Attentional Capture: Modeling Automatic Mechanisms and Top-Down Control

Andrew Lovett (andrew.lovett@nrl.navy.mil)
Will Bridewell (will.bridewell@nrl.navy.mil)
Paul Bello (paul.bello@nrl.navy.mil)
U.S. Naval Research Laboratory, 4555 Overlook Ave SW
Washington, DC 20375 USA

Abstract

We present a computational model of attentional capture in humans. The model distinguishes between automatic mechanisms that directly determine the focus of visual attention, and deliberate mental actions an individual can perform to influence these mechanisms. The automatic mechanisms select an object as the focus of attention and enhance its location and features, so that nearby or similar objects are likely to be selected in the future. The deliberate actions include engaging with a selected object to further enhance its features, and retrieving a previously selected object from memory. By performing these actions, the model is able to exert limited top-down control over capture, increasing the probability that task-relevant objects will be attended and irrelevant objects will be ignored. To evaluate the model, we conduct a simulation of a recent visual search study, demonstrating that the model can account for three established factors that are known to influence capture.

Keywords: visual attention; visual search; computational modeling

Introduction

What drives attentional capture? That is, when we view a scene, why is our attention drawn to one object, and not to another? This question is important because where we attend determines what information we represent. Whether we are reading a map, driving a car, or shopping at a store, we can perform the task more efficiently if we attend to objects that provide relevant information and ignore task-irrelevant objects.

Much of the debate over attentional capture concerns the role of top-down control (Folk, Remington, & Johnston, 1992; Müller, Reimann, & Krummenacher, 2003; Theeuwes, Reimann, & Mortier, 2006). To what extent can humans deliberately manipulate our own mental states, such that task-relevant objects are more likely to be attended? The evidence suggests that in many cases, task-relevant objects draw attention not because of deliberate control, but because they are similar to objects we have attended recently. For example, if a task involves looking for red objects, the act of finding a red object on previous trials will prime the viewer to find one more easily on future trials (Maljkovic & Nakayama, 1994; Theeuwes et al., 2006). However, in some cases participants appear to be able to strategically tune their attentional systems based on semantic information, such as a word describing the color of the object they should find next (Leonard & Egeth, 2008; Belopolsky & Awh, 2016).

To better understand how top-down goals affect attentional capture, it is helpful to model the specific mechanisms underlying attention. We previously developed a model of multiple-object tracking that relied on two attentional mechanisms: selection and enhancement (Lovett, Bridewell, & Bello, 2017). Selection picks out an item for further processing, and may be thought of as a generalized form of attentional capture, whereas enhancement increases sensitivity to stimuli at a particular location or with particular features. These two mechanisms are closely interwoven: after an object is selected, its location and features are enhanced, such that objects at the same location or with similar visual features are more likely to be selected in the future.

Here, we present a novel computational model that applies the selection and enhancement mechanisms to a visual search task, in which participants must find a blue or orange circle in a field of distractor circles and judge the orientation of a line inside it (Figure 1). Critically for the topic at hand, neither selection nor enhancement is directly controlled in the model. However, other deliberate actions can influence what gets enhanced, thereby biasing the model to select task-relevant objects. In particular, after an object is selected, if the object is task-relevant then the model can `engage` with it. Engagement is the act of maintaining focus on an object while reasoning about its features, for example, judging the orientation of a line inside an attended circle. Engagement leads to greater enhancement of an object's location and features, which supports sustained selection of that object but also causes objects with similar features to be selected in the future.

In the model, engagement also causes the object's representation to be stored in long-term memory, from which it can be retrieved at a later time. Thus, if the model later receives a cue, for example indicating that the next search target will be orange, it can deliberately `retrieve` a representation of a previously selected orange object from memory, allowing that representation to be selected and engaged with, so that orange objects are more likely to be selected.

In the following section, we describe three factors that affect attentional capture, and we argue that our model, which integrates selection and enhancement mechanisms with deliberate mental actions, can explain each factor. We then present the model and describe an evaluation in which it simulates human performance on a search task. We close by considering predictions of the model and directions for future research.

