientation, color, and size). They showed for example that the detection of a common right-tilted target was 60 ms slower in the cross-dimension relative to both the intra-dimension condition and the control condition. In similar experiments, Wolfe et al. (2003) had observers search a whole block of trials for a red target between green nontargets (i.e., color singleton) or for a vertical line between horizontal line segments (i.e., shape singleton). These blocked conditions were compared to mixed conditions consisting of blocks of trials in which the target could either be red, green vertical or horizontal. On the basis of these data, Wolfe et al. (2003) concluded “top-down information makes a substantial contribution to RT even for the simplest of feature searches. Fully mixed RTs are about 80 ms slower than are blocked RTs” (p. 485).

Even though these studies unequivocally show that knowing what the target is going to be has a large effect on search, these effects cannot necessarily be attributed to top-down set because in these

experiments blocked conditions in which the target stays the same during a whole block of trials are compared to mixed conditions in which the target changes from trial to trial. As discussed before (Pinto et al., 2005) when the target stays the same across a block of trials, there may be large inter-trial priming effects which have nothing to do with top-down set. In other words, observers are not faster in a blocked condition because they actively prepare for the upcoming target singleton (as a top-down approach would assume) but are faster because the target singleton on the current trial is simply the same as the one on the previous trial causing strong inter-trial bottom-up priming effects (see also Belopolsky et al., 2010).

This shortcoming was recognized by Müller et al. (2003). To rule out the possibility that the effects could be due to passive bottom-up priming they introduced a trial-by-trial cueing procedure. In Müller et al.'s (2003) experiments, observers had to indicate presence or absence of a feature singleton. This feature singleton could have a unique orientation (a tilted line segment between vertical line segments) or a unique color (a red or blue line segment between green segments). The dimension of the upcoming target singleton (if present) was cued with a verbal cue (the word "color", or "shape") with an 80% validity. There was also a condition with the verbal cue "neutral" in case observers did not know which dimension would be cued. The results showed a clear cueing effect: when the cue was valid (they were cued with the word "color" and a color singleton was present) observers were faster than in the neutral condition. Similarly, when the cue indicated the wrong dimension (the word "color" while the target was an orientation singleton) observers were slower. These data are considered as evidence that top-down knowledge can improve the selection of a feature singleton. Knowing what you are looking for improves visual search for the target and speeds up the attentional selection process.

To account for data like these, Müller and colleagues developed a 'dimension-weighting' account of visual selection (Found & Müller, 1996; Müller et al., 2003). For example, when observers know in advance the target dimension they are looking for, they can place more attentional weight on that dimension. This placing of more attentional weight is a top-down process and speeds up the selection of the target. The reasoning is that target detection involves "an attentional mechanism that modifies the processing system by allocating selection weight to the various dimensions that potentially define the target" (Müller et al., 2003; p. 1021). According to the dimension-weighting account, there is a limit to the total attentional weight available to be allocated at any one time to the various dimensions of the target object. It is assumed that potential target-defining dimensions are assigned weight in accordance with their instructed importance and their variability across trials. The greater the weight allocated to a particular dimension, the faster can the presence of a target defined in that dimension be discerned. It is argued that the weight assigned to a given dimension is modulated in a top-down way. The weight to the target dimension can be increased or the weight to the distractor can be attenuated. These processes enhance the salience signal produced by the target at the master map level, while reducing the activity that is generated by the distractor, so that the target is more likely to win the competition for focal attention. Dimensional Weighting is in fact similar to Guided Search (Wolfe, 1994) except that it focuses specifically on stimulus dimensions while Guided Search assumes that top-down set can also be assigned to stimulus features specific signals.

4.3.2. Alternative interpretation

Accounts like Guided Search (Wolfe, 1994) and Dimensional Weighting (Müller et al., 2003) are problematic for the stimulus-driven attentional capture account because these theories assume that selection of a singleton can be modulated in a top-down way. Moreover, the experiments discussed (Müller et al., 2003) in which a verbal cue is effective in speeding up response time are in inconsistent with the earlier discussed results of (Theeuwes et al., 2006, 2008; see

Section 3.2.1.2, Fig. 4) who used the same procedure (i.e., a verbal cue) but did not report an effect of cue validity on RT.

Even though these results appear to be conflicting Theeuwes et al. (2006) provided evidence that the cueing effects on feature search may not affect visual selection processes but rather processes following attentional selection. In one of the experiments of Theeuwes et al. (2006) a verbal cue was used in a task in which observers had to decide whether a singleton was present or not. When a shape singleton or color singleton was present observers had to respond "yes" by pressing one of two buttons. When all elements were identical (all were circles) they responded "no" by pressing the other button. Fig. 4 gives an example of a trial. In this way the task requirements were identical to Müller et al. (2003) and Theeuwes et al. (2006) basically replicated the main findings of Müller et al. (2003) As in Müller et al. there was a cue validity effect in the sense that observers were fast when the cue was valid and slow when it was invalid. However, in subsequent experiments, Theeuwes et al. (2006) kept the task identical except observers no longer had to respond to the presence or absence of the singleton but had to respond to the line segment placed inside the target singleton (see also Theeuwes & Van der Burg, 2007). Just by changing the response requirements the reliable cueing effect that was obtained in the feature detection experiment in which observers responded to the presence or absence of the target singleton was no longer present when it became a compound search task (see the data of Fig. 4). Cueing the singleton on the upcoming trial had no effect whatsoever (see Section 3.2.1.2 for a detailed description). Note however that Müller and Krummenacher (2006) did report dimensional cueing effects in a compound search task.

