Advanced Review

Stimulus-driven capture and contingent capture

Jan Theeuwes,* Christian N. L. Olivers and Artem Belopolsky

Whether or not certain physical events can capture attention has been one of the most debated issues in the study of attention. This discussion is concerned with how goal-directed and stimulus-driven processes interact in perception and cognition. On one extreme of the spectrum is the idea that attention capture is primarily stimulus driven and automatic. On the other end is the notion that attention capture is always contingent on the goals of the observer, and thus under top-down control. This review discusses the empirical evidence for each of these viewpoints and the theoretical consequences. In addition, there is a discussion of the issues that remain controversial within the debate between the two viewpoints. It is concluded that visual selection depends on the interaction between bottom-up and top-down processes with a special role for spatial attention as the top-down gatekeeper for attention capture. © 2010 John Wiley & Sons, Ltd. *WIREs Cogn Sci*

The amount of visual information entering our eyes is much greater than what our brain can fully process. It is therefore necessary that we can select information that is relevant and ignore information that is irrelevant for our tasks, particularly because irrelevant information can disrupt our ongoing behavior. Visual attention is the selection mechanism by which some visual events are prioritized, whereas others are excluded from processing. We speak of *attention capture* when attention is unintentionally drawn to the location of such an event.

A fundamental question that has spurred a heated debate over the last 15 years is whether we have full control over what we select from the environment. Selection may be controlled by the observer in a top-down way, or may be controlled by the properties of the stimulus field, in a stimulus-driven, bottom-up way. According to the *stimulus-driven capture* account, selection is initially determined by the physical salience of the objects in the environment. In this view, attention will go first to the most salient element in the environment regardless of top-down control settings. After this element has been selected, its location gets inhibited, and attention may move on to the next most salient object in the hierarchy.[1–11] This mechanism has recently been modeled in the Itti and Koch's[12] computational model of salience driven selection.

According to a different account known as the *contingent attention capture* hypothesis, attention capture is never stimulus driven, but always contingent on the top-down settings of the observer.[13–19] In other words, the 'contingent capture' model postulates that only stimuli that match the top-down control settings will capture attention; stimuli that do not match the top-down settings will be ignored. This paper reviews the evidence available for stimulus-driven and contingent capture.

VISUAL SELECTION

For several decades, there has been agreement that visual selection involves two functionally independent stages of processing.[20–22] An early visual stage, sometimes referred to as pre-attentive, operates in parallel across the visual field and a later stage, often referred to as attentive, can deal with only one (or a few items) at the same time. Even though it appears that 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 still present.[12,18,23–25] 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

*Correspondence to: J.Theeuwes@psy.vu.nl

Department of Cognitive Psychology, Vrije Universiteit Amsterdam, The Netherlands

DOI: 10.1002/wcs.83

passed on to the second stage of processing. In line with the two-stage approach, passing on an item to the second stage of processing implies that this item has been *selected* for further processing.[20,22]

This means that from all objects that are present in the visual field (and are available at the pre-attentive stage of processing), only the object that is passed onto the final stage of processing will affect decision-making and responding. This passing on from the initial stage of pre-attentive processing to attentive processing is what is considered to be *selection*. As mentioned earlier, this distinction between early parallel and later serial processing is present (explicitly or implicitly) in most classic and recent theories of visual selection. Note, however, that there are also theories that do not make a distinction between early and later processing and assume that everything is processed in parallel up to a high level.[26] These theories are based on the classic conception of 'late-selection'[27] and equate selection with the processes that are involved in decision-making and response selection.

Studies investigating initial attentional control typically use displays in which the target is defined as a feature singleton. When confronted with such a display (such as displays in which one element is red and the others are green), one is able to immediately detect this element without any effort. Typically, search time to determine whether such a feature singleton is present or not, is independent of the number of elements in the display. Finding these flat search functions is important because it implies that we are dealing with parallel pre-attentive search. Only when parallel, pre-attentive search is involved, we can determine whether the initial selection is controlled in a top-down or bottom-up way. If search is slow and effortful (as e.g., 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 attention capture.