Background

At least three factors govern which objects capture visual attention when viewing a scene: physical salience, selection history, and top-down goals (Awh, Belopolsky, & Theeuwes, 2012). Physical salience increases with the amount of contrast between an object and the rest of the scene, but decreases with the amount of contrast between the other objects in a scene; for example, a red circle will be strongly salient in a field of identical green squares (Duncan & Humphreys, 1989). Salience is determined by both local contrast (between an object and its immediate surroundings) and global contrast (between an object and the other objects throughout the visual scene) (Nothdurft, 1993; Madison, Lleras, & Buetti, 2018).

Whereas salience is a property of the visual stimuli, selection history relates to the viewer's mental state. An object will tend to draw attention if it is visually similar to objects that have been attended in the recent past. In search tasks, this effect often manifests as intertrial priming, where a target is found more easily if its features remain constant from one trial to the next (Maljkovic & Nakayama, 1994). Similarly, a target is found more easily if it is in the same location as a recently attended object (Folk et al., 1992).

Finally, top-down goals involve deliberate control over what object captures attention. This effect is demonstrated when viewers see a cue describing a target, rather than an object similar to the target, and then are able to find the target more readily. A spatially descriptive cue might be an arrow pointing to the region where the target will appear (Posner, 1980), whereas a featurally descriptive cue might be a word describing the target's distinguishing feature (e.g., "red") (Leonard & Egeth, 2008). The ability to use these cues suggests the viewer is making an adjustment that causes objects that match the description to draw attention.

Recently, Belopolsky and Awh (2016) examined the combined contributions of these three factors to attentional capture. They used a search task in which participants viewed six colored circles, found a target circle that could be either blue or orange, and reported whether the line inside the circle was horizontal or vertical (Figure 1). To explore the effect of salience, the colors of the distractor circles were varied: on half the trials, all the distractors were green, resulting in a salient target, whereas on the other half, the distractors were all different colors, resulting in a nonsalient target. To explore the effect of top-down goals, each search trial was preceded by a verbal cue, either the word "blue" or "orange," that predicted the upcoming target's color 80% of the time. Finally, to explore the effect of selection history, performance on repetition trials, where the target's color was the same as the color from the previous trial (e.g., the circle was orange for two trials in a row), was contrasted with performance on non-repetition trials.

Critically, in one study Belopolsky and Awh (2016) presented the search display for only 100 ms, after which the lines within each circle were masked. This brief display time has two major advantages: (1) there is no time to saccade to one of the circles, so eye movements cannot be a factor, and (2) if the first circle attended by the participants is neither blue nor orange, there is no time to look for another circle. Thus, the authors were able to isolate attentional capture from the separate task of assessing whether an attended object meets the search criteria.

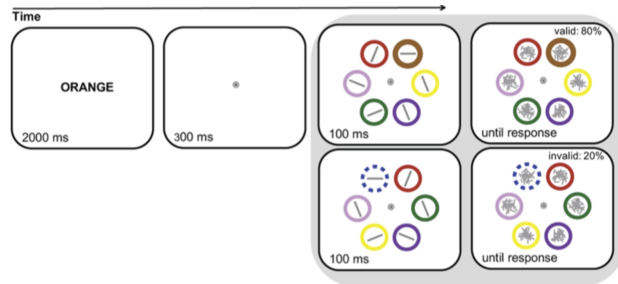


Figure 1: Examples of visual search task with a nonsalient target from Belopolsky and Awh (2016). The top row shows a valid trial with an orange target (appears brownish), whereas the bottom row shows an invalid trial with a blue target. The blue circle is dotted for illustration purposes.

Figure 2 shows the experiment results. Accuracy increased when the target was salient, when the cue was valid (e.g., the cue "orange" preceded an orange circle), or when a target color repeated, indicating that each of the three factors contributed to attentional capture. In addition, there were numerous interactions, notably, cue validity had a greater effect when the target was nonsalient, target repetition had a greater effect when the target was nonsalient, and there was a three-way interaction among the factors. We propose that these interactions are driven by a ceiling effect. As an example, when a target is salient, there is a high likelihood of attending to it during the critical 100 ms, and thus there is little room for additional improvement if the cue is valid.