Theeuwes et al. (2006) argued that changing the response requirement should not have affected the cue validity effect as long as the cue speeds up the selection the singleton. If the cue has an effect on visual selection in the sense that the pre-knowledge provided by the cue allows a better guidance of attention towards the location of the target singleton, as assumed by Dimensional Weighting (Müller et al., 2003) and Guided Search (Wolfe, 1994), then changing the response requirements should have no effect on visual search (for a replication using sensitivity as a dependent measure see Theeuwes & Van der Burg, 2007). Theeuwes et al. (2006) argued that in previous visual search studies in which observers responded to the presence or absence of the feature singleton as for example in Müller et al. (2003) and Wolfe et al. (2003), the cue validity effects reported do not represent effects that operate on visual selection but rather on response selection. Indeed, it is assumed that the knowledge that the cue provides does not speed up attentional guidance towards the singleton but affects the speed with which the response button can be pressed after the item had been selected. In terms of the attentional selection illustration shown in Fig. 2, this implies that the cue operates on the last process instead on the selection process as these theories assume.

A study by Mortier et al. (2005) provided compelling evidence for this idea. That study used also a cueing procedure indicating the dimension of the upcoming target and instead of responding to the line inside the singleton observers responded to the presence or absence of a shape or color singleton. Observers were cued with the word "color" or "shape" with a cue validity of 80%. The crucial manipulation in these experiments was that instead of a visual search experiment, observers responded to a single object presented in the center of the display. The target could either be a red or green circle (color dimension) or a gray triangle or square (shape dimension). In target absent trials a single gray circle was presented in the center of the display. Even though these was nothing to search for (there was one single object presented in the center), the same cueing effects were observed as in the visual search experiments of Müller et al. (2003). It was concluded that the cueing effects which are typically attributed to top-down guidance-of-search processes, also occur in conditions in which there is no search. Therefore, previous studies that have suggested top-down effect on visual search processes probably showed effects that are related to response selection. Given

this analysis, there is not much, if any, evidence that non-spatial information can guide the initial visual selection process (see also Section 3.2.1.2). Clearly, top-down effects do play a large role after the target singleton has been selected in the sense that it may speed up the response to the target singleton.

Recently, Geyer et al. (2008) and Müller, Geyer, Zehetleitner, & Krummenacher (2009) conducted a series of experiments which they interpreted as evidence for the Dimensional Weighting account. They employed the classic additional singleton paradigm and varied the probability of the presence of a distractor singleton. Note that in the original experiments of Theeuwes (1991a, 1992, 1994) the distractor singleton was present on each and every trial. Müller et al. (2009) varied the proportion of distractor singleton trials between 0% and 100% (from never a distractor to always a distractor). In addition, they varied the amount of practice. The results indicate that the interference caused by the distractor varied with the amount of practice and with the proportion of distractor singleton trials. Overall, interference (the size of the effect in RT) was larger when there was less practice and when the probability that a distractor was present on a given trial was low. It was concluded (Geyer et al., 2008; Müller et al., 2009) that these findings provide strong evidence against that the stimulus-driven capture account of Theeuwes (1991a, 1992, 1994a,b) because their findings show that distractor interference is top-down modulable. The findings are interpreted as evidence for the Dimensional Weighting account (Müller et al., 1995, 2003) because it is assumed that observers apply a top-down dimensional weight to block the interference of distractor dependent on the likelihood that a distractor will be present. In other words, if observers expect that a distractor is going to be present then they modulate the attentional weights in such a way that the salience of the distractor is reduced, not causing much, if any, interference.

Even though an explanation in terms of top-down attentional weights is possible, finding a modulation of the distractor effect as reported by Geyer et al. (2008) and Müller et al. (2009) is also completely compatible with stimulus-driven account. As outlined, the size of interference effect in RT ms in the additional singleton task is less determined by the amount of capture, but much more by the speed with which attention can be disengaged from the distractor location after it has been captured. If the distractor is relatively rare (as in Müller et al., 2009 experiments), the speed of disengagement will be relatively slow because an infrequent distractor introduces more uncertainty. Observers have to determine each trial whether they were captured by the distractor or by the target. This decision process takes more time when observers are unsure about the presence of the distractor singleton. Clearly, the stimulus-driven account can easily explain that there is more capture when the distractor is presented less frequently and/or when observers have less practice. Our interpretation fits nicely with the general notion that stimulus-driven capture and top-down disengagement go hand in hand.

It should be noted that the general idea of Dimensional Weighting (Müller et al., 1995, 2003) or Guided Search (Wolfe, 1994) in which weights are allocated to a particular dimension or features can be compatible with the view of stimulus-driven selection as long as it is assumed that the weights are shifted in bottom-up way mainly driven by priming. Like Dimensional Weighting and Guided Search, the stimulus-driven account also assumes that weights are shifted, but not by the person in a volitional way but by previous experience with the stimulus.

