



Commentary

Why salience is not enough: Reflections on top-down selection in vision

Howard E. Egeth^{a,*}, Carly J. Leonard^b, Andrew B. Leber^c^a Johns Hopkins University, United States^b University of California, Davis, United States^c University of New Hampshire, United States

The target article represents a distillation of nearly 20 years of work dedicated to the analysis of visual selection. Throughout these years, Jan Theeuwes and his colleagues have been enormously productive in their development of a particular view of visual selection, one that emphasizes the role of bottom-up processes. This work has been very influential, as there is substantial merit to many aspects of this research. However, this endeavor has also been provocative—the reaction to this work has resulted in a large body of research that emphasizes the role of top-down processes. Here we highlight recent work not covered in Theeuwes's review and discuss how this literature may not be compatible with Theeuwes's theoretical perspective. In our view this ongoing debate has been one of the most interesting and productive in the field. One can only hope that in time the ultimate result will be a complete understanding of how visual selection actually works.

1. Some general comments

It will be helpful to address at the outset a definitional issue. In his Introduction, Theeuwes (2010) labels as top-down effects those that are “completely under the control of the intentions of the observer” such that “a person can choose *at will* from the environment what to select” (p. 77) (italics in original). Note that if a symmetrically clear (and extreme) definition of bottom-up processing were adopted, it would require that bottom-up selection be due solely to featural information present in the current stimulus. However, as we see later in Theeuwes's paper, priming and the effect on visual search of information held in working memory are considered to be examples of bottom-up processing. It is surely worthwhile to debate the proper way to categorize these influences, but at this time it seems to us that they represent some of the clearest forms of real-life top-down guidance in visual processing. It is a particularly unusual stretch to construe biasing by information maintained (presumably “*at will*”) in working memory as bottom-up, given that leading models of attention feature this type of interaction as a mechanism of top-down control (e.g., Desimone & Duncan, 1995).

2. In defense of search modes

Theeuwes reviews a long series of studies implementing the additional singleton paradigm in which search for a shape singleton is slowed by the presence of an irrelevant color singleton (e.g., Theeuwes, 1992). Bacon and Egeth (1994) proposed that perhaps participants weren't really engaged in search for a specific shape; maybe they were looking for any target that differed markedly from its surrounding items. Such a processing mode would clearly leave participants vulnerable to a color singleton. We tested this by making efforts to prevent a singleton search strategy, and when we did, we found that a color singleton no longer captured attention.

This was meant as a modest proposal, with singleton detection mode best construed as an example of a top-down set (e.g., Folk, Remington, & Johnston, 1992). Although the proposal has been invoked to help account for various experimental outcomes, Theeuwes (2010) expresses skepticism, going so far to state that “the concept of a search mode doesn't explain much, if anything” (p. 90).

Theeuwes questions why, if participants are capable of avoiding capture via feature-search mode, do they not use it all the time. We have considered this issue, and we believe the answer could lie in the principle of satisficing. If we assume that establishing a feature-search mode requires the investment of effort, then participants are likely to settle for visual search strategies that yield adequate, if not optimal, performance (see Leber & Egeth, 2006). Consider that the capture effects discussed here are on the order of 25 ms; in a typical 30-minute experimental session that includes 250 distractor trials, the cumulative cost of briefly attending irrelevant singletons is just a matter of a few seconds. Further, as Theeuwes describes, existing data have suggested an overall behavioral slowing when participants engage in feature-search mode (i.e., on distractor-absent trials). Thus the degree to which overall speed, collapsed across distractor-present and distractor-absent trials, suffers as a result of adopting singleton detection mode could be negligible.

As for why the slowing occurs when feature-search mode is adopted, we have just begun to investigate this issue. Theeuwes's perspective is that salience dominates attentional orienting early on, due to a totally stimulus-driven initial wave of processing. He further argues that only after this initial wave subsides can top-down control be implemented. According to this viewpoint, stimulus-driven capture can be avoided only if observers delay their search until after the initial salience-dominated processing subsides (van Zoest, Donk &

* Corresponding author.

E-mail address: egeth@jhu.edu (H.E. Egeth).