The finding that upon presentation, a feature singleton can immediately be detected has led to the suggestion that feature singletons receive attentional priority independent of the intentions of the observer. In other words, when searching for a pre-specified target (such as a red circle between green circles) one may argue that selection occurs in a purely bottom-up way. This claim may not necessarily be correct. If the feature singleton is also the element that observers are instructed to look for, one cannot determine whether this immediate selection of the feature singleton is the result of bottom-up or top-down control. As pointed out by Yantis and Egeth,[28] one can only speak of selection in a purely stimulus-driven fashion when the stimulus feature in question is completely task-irrelevant, so that there is no incentive for the observer to attend to it deliberately. As expressed by Yantis and Egeth[28]: 'If an object with such an attribute captures attention under these conditions, then and only then can that attribute be said to capture attention in a purely stimulus-driven fashion' (p. 663). When objects or events receive priority independently of the observer's goals and beliefs one refers to *attention capture* when such an event only captures our attention[3,4] and one refers to *oculomotor capture* when such an event also inadvertently triggers a saccade to the location of the event.[29]

STIMULUS-DRIVEN CAPTURE

To investigate the contribution of top-down and bottom-up control in visual selection, in the early 1990s, Theeuwes[2–4] developed the so-called additional singleton task. This task was a visual search task in which observers search for one specific clearly defined salient singleton while another irrelevant singleton was simultaneously present. Figure 1 gives examples of the display. It is important to note that in this task the irrelevant color distractor was never the target, so there was no reason for observers to attend to it. If in this task the irrelevant singleton would capture attention, it would fulfill the condition set out by Yantis and Egeth[28] for pure stimulus-driven capture. Moreover, observers also had a clear search goal: throughout the whole experiment observers consistently looked for the same target (a green diamond) allowing them to adapt a clear top-down search set. It is important to note that in this task observers searched for a particular target shape (a green diamond) but responded to the orientation of the line segment inside the target shape. This makes it possible to disentangle factors affecting the selection of the target from those affecting the response selection. As is clear from Figure 1, search functions are flat suggesting that this task involves pop-out detection involving the first stage of pre-attentive processing.

The main finding of the additional singleton search task is that RTs in the condition in which a unique color irrelevant distractor is present are higher than when such distractor is not present (see Figure 1). Importantly, an irrelevant singleton only causes an RT increase when the distractor is more salient than the target. When the color distractor was made less salient [see Figure 1 (right panels)], its presence did not affect search for the diamond target anymore. These findings are important because they indicate that it is the bottom-up salience signal of the stimuli in the visual

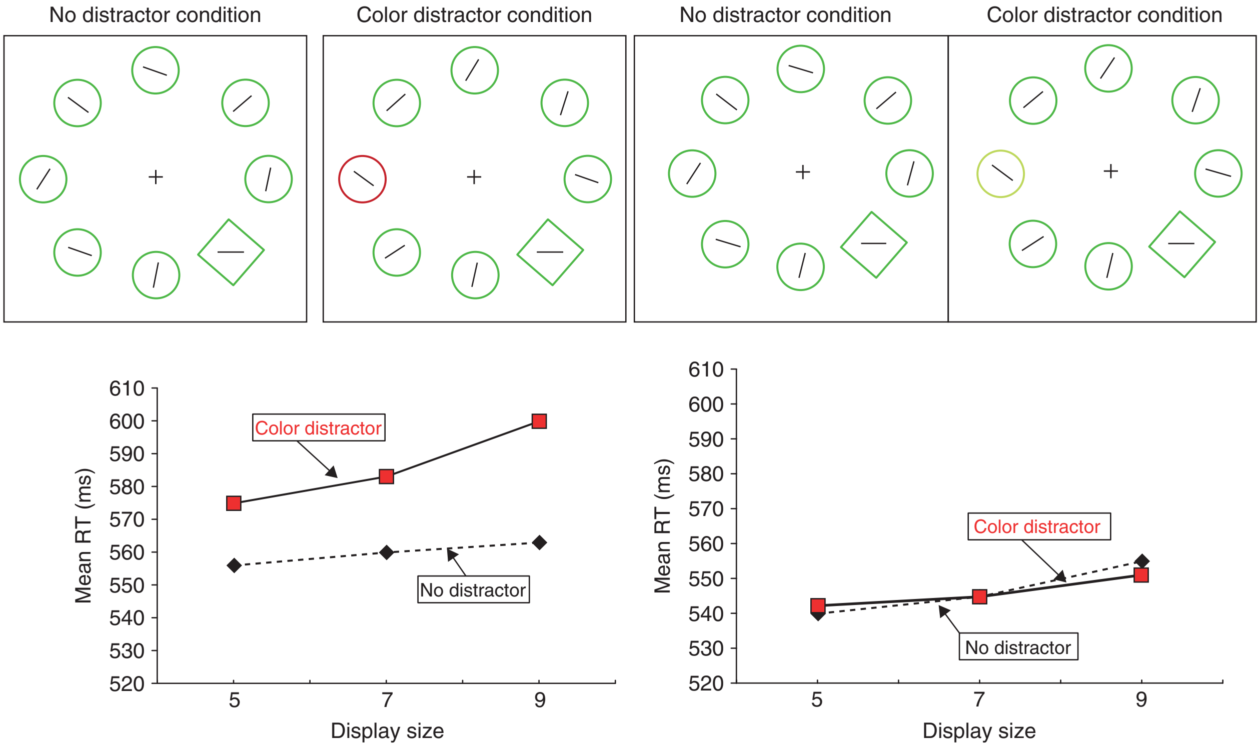

FIGURE 1 | Stimuli and data from Theeuwes.[3] Observers search throughout the whole experiment for a shape singleton, a green diamond presented among a variable number of circles. Observers respond to the orientation (horizontal or vertical) of the line segment presented within the target diamond shape. 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 stimulus-driven attention capture.