5. Overall conclusions

The present paper argues for the notion that during the first sweep of information through the brain (<150 ms) visual selection is completely stimulus-driven. Initial selection is based on the salience of objects present in the visual field. The salience of objects can be modulated by bottom-up priming. Only later in time (>150 ms), through massive recurrent feedback processing, volitional control based on expectancy and goal set will bias visual selection in a top-

down manner. Top-down knowledge regarding non-spatial features of the objects in the visual field (such as color, shape, luminance, etc.) cannot alter the initial selection priority. Only by adjusting the size of the attentional window, the initial sweep of information through the brain may be altered in a top-down way.

We have argued that alternative viewpoints that assume a large role for top-down selection may in fact have addressed those processes that occur later in time following initial selection. The speed with which one is able to disengage attention from a location to which attention has erroneously moved, can explain many findings that appear to say something about initial top-down selection but in fact reveal top-down processes following initial selection.

Acknowledgements

I would like to thank Hermann Müller, Claus Bundesen and Kyle Cave for their excellent comments on a previous version of this paper.

References

- Abrams, R. A., & Christ, S. E. (2003). Motion onset captures attention. *Psychological Science*, 14(5), 427–432.
- Anderson, B. A., & Folk, C. L. (2010). Variations in the magnitude of attentional capture: Testing a two-process model. *Attention Perception & Psychophysics*, 72, 342–352.
- Ashbridge, E., Walsh, V., & Cowey, A. (1997). Temporal aspects of visual search studied by transcranial magnetic stimulation. *Neuropsychologia*, 35(8), 1121–1131.
- Bacon, W. F., & Egeth, H. E. (1994). Overriding stimulus-driven attentional capture. *Perception & Psychophysics*, 55, 485–496.
- Becker, S. I. (2007). Irrelevant singletons in pop-out search: Attentional capture or filtering costs? *Journal of Experimental Psychology: Human Perception and Performance*, 33(4), 764–787.
- Belopolsky, A. V., & Theeuwes, J. (in press). No capture outside of the attentional window. *Vision Research*. doi:10.1016/j.visres.2010.08.023.
- Belopolsky, A. V., Theeuwes, J., & Kramer, A. F. (2005). Prioritization by transients in visual search. *Psychonomic Bulletin & Review*, 12(1), 93–99.
- Belopolsky, A. V., Zwaan, L., Theeuwes, J., & Kramer, A. F. (2007). The size of an attentional window modulates attentional capture by color singletons. *Psychonomic Bulletin & Review*, 14(5), 934–938.
- Belopolsky, A., Schreij, D., & Theeuwes, J. (2010). What is top-down about contingent capture? *Attention Perception & Psychophysics*, 72, 326–341.
- Bisley, J. W., & Goldberg, M. E. (2003). Neuronal activity in the lateral intraparietal area and spatial attention. *Science*, 299, 81–86.
- Broadbent, D. E. (1958). *Perception and communication*. Oxford: Pergamon Press.
- Brouwer, R. F. T., & van der Heijden, A. H. C. (1997). Identity and position: Dependence originates from independence. *Acta Psychologica*, 95(3), 215–237.
- Bundesen, C. (1990). A theory of visual attention. *Psychological Review*, 97, 523–547.
- Bundesen, C., Pedersen, L. F., & Larsen, A. (1984). measuring efficiency of selection from briefly exposed visual-displays – A model for partial report. *Journal of Experimental Psychology: Human Perception and Performance*, 10(3), 329–339.
- Bundesen, C., Habekost, T., & Kyllingsbaek, S. (2005). A neural theory of visual attention: Bridging cognition and neurophysiology. *Psychological Review*, 112, 291–328.
- Burnham, B. R. (2007). Displaywide visual features associated with a search display's appearance can mediate attentional capture. *Psychonomic Bulletin Review*, 14(3), 392–422.
- Buschman, T. J., & Miller, E. K. (2007). Top-down versus bottom-up control of attention in the prefrontal and posterior parietal cortices. *Science*, 315(5820), 1860–1862.
- Carrasco, M., Ling, S., & Read, S. (2004). Attention alters appearance. *Nature Neuroscience*, 7(3), 308–313.
- Cave, K. R., & Pashler, H. (1995). Visual selection mediated by location: Selecting successive visual objects. *Perception & Psychophysics*, 57(4), 421–432.
- Chelazzi, L., Duncan, J., Miller, E. K., & Desimone, R. (1998). Responses of neurons in inferior temporal cortex during memory-guided visual search. *Journal of Neurophysiology*, 80, 2918–2940.
- Chelazzi, L., Miller, E. K., Duncan, J., & Desimone, R. (2001). Responses of neurons in macaque area V4 during memory-guided visual search. *Cerebral Cortex*, 11(8), 761–772.
- Chen, P., & Mordkoff, J. T. (2007). Contingent capture at a very short SOA: Evidence against rapid disengagement. *Visual Cognition*, 15(6), 637–646.
- Connor, C. E., Egeth, H. E., & Yantis, S. (2004). Visual attention: Bottom-up versus top-down. *Current Biology*, 14(19), R850–R852.
- Constantinidis, C., & Steinmetz, M. A. (2005). Posterior parietal cortex automatically encodes the location of salient stimuli. *Journal of Neuroscience*, 25(1), 233–238.
- Corbetta, M., & Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nature Neuroscience*, 3, 201–215.
- Dalton, P., & Lavie, N. (2007). Overriding auditory attentional capture. *Perception & Psychophysics*, 69(2), 162–171.
- de Fockert, J., Rees, G., Frith, C., & Lavie, N. (2004). Neural correlates of attentional capture in visual search. *Journal of Cognitive Neuroscience*, 16(5), 751–759.
- Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual Review of Neuroscience*, 18, 193–222.