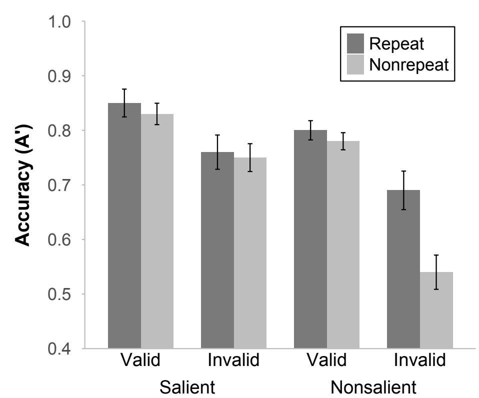


Figure 2: Visual search results from Belopolsky and Awh (2016). Error bars are ± 1 SE.

Selection and Enhancement

Both selection history and top-down goals may result from an interaction between selection and enhancement. After an object is selected, viewers show enhanced sensitivity to other objects in the same location or with the same visual features (Posner, 1980; Egly, Driver, & Rafal, 1994; Bichot, Rossi, & Desimone, 2005). Enhancement manifests as both a greater probability of selecting a stimulus among a field of distractors, and a shorter delay between stimulus onset and selection. Neural evidence suggests enhancement is rooted in modulation of the early visual cortex, for example, after a red object is selected, neurons will respond more strongly to red stimuli throughout the visual field (Somers, Dale, Seiffert, & Tootell, 1999; Saenz, Buracas, & Boynton, 2002).

Applying selection and enhancement to the Belopolsky and Awh (2016) study (discussed previously), the effect of selection history can be readily explained: participants should select an orange circle more quickly if the previous target was also orange because the recent selection would cause the orange color to be enhanced. Explaining top-down goals requires one further step—-after participants view a cue such as the word "orange," they must perform some mental action that produces a representation of an orange object, so that the representation can be selected and the color orange can be enhanced. We propose that participants retrieve a previous example of an orange object from memory. Such a retrieval should be easy, as participants are regularly engaging with orange circles throughout the experiment (note that one alternate hypothesis might be that participants perform mental imagery, imagining an orange circle).

In the next section, we describe a computational model of human performance on the Belopolsky and Awh (2016) search task.

Model

The model is based on three core claims about human attentional processing.

1) Selection picks out a single focus of attention, such as an object in the visual field. Objects are selected based on their *activation strength*, which is a combination of physical salience and spatial/featural enhancement. An object with a higher activation strength is more likely to be selected from among a field of other objects. In addition, an object with a higher activation strength will be selected more quickly after its onset.

2) Selecting an object enables constructing an object representation that can be stored in *visual short-term memory* (VSTM), which is a low-capacity store for representations of recently selected objects (Treisman & Gelade, 1980; Vogel, Woodman, & Luck, 2001). Once an object is represented in VSTM, the viewer can decide to `engage` or `disengage` with the object, depending on whether the object is task relevant. Engagement makes an object's features accessible for further reasoning and supports storing the object's representation in *long-term memory* (LTM), where it will be available for retrieval at a later time. In addition, engagement causes an object's location and features to be enhanced, which helps to maintain focus on the object, while also increasing the probability that nearby or similar objects will be selected. In contrast, disengaging from the object causes its location to be suppressed, so that a different object can be selected.

3) A viewer can `retrieve` an object representation matching a verbal cue (e.g., "orange") from LTM. If this retrieved object representation is selected, then it will be stored back in VSTM, and its features can be enhanced.

Model Framework

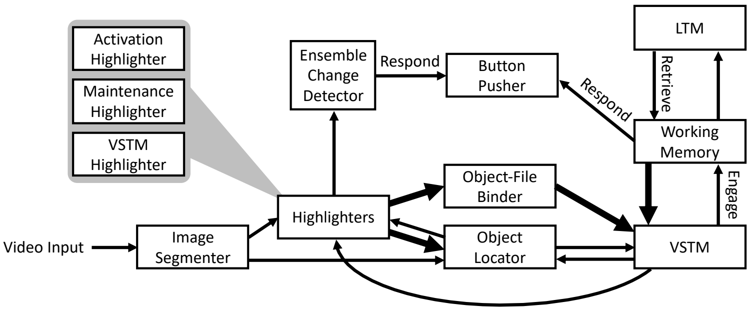
The model is implemented in ARCADIA (Bridewell & Bello, 2016), a computational framework developed to explore the relationships among attention, perception, cognition, and action. ARCADIA models operate over a sequence of cycles. On each cycle, a set of components work in parallel, processing input and generating output. One output item is selected as the focus of attention, and then the next cycle commences, with components receiving as input the output from other components on the previous cycle.