Theeuwes, 2004). This line of reasoning is interesting, although we offer a speculative alternative: during periods of high goal-driven control, participants also adopt a more conservative responding strategy. That is, while greater cognitive control reduces vulnerability to distraction through greater attentional selectivity, it could be accompanied by a greater commitment to producing the correct response (akin to a speed-accuracy tradeoff). It will be important to test this alternative empirically.

Thus far we have discussed why observers might not mind interference from irrelevant stimuli, but to be clear, we do not wish to imply that singleton-detection mode is the default mode of processing. Indeed, there is ample evidence that feature-search mode is routinely engaged to reduce distraction from known singletons to some extent, even if not always optimally so.

As Theeuwes noted, capture effects are much smaller when both the target and distractor are known in advance, suggesting that some degree of top-down control is present in these studies. Pinto, Olivers and Theeuwes (2005) attributed at least some of the reduction in capture to priming, due to intertrial target-feature repetitions. However, we carried out a similar analysis and found that priming only accounts for part of the reduction in capture; the remaining difference is likely the result of some engagement of feature search (Lamy, Carmel, Egeth & Leber, 2006).

3. Size of the attentional window

In the face of strong evidence of top-down feature-based influences on the guidance of attention, Theeuwes and his colleagues have suggested that top-down changes in the attentional window between search modes may be the real explanation. As researchers who have argued for the existence of top-down selection we should be delighted to have Theeuwes acknowledge at least one source of top-down control, albeit not quite of the flavor we have promoted. We think this may well be an important factor, possibly closely related to the distinction between singleton detection mode and feature-search mode. Thus it may seem almost churlish to point out some problems with the work on the attentional window.

The idea was largely circular until the paper by Belopolsky, Zwaan, Theeuwes, and Kramer (2007) in which the size of the window was explicitly manipulated. The results of that study led to the conclusion that salience computations are not carried out outside the attentional window. Belopolsky et al. (2007) point out that this conclusion is supported by the results of Joseph, Chun, and Nakayama (1997) who found that a singleton detection task could not be carried out successfully in the periphery when participants were engaged in an RSVP letter-identification task at fixation. More specifically, those authors found that orientation discrimination suffered from an attentional blink when presented soon after the target letter in the RSVP stream.

The problem with this conclusion is that it does not comport with a wide variety of other results in the literature. For one, Egeth, Leonard, and Palomares (2008) carried out a study somewhat similar to that of Joseph et al. (1997) and found no attentional blink when observers had to indicate if zero or one target-colored patches were shown among distractors in the periphery. There are, of course, numerous other studies that show that participants performing a difficult task at fixation can also detect singletons in the periphery (e.g., Braun & Sagi, 1990). However, all of the foregoing studies in this section are examples of dual-task paradigms in which it would make sense for participants to devote some attention to the periphery.

In experiments by Folk, Leber, and Egeth (2002), participants attended to a rapid stream of letters at fixation and reported the identity of the letter presented in the target color. Task-irrelevant number signs were presented in the periphery. When one of these number signs matched the target color, a strong “spatial blink” occurred, even though, to repeat, the peripheral number signs (and

the locations they occurred at) were task-irrelevant. As the central task was demanding, this suggests that salience computations are indeed carried out beyond the confines of the attentional window (see Lamy, Leber, & Egeth, 2004, Exp. 3, for a more detailed discussion of the role of salience in this paradigm).

The emphasis we see here on the importance of the spatial distribution of attention is consistent with Theeuwes's belief in the primacy of space over features. We don't want to do the equivalent of throwing the baby out with the bathwater. The notion that the size of the attentional window may be set top-down, and may, in turn, modulate the extent to which irrelevant singletons capture attention is an interesting one that deserves, indeed requires, further study. However, evidence for this account is not nearly convincing enough to surmount strong findings showing that feature-based processing is important in the earliest allocation of attention (this issue is discussed further in the section on physiological evidence).

4. Feature weighting

We also find problematic the discussion in section 4.3 of top-down weighting of stimulus features or dimensions, which argues against the ability of top-down settings to affect the selection of a feature singleton. Several investigators have shown positive effects (e.g., Müller, Reimann, & Krummenacher, 2003). Theeuwes suggests that the effects observed may not be due to changes in the initial allocation of attention but rather to processes subsequent to attentional selection, as suggested by the pattern of results obtained by Theeuwes, Riemann, and Mortier (2006). In a singleton detection task, advance cueing of the dimension of the upcoming singleton resulted in both costs and benefits. When the same singleton search was used but the task now required identifying the orientation of a line within the singleton, the costs and benefits due to advanced cueing were eliminated. Theeuwes et al. (2006) argue that their study shows that changing the response requirement to a compound search task eliminated the validity effect, suggesting that the top-down cueing really is affecting subsequent processes such as response selection. They conclude that in singleton search only bottom-up priming is effective; top-down knowledge cannot guide the allocation of spatial attention to a featural singleton.

