

# Irrelevant Singletons Capture Attention

Jan Theeuwes

## ABSTRACT

When searching for a particular target, irrelevant distractors that happen to be more salient than the target will capture our attention against our intentions. We show that (1) spatial attention is first directed to the location of the most salient object in the field; (2) after attention is captured by the location of the salient singleton, this location becomes inhibited to allow disengagement of attention; (3) disengagement of attention from an inhibited location can proceed very fast (within 150 ms); (4) salient singletons may not only capture our attention but may also capture our eyes. We conclude that the visual system encodes differences in features in parallel across the visual field irrespective of the top-down goals of the observer and attention is directed automatically to the location of the highest saliency.

## I. INTRODUCTION

A crucial research question is the extent to which we are able to exert attentional control over what we select from the visual environment. Selection may either be controlled by the properties of the stimulus field or by intentions of the observer. When we intentionally select only those objects and events needed for our current tasks, selection is said to occur in a voluntary, goal-directed manner. When, irrespective of our goals and beliefs, specific properties present in the visual field determine what we select, this selection is said to occur in an involuntary, stimulus-driven manner. It is as if attention is captured by the stimulus (e.g., Yantis, 2000). Often such attentional capture is accompanied by an exogenous saccade to the location of the object

or event, a phenomenon termed *oculomotor capture* (Theeuwes et al., 1998).

When confronted with a display in which one element is unique in a basic visual dimension (such as a red element surrounded by green elements), one is able to immediately detect this element without effort. Search time to determine whether such a salient target is present or not is independent of the number of elements in the display. Elements that pop out from the display are referred to as *feature singletons* or simply *singletons*. Typically, search functions with slopes of less than 5 or 6 ms per item are considered to reflect parallel search (e.g., Treisman and Gormican, 1988). Such a "pop-out effect" is used as a diagnostic for determining whether visual information is available at the preattentive parallel level, which is assumed to segment the visual field into objects and background. Although the preattentive parallel stage can segment the visual field, it is assumed that a second stage of focused attention is necessary before one can act upon segmented objects.

Given the observation that upon presentation, a feature singleton is detected immediately, one may believe that feature singletons receive attentional priority independent of the intentions of the observer. In other words, when searching for a prespecified target (such as a red circle between green circles) one may argue that attention is captured in a bottom-up way by the uniquely colored element. However, this argument is not necessarily incorrect because the element that pops out is the target one is looking for and thus possibly part of the observer's top-down attentional set. To investigate whether salient singletons capture attention in a purely stimulus driven manner Theeuwes (1991b, 1992, 1994) developed a paradigm referred to as the *additional singleton* paradigm (e.g.,

Simons, 2000). In this visual search task, two feature singletons are simultaneously present in the visual field; one of the singletons is the target one is looking for while the other is a distracter that has to be ignored.

## II. FEATURE SINGLETONS CAPTURE ATTENTION

The logic underlying the irrelevant singleton task is simple: Participants were asked to perform a visual search task, and one item in the search display is a unique salient feature singleton that is completely unrelated and completely irrelevant to the search task. The feature singleton is never the search item, and should therefore play no role in the observer's attentional set. This condition is compared to a condition in which an irrelevant featural singleton is not present. Figure 69.1 presents an example of typical stimulus displays and results (from Theeuwes, 1992). Participants with displays consisting of colored circles or diamonds appear on the circumference of an imaginary circle. Line segments of different orientations appear in the circles and diamonds. Participants had to determine the orientation of the line segment appearing in the target shape. The target shape that participants

searched for was a singleton because it was the only diamond present in the display. In the distracter condition, an irrelevant color distracter singleton was also present in the display. Time to find the shape singleton increased when an irrelevant color singleton was present (i.e., one of the circles was red).

Even though participants had a clear top-down set to search for the shape singleton (i.e., the single green diamond), the presence of an irrelevant singleton (i.e., the single red circle) caused interference. It should be noted that search was performed by preattentive parallel search as the search slopes were basically flat (see Fig. 69.1). In subsequent experiments, it was shown that selection depended on the relative saliency of the stimulus attributes: When the color singleton was made less salient than the shape singleton (by reducing the color difference between the target and the nontarget elements), the shape singleton interfered with search for the color singleton while the color singleton no longer interfered with the search for the shape singleton.