field that determines the selection order. Even though observers knew that the red singleton was never the target they could not apply sufficient top-down control to prevent its selection. In one experiment, Theeuwes[3] showed that the interference effect caused by the irrelevant distractor remained present even after 1800 trials of training. Thus even extensive practice cannot induce sufficient top-down control to overcome the interference caused by a salient distractor.

The increase in search time in conditions in which an irrelevant singleton was present was explained in terms of attention capture.[2–4] Because the irrelevant color singleton was selected exogenously and captured attention, it required more time before the target singleton could be selected and a response could be generated. Given the observation that attention capture completely depended on the relative salience of the singleton target and the distractor singleton, it was argued that early visual pre-attentive processing is only driven by bottom-up factors. This implies that early on, during the first sweep of information through the brain, the competition between the two salient objects is resolved in accordance with the relative strength of bottom-up salience signals. It was concluded that top-down control over feature selection (i.e., knowing that the target is a diamond and not a red circle) and extensive training were not able to affect this pure stimulus-driven capture.

Note that *after* attention is captured by the salient singleton, there is ample opportunity for top-down control. For example, when the singleton selected is a red circle and observers know that they are searching for a green diamond (see Figure 1), top-down processing allows a fast and efficient disengagement of attention from the erroneously selected singleton. This was demonstrated in a study in which the color distractor singleton was presented 150 ms before the presentation of the target. In that study, there was no behavioral evidence that the irrelevant color singleton captured attention[8] suggesting that the erroneous capture by a distractor singleton can be overcome within 150 ms. Clearly, attention capture is short-lived and after a distractor is selected, top-down (possibly recurrent) processing disengages the processing from the distractor location enabling the selection of the target. This process of disengaging attention from the distractor location and engaging attention at the target location only takes 100–150 ms (see also Ref 30) for a similar result.

Since its introduction in 1991, the basic findings of additional singleton paradigm (and variations of it) have been replicated by many labs using reaction time,[30–34] d-prime[35,36] and saccadic eye movements.[11] Even though the basic finding is undisputed, since its introduction there has been a large controversy about the interpretation of performance decrement caused by the irrelevant distractor—that is, whether it reflects attention capture at all.

CONTINGENT ATTENTION CAPTURE

Opposing the idea of stimulus-driven attention capture is the hypothesis that capture is contingent on top-down control setting. According to this view known as the contingent capture hypothesis[14,15] selection depends critically on the explicit or implicit perceptual goals held by the observer at any given time. When performing a visual task such as searching for a traffic light while approaching a busy intersection, it is assumed that the activation of a search template (e.g., here red or green) ensures that our attention only gets captured by objects that look like they may be relevant for the task at hand. In this example, this might be the traffic light, but attention may also be captured by an irrelevant object that matches what one is looking for, such as the onset of a brake light on the car in front of you, or the changing to green of the pedestrian light.

In their original series of experiments, Folk et al.[14] showed that visual selection depends critically on the top-down attentional set of the observer. Folk et al.[14] used a spatial cueing paradigm in which a cue display was followed in rapid succession by a target display. There were four elements in the target display.

In the target display, observers were required to identify the unique element. There were two types of target display: In the *color* display, the target was red, whereas 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 with an abrupt onset. Immediately preceding the target display at 150 ms stimulus onset asynchrony (SOA), a cue display was presented. Similar to the target displays, cue displays consisted of two types: the cue was either defined by color (in which one location was surrounded by red dots and the other three locations were surrounded by white dots) or by onset (in which one location was surrounded by an abrupt onset of white dots and the remaining locations remained empty). In the critical experiment, the cue that preceded the search display could either be valid (i.e., it appeared at the same location as the target), or invalid (i.e., it appeared at the same location as the target). Among the four potential target locations, the cue was valid on 25% of the trials and invalid on the remaining trials. Thus, consistent with the criterion set by Yantis and Egeth,[28] there was no incentive for observers to deliberately attend to the cue fulfilling the criterion for pure stimulus driven capture. The important finding was that when observers were looking for a target defined as an abrupt onset, observers were fast on valid trials and relatively slow on invalid trials, but only when the cue was defined by an onset too. The same was found for the conditions in which they were looking for a target defined by color: observers were fast on valid trials and relatively slow on invalid trials, but only when the cue was defined by color. When the cue was defined by the other feature (i.e., color when looking for onset or the other way around), no effect of cue validity was found (see Figure 2). 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, the cue captured 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 clearly indicated that the top-down attentional set determines the selection priority: when set for a particular feature singleton, one will select each element that matches this top-down set; feature singletons that do not match top-down attentional sets will not be selected and will simply be ignored. Obviously, according to this view the attentional readiness adopted by the observer determines selection.