Models built in ARCADIA consist of (1) a set of components; (2) an *attentional strategy*, which sets out the priorities for which component's output will be selected as the focus of attention after each cycle, and (3) optionally, a set of *stimulus-response links*, which indicate that once certain conditions are met, an action should be taken.

Model Runthrough

Figure 3 presents the model's components and illustrates the flow of information. Thin arrows indicate information that flows on every cycle, thick arrows indicate information that flows only when it is selected as the focus of attention, and arrows accompanied by words indicate information that flows only when an action is taken. In the following sections, we shall describe the components and the flow of information in detail, using the search task in Figure 1 as a running example. Note that the model is designed to run on stimuli identical to those shown to humans, with one exception: because the model lacks reading comprehension, the verbal cues "orange" and "blue" are replaced with horizontal and vertical rectangles, respectively. Thus, a horizontal rectangle indicates that the next target will likely be orange.

Figure 3 also provides the model's stimulus-response links, which indicate the conditions under which the model should `engage` with an object, `retrieve` an object representation from memory, or `respond` by pressing a virtual button to end a trial. Whereas many of the model's components perform general-purpose visual processing and have been used in other task models (Bridewell & Bello, 2016; Lovett et al., 2017), the stimulus-response links encode task-specific knowledge about when actions should be performed (e.g., a horizontal rectangle indicates an orange object should be retrieved from memory). These actions provide a means for the model to influence the selection and enhancement mechanisms, and thereby increase the likelihood of task-relevant



Component	Conditions	Action
VSTM	Object is orange, blue, horizontal, or vertical	`Engage` with object
Working Memory	Object is horizontal	`Retrieve` an orange object
Working Memory	Object is vertical	`Retrieve` a blue object
Working Memory	Object is orange or blue	`Respond` “horizontal” or “vertical”
Ensemble Change Detector	Features pooled across the image have changed, and there are multiple objects on screen	`Respond` “guess”

Figure 3: Left: Flow of information between model components. Right: Stimulus-response links for the model.

objects being selected.

Returning to the model’s components, processing begins with the Image Segmenter, which takes each frame from an input video and segments it into regions representing possible objects. In the current example, each video begins with a horizontal or vertical rectangle, so the component identifies only one region of interest. Later, when there is a fixation circle surrounded by six larger circles, the component identifies seven regions.

Regions of interest are quickly forgotten unless they are selected as the focus of attention. To this end, components known as *highlighters* suggest particular regions as candidates for attention. In the present model, the Activation Highlighter suggests a region if its combined salience and enhancement (discussed in greater detail later) exceeds a threshold. In contrast, the Maintenance Highlighter suggests the region whose location matches the current focus of attention; this component supports maintaining focus on an object over time. Finally, the VSTM Highlighter suggests a region whose location matches any object represented in VSTM; this component supports returning focus to a recently selected object. Note that the model’s attentional strategy gives the highest priority to the Activation Highlighter and the lowest priority to the VSTM Highlighter. This means the model will (1) focus on an object with a sufficiently high activation strength, or if none have a sufficiently high strength, (2) maintain focus on the currently selected object, or if no current object is selected, (3) return focus to a recently selected object.

In the visual search example, when only the fixation circle is visible, it will be selected. When the six outer circles appear around the fixation circle, focus will be maintained on the fixation circle until the activation strength of one of the outer circles exceeds the threshold.

After a region is selected, the Object-File Binder constructs an object representation describing what is found at that region, while at the same time the Object Locator records the region’s location and begins tracking the object. The object representation includes the object’s physical dimensions and visual features (color, orientation, and brightness). In the current example, the representation contains the necessary information for determining whether a rectangle is vertical or horizontal, determining the color of a circle, or determining whether the line inside a circle is horizontal or vertical.