Theeuwes (1991b, 1992, 1994, 1996) explained the increase in search time in conditions in which an irrelevant singleton was present in terms of attentional capture. Because attention was automatically captured by the distracter singleton (the most salient element in the display), it took longer before attention could be redirected to the location of the target singleton and a response could be emitted. Given the observation that selectivity depended completely on the relative saliency of target and distracter singleton, it was suggested that early visual preattentive parallel processing is driven only by bottom-up factors such as saliency. Irrespective of the top-down goals of the observer, spatial attention is automatically and involuntarily captured by the most salient singleton. The shift of spatial attention to the location of the singleton implies that the singleton is selected for further processing. If this singleton is the target, a response is made. If it is not the target, attention is directed to the next most salient singleton. The initial shift of attention to the most salient singleton is thought to be the result of relatively inflexible, hardwired mechanisms that are triggered by the presence of a feature difference signals. It is assumed that at each location in the visual field a local feature contrast is calculated that represents how different that object is within a particular primitive feature dimension (e.g., color, shape, movement, etc.).

The notion suggested by Theeuwes is similar to that of Koch and Ullman (1985), who introduced the notion of a saliency map to accomplish preattentive selection. This map is a two-dimension topographical map that encodes the saliency of objects in their visual environ-

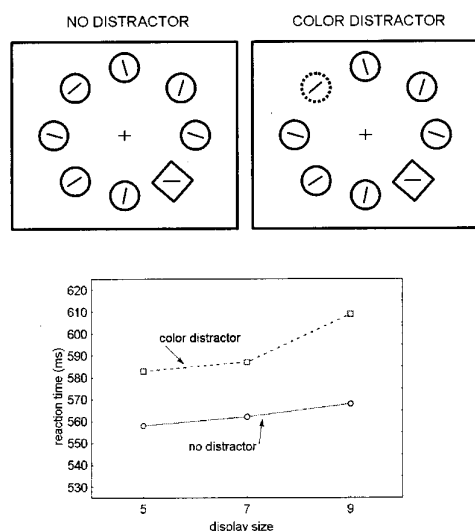


FIGURE 69.1 Stimuli and data from Theeuwes (1992). Top: A vertical and horizontal bar appeared within a green diamond (the other bars were oblique). Top left: all circles are green (solid lines); top right: one of the circles is red (dashed lines). Participants were to report the orientation of the bar in the diamond. Bottom: The presence of the irrelevant color singleton slowed responses to the bar (from Theeuwes, 1992; Experiment 1).

ment. Neurons in this map compete among each other, giving rise to a single winning location (winner take all) that contains the most salient element. The saliency map is the result of preattentive parallel encoding across the visual field calculating differences in simple visual features such as intensity, contrast, color, and orientation. Focused spatial attention simply scans the locations of decreasing activation (saliency). If a location is inhibited, the next salient location will receive spatial attention (see also Itti and Koch, 2001; Sagi and Julesz, 1985; Nothdurft, 2000).

It has been argued that preattentive computation simply calculates differences in features within dimensions resulting in a pattern of activations at different locations (Theeuwes, 1992, 1994). Given the previous example (see Fig. 69.1), at the location of the red singleton a large difference signal arises because the singleton differs from all other nontargets within the color dimension. At the location of the circle singleton, a large difference signal arises because the circle differs from all other elements within the shape dimension. Focal attention is automatically and unintentionally shifted to locations in the display containing large local feature differences, regardless of the dimension in which this feature difference occurs. The source of the preattentively calculated difference signal (whether it is caused by a color singleton or a shape singleton) is not available preattentively and can be determined only after focused attention has moved to the location of the difference signal. In other words, an observer knows whether the singleton was the target only after selecting the location with the large difference signal. In this view, the saliency of the singleton, and not its identity, its color, its shape, its brightness, and so on, will determine which element captures attention. Obviously, given this model, selection operates purely bottom-up irrespective of the task demands. The automatic shifts of attention are considered to be the result of relatively inflexible, hardwired mechanisms that are triggered by the presence of these difference signal interrupts. It is assumed that the parallel process can perform only a *local-mismatch* detection followed by a serial stage in which the most mismatching areas are selected for further analysis.

