Goal-Directed Deployment of Attention in a Computational Model: A Study in Multiple-Object Tracking

Andrew Lovett (andrew.lovett.ctr@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

Abstract

We present a computational model exploring goal-directed deployment of attention during object tracking. Once selected, objects are tracked in parallel, but serial attention can be directed to an object that is visually crowded and in danger of being lost. An attended object's future position can be extrapolated from its past motion trajectory, allowing the object to be tracked even when it is briefly occluded. Using the model, we demonstrate that the difficulty of tracking through occlusions increases with the number of objects because they compete for serial attention.

Keywords: attention; perception; cognitive model; multiple-object tracking; visual cognition

Introduction

Our visual experience of the world is rich and exhilarating, full of a wide variety of objects that move either predictably or erratically. Making sense of what we see requires an ability to follow these objects over time, sometimes tracking two, three, or more at once. The criticality of this capability is reflected in the existence of low-level mechanisms in the vision system that can follow multiple objects in parallel, seemingly without explicit attention. However, when the paths of objects intersect or when one object occludes another, these mechanisms are insufficient, requiring that we attend to an object to disambiguate it from others.

What guides our attention toward any one particular object in the visual field? There is a considerable amount of literature that seeks to answer this question by appealing to *visual salience* (Borji & Itti, 2013). That work emphasizes contrast effects in the early visual system that draw our eyes to regions of fast motion, bright lights, and flashes of color. The target phenomenon described by that literature is goal-free, visual perception—that is, where a person would look if told to freely examine an image. The corresponding results say little about top-down influences on visual attention, such as when a person adopts a goal to track an object over time.

Fortunately, there is a task that lets us study goal-directed, visual attention under exactly these conditions. Experiments on multiple-object tracking (MOT) have people distinguish one or more *targets* from a larger set of identical *distractors* as they move across an empty background (Pylyshyn & Storm, 1988). In these studies, the only identifying characteristic of each object is its motion history. As a result, the low-level mechanisms that maintain the identity of objects in parallel are taxed, which lets researchers study their limits and therefore determine the conditions that require visual attention.

We know, for instance, that when targets come close to other objects, they tend to draw attention (Iordanescu, Grabowekcy, & Suzuki, 2009; Zelinsky & Todor, 2010). This effect of *crowding* does not hold when distractors are clustered together, so the goal to track the targets must play a role in protecting them from possible confusion with the other objects. It also appears that the difficulty of MOT increases when multiple targets are crowded simultaneously (Srivastava & Vul, 2016). These findings suggest that visual attention is deployed serially and can only disambiguate one threatened target at a time.

In this paper, we present a computational model that accounts for one of the roles that attention plays in object tracking. Our research builds on previous work by Bello, Bridewell, and Wasylyshyn (2016) that assumed attention is serially deployed to initially encode targets, after which a parallel process that does not require attention exclusively handles object tracking. In their model, the interaction between attention and the tracking goal was limited to keeping visual attention on the targets. The model's ability to track targets broke down when the targets' previous positions were insufficient for distinguishing them from other nearby objects (e.g., when the objects moved quickly or were close to each other). In the updated model presented here, the processes for serial shifts of attention are refined and contribute throughout the task.

Specifically, the updated model detects when objects flagged as targets are visually crowded and, in response, directs attention to them. Sustained attention on an object enables the construction of its motion trajectory, which can be used to predict its future position. This extra information gives the model the ability to follow a target through an *occlusion event*, where another object overlaps or covers the target. Because attention is deployed serially, only one target can be tracked in this way at a time. As a result, when two targets are crowded, they vie for attention and the one that is not selected remains in danger of being lost by the parallel, tracking process.