After an object representation is constructed, the attentional strategy prioritizes selecting it as the focus of attention, so that it can be stored in VSTM (visual short-term memory) which holds representations of the four most recently selected objects. At this point, if the object is task-relevant (a blue or orange target circle, or a vertical or horizontal rectangular cue), the model’s stimulus-response links trigger an `engage` action (Figure 3, right side). Engaging with an object causes its representation to move into Working Memory, where it is accessible to other components. In addition, engaging causes the object’s location and features to be enhanced (in the present model, the location is enhanced only when the object is visible, and the only feature that can be enhanced is color). For simplicity, if the model does not `engage` with an object, then the model behaves as if it had `disengaged` with the object: the object’s location is suppressed, which encourages selection of other objects. Note that all enhancement and suppression effects last only while the object is remembered in VSTM.

In the model, Working Memory functions as a conduit between VSTM and LTM. After a representation is copied from VSTM to Working Memory, it is stored in LTM, which has a greater capacity than VSTM. Later, if the model performs a `retrieve` action, an object representation is copied back into Working Memory, where it can be selected and stored in VSTM. The model’s stimulus-response links specify that it should retrieve an orange object after engaging with a horizontal rectangle, or retrieve a blue object after engaging with a vertical rectangle.

In the visual search example, the interactions between VSTM, Working Memory, and LTM give rise to effects of selection history and top-down goals. Suppose two sequential trials each involve an orange circle, and suppose the model successfully selects the orange circle on the first trial. Beginning with this first orange circle, the model will perform the following sequence of selections:

1. Select the first orange circle and generate a response. This ends the first trial.

2. Select the rectangular cue at the beginning of the next trial.
3. Based on the cue, retrieve an object representation from LTM and store it in Working Memory. Select this object representation.
4. Select the fixation circle that precedes the critical 100 ms.

As each of these object or object representations is selected, it will be stored in VSTM. Because VSTM has a capacity for four objects representations, all four will remain in VSTM at the beginning of the second trial's critical 100 ms. Because the circle from the previous trial is in VSTM, its color will be enhanced, resulting in a selection history effect; and because the circle retrieved from LTM is in VSTM, its color will be enhanced, resulting in a top-down goal effect.

Finally, the Button Pusher is passed one of three responses: "vertical" or "horizontal" if the model engages with a target (blue or orange) circle and determines the orientation of its inner line, or a "guess" response if the masks cover the circles before the model engages with a target circle. The appearance of the masks is detected by the Ensemble Change Detector, which responds to large-scale changes to the image.

Overall, the model succeeds at the search task if it selects and engages with the target circle during the 100 ms before the masks appear, enabling it to generate the appropriate response. It fails if either it selects the target circle after the masks appear, in which case the response may be incorrect; or it never selects the target circle, in which case it generates a "guess" response.

Activation Highlighter

The Activation Highlighter integrates salience and enhancement to determine each region's activation strength. Salience is computed via a novel algorithm based on Itti, Koch, and Niebur's (1998) classic computational approach. Operating over the color, orientation, and brightness dimensions, the algorithm computes local contrast throughout the image, and then computes global contrast for each region of interest. A region's salience varies from 0 to 1, where 1 indicates the region strongly stands out on one dimension (e.g., its color is unique, whereas the other regions all have similar colors), or moderately stands out on multiple dimensions.

Spatial enhancement is computed based on whether the region overlaps the location of an object in VSTM. For simplicity, we assign a score of 1 if it overlaps an enhanced object, -1 if it overlaps a suppressed object, and 0 if it does not overlap an object.

Featural enhancement, currently computed only for color, is based on the similarity between colors within a region and colors of objects being enhanced. A region will receive a score of 1 if it perfectly matches the colors of all enhanced objects. Note that in some cases, two different colors may be enhanced—for example, if the previous trial involved a blue circle, but the model just retrieved an orange circle from memory. In these cases, a region will receive a score based on the average of its color match to the two enhanced objects.