### A. Spatial Attention Is Focused on the Most Active (Winner) Location

The increase in reaction time for those conditions in which the irrelevant singleton was present has been explained in terms of attentional capture: focused attention moved to the location of the salient singleton before it could move to the location of the (less salient)

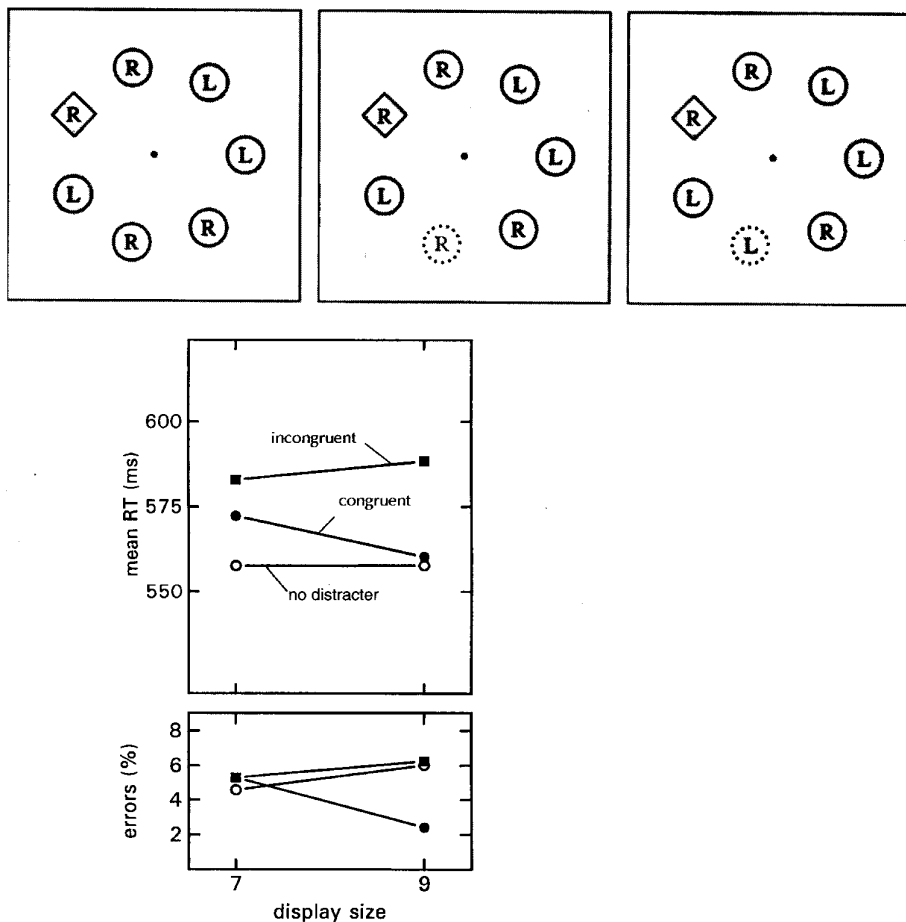
target singleton. However, one may argue that the presence of an irrelevant singleton increased RT not because of an additional shift of spatial attention to the distracter location but because the preattentive computation of the target feature took longer in the presence of a distracter singleton. In other words, simply because another irrelevant singleton was present, computing the target feature singleton and directing attention to it may take more time than when no such irrelevant singleton was present. Note that this view does *not* entail a shift of spatial focused attention to the most active location.

To address this issue Theeuwes (1996) used the same additional singleton search task but changed the elements inside the search items from line segments to large letter characters. When the letter R was presented in the shape singleton (i.e., the diamond) participants had to respond with their right hand. When the letter L was presented in the shape singleton participants responded with the left hand. Importantly, Theeuwes (1996) manipulated the congruency of the character at the location of the irrelevant distracter (see Fig. 69.2). In half of the trials the character at the distracter location was associated with the same response as was required by the target (e.g., an R in the target singleton and an R in the distracter singleton); on the other half it was the opposite of what was required by the target (e.g., an R in the target singleton and an L in the distracter singleton).