The findings of Folk et al. have been replicated many times using various modifications of the classic paradigm[15,37–40] (but see Ref 41).

ONLY ONSETS CAPTURE ATTENTION

In the extreme, stimulus-driven capture and contingent capture represent opposite viewpoints: stimulus driven capture claims that capture is always bottom-up, whereas contingent capture claims that capture is always dependent on top-down settings. A viewpoint

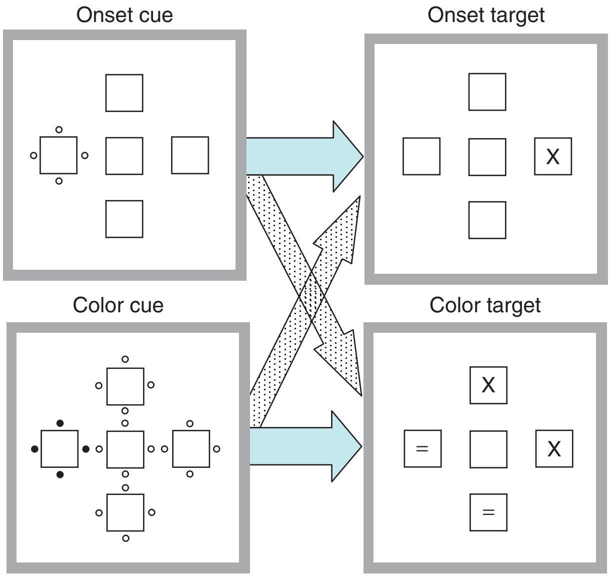

FIGURE 2 | The contingent capture paradigm of Folk et al.[14] Observers had to respond to a target singleton (either an 'X' or an '='). The target was defined as singleton which had a unique color ('color target' condition, bottom-right) or was the only element presented as an onset ('onset target' condition, 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 color cue) only captured attention when observers were set to look for it.

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 Jonides,[42] Jonides and Yantis[43] (see also Yantis & Egeth, 28), conducted several studies investigating whether feature singleton receive attentional priority. Yantis and colleagues adopted a visual search task, such that the target of search was a nonsingleton letter. This type of search is not efficient revealing search times that increase with the number of elements present in the display. In each search 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. Because the salient element was the target at chance level, there was no incentive to deliberately start searching at the salient singleton (see Figure 3). Jonides and Yantis[43] showed that subjects 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 nonunique 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 nonsalient 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[43] 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[43] (see also Yantis and Egeth[28]) 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 and Egeth[28] claimed that selection is under top-down control except for elements that are presented with abrupt onset constituting a new object.

Recently, Schreij et al.[44] confirmed the special role for onsets in a paradigm similar to that of Folk et al.[14] In this study, it was shown that even when observers have an attentional set for color, an irrelevant new object presented with an abrupt onset captured attention. It was concluded that abrupt onsets can override a top-down set for color. The conclusions of this study were recently challenged by Folk et al.[45] arguing that the inference effect of the abrupt onset may not necessarily be because of attention capture but could represent nonspatial filtering costs. This and other controversies will be discussed next.

CONTROVERSIES

Over the past 20 years or so, there has been a heated discussion on the extent to which visual selection is under top-down control. On the one hand, there is the stimulus-driven capture view of Theeuwes suggesting that there is no top-down control in early vision and that the initial sweep of information is basically bottom-up; while on the other hand, there are views that assume full top-down control[14] or views that assume top-down control under specific conditions (such as Refs 31,43). Even though the debate started more than 20 years ago, with many attempts to reconcile the differences, there is still no clear answer to the question.