Theeuwes's (2010) argument ignores an experiment by Leonard and Egeth (2008) directed at this question, with a design very similar to that of Theeuwes et al. (2006). They also used a compound discrimination task and gave participants verbal cues concerning the color of the upcoming singleton. They investigated whether intertrial priming could really account for what on the surface would seem like top-down guidance to a singleton target. Perhaps the most important observation is that even when the target color differed from that of the previous trial, significant facilitation of attentional allocation occurred when top-down guidance was available. That is, the effect exists even in the absence of priming from the previous trials. On the face of it, these results showing a strong top-down cueing effect with a compound discrimination task would seem directly contrary to those of Exp. 2 of Theeuwes et al. (2006). Leonard and Egeth (2008) pointed out that the difference in outcomes may be due to the difference in display sizes (3, 5, and 7 in their case vs. 9 for Theeuwes et al.). As display size increased, the target became more salient and reaction times became faster for those trials in which search relied on bottom-up salience. However, when an informative cue allowed for selection to be guided by top-down factors, the reaction times were consistently faster, regardless of the target's salience. At Theeuwes's display size 9, the target is highly salient and the response generated may be of the same latency as the speed possible with top-down guidance. This could be construed as consistent with Theeuwes's (2010) argument that attentional effects should become more prominent when responses are slowest. However, it is also equally compatible with the notion that top-down

effects are real even early in perception, but that it becomes difficult as a practical matter to discern them when reaction times approach a “floor” or when similar RTs can result from different guidance mechanisms.

5. Physiological evidence

Theeuwes focuses much discussion on single-cell recording experiment by Ogawa and Komatsu (2004), offering it as a pillar of support for his perspective. This is understandable, as the stimuli and design were much like that of the additional singleton paradigm (and, we must point out, open to the use of singleton detection mode). The chief result was that V4 neurons differed in their responses depending on whether the monkey was in a color search or a shape search condition—but that difference only emerged after 195 ms. Theeuwes (2010) claims this as support for his current model which has a first sweep through the brain of visual information unaffected by attention. That is, the neural responses up through 195 ms. reflect the first sweep, while the subsequent divergence in the functions reflects reentrant feedback from higher centers to V4.

In line with his theory, Theeuwes (2010) also emphasizes that spatial attention can modulate the feed-forward sweep of visual processing (e.g., Hillyard & Munte, 1984), as evidenced by amplitude modulations in sensory-specific ERP components such as the P1. Furthering his proposal, he cites work showing that the direction of attention to non-spatial features does not occur until much later.

However, these citations do not tell the whole story. Zhang and Luck (2009) have recently shown clear evidence that feature-based attention to a particular color can indeed increase early sensory processing for stimuli matching the current top-down attentional set. In this study, participants maintained fixation while monitoring a target-colored subset of moving dots presented peripherally at an attended spatial location. An irrelevant set of dots was presented occasionally in the unattended visual field. When these dots matched the attended color, there was a significant increase in the amplitude of the P1 component. In addition to showing modulation by feature-based attention within the first 100 ms of cortical visual processing, this result is also consistent with contingent capture accounts in showing that this type of top-down biasing occurs across the visual field.

A recent fMRI study has strikingly confirmed the point that top-down influence can occur across the visual field, and can occur outside the region on which the window of spatial attention is focused (Serences & Boynton, 2007). Indeed, in this study the “feature-specific attention effects spread across the visual field—even to regions of the scene that do not contain a stimulus” (Serences & Boynton, 2007, p. 301). Those authors quite plausibly thought that this spread of feature-based attention to empty regions of space could “facilitate the perception of behaviorally relevant stimuli by increasing sensitivity to attended features at all locations in the visual field” (p. 301).