Theeuwes (1996) reasoned that if focal attention never went to the location of the irrelevant singleton, the identity of its character would have no effect on responding. However, a clear congruency effect was found (see Fig. 69.2), which provided evidence that before a response was given, attention was at the location of the irrelevant singleton. This finding is completely in line with the notion that *spatial* focused attention is directed automatically to the most salient location in the scene.

### B. Attentional Selection and Inhibition of Return

The saliency map guides the attentional focus to the most salient location in the scene. But how can one prevent attention from permanently focusing onto the most active location? Within their model Itti and Koch (2001) suggested that after a location is attended it will be suppressed allowing the winner-take-all network to focus on the next most salient location. The phenomenon labeled as inhibition of return (Posner and Cohen, 1984) represents such an inhibitory tagging mechanism: after attention has been



**FIGURE 69.2** Stimuli and data from Theeuwes (1996). Top: Sample stimulus display (display size 7). In the no-distractor condition (top left) the green diamond shape appears among green circles. In the congruent condition (top middle), the letter inside the green diamond target shape (in this case the letter R) is identical to the letter inside the red circle distractor. In the incongruent condition (top right), the letter inside the green diamond target shape is different from the letter inside the red circle distractor. Solid lines indicate green and dotted lines indicate red. Bottom: The incongruent condition is significantly slower than the congruent condition suggesting that the letter at the location of the to-be-ignored singleton was processed (from Theeuwes, 1996; Experiment 1).

shifted to a location in space, there is delayed responding to stimuli subsequently displayed at that location (see Klein [2000] for a review). Theeuwes and Godijn (2002) provided direct evidence for the occurrence IOR at a salient distracter singleton location: reaction times for stimuli presented at the location of a salient singleton were higher than for stimuli presented at a non-salient distracter location. Based on the agreement that IOR is the result of an early, reflexive, involuntary orienting system (e.g., Posner and Cohen, 1984), Theeuwes and Godijn (2002) argued that these results could only be interpreted as evidence for exogenous bottom-up attentional capture.

### C. Speed of Disengaging Spatial Attention

The bottom-up model assumes that attention is directed automatically to the most salient location regardless of whether it is a target or distractor (Theeuwes, 1992, 1994). The preattentive mechanism that extracts visual features calculates difference signals; only after focused attention has been directed to such a salient location, the exact feature becomes available (e.g., it is a color singleton, shape singleton, etc.). Therefore, top-down knowledge cannot guide the focused attentional selection mechanism. Knowing for example that the target is a red and squared cannot

help selection but only processes that occur after an element has been selected for further processing. One crucial question is how long attention stays focused at a salient location when it is clear that the location attended does not contain the target one is looking for.

Theeuwes et al. (2000) showed that once attention is directed to the location of an irrelevant singleton it takes only a very brief time to disengage attention from that location. Theeuwes et al. (2000) used a visual search task similar to that of Theeuwes (1992; see Fig. 69.1) in which participants searched for a shape singleton. Prior to the presentation of the target display, at different times (50, 100, 150, 200, 250, and 300 ms) a color singleton was presented. Theeuwes et al. (2000) showed that the presence of an irrelevant salient distractor had an effect only when the target and distractor singleton were presented in close succession (at times of 50 and 100 ms). When the distractor singleton was presented considerable time (times of 150 to 300 ms) before the presentation of the target singleton, there was no evidence that attention was focused on the salient distractor. These results indicate that bottom-up attention is always directed to the most salient singleton; however, when it is clear that the element that captured attention in a bottom-up way is not the target, within 150 ms attention may be redirected to another location. Note that top-down processing plays a crucial role in the speed with which one is able to disengage attention. Thus, if the singleton that captures attention in a bottom-up way resembles the target one is looking for, the speed of disengagement may be much slower. As noted earlier, in order to disengage attention one may require the suppression of the (features at the) attended location.

