Title
Selection, Engagement, & Enhancement: A Framework for Modeling Visual Attention

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 43(43)

ISSN
1069-7977

Authors
Lovett, Andrew
Bridewell, Will
Bello, Paul

Publication Date
2021

Peer reviewed

Selection, Engagement, & Enhancement: A Framework for Modeling Visual Attention

Andrew Lovett (andrew.lovett@nrl.navy.mil)
Will Bridewell (will.bridewell@nrl.navy.mil)
Paul Bello (paul.bello@nrl.navy.mil)
US Naval Research Laboratory
Washington, DC USA

Abstract

This paper presents a theoretical framework for modeling human visual attention. The framework's core claim is that three mechanisms drive attention: selection, which picks out an item for further processing; engagement, which tags a selected item as relevant or irrelevant to the current task; and enhancement, which increases sensitivity to task-relevant items and decreases sensitivity to task-irrelevant items. Building on these mechanisms, the framework is able to explain human performance on attentionally demanding tasks like visual search and multiple object tracking, and it supports a broad range of predictions about the interactions between such tasks.

Keywords: cognitive modeling; visual attention

Introduction

Visual attention plays a critical role in human cognition, allowing us to focus mental processing on task-relevant objects in the world around us. Unfortunately, despite its obvious importance, attention remains an elusive concept—researchers cannot even agree on whether it is a process, a resource, or an effect. To address this challenge, we previously argued for setting aside 'attention' as a holistic concept and instead exploring the specific mechanisms that support attentive processing. We illustrated our approach in computational models of two attentionally demanding tasks: multiple object tracking (Lovett, Bridewell, & Bello, 2019b) and visual search (Lovett, Bridewell, & Bello, 2019a).

Here, we present a framework that generalizes over our previous models and lays the groundwork for both modeling and predicting human performance on a broad range of tasks. The framework builds on the core claim that three key mechanisms support visual attention in humans: selection, enhancement, and engagement. Selection picks out an item for further processing, as when participants find a target during visual search, or when a highly salient distractor captures attention (Wolfe, 2007; Belopolsky & Awh, 2016). Enhancement increases sensitivity to either a location or a visual feature, as when a visual cue at a particular location primes participants to respond more quickly to future stimuli at that location (Egly, Driver, & Rafal, 1994; Posner, 1980).

Selection and enhancement are closely interconnected: after an object is selected, its features and location may become enhanced, such that similar or nearby objects are likely to be selected in the future. However, this connection is modulated by engagement, the third mechanism, which provides a means for individuals to exert top-down control over attentive processing (Lovett et al., 2019a). After an object is selected, if the object is task-relevant, then an individual may choose to engage with it, triggering enhancement of the object's location and task-relevant features to support further processing of that object. Alternatively, if the object is task-irrelevant, the individual may disengage from it, triggering suppression of its location so that a new object can be quickly selected. In this way, individuals can influence their attentive mechanisms so that task-relevant objects are more likely to be selected and processed in the future.

Our framework distinguishes itself from other theories of attention in that it combines specificity with generality. Firstly, we forego the poorly defined term *attention* to focus on three specific mechanisms, each of which can be captured concretely in computational models such as those we've developed previously. Secondly, we propose a general role for those mechanisms across a broad range of attentionally demanding tasks. Notably, enhancing an object's location or features may (1) cause the object to be selected more frequently and more quickly (Belopolsky & Awh, 2016), (2) allow the object to be tracked as it moves (Franconeri, Jonathan, & Scimeca, 2010), (3) aid in segmenting the object out from its background (Gheri, Morgan, & Solomon, 2007), and (4) cause the object to receive more weight during ensemble perception, when summary statistics (for example, average size) are computed across all objects in the visual field (De Fockert & Marchant, 2008). By making this strong claim about enhancement's generality, we are able to maximize our framework's predictive power, as it applies both to individual tasks and to interactions among the tasks. Further research will be required to test the framework's predictions and revise it if necessary.

In the following section, we provide further background on our three attentive mechanisms, including describing two previous computational models that incorporated these mechanisms. Afterwards, we present a theoretical framework in which selection, engagement, and enhancement direct visual processing. This framework lays the groundwork for future computational models that can explain human performance on a range of attentionally demanding tasks. We close by considering predictions derived from the framework and directions for future work.