- Donk, M., & Theeuwes, J. (2001). Visual marking beside the mark: Prioritizing selection by abrupt onsets. *Perception & Psychophysics*, 63(5), 891–900.
- Donk, M., & Theeuwes, J. (2003). Prioritizing selection of new elements: Bottom-up versus top-down control. *Perception & Psychophysics*, 65(8), 1231–1242.
- Donk, M., & van Zoest, W. (2008). Effects of salience are short-lived. *Psychological Science*, 19(7), 733–739.
- Duncan, J. (1985). Visual search and visual attention. In M. I. Posner, & O. S. M. Marin (Eds.), *Attention and performance*, Vol. XI. (pp. 85–106) Hillsdale, NJ: Erlbaum.
- Duncan, J., & Humphreys, G. W. (1989). Visual search and stimulus similarity. *Psychological Review*, 96, 433–458.
- Eimer, M. (1995). Event-related potential correlates of transient attention shifts to color and location. *Biological Psychology*, 41(2), 167–182.
- Eimer, M., & Kiss, M. (2007). Attentional capture by task-irrelevant fearful faces is revealed by the N2pc component. *Biological Psychology*, 74(1), 108–112.
- Eimer, M., & Kiss, M. (2008). Involuntary attentional capture is determined by task set: Evidence from event-related brain potentials. *Journal of Cognitive Neuroscience*, 20(8), 1423–1433.
- Eriksen, B. A., & Eriksen, C. W. (1974). Effects of noise letters upon the identification of a target letter in a nonsearch task. *Perception & Psychophysics*, 16, 143–149.
- Folk, C. L., & Remington, R. W. (1998). Selectivity in distraction by irrelevant featural singletons: Evidence for two forms of attentional capture. *Journal of Experimental Psychology: Human Perception and Performance*, 24(3), 847–858.
- Folk, C. L., & Remington, R. (2006). Top-down modulation of preattentive processing: Testing the recovery account of contingent capture. *Visual Cognition*, 14(4–8), 445–465.
- Folk, C. L., Remington, R. W., & Johnston, J. C. (1992). Involuntary covert orienting is contingent on attentional control settings. *Journal of Experimental Psychology: Human Perception and Performance*, 18(4), 1030–1044.
- Folk, C. L., Remington, R. W., & Wright, J. H. (1994). The structure of attentional control: Contingent attentional capture by apparent motion, abrupt onset, and color. *Journal of Experimental Psychology: Human Perception and Performance*, 20, 317–329.
- Folk, C. L., Remington, R. W., & Wu, S. C. (2009). Additivity of abrupt onset effects supports nonspatial distraction, not the capture of spatial attention. *Attention Perception & Psychophysics*, 71(2), 308–313.
- Found, A., & Muller, H. J. (1996). Searching for unknown feature targets on more than one dimension: Investigating a “dimension-weighting” account. *Perception & Psychophysics*, 58(1), 88–101.
- Franconeri, S. L., Hollingworth, A., & Simons, D. J. (2005). Do new objects capture attention? *Psychological Science*, 16(4), 275–281.
- Geyer, T., Muller, H. J., & Krummenacher, J. (2008). Expectancies modulate attentional capture by salient color singletons. *Vision Research*, 48(11), 1315–1326.
- Gibson, B. S., & Amelio, J. (2000). Inhibition of return and attentional control settings. *Perception & Psychophysics*, 62(3), 496–504.
- Gibson, B. S., & Bryant, T. A. (2008). The identity intrusion effect: Attentional capture or perceptual load? *Visual Cognition*, 16(2–3), 182–199.
- Gibson, B. S., & Kelsey, E. M. (1998). Stimulus-driven attentional capture is contingent on attentional set for displaywide visual features. *Perception & Psychophysics*, 24(3), 699–706.
- Gibson, B. S., & Peterson, M. A. (2001). Inattention blindness and attentional capture: Evidence for attention-based theories of visual salience. In C. L. F. B. S. Gibson (Ed.), *Attraction, distraction and action: Multiple perspectives on attentional capture* (pp. 51–76). New York: Elsevier.
- Godijn, R., & Theeuwes, J. (2002). Programming of endogenous and exogenous saccades: Evidence for a competitive integration model. *Journal of Experimental Psychology: Human Perception and Performance*, 28(5), 1039–1054.