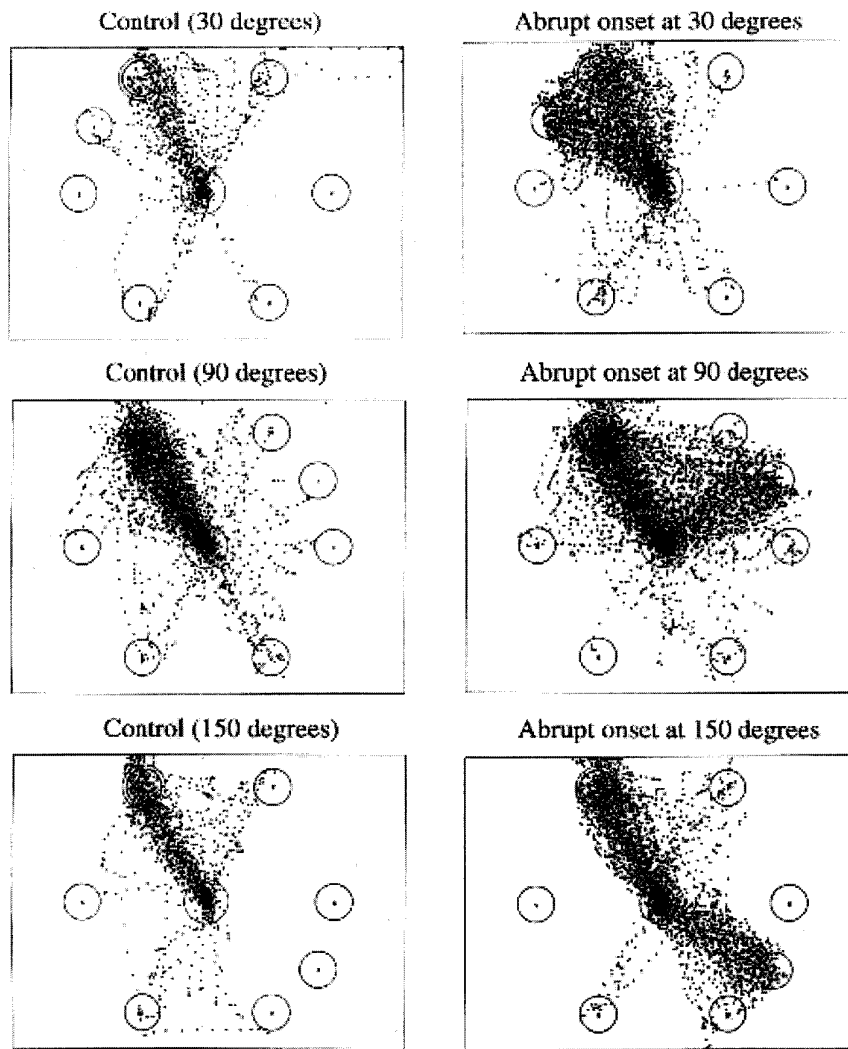
### III. ATTENTION AND EYE MOVEMENTS

It is generally agreed that there is an obligatory and selective coupling between saccade execution and visual attention to one common target object (e.g., Deubel and Schneider, 1996). When a saccade is programmed, attention precedes the eyes to the saccade target location. To determine whether salient singletons not only capture attention but also eye movements, Theeuwes et al. (1998) developed a paradigm

referred to as the oculomotor capture paradigm, which uses the same logic as the additional singleton paradigm. Instead of inferring capture on the basis of a slowed response to the target, capture is reflected by an inappropriate eye movement toward the irrelevant item. Participants had the explicit instruction to search and make a saccadic eye movement toward the only gray element in the display. On some trials, an irrelevant singleton (an abrupt onset) was added to the display. Participants knew the onset was irrelevant and also knew that they had to ignore it. The condition in which a to-be-ignored onset was presented somewhere in the visual field was compared to a control condition in which there was no onset added to the display. The results showed that when no onset was added to the display, observers made saccades that generally went directly to the uniquely colored circle. However, in those trials in which an onset was added to the display, in about 30 to 40 percent of the trials the eye went in the direction of the onset, stopped briefly, and then went on to the target. Figure 69.3 shows the results. The graphs on the left side depict the control condition without the onset; the graphs on the right side depict the condition in which an onset was presented. Note that in the condition with the onset, the eye often went to the distracter. This occurred even when the onset appeared at a side opposite to that of the target circle (see Fig. 69.3 bottom panels: an abrupt onset is presented at 150 degrees separation). These findings suggest that salient singletons may not only capture our attention but may also capture our eyes.