Background

Although visual attention may be poorly defined, there are well-studied mechanisms that play a clear role in supporting it. Firstly, selection focuses visual processing on a particular item—often an object—in the visual field, allowing individuals to identify and respond to the object's location and features. This mechanism plays a central role in visual search tasks where participants must determine if a target is present (Treisman & Gelade, 1980; Wolfe, 2007), as well as discrimination tasks where participants must determine whether a target has a particular feature (Belopolsky & Awh, 2016). Despite its importance, selection is not directly under the viewer's control; rather, it often picks out task-irrelevant but highly salient distractor objects in a scene (Awh, Belopolsky, & Theeuwes, 2012), giving rise to the term *attentional capture.*

Critically, selection is not purely driven by salience. It is also influenced by a second mechanism, enhancement, which increases (or decreases) sensitivity to stimuli. We use the broad term enhancement to cover both spatial enhancement (commonly called space-based or object-based attention) and feature enhancement (commonly called feature-based attention). For example, after viewing a cue at a particular location, participants respond more quickly to stimuli that appear at that location (Posner, 1980; Eriksen & St. James, 1986). Furthermore, if the cue appears within the contour of an object, participants respond more quickly to stimuli appearing within that same object's contour, suggesting that spatial enhancement respects the contours of available objects (Egly et al., 1994). Finally, after viewing a target with a particular feature (for example, a color), participants respond more quickly to other stimuli with that feature across the entire visual field (Theeuwes, 2013).

We view selection and enhancement as interconnected in a cyclic relationship: selecting an object causes its location and features to be enhanced, such that nearby or visually similar objects are more likely to be selected in the future. However, it is critically important that people be able to exert some control over this relationship, so that they can increase the chances that task-relevant objects will be selected. Therefore, we previously proposed that an engagement mechanism lies between selection and enhancement (Lovett et al., 2019a). After an item is selected, participants may choose to engage with the object if it is task relevant. Engagement triggers enhancing the object's location and task-relevant features, so that the selection mechanism will maintain focus on the object. In contrast, participants may choose to disengage from a task-irrelevant object, which triggers suppressing the object's location so that another object can be selected. Critically, engagement and disengagement do not only influence processing of the currently selected object—they also affect processing of future objects that match the selected object's location or features.

In the following sections, we describe a computational model of multiple object tracking that uses selection and enhancement, and a model of visual search that uses all three attentive mechanisms. We next consider how one mechanism, enhancement, may influence visual processing more broadly.

Modeling Multiple Object Tracking

Multiple object tracking (MOT) is a well-established task that tests the limits of visual attention in dynamic scenes (Pylyshyn & Storm, 1988). In the task, participants must track a set of moving targets while distinguishing them from identical-looking distractors. Typically, participants can track 4–5 or more moving targets at once, suggesting that attention can be distributed among these objects, but without a clear theory of what 'attention' is, it is difficult to say what is being distributed.

In developing a MOT model (Lovett et al., 2019b), we suggested that selection is a serial process that picks out one target at a time, whereas spatial enhancement is a parallel process that can encode the locations of multiple previously selected objects (e.g., all of the targets in the tracking task). Furthermore, we claimed enhancement is supported by a parallel updating mechanism that tracks all target locations as the targets move by repeatedly matching each target's enhanced region to the closest visible object. In this way, multiple target locations can be tracked in parallel even when the targets are not selected. However, individual targets can be selected to identify additional information beyond the target's location, for example using its motion direction to predict where it will emerge after traveling behind an occluder. Note that if a participant is performing the MOT task well, distractors are never selected, and so their locations are never enhanced.

Using our model, we were able to simulate human performance in two classic MOT experiments and to explain the ways targets compete for attention: (1) targets compete for space because spatially enhancing a target's location results in suppressing the surrounding area (Tsotsos, 1995; Desimone & Duncan, 1995), which can interfere with processing of nearby targets (Franconeri et al., 2010) (Figure 1); (2) targets compete for time because only one target can be selected at a time to compute additional information about the target, such as the direction it is traveling.

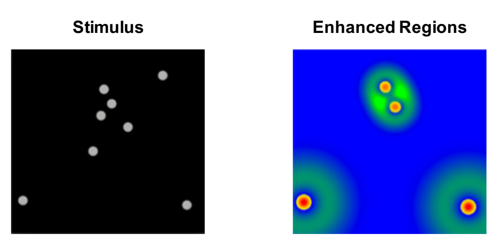

Figure 1: Enhanced (red and yellow) and surrounding suppressed (green) regions for four tracked targets. The two neighboring targets mutually suppress each other.