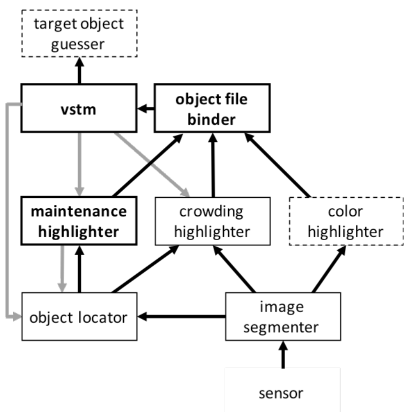


Figure 1. Flow of information between components, both bottom-up (black arrows) and top-down (gray arrows). Components in bold respond to the focus of attention. Components with dashed borders are task-specific.

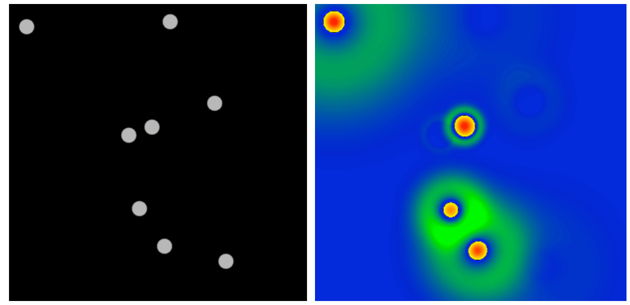
We claim that tracking through occlusion is facilitated by goal-directed deployment of attention to the target involved. We support this claim by providing a computational model of visual attention described in the next section. Briefly, this model requires that crowded targets compete for attention and its associated computational benefits. To test the model, we apply it to stimuli drawn from the work by Luu and Howe (2015) showing that people are better at predicting target positions from past trajectories when there are fewer targets. We find that the model accounts for these results and is in accordance with a broader range of findings in the literature.

Computational Model of Visual Attention

The computational model is implemented using ARCADIA (Bridewell & Bello, 2016a), a cognitive system designed for exploring the role of attention. The system operates in cycles that correspond to 25 ms of activity in human perception. On each cycle *components*, which carry out all the computation in a model, place their results in a location called *accessible content*. ARCADIA uses an *attentional strategy* to select one of these results as a *focus of attention*, which directs processing in a subset of components. On the subsequent cycle, the components receive sense data (e.g., a video frame), accessible content, and the focus of attention as input and produce the next collection of accessible content as output.

Like other models built using ARCADIA, this model of visual attention consists of a set of components and an attentional strategy. Many of the components included in the current model were previously described by Bello, Bridewell, and Wasylyshyn (2016). Looking at Figure 1, these include *image segmenter, object locator, object-file binder,* and *vstm* (which implements visual short-term memory). In the rest of this section, we summarize these components, mention changes to *object locator*, detail the new components, and discuss the attentional strategy.

Beginning at the bottom of Figure 1, *image segmenter* polls a sensor that provides one frame of video input each



A **B**

Figure 2: **A:** An image with eight objects. **B:** A priority map in which four of the objects are tracked.

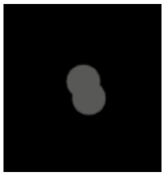
cycle. The component then outputs a set of *proto-objects*, hypotheses for the locations of objects (Rensink, 2000), based on closed-contour regions in the frame. For proto-objects to lead to object representations, they must receive attention. To this end, a set of *highlighters*, described later in this section, proposes one or more candidate proto-objects for the focus of attention. If ARCADIA focuses on one of these candidates, then the *object file binder* constructs an *object file*, which is based on the ideas of Treisman and Gelade (1980) and is a representation that binds together any visual features found at the proto-object's location (e.g., color profile, size). If that object file receives attention, then the *vstm* component stores it in memory.

Improvements to the Parallel Aspects of Tracking

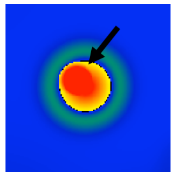
If tracking objects always required attention, then it would take four ARCADIA cycles (100 ms) to go from visual input to representing a single object in *vstm*. To update that object's location would take the same amount of time. Even accounting for the ability to pipeline parts of the process, two cycles (50 ms) are required for shifting covert attention and representing an object. The timing needed to serially update the location of multiple objects, which is based on evidence from visual search in humans (Wolfe, 2003), is unrealistic. Therefore, the model needs a way to track objects in parallel.