- Handy, T. C., Jha, A. P., & Mangun, G. R. (1999). Promoting novelty in vision: Inhibition of return modulates perceptual-level processing. *Psychological Science*, 10(2), 157–161.
- Heslenfeld, D. J., Kenemans, J. L., Kok, A., & Molenaar, P. C. M. (1997). Feature processing and attention in the human visual system: An overview. *Biological Psychology*, 45(1–3), 183–215.
- Hickey, C., & Theeuwes, J. (2008). Abstract. *Journal of Vision*, 8(6).
- Hickey, C., & Theeuwes, J. (under review). Context and competition in the capture of visual attention.
- Hickey, C., McDonald, J. J., & Theeuwes, J. (2006). Electrophysiological evidence of the capture of visual attention. *Journal of Cognitive Neuroscience*, 18(4), 604–613.
- Hillyard, S. A., & Munte, T. F. (1984). Selective attention to color and location – An analysis with event-related brain potentials. *Perception & Psychophysics*, 36(2), 185–198.
- Hochstein, S., & Ahissar, M. (2002). View from the top: Hierarchies and reverse hierarchies. *Neuron*, 36, 791–804.
- Hodsoll, J., Mevorach, C., & Humphreys, G. W. (2009). Driven to less distraction: rTMS of the right parietal cortex reduces attentional capture in visual search. *Cerebral Cortex*, 19(1), 106–114.
- Hunt, A. R., von Muhlenen, A., & Kingstone, A. (2007). The time course of attentional and oculomotor capture reveals a common cause. *Journal of Experimental Psychology: Human Perception and Performance*, 33(2), 271–284.
- Itti, L., & Koch, C. (2001). Computational modelling of visual attention. *Nature Reviews Neuroscience*, 2(3), 194–203.
- Jonides, J., & Yantis, S. (1988). Uniqueness of abrupt visual onset in capturing attention. *Perception & Psychophysics*, 43, 346–354.
- Kahneman, D., Treisman, A., & Burkell, J. (1983). The cost of visual filtering. *Journal of Experimental Psychology: Human Perception and Performance*, 9(4), 510–522.
- 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.
- Kim, M. -S., & Cave, K. R. (1995). Spatial attention in visual search for features and feature conjunctions. *Psychological Science*, 6(6), 376–380.
- Kim, M. S., & Cave, K. R. (1999). Top-down and bottom-up attentional control: On the nature of interference from a salient distractor. *Perception & Psychophysics*, 61, 1009–1023.
- Kiss, M., Van Velzen, J., & Eimer, M. (2008). The N2pc component and its links to attention shifts and spatially selective visual processing. *Psychophysiology*, 45(2), 240–249.
- Klein, R. (2000). Inhibition of return. *Trends in Cognitive Sciences*, 4(4), 138–147.
- Koch, C., & Ullman, S. (1985). Shifts in selective visual attention: Towards the underlying neural circuitry. *Human Neurobiology*, 4, 219–227.
- Kristjánsson, Á., Wang, D., & Nakayama, K. (2002). The role of priming in conjunctive visual search. *Cognition*, 85, 37–52.
- Kumada, T. (1999). Limitations in attending to a feature value for overriding stimulus-driven interference. *Perception & Psychophysics*, 61, 61–79.
- LaBerge, D. (1995). *Attentional processing: The brain's art of mindfulness*. Cambridge, MA: Harvard University Press.
- Lamme, V. A. F., & Roelfsema, P. R. (2000). The distinct modes of vision offered by feedforward and recurrent processing. *Trends in Neuroscience*, 23, 571–579.
- Lamy, D., & Egeth, H. E. (2003). Attentional capture in singleton-detection and feature-search modes. *Journal of Experimental Psychology: Human Perception and Performance*, 29(5), 1003–1020.
- Lamy, D., & Tsai, Y. (2001). On the status of location in visual attention. *European Journal of Cognitive Psychology*, 13(3), 305–342.
- Lamy, D., Tsai, Y., & Egeth, H. E. (2003). Does a salient distractor capture attention early in processing? *Psychonomic Bulletin & Review*, 10(3), 621–629.
- Lamy, D., Leber, A., & Egeth, H. E. (2004). Effects of task relevance and stimulus-driven salience in feature-search mode. *Journal of Experimental Psychology: Human Perception and Performance*, 30(6), 1019–1031.