Modeling Visual Search

In visual search, participants must identify a target object in a field of distractors. The task is useful for determining which factors influence selection, making it easier or harder to pick out the target. In an influential review of the field, Awh, Belopolsky, and Theeuwes (2012) claimed that three factors drive this selection: the bottom-up salience of the target, the top-down goals of the searcher, and the history of recent selections. This claim was supported by a recent study that showed all three factors influencing the same search task (Belopolsky & Awh, 2016). In the study (Figure 2), participants had just 100 ms to find a target circle that could be either blue or orange and to determine the orientation of a line within it. Participants performed the task more accurately when all the distractor circles were the same color (bottom-up salience), they were told which target color to expect before the trial started (top-down goals), or the target color was a repeat of the previous trial's color (selection history).

In developing a visual search model (Lovett et al., 2019b), we suggested that the select-engage-enhance sequence could explain the influences of both top-down goals and selection history. The explanation for selection history is straightforward: when participants select a target, its color becomes enhanced, such that a target with the same color may be selected more easily in the future—notably, if a distractor is selected by mistake, participants quickly disengage from it, so the distractor's color does not become enhanced. To explain top-down goals, we proposed that when participants are told to expect a particular color, they recall a previous target that possessed the same color from memory. This recalled object representation can be selected just as new objects in the world are selected, and its color can be enhanced. By combining this account of selection and enhancement with a simple salience model, we were able to simulate human performance on the Belopolsky (2016) search task.

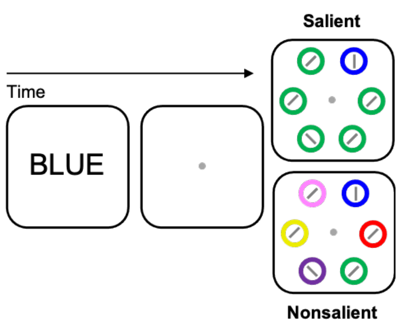

Figure 2: Visual search example adapted from Belopolsky and Awh (2016).

Generalizing to Other Tasks

The previous examples show how the select-engage-enhance sequence can influence object tracking and visual search. However, neuroscience research has found that both spatial and feature enhancement increase neural sensitivity in the early visual cortex (Somers, Dale, Seiffert, & Tootell, 1999; Roelfsema, Lamme, & Spekreijse, 1998; Saenz, Buracas, & Boynton, 2002; Zhang & Luck, 2009), suggesting that these mechanisms may play a far broader role across all of vision. Here, we consider two tasks that may be affected: figure-ground segmentation and ensemble perception.

Figure-ground segmentation involves picking out figures representing possible objects from the background. This process can be immensely complex in cluttered environments, where the background itself may be full of overlapping objects. In such cases, it would be valuable if enhancement not only aided in selecting a task-relevant object from among the available figures but also aided in segmenting those figures from the ground. For example, enhancing red could aid in picking out a red pepper from a bag of fruit, whereas enhancing a particular motion direction could aid in segmenting out a person walking through a crowd in which other people are walking in different directions. To take an example from psychological research, consider the crowding phenomenon, in which perception of a target object's features is hampered by the presence of nearby objects. This effect becomes less pronounced when the nearby objects are less similar to the target (Gheri et al., 2007; Kooi, Levi, Tripathy, & Toet, 1994). For example, in Figure 3, it is easier to perceive the central object's orientation when its color is different from the surrounding objects. This similarity effect could be driven by enhancement—by enhancing the color white, participants might more precisely segment out the target object from the non-white surrounding objects, and thus be able to judge the target object's orientation more effectively.

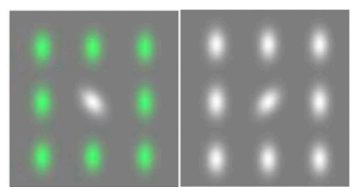

Figure 3: Crowding stimuli from Gheri et al. (2007). The task is to judge the central figure's orientation.