To address this need, the model includes tracking functionality in *object locator*. In earlier ARCADIA models, this component kept location information up-to-date by matching object files in *vstm* to the proto-objects nearest to each object's last known location. This approach was inspired by Pylyshyn's (1989) proposal that roughly four objects can be tracked in parallel using visual indices and by Dawson's (1991) work that identified a nearest-neighbor constraint in apparent motion, which is likely related to tracking.

In this model, we refine *object locator* to provide an account of Dawson's constraint based on newer results in visual processing. This new implementation generates a two-dimensional priority map (Fecteau & Munoz, 2006; Bisley & Goldberg, 2010), with enhanced regions at each tracked object's last known location and suppressed regions around them. Evidence for this treatment of spatial regions



A



B

Figure 3. **A:** An occlusion event. **B:** Part of a priority map, updated with an attended object's predicted location, shown as the off-center darker red circle.

comes from early work on multifocal attention (Castiello & Umiltà, 1992) and center-surround suppression (Tsotsos et al., 1995; Desimone & Duncan, 1995). To track objects in parallel, each object file in *vstm* is matched to the proto-object that most overlaps its corresponding enhanced region on the priority map. Figure 2B shows an example of such a map with enhanced red and yellow circles and suppressed outer green rings. Importantly, if two tracked objects are near each other, one object's suppressed region may overlap another object's enhanced region (see the two lower circles in Figure 2B), resulting in a smaller enhanced region and a greater chance of a tracked object being lost.[1]

Goal-Directed Attention in Tracking

Adopting a goal to track specific moving objects, or targets, alters how attention is deployed. In particular, attention can be drawn to a target when there is a risk that the parallel, tracking mechanisms could fail for that object. For instance, as suggested by Figure 3A, when multiple objects overlap, they look like a single proto-object. After those objects move apart, it is unclear which one, if any, was previously a target. This problem arises because following a target through an occlusion event requires more information than only its previous location. On these occasions, the model uses an attended target's recent motion history to extrapolate its future position in order to track it through occlusions. We conjecture that serial attention is required for this process because it involves binding trajectory information and the corresponding extrapolated position to a particular object file.

Goal-directed deployment of attention is assisted by the highlighters mentioned earlier in this section. Recall that these components propose proto-objects as candidates for attention and therefore determine which objects will be stored in *vstm* and tracked by *object locator*. There are three highlighters, one of which is task specific and the other two are generally important for tracking. First, *color highlighter* is used to identify targets and queries about objects in the multiple-object tracking videos, indicated by objects changing color in the videos.

The other two highlighters propose proto-objects corresponding to currently tracked objects. The *crowding highlighter* proposes each tracked object as a candidate for attention and includes as information the distance from each one to the nearest other proto-object. This value provides a measure of crowding and is based on the finding that tracked objects draw attention when they are visually crowded and in danger of being lost (Iordanescu, Grabowekcy, & Suzuki, 2009; Zelinsky & Todor, 2010).

The *maintenance highlighter* proposes maintaining attention on the object that was last in focus. If attention remains on the same object over a period of time, this component computes its motion trajectory from location changes over a window of two to three cycles. Additionally, *maintenance highlighter* detects *occlusion events*, where the focused object is partially or completely occluded by another object (e.g., Figure 3A). When the attended object is occluded, the component predicts the focused object's position based on its recorded trajectory. This information lets *object locator* update its priority map to enhance the object's predicted location (the off-center, red circle in Figure 3B), improving its ability to continue tracking the object after the occlusion event ends.

The final component, *target object guesser,* records the model's responses in the multiple-object tracking task. This component reports whether the model considers a probed item to be a target (tracked) or a distractor.