- Lamy, D., Carmel, T., Egeth, H. E., & Leber, A. B. (2006). Effects of search mode and intertrial priming on singleton search. *Perception & Psychophysics*, 68(6), 919–932.
- Lavie, N., & De Fockert, J. (2005). The role of working memory in attentional capture. *Psychonomic Bulletin & Review*, 12(4), 669–674.
- Lavie, N., & de Fockert, J. (2006). Frontal control of attentional capture in visual search. *Visual Cognition*, 14(4–8), 863–876.
- Leber, A. B. (2009). The divided self: fMRI reveals within-subject fluctuations in the resistance to attention capture over time. *Journal of Vision*, 9(8), 104 Abstract.
- Leber, A. B., & Egeth, H. E. (2006). It's under control: Top-down search strategies can override attentional capture. *Psychonomic Bulletin & Review*, 13(1), 132–138.
- Leblanc, E., Prime, D. J., & Jolicoeur, P. (2008). Tracking the location of visuospatial attention in a contingent capture paradigm. *Journal of Cognitive Neuroscience*, 20(4), 657–671.
- Li, Z. (2002). A saliency map in primary visual cortex. *Trends in Cognitive Sciences*, 6(1), 9–16.
- Lu, S. N., & Han, S. H. (2009). Attentional capture is contingent on the interaction between task demand and stimulus salience. *Attention Perception & Psychophysics*, 71(5), 1015–1026.
- Luck, S. J., & Hillyard, S. A. (1994). Electrophysiological correlates of feature analysis during visual-search. *Psychophysiology*, 31(3), 291–308.
- Luck, S. J., Woodman, G. F., & Vogel, E. K. (2000). Event-related potential studies of attention. *Trends in Cognitive Sciences*, 4(11), 432–440.
- Ludwig, C. J. H., & Gilchrist, I. D. (2002). Stimulus-driven and goal-driven control over visual selection. *Journal of Experimental Psychology: Human Perception and Performance*, 28(4), 902–912.
- Maljkovic, V., & Nakayama, K. (1994). Priming of pop-out: I. Role of features. *Memory & Cognition*, 22(6), 657–672.
- Mangun, G. R. (1995). Neural mechanisms of visual selective attention. *Psychophysiology*, 32(1), 4–18.
- Martinez, A., Anllo-Vento, L., Sereno, M. I., Frank, L. R., Buxton, R. B., Dubowitz, D. J., et al. (1999). Involvement of striate and extrastriate visual cortical areas in spatial attention. *Nature Neuroscience*, 2(4), 364–369.
- Mathôt, S., Hickey, C., & Theeuwes, J. (2010). From Reorienting of Attention to Biased Competition: Evidence from Hemifield Effects. *Attention, Perception & Psychophysics*, 72, 651–657.
- Moore, C. M., & Egeth, H. (1998). How does feature-based attention affect visual processing? *Journal of Experimental Psychology: Human Perception and Performance*, 24(4), 1296–1310.
- Mortier, K., Theeuwes, J., & Starreveld, P. (2005). Response selection modulates visual search within and across dimensions. *Journal of Experimental Psychology: Human Perception and Performance*, 31(3), 542–557.
- Mounts, J. R. W. (2000). Evidence for suppressive mechanisms in attentional selection: Feature singletons produce inhibitory surrounds. *Perception & Psychophysics*, 62, 969–983.
- Mulckhuyse, M., Van der Stigchel, S., & Theeuwes, J. (2009). Early and late modulation of saccade deviations by target distractor similarity. *Journal of Neurophysiology*, 102(3), 1451–1458.
- Müller, H. J., & Findlay, J. M. (1988). The effect of visual attention on peripheral discrimination thresholds in single and multiple element displays. *Acta Psychologica*, 69, 129–155.
- Muller, H. J., & Krummenacher, J. (2006). Locus of dimension weighting: Pre-attentive or post-selective? *Visual Cognition*, 14, 490–513.
- Müller, H. J., Heller, D., & Ziegler, J. (1995). Visual search for singleton feature targets within and across feature dimensions. *Perception & Psychophysics*, 57, 1–17.
- Müller, H. J., Reimann, B., & Krummenacher, J. (2003). Visual search for singleton feature targets across dimensions: Stimulus and expectancy-driven effects in dimensional weighing. *Journal of Experimental Psychology: Human Perception and Performance*, 29, 1021–1035.