Ensemble perception involves viewing a set of objects and estimating some summary statistic, for example, the average size, color, or orientation of the objects (Ariely, 2001; Chong & Treisman, 2003). This process appears to be influenced by attention (Li & Yeh, 2017), with recently attended objects receiving more weight in the summary computation (De Fockert & Marchant, 2008). Generalizing from this finding, we propose that when summary statistics are computed, objects receive more weight if their locations or features are enhanced due to a person's recent selection history.

Modeling Framework

Figure 4 depicts our proposed modeling framework, which builds on the previous computational models and explores the roles of selection, engagement, and enhancement in visual processing. We view this as an implementation-agnostic framework that could be instantiated in different computational architectures. In our own research, we are implementing the framework in ARCADIA, an architecture designed for

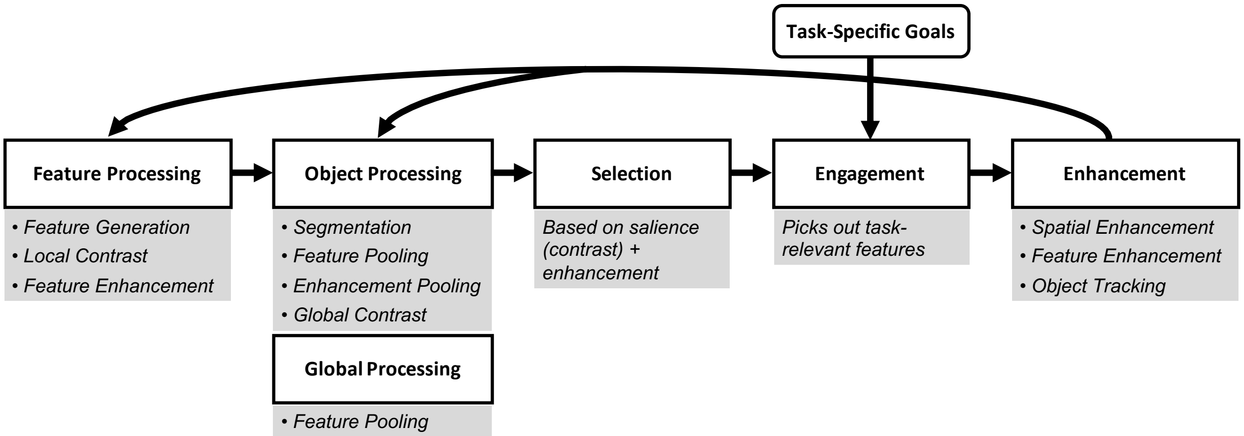

Figure 4: Framework for modeling visual attention. Arrows indicate the flow of information.

modeling the interactions among attention, perception, cognition, and action (Bridewell & Bello, 2016).

In the following sections, we describe the framework using the visual search example (Figure 2) as a running example.

Feature Processing

Feature processing computes feature information—including color, brightness, edge orientation, and motion direction—at every point in the visual field. There are three types of information: feature values (e.g., how blue a point is, how bright it is, or how quickly it is moving in a rightward direction), local contrast (e.g., how much more blue than its surroundings a point is), and feature enhancement, which is based on similarity to the currently enhanced feature values (Treue & Trujillo, 1999). For example, in Figure 2, enhancing blue will strongly affect points along the blue circle, weakly affect points along the purple circle, not affect points with no bluish hue. All of this processing is computed across the image at multiple scales, consistent with a classic computational approach to visual salience (Itti, Koch, & Niebur, 1998).

Object Processing

Object processing identifies regions of interest corresponding to possible objects in the visual scene and encodes a *proto-object* representation (Rensink, 2000) to describe each region. These proto-objects serve as candidates for selection—if a proto-object is selected, then a full-fledged object representation will be constructed to describe the corresponding object. In the current example, the six colored circles would each have an associated proto-object, making each circle a candidate for selection.

Proto-objects are identified through figure-ground segmentation, which picks out maximally large, continuous regions of constant color in the image (Palmer & Rock, 1994). Notably, this process can operate on any of the information produced by feature processing, for example segmenting out figures that are brighter, redder, or faster than their immediate surroundings. Furthermore, after an object has been selected and its features have been enhanced, segmentation can operate directly over the feature enhancement map, segmenting out figures that closely match the enhanced feature profile. For examine, consider again the case of a red pepper in a bowl of fruit. Normally, it would be difficult to segment out the red pepper due to the large number of overlapping objects. But in a feature enhancement map where the color red was enhanced, there would be only two regions—the enhanced red region, and everything else—making segmentation easy.