The Time Course of Attention Capture

An important question that needs to be answered is why in Folk et al.'s spatial cueing paradigm

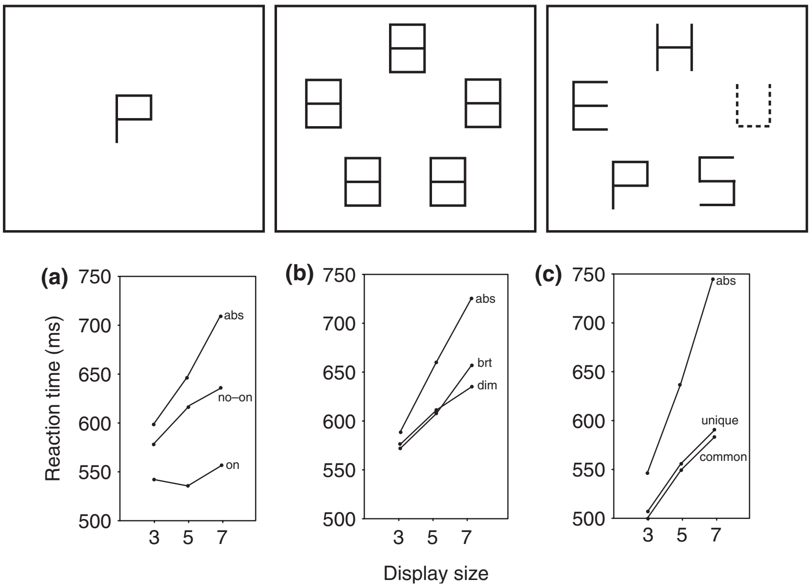

FIGURE 3 | Paradigm and data from Jonides and Yantis.[43] In the first display, a target letter was displayed for 1000 ms (in this case the letter P) followed by a premask display for 1000 ms. In the search display, one letter had a unique color (dotted lines). At chance level, this letter could be the 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 do 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.

(see Figure 2) there is full top-down control over selection and why there appears to be no top-down control in Theeuwes irrelevant singleton paradigm (see Figure 1). The answer to this question may lie in the procedural differences between the paradigms. As noted, Folk et al. use a spatial cueing paradigm in which observers have to ignore a 'cue' that appears 150 ms prior to the presentation of the target display (see, e.g., Ref 14). Observers respond to a character shape (X vs. =), which had either a unique color or a unique abrupt onset. When the search display was preceded by a to-be-ignored featural singleton (the 'cue') that matches the singleton for which they 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' did not match the singleton for which they are searching, its appearance apparently 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 zero effect), whereas a cue that matches the search goal has an effect on RT.

Contrary to the spatial cueing paradigm, Theeuwes used the above described irrelevant singleton search task in which the target and distractor singleton were *simultaneously* present. He showed that, independent of any top-down goal, an irrelevant singleton that was more salient than the singleton target interfered with search. As Theeuwes et al.[8] have argued, it is quite feasible that the irrelevant cue did capture attention in Folk et al.'s spatial cueing paradigm. Because there was a delay of 150 ms between the presentation of the cue and the search display, observers may have been able to overcome the attention capture by the time the search display was presented (see also Ref 4). Disengagement of attention from the cue may have been relatively fast when the cue and target did not share the same defining properties (e.g., the cue is red and the target is an onset), whereas disengagement from the cue may have been relatively slow in cases in which the cue and target share the same defining properties (e.g., both were red).

Recently, Chen and Mordkoff[46] tested the idea whether the 150 ms SOA in the Folk et al. paradigm masked attention capture by irrelevant singletons because of the relatively long interval between cue and target display. Instead of a 150 ms SOA, Chen and Mordkoff[46] used a 35 ms SOA and still found no evidence for attention capture by irrelevant singletons. This study provides evidence against the idea that the absence of capture in Folk et al.'s paradigm is because of the relatively long delay between the presentation of the cue and the search display (see also Refs 37,47).

Filtering Costs

Theeuwes observed an increase in reaction time in conditions in which the irrelevant singleton was present (see Figure 1). The increase in RT was explained in terms of attention capture: attention moved exogenously to the location of the salient singleton before it could move to the location of the (less salient) singleton target. Folk and Remington[16] offered an alternative explanation for the increase in RT in conditions, in which a distractor was present.

They argued that 'filtering costs'[48] lead to an increase in search time caused by the irrelevant singleton. The idea of filtering costs is that the presence of an irrelevant singleton may slow the deployment of attention to the target item by requiring an effortful and time-consuming filtering operation. According to this line of reasoning, attention is employed in a top-down way and goes directly to the singleton target; simply because another irrelevant singleton is present, directing attention to the target may take more time than when no such irrelevant singleton is present. Note that this view does *not* entail a shift of spatial attention to the location of the irrelevant singleton (i.e., no attention capture), although it remains unclear what the exact filtering mechanism is. The controversy regarding the viability of filtering to explain attention capture is still not resolved (see recent papers of Schreij et al.[44] and Folk et al.[45]). Without going into further details, it has become a discussion about what the concept of filtering entails. If filtering has a spatial component, then conceptually there may be no difference anymore between a fast shift of attention to the singleton location followed by a fast disengagement of attention and the concept of filtering.