- Müller, H. J., Geyer, T., Zehetleitner, M., & Krummenacher, J. (2009). Attentional capture by salient color singleton distractors is modulated by top-down dimensional set. *Journal of Experimental Psychology: Human Perception and Performance*, 35(1), 1–16.
- Munneke, J., Heslenfeld, D. J., & Theeuwes, J. (2008). Directing attention to a location in space results in retinotopic activation in primary visual cortex. *Brain Research*, 1222, 184–191.
- Neisser, U. (1967). *Cognitive psychology*. New York: Appleton Century Crofts.
- 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(1), 15–34.
- Ogawa, T., & Komatsu, H. (2004). Target selection in area V4 during a multidimensional visual search task. *Journal of Neuroscience*, 24(28), 6371–6382.
- Olivers, C. N. L., Meijer, F., & Theeuwes, J. (2006). Feature-based memory-driven attentional capture: Visual working memory content affects visual attention. *Journal of Experimental Psychology: Human Perception and Performance*, 32(5), 1243–1265.
- Pinto, Y., Olivers, C. N. L., & Theeuwes, J. (2005). Target uncertainty does not lead to more distraction by singletons: Intertrial priming does. *Perception & Psychophysics*, 67(8), 1354–1361.
- Posner, M. I. (1980). Orienting of attention, the VIIth Sir Frederic Bartlett lecture. *Quarterly Journal of Experimental Psychology*, 32, 3–25.
- Posner, M. I., & Cohen, Y. (1984). Components of visual orienting. In H. Bouma, & D. G. Bouwhuis (Eds.), *Attention and performance X: Control of language processes* (pp. 531–556). Hillsdale, NJ: Lawrence Erlbaum.
- Posner, M. I., Davidson, B. J., & Snyder, C. R. R. (1980). Attention and the detection of signals. *Journal of Experimental Psychology: General*, 109, 160–174.
- Pratt, J., Sekuler, A. B., & McAuliffe, J. (2001). The role of attentional set on attentional cueing and inhibition of return. *Visual Cognition*, 8, 33–46.
- Prinzmetal, W., McCool, C., & Park, S. (2005). Attention: Reaction time and accuracy reveal different mechanisms. *Journal of Experimental Psychology: General*, 134(1), 73–92.
- Proulx, M. J., & Egeth, H. E. (2006). Target-nontarget similarity modulates stimulus-driven control in visual search. *Psychonomic Bulletin & Review*, 13(3), 524–529.
- Rauschenberger, R. (2003). Attentional capture by auto- and allo-cues. *Psychonomic Bulletin & Review*, 10(4), 814–842.
- Remington, R. W., Folk, C. L., & McLean, J. P. (2001). Contingent attentional capture or delayed allocation of attention? *Perception & Psychophysics*, 63(2), 298–307.
- Reynolds, J. H., & Chelazzi, L. (2004). Attentional modulation of visual processing. *Annual Review of Neuroscience*, 27, 611–647.
- Schrei, D., Owens, C., & Theeuwes, J. (2008). Abrupt onsets capture attention independent of top-down control settings. *Perception & Psychophysics*, 70(2), 208–218.
- Schrei, D., Theeuwes, J., & Olivers, C. N. L. (2010). Abrupt onsets capture attention independent of top-down control settings II: Additivity is no evidence for filtering. *Attention, Perception & Psychophysics*, 72, 672–682.
- Schrei, D., Theeuwes, J., & Olivers, C. N. L. (in press). Irrelevant onsets cause inhibition of return regardless of attentional set. *Attention, Perception & Psychophysics*.
- Schubö, A. (2009). Salience detection and attentional capture. *Psychological Research Psychologische Forschung*, 73(2), 233–243.
- Serences, J. T., & Yantis, S. (2006). Selective visual attention and perceptual coherence. *Trends in Cognitive Sciences*, 10(1), 38–45.
- Starreveld, P. A., Theeuwes, J., & Mortier, K. (2004). Response selection in visual search: The influence of response compatibility of nontargets. *Journal of Experimental Psychology: Human Perception and Performance*, 30(1), 56–78.
- Sternberg, S. (1969). The discovery of processing stages: Extensions of Donders' method. *Acta Psychologica*, 30, 276–315.
- Theeuwes, J. (1989). Effects of location and form cueing on the allocation of attention in the visual-field. *Acta Psychologica*, 72(2), 177–192.
- Theeuwes, J. (1990). Perceptual selectivity is task dependent – Evidence from selective search. *Acta Psychologica*, 74(1), 81–99.
- Theeuwes, J. (1991). Cross-dimensional perceptual selectivity. *Perception & Psychophysics*, 50(2), 184–193.
- Theeuwes, J. (1991). Exogenous and endogenous control of attention – The effect of visual onsets and offsets. *Perception & Psychophysics*, 49(1), 83–90.
- Theeuwes, J. (1992). Perceptual selectivity for color and form. *Perception & Psychophysics*, 51(6), 599–606.
- Theeuwes, J. (1994). Endogenous and exogenous control of visual selection. *Perception*, 23(4), 429–440.
- Theeuwes, J. (1994). Stimulus-driven capture and attentional set – Selective search for color and visual abrupt onsets. *Journal of Experimental Psychology: Human Perception and Performance*, 20(4), 799–806.
- Theeuwes, J. (1995). Abrupt luminance change pops out – Abrupt color-change does not. *Perception & Psychophysics*, 57(5), 637–644.
- Theeuwes, J. (1995). Perceptual selectivity for color and form: On the nature of the interference effect. In A. F. Kramer, M. G. H. Coles, & G. D. Logan (Eds.), *Converging operations in the study of visual attention* (pp. 297–314). Washington, DC: American Psychological Association.
- Theeuwes, J. (1995). Temporal and spatial characteristics of preattentive and attentive processing. *Visual Cognition*, 2(1), 221–233.
- Theeuwes, J. (2004). Top-down search strategies cannot override attentional capture. *Psychonomic Bulletin & Review*, 11(1), 65–70.
- Theeuwes, J. (2005). Irrelevant singletons capture attention. In L. Itti, G. Rees, & J. Tsotsos (Eds.), *Neurobiology of attention* (pp. 418–424). San Diego: Elsevier.