To ensure some randomness, Gaussian noise is added to the activation strength, according to the following formula:

$$Gaussian(gaussian\text{-}width) + weight_{sal} * Salience +$$

$$(1 - weight_{sal}) * 0.5(Enhancement_{space} + Enhancement_{features})$$

Finally, the Activation Highlighter computes the average activation strength over the past five cycles and compares this average to an *activation-threshold* to determine whether a region has a sufficiently high score to be selected. Averaging over five cycles achieves the desired effect that objects with more salience or enhancement will be selected more quickly, as it will take fewer cycles after onset for the running average to exceed *activation-threshold*.

Note that there are three free parameters: $weight_{sal}$ the weight given to salience, relative to enhancement; *gaussian-width* the width of the Gaussian noise; and *activation-threshold*. For now, we set $weight_{sal}$ to 0.2 (meaning salience receives one quarter the weight of enhancement), and we shall use the simulation that follows to explore possible noise and threshold values.

Evaluation

To simulate the Belopolsky and Awh (2016) search task, we generated input videos that match the original study's stimuli exactly, with two exceptions: (1) as discussed previously, the verbal cues "orange" and "blue" were replaced with horizontal and vertical rectangles; (2) some portions of each trial were sped up to save processing time, but the critical 100 ms display time went unchanged.

In the original experiment, 24 participants each viewed a large number of practice trials, followed by 600 search trials. For the simulation, five virtual participants each viewed 40 practice trials, followed by 600 search trials. Because the virtual participants were all the same model, and they differed only in the particular trials they viewed, we combined the 3000 (5×600) results and analyzed by item. To reduce variance, "guess" responses were treated as 0.5 correct.

We ran the simulation across a range of *activation-threshold* and *gaussian-width* values. Figure 4 presents the results with a low (0.04) or medium (0.11) threshold, and with no or moderate (0.1) noise. Overall, it appears that a medium threshold and some noise were needed to achieve human-like performance; without these, the model performed at or near ceiling for all salient targets. The rightmost graph in Figure 4 closely matches the human results (Figure 2, note that the units are different), but a qualitative comparison suggests that repetition provides a stronger benefit to the model than to humans. In the model, repetition and valid cues provide similar benefits, but perhaps the benefit from repetition should be weaker because viewers stop engaging with a target circle after a trial ends.

To examine the benefits of target salience, cue validity, and repetition, we conducted an ANOVA for each simulation run. These analyses confirmed that all three factors contributed

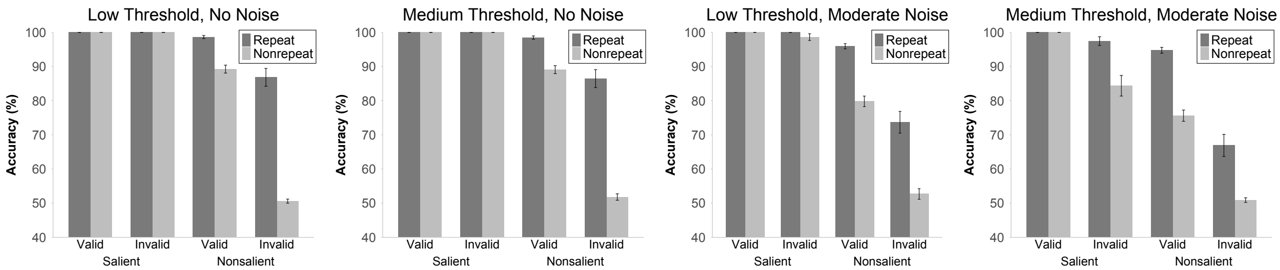


Figure 4: Model simulation results. Error bars are ± 1 SE.

significantly to accuracy (all ps $< .05$), and additionally found that most interactions between factors were significant. As we discussed when considering the human results, we believe these interactions are driven by a ceiling effect—note that performance is at 100% for some conditions.

Conclusion

Our computational model is able to perform a visual search task, while demonstrating how salience, selection history, and top-down goals influence attentional capture. In particular, the acts of `engaging` with task-relevant objects and `retrieving` previously selected objects influence which features become enhanced, thereby causing relevant objects to be selected more easily in the future.