The model's attentional strategy is a priority list over the elements in accessible content. The highest priority is to focus on new object-files for storage in *vstm*. Below that, the strategy prefers proto-objects, which enables encoding them into object files. The preferences for proto-objects are ordered with *color highlighter* first*,* which ensures that targets are initially encoded and that probes are noticed when objects change color. The next highest priority is to maintain attention on a crowded target, one whose distance to the nearest other proto-object has fallen below a *crowding distance threshold*. The third highest priority is to attend to whichever target is the most crowded—the one with the lowest crowding distance. This ordering enables goal-directed deployment of attention to objects that are in danger of being lost, and it handles competition between simultaneously crowded objects. Once an attended object is endangered, attention will stay on it until the distance to nearby proto-objects exceeds the crowding threshold even if other targets are also in danger.

In summary, the model consists of eight components (Figure 1), four of which are new and one of which was substantially changed. Two components are task specific: *target object guesser* and *color highlighter*. The model includes three free parameters, two in *object locator*[1] and

[1] *Object locator* constructs a priority map in three steps. First, following Bouma's law for visual crowding (Whitney & Levi, 2011), *object locator* generates Marr wavelets centered at each tracked object's location, scaled so that the sizes of the suppressive fields increase as those objects enter the periphery. Second, since untracked stimuli produce visual crowding at a weaker rate than tracked stimuli (Whitney & Levi, 2011), for each proto-object, we include the negative component of a wavelet whose amplitude is set to 20% of that for tracked objects. Third, Holcombe, Chen, and Howe (2014) report a general cost for having more tracked objects. We account for this effect with long-range suppression in the visual field, implemented as a constant value (0.04 in the model) subtracted from the entire visual field outside of each tracked object's enhanced region.

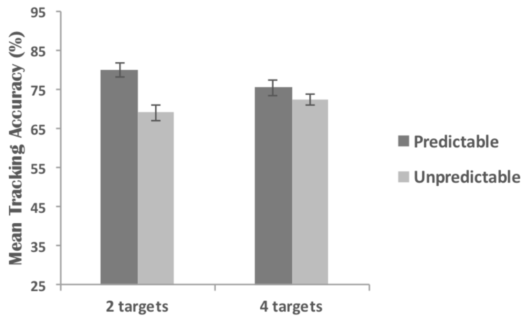


Figure 4. Simulated accuracy across conditions. Error bars are standard error. Speeds were calibrated separately for 2 and 4 targets to achieve about 75% accuracy.

the *crowding distance threshold*, which indicates when targets are too close to other proto-objects. In the next section, we report an experiment that supports the validity of this model in the context of multiple-object tracking.

Evaluation

We evaluated the computational model by running it on MOT videos similar to those used in Luu and Howe's (2015) Experiment 1. In that work, participants tracked either two or four targets with either predictable or unpredictable motion trajectories. Luu and Howe's key finding was that people more accurately track objects with predictable trajectories than with unpredictable trajectories, but only in the two target condition. The model in this paper accounts for this effect, showing that goal-directed deployment of attention can be used to predict a target's location from its past trajectory. This ability enables tracking a single target through an occlusion, and when multiple targets are simultaneously crowded, they compete for attention. This competition for resources means that task difficulty increases with the number of targets.

Experiment

In each trial of Luu and Howe's experiment, two or four out of eight total disks were highlighted in red to indicate that they were the targets. Afterwards, all disks turned black and the disks moved for 5.5 s while participants fixated on a center cross. During this time the disks could occlude (i.e., pass through) each other. At the end, two disks were highlighted in sequence, and participants indicated whether each one was a target. Each highlighted disk had a 50% chance of being a target, and participants needed to respond correctly on both for the trial to be coded as *correct*.

There were two movement conditions for the experiment. In the first condition, every disk moved predictably in straight lines and changed direction only after bouncing off the edge of the display. In the second condition, the disks moved similarly, but every 300–600 ms, they would randomly change direction. This unpredictable movement was expected to reduce the reliability of any effort to compute and utilize motion trajectories.