Although such steps are unnecessary in the current example, they would be helpful in a more natural search scenario, for example enhancing the color red to segment out a red pepper from the other fruit in a bowl, where the large number of overlapping objects would otherwise make it difficult to segment figure from ground.

After proto-objects are identified, a pooling process computes average feature information over all the points lying within each proto-object's contour. A similar pooling process computes the total spatial and featural enhancement for each proto-object. These pooling processes provide access to an object's average feature values—for example its average hue and brightness—as well as its overall degree of enhancement. In the current example, the proto-object for the blue circle would represent its color and indicate that it is has received strong feature enhancement.

Finally, each proto-object's visual salience—the degree to which it stands out from other items in the scene—is computed via three steps: 1) local contrast within a proto-object is pooled, indicating for example how blue a proto-object is compared to its surroundings (Nothdurft, 1993); 2) global contrast is computed by comparing pooled contrast values across all proto-objects, indicating for example how blue a proto-object is compared to all other proto-objects (Madison, Lleras, & Buetti, 2018); 3) salience is computed by integrating across all features within a dimension, indicating for example whether a single proto-object is the only one whose color stands out (Itti et al., 1998). If we consider the two rightmost images in Figure 2, in each case the blue circle is the bluest object in the scene. However, in the bottom example the blue circle is nonsalient because other objects are unique along other color dimensions, for example, one circle is the reddest and another is the greenest. In contrast, in the top example the blue circle is salient because no other circle's color stands out.

Global Processing

Operating in parallel with object processing, global processing models ensemble perception by computing average feature and spatial information across the entire visual scene. This step uses a feature pooling process similar to the one in object processing, but points are weighted by their degree of spatial and feature enhancement.

Although the visual search task does not require ensemble perception, an alternate task that did would be biased by the pattern of enhancement. For example, an estimate of the average location of all circles would be biased towards the enhanced, blue circle's location.

Selection

Selection picks out one of the proto-objects to make it available for further processing. To facilitate this process, proto-objects are scored based on the sum of their salience and their (spatial and feature) enhancement. To prevent proto-objects from being selected the moment they appear, each proto-object's score accumulates over time until it exceeds a threshold, with the accumulation rate depending on the overall score. For example, the salient blue circle has a higher overall score than the nonsalient blue circle (it benefits from both salience and feature enhancement); therefore, its score accumulates at a greater rate, and it will be selected more quickly after onset.

After a proto-object is selected, a corresponding object representation is encoded in visual short-term memory (VSTM), which stores records of recently selected objects. VSTM makes object representations available for further reasoning, for example confirming that the selected circle is blue and identifying the orientation of the line inside it.

Engagement

Engagement can be handled through task-specific rules, which associate particular object properties with an *engage* or *disengage* action. These rules trigger when an object is encoded in VSTM. For example, a rule might indicate that when a blue or orange circle is selected, the system should engage with that circle, whereas if any other circle is selected, the system should disengage from it. An addition rule might indicate that the system should disengage from the gray fixation circle in the middle of the display, to avoid being distracted by it.

Engage actions can specify which particular feature dimensions are task relevant. In the current example, the color and orientation dimensions are relevant, because color is used to evaluate whether a correct circle has been selected, and orientation is used to judge the orientation of the line within the circle.

Enhancement

Finally, enhancement responds to engage/disengage actions by either enhancing or suppressing an object's location and enhancing the object's task relevant features. This mechanism operates over all recently selected objects stored in VSTM, but the strength of the enhancement decays over time unless an object is reselected. In the current example, after the blue circle is selected and engaged with, its location and color will be enhanced, resulting in two effects: 1) the blue circle will continue to be selected while it remains visible; 2) other blue circles will be selected more frequently and more speedily in the immediate future.

Enhancement interfaces with a parallel updating mechanism that tracks the locations of all VSTM objects as they move, even when they are not currently selected. However, this mechanism has no effect on the current example task because all objects are stationary.