Ultimately, the model suggests that humans possess only limited top-down control over attentional capture. For example, the model predicts that a verbal cue will be effective only when viewers are able to act on it. Suppose that after many trials of the visual search experiment, viewers are presented with a novel verbal cue, such as "red." This cue should provide little benefit because viewers have not been engaging with red circles, and thus red circles are unavailable for retrieval. In contrast, a novel *visual* cue, such as an image of a red circle preceding the search trial, should provide an immediate benefit because selecting the red circle causes its features to be enhanced.

In developing this model, we drew inspiration from previous models of visual search and attentional capture. Notably, most models explain the influence of salience (Itti et al., 1998), top-down goals (Wischnewski, Steil, Kehrer, & Schneider, 2009), or both (Tsotsos, Kotseruba, & Wloka, 2016). We believe our model is unique in explaining the influences of salience, top-down goals, and selection, while making explicit claims about the limits of top-down control.

Moving forward, we plan to evaluate our model and the parameters that have been calibrated on the present task by simulating additional search tasks. These will include conjunctive searches, in which there is benefit to enhancing multiple feature dimensions (color, orientation, curvature, etc) in parallel (Wolfe, 2007). In addition, these will include longer searches in which there is time to move the eyes. Eye movement—which can be simulated in ARCADIA—is another deliberate action that influences selection and enhancement. Thus, viewers can optimize their search performance through strategic control of their looking patterns (Pomplun, Garaas, & Carrasco, 2013). By modeling the actions and strategic decisions that affect attentional capture, we hope to better understand how people can effectively extract important, task-relevant information from the world around them.

Acknowledgments

This research was performed while the first author held an NRC Research Associateship award at the U.S. Naval Research Laboratory. The authors would like to acknowledge support from the Office of Naval Research. 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

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

Belopolsky, A. V., & Awh, E. (2016). The role of context in volitional control of feature-based attention. *JEP: Human Perception & Performance*, *42*(2), 213–224.

Bichot, N. P., Rossi, A. F., & Desimone, R. (2005). Parallel and serial neural mechanisms for visual search in macaque area V4. *Science*, *308*(5721), 529–534.

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

Duncan, J., & Humphreys, G. W. (1989). Visual search and stimulus similarity. *Psych Review*, *96*(3), 433–458.

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.

Folk, C. L., Remington, R. W., & Johnston, J. C. (1992). Involuntary covert orienting is contingent on attentional control settings. *JEP: Human Perception & Performance*, *18*(4), 1030–1044.

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.

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.

Lovett, A., Bridewell, W., & Bello, P. (2017). Goal-directed deployment of attention in a computational model: A study in multiple-object tracking. In *Proceedings of the 39th annual meeting of the cognitive science society* (pp. 2640–2645). London.

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.

Maljkovic, V., & Nakayama, K. E. N. (1994). Priming of pop-out : I. Role of features. *Memory & Cognition*, *22*(6), 657–672.

Müller, H. J., Reimann, B., & Krummenacher, J. (2003). Visual search for singleton feature targets across dimensions: Stimulus- and expectancy-driven effects in dimensional weighting. *JEP: Human Perception & Performance*, *29*(5), 1021–1035.

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

Pomplun, M., Garaas, T. W., & Carrasco, M. (2013). The effects of task difficulty on visual search strategy in virtual 3D displays. *Journal of Vision*, *13*(3), 1–22.

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

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., 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.

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

Tsotsos, J. K., Kotseruba, I., & Wloka, C. (2016). A focus on selection for fixation. *Journal of Eye Movement Research*, *9*(5), 1–34.

Vogel, E. K., Woodman, G. F., & Luck, S. J. (2001). Storage of features, conjunctions, and objects in visual working memory. *JEP: Human Perception & Performance*, *27*(1), 92–114.

Wischnewski, M., Steil, J. J., Kehrer, L., & Schneider, W. X. (2009). Integrating inhomogeneous processing and proto-object formation in a computational model of visual attention. In H. Ritter, G. Sagerer, R. Dillmann, & M. Buss (Eds.), *Human centered robot systems: Cognition, interaction, technology* (pp. 93–102). Berlin: Springer.

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.