- Theeuwes, J., & Belopolsky, A. (2010). Top-down and bottom-up control of visual selection: controversies and debate. In V. Coltheart (Ed.), *Tutorials in visual cognition* (pp. 67–92). New York: Psychology Press.
- Theeuwes, J., & Burger, R. (1998). Attentional control during visual search: The effect of irrelevant singletons. *Journal of Experimental Psychology: Human Perception and Performance*, 24(5), 1342–1353.
- Theeuwes, J., & Chen, C. Y. D. (2005). Attentional capture and inhibition (of return): The effect on perceptual sensitivity. *Perception & Psychophysics*, 67(8), 1305–1312.
- Theeuwes, J., & Godijn, R. (2001). Attention and oculomotor capture. In C. L. Folk, & B. S. Gibson (Eds.), *Attraction, distraction, and action: Multiple perspectives on attentional capture*. New York: Elsevier Science.
- Theeuwes, J., & Godijn, R. (2002). Irrelevant singletons capture attention: Evidence from inhibition of return. *Perception & Psychophysics*, 64(5), 764–770.
- Theeuwes, J., & Godijn, R. (2004). Inhibition-of-return and oculomotor interference. *Vision Research*, 44(12), 1485–1492.
- Theeuwes, J., & Van der Burg, E. (2007). The role of spatial and nonspatial information in visual selection. *Journal of Experimental Psychology: Human Perception and Performance*, 33(6), 1335–1351.
- Theeuwes, J., & van der Burg, E. (2008). The role of cueing in attentional capture. *Visual Cognition*, 16(2–3), 232–247.
- Theeuwes, J., Kramer, A. F., Hahn, S., & Irwin, D. E. (1998). Our eyes do not always go where we want them to go: Capture of the eyes by new objects. *Psychological Science*, 9(5), 379–385.
- Theeuwes, J., Kramer, A. F., Hahn, S., Irwin, D. E., & Zelinsky, G. J. (1999). Influence of attentional capture on oculomotor control. *Journal of Experimental Psychology: Human Perception and Performance*, 25(6), 1595–1608.
- Theeuwes, J., Atchley, P., & Kramer, A. F. (2000). On the time course of top-down and bottom-up control of visual attention. In S. M. J. Driver (Ed.), *Attention & performance*, vol. 18. (pp. 105–125). Cambridge: MIT Press.
- Theeuwes, J., Kramer, A. F., & Kingstone, A. (2004). Attentional capture modulates perceptual sensitivity. *Psychonomic Bulletin & Review*, 11(3), 551–554.
- Theeuwes, J., Reimann, B., & Mortier, K. (2006). Visual search for featural singletons: No top-down modulation, only bottom-up priming. *Visual Cognition*, 14(4–8), 466–489.
- Theeuwes, J., Van Der Burg, E., & Belopolsky, A. (2008). Detecting the presence of a singleton involves focal attention. *Psychonomic Bulletin & Review*, 15(3), 555–560.
- Theeuwes, J., Olivers, C. N. L., & Belopolsky, A. V. (in press). Stimulus-driven capture and contingent capture. *Wiley Interdisciplinary Reviews Cognitive Science*. doi:10.1016/j.visres.2010.08.023.
- Treisman, A. M. (1988). Features and objects: The fourteenth Bartlett memorial lecture. *The Quarterly Journal of Experimental Psychology*, 40A, 201–237.
- Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, 12, 97–136.
- Treue, S. (2003). Visual attention: The where, what, how and why of saliency. *Current Opinion in Neurobiology*, 13(4), 428–432.
- Tsal, Y., & Lavie, N. (1988). Attending to color and shape: The special role of location in selective visual processing. *Perception & Psychophysics*, 44, 15–21.
- van Zoest, W., Donk, M., & Theeuwes, J. (2004). The role of stimulus-driven and goal-driven control in saccadic visual selection. *Journal of Experimental Psychology: Human Perception and Performance*, 30(4), 746–759.
- vanderHeijden, A. H. C., Kurvink, A. G., elange, L., deLeeuw, F., & vanderGeest, J. N. (1996). Attending to color with proper fixation. *Perception & Psychophysics*, 58(8), 1224–1237.
- VanRullen, R., & Koch, C. (2003). Visual selective behavior can be triggered by a feed-forward process. *Journal of Cognitive Neuroscience*, 15(2), 209–217.
- von Wright, J. M. (1970). On selection in visual immediate memory. *Acta Psychologica*, 33, 280–292.
- Wolfe, J. M. (1994). Guided Search 2.0. A revised model of visual search. *Psychonomic Bulletin & Review*, 1, 202–238.
- 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*, 29(2), 483–502.
- Woodman, G. F., & Luck, S. J. (1999). Electrophysiological measurement of rapid shifts of attention during visual search. *Nature*, 400(6747), 867–869.
- Wykowska, A., & Schubö, A. (in press). Irrelevant singletons in visual search do not capture attention but can produce nonspatial filtering costs. *Journal of Cognitive Neuroscience*.
- Yantis, S. (1993). Stimulus-driven attentional capture and attentional control settings. *Journal of Experimental Psychology: Human Perception and Performance*, 19, 676–681.
- Yantis, S., & Egeth, H. E. (1999). On the distinction between visual salience and stimulus-driven attentional capture. *Journal of Experimental Psychology: Human Perception and Performance*, 25, 661–676.
- Yantis, S., & Hillstrom, A. P. (1994). Stimulus-driven attentional capture: Evidence from equiluminant visual objects. *Journal of Experimental Psychology: Human Perception and Performance*, 20, 95–107.
- Yantis, S., & Johnston, J. C. (1990). On the locus of visual selection: Evidence from focused attention tasks. *Journal of Experimental Psychology: Human Perception and Performance*, 16, 135–149.
- Yantis, S., & Jonides, J. (1984). Abrupt visual onsets and selective attention: Evidence from visual search. *Journal of Experimental Psychology: Human Perception and Performance*, 10, 601–621.
- Yantis, S., & Jonides, J. (1990). Abrupt visual onsets and selective attention: Voluntary versus automatic allocation. *Journal of Experimental Psychology: Human Perception and Performance*, 16, 121–134.
- Yeh, S. L., & Liao, H. I. (2008). On the generality of the contingent orienting hypothesis. *Acta Psychologica*, 129(1), 157–165.