The Serences and Boynton study is consistent with other evidence that feature-based attention can affect neural responsiveness not just before 195 ms, but *before the stimulus is even presented*. Such prestimulus effects are known as baseline shifts and have been shown in single-cell recordings and fMRI studies when not only spatial, but featural attention has been manipulated (e.g., Chawla, Rees, & Friston, 1999; Hayden & Gallant, 2005).

The feature-based modulation of neural activity in the absence of direct visual stimulation (including baseline effects) is akin to what psychologists have long called “attentional set.” The fuller picture that emerges from demonstrations of such effects is that it is possible for top-down attentional set to modulate neural firing in ways other

than on-the-fly reentrant processing of the sort presumably underlying the Ogawa and Komatsu (2004) data.

6. Conclusion

The debate over the contribution of top-down processes to visual selection is by no means over. We look forward in the expectation that further research will elucidate the mechanisms underlying selection, be they top-down, bottom-up, or somewhere in between.

Acknowledgments

H. E. was supported in part by a grant from the Office of Naval Research (N000141010278); C. L. was supported in part by grants from the National Institute of Mental Health (R01MH076226 and R01MH065034).

References

- Bacon, W. F., & Egeth, H. (1994). Overriding stimulus-driven attentional capture. *Perception & Psychophysics*, *55*, 485–496.
- 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.
- Braun, J., & Sagi, D. (1990). Vision outside the focus of attention. *Perception & Psychophysics*, *48*(1), 45–58.
- Chawla, D., Rees, G., & Friston, K. J. (1999). The physiological basis of attentional modulation in extrastriate visual areas. *Nature Neuroscience*, *2*(7), 671–676.
- Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual Review of Neuroscience*, *18*, 193–222.
- Egeth, H., Leonard, C. J., & Palomares, M. (2008). The role of attention in subitizing: is the magical number 1? *Visual Cognition*, *16*(4), 463–473.
- Folk, C. L., Leber, A. B., & Egeth, H. E. (2002). Made you blink! Contingent attentional capture produces a spatial blink. *Perception & Psychophysics*, *64*, 741–753.
- 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*, 1030–1044.
- Hayden, B. Y., & Gallant, J. (2005). Time course of attention reveals different mechanism for spatial and feature-based attention in area V4. *Neuron*, *47*(5), 637–653.
- Hillyard, S. A., & Münte, T. F. (1984). Selective attention to color and location: An analysis with event-related brain potentials. *Perception & Psychophysics*, *36*(2), 185–198.
- Joseph, J. S., Chun, M. M., & Nakayama, K. (1997). Attentional requirements in a “preattentive” feature search task. *Nature*, *387*, 805–808.
- Lamy, D., Carmel, T., Egeth, H., & Leber, A. B. (2006). Effects of search mode and intertrial priming on singleton search. *Perception & Psychophysics*, *68*(6), 919–937.
- 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*, 1019–1031.
- Leber, A. B., & Egeth, H. E. (2006). Attention on autopilot: Past experience and attentional set. *Visual Cognition*, *14*, 565–583.
- Leonard, C. J., & Egeth, H. E. (2008). Attentional guidance in singleton search: An examination of top-down, bottom-up, and intertrial factors. *Visual Cognition*, *16*(8), 1078–1091.
- Muller, H. J., Reimann, B., & Krummenacher, J. (2003). Visual search for singleton feature targets across dimensions: Stimulus- and expectancy-driven effects in dimensional weighting. *Journal of Experimental Psychology: Human Perception and Performance*, *29*, 1021–1035.
- Ogawa, T., & Komatsu, H. (2004). Target selection in area V4 during a multidimensional visual search task. *The Journal of Neuroscience*, *24*(28), 6371–6382.
- 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*, 1354–1361.
- Serences, J. T., & Boynton, G. M. (2007). Feature-based attentional modulation in the absence of direct visual stimulation. *Neuron*, *55*, 301–312.
- Theeuwes, J. (1992). Perceptual selectivity for color and form. *Perception & Psychophysics*, *51*(6), 599–606.
- Theeuwes, J. (2010). Top-down and bottom-up control of visual selection. *Acta Psychologica*, *135*, 77–99 (this issue).
- 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.
- 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.
- Zhang, W., & Luck, S. J. (2009). Feature-based attention modulates feedforward visual processing. *Nature Neuroscience*, *12*, 24–25.