At the beginning of the study, the motion speeds for each participant were calibrated to determine the speed where the participant achieved 75% accuracy. Calibration occurred separately for two and four targets and used only predictable motions. Afterwards, participants were tested over 120 randomly generated trials, 30 in each condition (number of targets × motion predictability), with conditions interleaved.

Luu and Howe reported data from 15 participants. Their results found, unsurprisingly, that tracking two targets was easier than tracking four, as indicated by a much higher speed when calibrating for two targets. Importantly, they observed a significant interaction between the number of targets and motion predictability. Pairwise comparisons indicated that predictable motions were easier than unpredictable motions for two targets but not for four targets. Luu and Howe proposed that object tracking is sensitive to motion trajectories for two targets, but less so for four targets, which is in line with findings by Fencsik, Klieger, and Horowitz (2007).

Model

To evaluate the computational model using Luu and Howe's experiment, we randomly generated 120 trial videos each for 15 virtual "participants" (the model was the same in each case, so only the trial videos varied). The videos matched the description in the paper as closely as possible with five minor exceptions.

(1) There was no fixation cross, but center fixation was enforced in the model.
(2) It was impossible to match to the original study's display size (15° x 15°) because the model does not perceive the display from a quantifiable viewing distance. However, the study's proportion of disk size to display size was maintained.
(3) Videos were constrained to begin and end with all disks at least one radius apart (such constraints are common but were not mentioned in the paper).
(4) To save simulation time, disks were highlighted for a shorter duration.
(5) Disk colors differed from the original, which was incidental.

The model's *crowding distance threshold* was 1.6 diameters, meaning an attended target would need to be at least this distance away from all other disks before the model could swap attention to another target.

Results

The calibration phase of the experiment differed slightly from Luu and Howe's approach. Because the model was held constant across virtual participants, we calibrated the speeds only once. We found that to ensure roughly equivalent accuracy close to 75%, the speeds were eight pixels per cycle for two targets and four pixels per cycle for four targets.

Figure 4 displays the results for each condition. A two-way ANOVA with set size (two vs. four targets) and predictability was conducted. There was a significant main effect of predictability, $F(1,56) = 14.3$, $p < .001$, indicating that accuracy was higher with predictable trajectories. There was also a significant interaction between set size and predictability, $F(1,56) = 4.8$, $p = .032$, indicating that the effect of predictability was greater for two targets. Unpaired comparisons confirmed that predictability had a significant effect on accuracy for two targets, $M = 79.8\%$ (predictable) vs. 68.9% (unpredictable), $t(28) = 3.91$, $p < .001$, but not for four targets, $M = 75.3\%$ (predictable) vs. 72.4% (unpredictable), $t(28) = 1.23$, $p = .230$.

Discussion

The model's results matched the human data, which suggests that the model accounts for two key findings. First, as evidenced in the speed discrepancies during the calibration phase, tracking two targets was easier for the model and for people than tracking four targets. Notably, increasing the number of targets increases both the number of possible occlusion events and the potential for simultaneous crowding. These effects are important for the model, which explains errors as resulting in part from failures to attend to targets during occlusion. As a result, slowing object movement reduces the number of occlusions and contributes to the ability to successfully track targets.

The second and more important finding is that the model more accurately tracked objects that moved predictably than those that moved unpredictably, but only for two targets. To understand this, we have to describe why the model could track some objects through occlusion events when the trajectories were unpredictable. Recall that objects changed direction only every 300–600 ms, or 12–24 cycles in ARCADIA, and that the model calculates motion trajectories over a 2 cycle window. As long as a target maintains course through the occlusion and the two cycles before it, tracking should work perfectly. In practice, this means that the unpredictable trajectories only disrupt a small proportion of occlusion events.