Search Modes

Bacon and Egeth[31] challenged the stimulus-driven view of Theeuwes by suggesting that in the Theeuwes' paradigm observers choose a search mode (the so-called singleton detection mode), which causes attention capture. If observers would choose another mode (the so-called feature-search mode) there would be no attention capture anymore. By definition, the act of choosing a specific search mode implies the involvement of top-down control and not bottom-up capture.

Bacon and Egeth first replicated Theeuwes'[3] experiment described above in which a color singleton interfered with search for a shape singleton. In the following experiment, they added additional shapes (i.e., squares and triangles) to the display so that the shape singleton was no longer unique. In this condition the color singleton did not interfere anymore. Bacon and Egeth[31] 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 once the feature search mode was set it was used throughout a whole block of trials. Bacon and Egeth[31] concluded that 'goal-directed selection of a specific known featural singleton identity may override stimulus-driven capture by salient singletons' (p. 493). 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 selection is very much under top-down control.

The idea of different search modes is an attractive one but Theeuwes[49] challenged the Bacon and Egeth[31] by suggesting that their results may have been caused by the heterogeneity of the search displays (i.e., the stimulus) rather than the attentional state of the observers (i.e., top-down settings). More complex stimuli may force subjects to search in a more focused fashion, which may partly preclude attention capture similar to the way it did in the study by Jonides and Yantis.[43] Indeed, when Theeuwes[49] used the same displays as Bacon and Egeth[31] with three different shape singletons (thus forcing observers to engage in a feature-search mode), but at the same time increased the salience of target and distractor singleton (by adding more nontarget elements), search became more efficient, and the interference caused by the color singleton re-emerged. However, Leber and Egeth[50] challenged Theeuwes[49] findings and claimed that extensive training can reinstate top-down control over attention capture.

Attentional Window

Given the notion that inefficient search eliminates attention capture, Theeuwes[49] suggested that the size of the '*attentional window*' of observers could be one of the factors explaining why in some experiments salient color singletons fail to capture attention (as e.g., Bacon & Egeth[31] and Jonides & Yantis[43]). In studies which do not find capture by a color singleton visual search often occurs in a serial or partly serial fashion, such that the search elements are examined individually or in small clusters. According to this hypothesis, when observers expect a difficult search task the size of the attentional window is reduced such that the window does not encompass the whole display. This increases the chance that the unique element is not included in the salience computations and does not capture attention. However, when the target is a unique object, as in the task used by Theeuwes,[3,4] the optimal strategy is to attend to the whole display at once to find the target. As a consequence, the uniquely colored item falls inside the attentional window, is processed preattentively and captures attention. This idea is supported by a well-known finding that when a target location is known in advance, even abrupt onsets do not capture attention.[51,52]

In a recent study, Belopolsky et al.[53] tested this idea. They used a design similar to Jonides and Yantis,[43] but in addition manipulated the size of the attentional window of observers. To ensure that observers spread their attention across the whole display, before they could start searching they had to make a judgment regarding the spatial lay-out of all the elements. Specifically, before searching the display observers had to decide whether the elements formed an upward-pointing triangle. Only then they could start their search. 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 the observers allowed to start searching the display.

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.[43] However, when attention was initially diffused over the global stimulus arrangement (diffuse attention condition), in many trials the irrelevant color singleton was selected first.

Belopolsky et al.'s findings suggest that the size of attentional window is an important factor in determining whether an irrelevant color singleton will capture attention (but see Ref 50). When attention is spread, visual search may be conducted across all items in the visual field in parallel (as in the irrelevant singleton task of Refs 2,3), at the expense that an irrelevant salient singleton would also be selected automatically. However when in anticipation of a difficult search, the attentional window is set to a small size, visual search will become serial (as in Ref 43), or partly serial (as in Ref 31). In these conditions, salient singleton falls outside of the attentional window will not capture attention. The attentional window provides a way to reconcile different views on the extent of top-down control, without assuming different search modes. While the size of the attentional window can be under top-down control, within the attentional window top-down control cannot preclude attention from being captured by the most salient feature.