Predictions

Because the framework posits a general role for the three attentional mechanisms, it supports novel predictions about how different attentionally demanding tasks should interact. These predictions would not be possible with a theory that focused on attention's role in a specific task, such as multiple-object tracking or visual search. Here, we consider two such predictions related to *inhibitory tagging*, a phenomenon in which, after performing a visual search task, participants respond more slowly to stimuli that appear at the locations of former search items (Klein, 1988; Thomas & Lleras, 2009). According to our framework, this phenomenon occurs because visual search involves selecting and quickly disengaging from search items (or groups of items) until a target is discovered. This disengagement process causes the search items' locations to be suppressed, to prevent reselecting a previously searched location. Because this suppression may linger after the task concludes, participants will be slower to select objects that appear at former search locations.

For our first prediction, we claim that if visual search were replaced with a task that encourages engaging with stimulus locations, then the inhibitory tagging effect would be reversed, creating a *facilitory tagging* effect. For example, suppose identical stimuli were used, but the first task was to monitor all objects and respond if any one of them changes. Unlike visual search, this task would require engaging with the stimuli, resulting in their locations being enhanced. As a result, participants should respond more quickly to stimuli that appear at the locations of formerly monitored items.

For our second prediction, we claim that inhibitory and facilitory tagging should not only influence how quickly a stimulus is selected but should also influence how strongly stimuli are weighted during ensemble perception. After a visual search task, if participants are instructed to determine an average feature value over a stimulus array, then stimuli at former search locations should receive less weight. But after a monitoring task, stimuli at formerly monitored locations should receive more weight. Hence, the task a person is performing should have far-reaching effects, influencing attention and perception on both that task and future tasks.

Future Work

Despite its explanatory and predictive power, the proposed framework is far from complete. Two important directions for future work are modeling perceptual actions and exploring cross-modal interactions.

First, perceptual actions allow an individual to exert control over perception and attention, directing processing towards task-relevant objects. The present framework includes two such actions—engage and disengage—but there are many others, including mental actions like scanning along a path (Ullman, 1984) and physical actions like eye and head movements. By integrating more of these actions into our framework, we can capture the active, strategic nature of perception (Tsotsos & Kruijne, 2014).

Second, vision is not the only modality that informs visual attention. Information from other modalities should be able to influence vision, as when you hear an approaching car and look in its direction. To explore this interaction, we are currently developing a cross-modal interface in which, after an object captures attention in the auditory modality, eye movements and spatial enhancement direct visual processing to the object's corresponding visual location.

Conclusion

Visual attention, as a holistic concept, is difficult to define, presenting a challenge to researchers who want to study it, model it, or write papers about it. Here, we have presented an alternative approach, in which we identify a set of attentive mechanisms and explore how they interact during attentionally demanding visual tasks, including multiple object tracking, visual search, feature discrimination in crowded environments, and ensemble perception. This approach gives rise to a computational framework that can explain a broad set of psychological results while producing testable predictions about visual tasks and the interactions between them. We hope this framework will aid other researchers in developing theories and models of attention, both in vision and across cognition.

Acknowledgements

The authors would like to acknowledge support from the Office of Naval Research and the Naval Research Laboratory. The views expressed in this paper are solely the authors' and should not be taken to reflect any official policy or position of the United States Government or the Department of Defense.

References

Ariely, D. (2001). Seeing sets: Representation by statistical properties. *Psychological Science*, *12*(2), 157–162.

Awh, E., Belopolsky, A. V., & Theeuwes, J. (2012). Top-down versus bottom-up attentional control: A failed theoretical dichotomy. *Trends in Cognitive Sciences*, *16*(8), 437–443.

Belopolsky, A. V., & Awh, E. (2016). The role of context in volitional control of feature-based attention. *Journal of Experimental Psychology: HPP*, *42*(2), 213–224.

Bridewell, W., & Bello, P. (2016). A theory of attention for cognitive systems. In *Fourth annual conference on advances in cognitive systems.*

Chong, S. C., & Treisman, A. (2003). Representation of statistical properties. *Vision Research*, *43*(4), 393–404.

De Fockert, J. W., & Marchant, A. P. (2008). Attention modulates set representation by statistical properties. *Perception and Psychophysics*, *70*(5), 789–794.

Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual Review of Neuroscience*, *18*(1), 193–222.

Egly, R., Driver, J., & Rafal, R. D. (1994). Shifting visual attention between objects and locations: Evidence from normal and parietal lesion subjects. *JEP: General*, *123*(2), 161–177.

Eriksen, C. W., & St. James, J. D. (1986). Visual attention within and around the field of focal attention: A zoom lens model. *Perception & Psychophysics*, *40*(4), 225–240.