### IV. CONCLUSION

In conclusion, there is strong evidence that salient singletons capture spatial attention in a purely bottom-up way. The most parsimonious interpretation of the current findings is that top-down control during early preattentive segmentation is not possible. Selectivity is determined by the saliency of objects in the visual field; that is, the most salient singleton gets attention first. After this location is inhibited the next salient location will receive attention.



**FIGURE 69.3** Oculomotor Capture. Data from Theeuwes et al. (1999). Eye movement behavior in the condition in which an abrupt onset was presented simultaneously with the target. The results are collapsed over all eight participants and normalized with respect to the position of target and onset. Sample points (every 4 ms) are taken from only the first saccade. Left panels: Eye movement behavior in the control condition in which no abrupt onset was presented. Right panels: Eye movement behavior in the condition in which an abrupt onset was presented; Either close to the target (top) somewhat away from the target (middle) and/or at the opposite side from the target (bottom). From Theeuwes et al. (1999).

## References

- Deubel, H., and Schneider, W. X. (1996). Saccade target selection and object recognition: Evidence for a common attentional mechanism. *Vision Research* 6, 1827–1837.
- Itti, L., and Koch, C. (2001). Computational modeling of visual attention. *Nature Reviews Neuroscience* 2, 194–203.
- Koch, C., and Ullman, S. (1985). Shifts in selective visual attention: towards the underlying neural circuitry. *Human Neurobiology* 4, 219–227.
- Klein, R. M. (2000). Inhibition of return. *Trends in Cognitive Science* 4, 138–147.
- Posner, M. I., and Cohen, Y. (1984). Components of visual orienting. In H. Bouma and D. Bouwhuis (Eds.), *Attention & Performance X*, pp. 531–556. Erlbaum, Hillsdale, NJ.
- Sagi, D., and Julesz, B. (1985). Detection versus discrimination of visual orientation. *Perception* 14, 619–628.
- Simons, D. J. (2000). Attentional capture and inattention blindness. *Trends in Cognitive Science* 4, 147–155.
- Theeuwes, J. (1991). Cross-dimensional perceptual selectivity. *Perception & Psychophysics* 50, 184–193.
- Theeuwes, J. (1992). Perceptual selectivity for color and form. *Perception & Psychophysics* 51, 599–606.

- 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**, 799–806.
- Theeuwes, J. (1996). Perceptual selectivity for color and form: On the nature of the interference effect. In A. F. Kramer, M. G. H. Coles, and G. D. Logan (Eds.), *Converging Operations in the Study of Visual Attention*, pp. 297–314. American Psychological Association, Washington DC.
- Theeuwes, J., Kramer, A. F., Hahn, S., and 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**, 379–385.
- Theeuwes, J., Atchley, P., and Kramer, A. F. (2000). On the time course of top-down and bottom-up control of visual attention. In S. Monsell and J. Driver (Eds.), *Attention & Performance (Vol 18)*. MIT Press; Cambridge.
- Theeuwes, J., and Godijn, R. (2002). Irrelevant singletons capture attention: evidence from inhibition of return. *Perception & Psychophysics* **64**, 764–770.
- Treisman, A. M., and Gormican, S. (1988). Feature search in early vision: Evidence from search asymmetries. *Psychological Review* **95**, 15–48.
- Yantis, S. (2000). Goal directed and stimulus driven determinants of attentional control. In S. Monsell and J. Driver (Eds.), *Attention & Performance (Vol 18)*. MIT Press; Cambridge.