As an explanation, the model suggests that there are two sources of error: missed occlusion events and unpredictable trajectories for attended occlusions. With four targets there are more missed occlusion events due to simultaneous crowding than with two, so proportionally that has a larger effect on the error rate than the unpredictable trajectories. This difference explains why unpredictable trajectories are more harmful with two targets than with four, and the combination of this with the overall small proportion of occlusion events disrupted by unpredictable trajectories explains the lack of a significant effect with four targets.

General Discussion

The model demonstrates the critical role of goal-directed visual attention in object tracking. Although attention is not always needed to update target locations, it provides key information to aid in tracking targets that are in danger of being lost due to visual crowding. In the reported model, attention provides a target's motion trajectory, which enables tracking through occlusions.

One explanation for how people track multiple objects is provided by the *multifocal* view of attention (Cavanagh & Alvarez, 2005). Proponents of this view have argued for two theoretical limits on attention. First, attention may be a limited resource that must be distributed among targets (Holcombe & Chen, 2012), which makes tracking more difficult when targets are crowded simultaneously and must compete for attention (Srivastava & Vul, 2016). Second, attention may be subject to spatial interference between neighboring targets (Franconeri, Jonathan, & Scimeca, 2010), which makes tracking more difficult when targets are nearer to each other (Shim, Alvarez, & Jian, 2008; Holcombe, Chen, & Howe, 2014).

The reported model offers a competing explanation that distinguishes between serial attention to a single object, which is used to bind features and compute motion trajectories; and parallel enhancement of multiple object-locations, which is used to track objects. These separate mechanisms account for both apparent limits described above. First, tracking difficulty increases when targets are simultaneously crowded because they compete for the serial focus of attention. Second, difficulty increases when targets are near each other because the parallel tracking process uses center-surround suppression, with an enhanced region and a surrounding suppressed region at each target's location. When two targets are close such that one's suppressed region overlaps the other's enhanced region, the enhanced region shrinks and there is a greater chance of losing the target. Additionally, difficulty increases with the number of targets and with object speed (Alvarez & Franconeri, 2007) because these manipulations increase the frequency of events where targets are simultaneously crowded or targets interfere with each other.

Although there are other computational models that have been applied to MOT, the reported model provides a novel explanation. Oksama and Hyönä (2008) relied solely on serial attention and Kazanovich and Borisyuk (2006) relied entirely on multifocal attention. Srivastava and Vul's (2016) Bayesian, multifocal model is similar to ours in that it distributes attention to visually crowded targets, which lets it predict greater tracking difficulty when targets are crowded simultaneously. However, their model makes no link between attentional distribution and computing motion trajectories. Additionally, the model cannot account for spatial interference between targets. Finally, their model is disconnected from video input, and it abstracts away the underlying correspondence problem.

In this paper, we demonstrated the role that goals may play in object tracking. In particular, the model's implicit goal to track targets enhances its ability by influencing where it attends. That is, selecting an object as a target recruits processes that monitor crowding and maintain focus when that target is endangered. Additionally, we note that the information made available by attending to an object is a form of indirect influence by the goal on visual processing (e.g., during the creation of the priority map). In the future we intend to explore other cases where the goal-directed deployment of attention interacts with perception and eventually with motor control.

Acknowledgments

This research was performed while the first author held an NRC Research Associateship award at the U.S. Naval Research Laboratory. The authors acknowledge support from the Office of Naval Research under grants N0001416WX01804 and N0001417WX00153. 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

Alvarez, G. A., & Franconeri, S. L. (2007). How many objects can you track? Evidence for a resource-limited attentive tracking mechanism. *Journal of Vision*, *7*(13):14.

Bello, P., Bridewell, W., & Wasylyshyn, C. (2016). Attentive and pre-attentive processes in multiple object tracking: a computational investigation modeling object construction and tracking. In *Proceedings of the 38th Annual Meeting of the Cognitive Science Society*. Philadelphia, PA.

Bisley, J. W., & Goldberg, M. E. (2010). Attention, intention, and priority in the parietal lobe. *Annual Review of Neuroscience*, *33*, 1–21.