CONCLUSION

In the present review we described the major controversies that exist in the field of attentional capture. As can be seen the debate regarding the extent and the mechanisms of top-down control over visual selection is still ongoing. However, a lot of progress has been made in understanding the intricate interplay between the top-down and bottom-up mechanisms. It is clear that our control over what gets selected is far from perfect; however, it is not nonexistent. As shown by the contingent capture hypothesis top-down control can strongly bias visual selection. At the same time, it has been unequivocally demonstrated that completely irrelevant, but salient stimuli interfere with our selection mechanisms, against our will. What the exact nature of this interference is (whether capture, filtering, or both) remains a matter of debate and awaits more specific models of attention. Even when salient objects inadvertently interfere, control over what gets selected can be quickly regained by disengaging attentional resources from the stimulus and redirecting them to the target, or by reducing the size of the attentional window and therefore restricting the area of parallel processing.

REFERENCES

1. Hickey C, McDonald JJ, Theeuwes J. Electrophysiological evidence of attention capture. *J Cogn Neurosci* 2006, 18:604–613.
2. Theeuwes J. Cross-dimensional perceptual selectivity. *Percept Psychophys* 1991, 50:184–193.
3. Theeuwes J. Perceptual selectivity for color and form. *Percept Psychophys* 1992, 51:599–606.
4. Theeuwes J. Stimulus-driven capture and attentional set: selective search for color and visual abrupt onsets. *J Exp Psychol* 1994, 20:799–806.
5. Theeuwes J. Endogenous and exogenous control of visual selection. *Percept Psychophys* 1994, 23:429–440.
6. Theeuwes J. Perceptual selectivity for color and form: on the nature of the interference effect. In: Kramer AF, Coles MGH, Logan GD, eds. *Converging Operations in the Study of Visual Attention*. Washington, DC: American Psychological Association; 1996, 297–314.
7. Theeuwes J. Temporal and spatial characteristics of preattentive and attentive processing. *Visual Cognition* 2005, 2:221–233.

8. Theeuwes J, Atchley P, Kramer AF. On the time course of top-down and bottom-up control of visual attention. In: Monsell S, Driver J, eds. *Attention and Performance*, vol. 18. Cambridge: MIT Press; 2000.
9. Theeuwes J, Godijn R. Irrelevant singletons capture attention: evidence from inhibition of return. *Percept Psychophys* 2002, 64:764–770.
10. Theeuwes J, Reimann B, Mortier K. Visual search for featural singletons: no top-down modulation, only bottom-up priming. *Vis Cogn* 2006, 14:466–489.
11. Van Zoest W, Donk M, Theeuwes J. The role of stimulus-driven and goal-driven control in visual selection. *J Exp Psychol* 2004, 30:746–759.
12. Itti L, Koch C. Computational modeling of visual attention. *Nat Rev Neurosci* 2001, 2:194–203.
13. Egeth HE, Yantis S. Visual attention: Control, representation, and time course. *Ann Rev Psychol* 1997, 48:269–297.
14. Folk CL, Remington RW, Johnston JC. Involuntary covert orienting is contingent on attentional control settings. *J Exp Psychol* 1992, 18:1030–1044.
15. Folk CL, Remington RW, Wright JH. The structure of attentional control: contingent attention capture by apparent motion, abrupt onset, and color. *J Exp Psychol* 1994, 20:317–329.
16. Folk CM, Remington RW. Selectivity in distraction by irrelevant featural singletons: evidence for two forms of attention capture. *J Exp Psychol* 1998, 24:847–858.
17. Müller HJ, Reimann B, Krummenacher J. Visual search for singleton feature targets across dimensions: stimulus- and expectancy-driven effects in dimensional weighting. *J Exp Psychol* 2003, 29:1021–1035.
18. Wolfe JM. Guided search 2.0: a revised model of visual search. *Psychon Bull Rev* 1994, 1:202–238.
19. Wolfe JM, Butcher SJ, Lee C, Hyle M. Changing your mind: on the contributions of top-down and bottom-up guidance in visual search for feature singletons. *J Exp Psychol* 2003, 29:483–502.
20. Broadbent DE. *Perception and communication*. London: Pergamon Press; 1958.
21. Neisser U. *Cognitive Psychology*. Englewood Cliffs, NJ: Prentice-Hall; 1967.
22. Treisman A, Gelade G. A feature integration theory of attention. *Cogn Psychol* 1980, 12:97–136.
23. Koch C, Ullman S. Shifts in selective visual attention: towars the underlying neural circuity. *Human Neurobiol* 1985, 4:219–227.
24. Li Z. A saliency map in primary visual cortex. *Trends Cogn Sci* 2002, 6:9–16.
25. Treue S. Climbing the cortical ladder from sensation to perception. *Trends Cogn Sci* 2003, 7:469–471.
26. Bundesen C. A theory of visual attention. *Psychol Rev* 1990, 97:523–547.
27. Deutsch JA, Deutsch D. Attention: some theoretical considerations. *Psychol Rev* 1963, 70:51–61.
28. Yantis S, Egeth HE. On the distinction between visual salience and stimulus-driven attention capture. *J Exp Psychol* 1999, 25:661–676.
29. Theeuwes J, Kramer AF, Hahn S, Irwin DE. Our eyes do not always go where we want them to go: capture of the eyes by new objects. *Psychol Sci* 1998, 9:379–385.
30. Kim MS, Cave KR. Top-down and bottom-up attentional control: on the nature of interference from a salient distractor. *Percept Psychophys* 1999, 61:1009–1023.
31. Bacon WF, Egeth HE. Overriding stimulus-driven attention capture. *Percept Psychophys* 1994, 55:485–496.
32. Kawahara J, Toshima T. Stimulus-driven control of attention: evidence from visual search for moving target among static nontargets. *Jpn J Psychon Sci* 1996, 15:77–87.
33. Kumada T. Limitations in attending to a feature value for overriding stimulus-driven interference. *Percept Psychophys* 1999, 61:61–79.
34. Mounts JRW. Attention capture by abrupt onsets and feature singletons produces inhibitory surrounds. *Percept Psychophys* 2000, 62:1485–1493.
35. Theeuwes J, Kramer AF, Kingstone A. Attention capture modulates perceptual sensitivity. *Psychon Bull Rev* 2004, 11:551–554.
36. Theeuwes J, Chen CYD. Attention capture and inhibition (of return): the effect on perceptual sensitivity. *Percept Psychophys* 2005, 67:1305–1312.
37. Folk CL, Remington R. Top-down modulation of preattentive processing: testing the recovery account of contingent capture. *Vis Cogn* 2006, 14:445–465.
38. Gibson BS, Amelio J. Inhibition of return and attentional control settings. *Percept Psychophys* 2000, 62:496–504.
39. Remington RW, Folk C, McLean JP. Contingent attention capture or delayed allocation of attention? *Percept Psychophys* 2001, 63:298–307.
40. Pratt J, Sekuler A, McAuliffe J. The role of attentional set on attentional cueing and inhibition of return. *Vis Cogn* 2001, 8:33–46.
41. Yeh SL, Liao HI. On the generality of the contingent orienting hypothesis. *Acta Psychol* 2008, 129:157–165.
42. Yantis S, Jonides J. Abrupt visual onsets and selective attention: evidence from selective search. *J Exp Psychol* 1984, 10:601–621.
43. Jonides J, Yantis S. Uniqueness of abrupt visual onset in capturing attention. *Percept Psychophys* 1988, 43:346–354.
44. Schreij D, Owens C, Theeuwes J. Abrupt onsets capture attention independent of top-down control settings. *Percept Psychophys* 2008, 70:208–218.