Franconeri, S. L., Jonathan, S. V., & Scimeca, J. M. (2010). Tracking multiple objects is limited only by object spacing, not by speed, time, or capacity. *Psychological Science*, *21*(7), 920–925.

Gheri, C., Morgan, M. J., & Solomon, J. A. (2007). The relationship between search efficiency and crowding. *Perception*, *36*(12), 1779–1787.

Itti, L., Koch, C., & Niebur, E. (1998). A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *20*(11), 1254–1259.

Klein, R. (1988, aug). Inhibitory tagging system facilitates visual search. *Nature*, *334*(6181), 430–431.

Kooi, F. L., Levi, D. M., Tripathy, S. P., & Toet, A. (1994). The effect of similarity and duration on spatial interaction in peripheral vision. *Spatial Vision*, *8*(2), 255–279.

Li, K. A., & Yeh, S. L. (2017). Mean size estimation yields left-side bias: Role of attention on perceptual averaging. *Attention, Perception, and Psychophysics*, *79*(8), 2538–2551. doi: 10.3758/s13414-017-1409-3

Lovett, A., Bridewell, W., & Bello, P. (2019a). Attentional capture: Modeling automatic mechanisms and top-down control. In *Proceedings of the 41st annual meeting of the cognitive science society.*

Lovett, A., Bridewell, W., & Bello, P. (2019b). Selection enables enhancement: An integrated model of object tracking. *Journal of Vision*, *19*(23).

Madison, A., Lleras, A., & Buetti, S. (2018). The role of crowding in parallel search: Peripheral pooling is not responsible for logarithmic efficiency in parallel search. *Attention, Perception, and Psychophysics*, *80*, 352–373.

Nothdurft, H. C. (1993). Saliency effects across dimensions in visual search. *Vision Research*, *33*(5-6), 839–844.

Palmer, S., & Rock, I. (1994). Rethinking perceptual organization: The role of uniform connectedness. *Psychonomic Bulletin & Review*, *1*(1), 29–55.

Posner, M. I. (1980). Orienting of attention. *The Quarterly Journal of Experimental Psychology*, *32*(1), 3–25.

Pylyshyn, Z., & Storm, R. W. (1988). Tracking multiple independent targets: Evidence for a parallel tracking mechanism. *Spatial Vision*, *3*(3), 179–197.

Rensink, R. A. (2000). The dynamic representation of scenes. *Visual Cognition*, *7*, 17–42.

Roelfsema, P. R., Lamme, V. A., & Spekreijse, H. (1998). Object-based attention in the primary visual cortex of the macaque monkey. *Nature*, *395*(6700), 376–381.

Saenz, M., Buracas, G. T., & Boynton, G. M. (2002). Global effects of feature-based attention in human visual cortex. *Nature Neuroscience*, *5*(7), 631–632.

Somers, D. C., Dale, A. M., Seiffert, A. E., & Tootell, R. B. (1999). Functional MRI reveals spatially specific attentional modulation in human primary visual cortex. *PNAS*, *96*(4), 1663–8.

Theeuwes, J. (2013). Feature-based attention: It is all bottom-up priming. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *368*(20130055), 1–11.

Thomas, L. E., & Lleras, A. (2009, aug). Inhibitory tagging in an interrupted visual search. *Attention, Perception, & Psychophysics*, *71*(6), 1241–1250.

Treisman, A., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, *12*(1), 97–136.

Treue, S., & Trujillo, J. C. M. (1999). Feature-based attention influences motion processing gain in macaque visual cortex. *Nature*, *399*(6736), 575–579.

Tsotsos, J. K. (1995). Modeling visual attention via selective tuning. *Artificial Intelligence*, *78*(1-2), 507–545.

Tsotsos, J. K., & Kruijne, W. (2014). Cognitive programs: Software for attention's executive. *Frontiers in Psychology*, *5*, 1–16.

Ullman, S. (1984). Visual routines. *Cognition*, *18*, 97–159.

Wolfe, J. M. (2007). Guided search 4.0: Current progress with a model of visual search. In W. Gray (Ed.), *Integrated models of cognitive systems* (pp. 99–119). New York: Oxford.

Zhang, W., & Luck, S. J. (2009). Feature-based attention modulates feedforward visual processing. *Nature Neuroscience*, *12*(1), 24–25.