Borji, A., & Itti, L. (2013). State-of-the-art in visual attention modeling. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *35*, 185–207.

Bouma, A. (1970). Interaction effects in parafoveal letter recognition. *Nature, 226*, 177-178.

Bridewell, W., & Bello, P. (2016a). A theory of attention for cognitive systems. In *Fourth Annual Conference on Advances in Cognitive Systems.* Evanston, IL.

Castiello, U., & Umiltà, C. (1992). Splitting focal attention. *Journal of Experimental Psychology: Human, Perception, & Performance, 18*, 837-848.

Cavanagh, P., & Alvarez, G. A. (2005). Tracking multiple targets with multifocal attention. *Trends in Cognitive Sciences*, *9*, 349–354.

Dawson, M. (1991). The how and why of what went where in apparent motion: modeling solutions to the motion correspondence problem. *Psychological Review, 98*, 561-603.

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

Fecteau, J. H., & Munoz, D. P. (2006). Salience, relevance, and firing: a priority map for target selection. *Trends in Cognitive Sciences*, *10*, 382–390.

Fencsik, D., Klieger, S., & Horowitz, T. (2007). The role of location and motion information in the tracking and recovery of moving objects. *Attention, Perception, & Psychophysics*, *69*, 567–577.

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*, 920–925.

Holcombe, A. O., & Chen, W.-Y. (2012). Exhausting attentional tracking resources with a single fast-moving object. *Cognition*, *123*, 218–228.

Holcombe, A. O., Chen, W.-Y., & Howe, P. D. L. (2014). Object tracking: absence of long-range spatial interference supports resource theories. *Journal of Vision*, *14*(6), 1.

Iordanescu, L., Grabowecky, M., & Suzuki, S. (2009). Demand-based dynamic distribution of attention and monitoring of velocities during multiple-object tracking. *Journal of Vision*, *9*(4), 1.

Kazanovich, Y., & Borisyuk, R. (2006). An oscillatory neural model of multiple object tracking. *Neural Computation*, *18*, 1413–1440.

Kramer, A. F., & Hahn, S. (1995). Splitting the beam: distribution of attention over noncontinguous regions of the visual field. *Psychological Science, 6*, 381-386.

Luu, T., & Howe, P. D. L. (2015). Extrapolation occurs in multiple object tracking when eye movements are controlled. *Attention, Perception, & Psychophysics*, *77*, 1919–1929.

Oksama, L., & Hyönä, J. (2008). Dynamic binding of identity and location information: a serial model of multiple identity tracking. *Cognitive Psychology*, *56*, 237–283.

Pylyshyn, Z. (1989). The role of location indexes in spatial perception: a sketch of the FINST spatial-index model. *Cognition*, *32*, 65–97.

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

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

Shim, W. M., Alvarez, G. A., & Jiang, Y. V. (2008). Spatial separation between targets constrains maintenance of attention on multiple objects. *Psychological Bulletin Review*, *15*, 390–397.

Srivastava, N., & Vul, E. (2016). Attention modulates spatial precision in multiple-object tracking. *Topics in Cognitive Science*, *8*, 335–348.

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

Tsotsos, J. K., Culhane, S. M., Kei Wai, W. Y., Lai, Y., Davis, N., & Nuflo, F. (1995). Modeling visual attention via selective tuning. *Artificial Intelligence*, *78*, 507–545.

Whitney, D., & Levi, D. M. (2011). Visual crowding: a fundamental limit on conscious perception and object recognition. *Trends in Cognitive Sciences*, *15*, 160–168.

Wolfe, J. M. (2003). Moving towards solutions to some enduring controversies in visual search. *Trends in Cognitive Sciences*, 7, 70–76.

Zelinsky, G. J., & Todor, A. (2010). The role of "rescue saccades" in tracking objects through occlusions. *Journal of Vision*, *10*(14), 29.