45. Folk CL, Remington RW, Wu SC. Additivity of abrupt onset effects supports nonspatial distraction, not the capture of spatial attention. *Percept Psychophys* 2009, 71:308–313.
46. Chen P, Mordkoff JT. Contingent capture at a very short SOA: evidence against rapid disengagement. *Vis Cogn* 2007, 15:637–646.
47. Lamy D, Egeth HE. Attention capture in singleton-detection and feature-search modes. *J Exp Psychol* 2003, 29:1003–1020.
48. Kahneman D, Treisman A, Burkell J. The costs of visual filtering. *J Exp Psychol* 1983, 9:510–522.
49. Theeuwes J. Top-down search strategies cannot override attention capture. *Psychon Bull Rev* 2004, 11:65–70.
50. Leber AB, Egeth HE. It's under control: top-down search strategies can override attention capture. *Psychon Bull Rev* 2006, 13:132–138.
51. Yantis S, Jonides J. Abrupt visual onsets and selective attention: voluntary versus automatic allocation. *J Exp Psychol* 1990, 16:121–134.
52. Theeuwes J. Exogenous and endogenous control of attention: the effect of visual onsets and offsets. *Percept Psychophys* 1991, 49:83–90.
53. Belopolsky AV, Zwaan L, Theeuwes J, Kramer AF. The size of attentional window modulates attention capture by color singletons. *Psychon Bull Rev* 2007, 14:934–938.

FURTHER READING

Burnham BR. Displaywide visual features associated with a search display's appearance can mediate attention capture. *Psychon Bull Rev* 2007, 14:392–422.

Corbetta M, Shulman GL. Control of goal-directed and stimulus-driven attention in the brain. *Nature Rev Neurosci* 2002, 3:201–215.

Serences JT, Yantis S. Selective visual attention and perceptual coherence. *Trends Cogn Sci* 2006, 10:38–45.