Investigating attentional theories of multiple object tracking using sparse displays

Miranda Johnson

John Palmer

Geoffrey M. Boynton

Department of Psychology,

University of Washington

G.B.'s Revision
11/30/25

Key words: attention, divided attention, multiple object tracking

Last Revised 11/30/2025

Intended for Attention, Perception, and Psychophysics

JP comments
12/9/25

Abstract

Tracking one slow-moving object is easy, but as the number of objects increases, our ability to track deteriorates. We investigate two competing attentional theories for the limits on tracking multiple objects. In serial switching theory, attention is switched from one object to the next during tracking. We focus on the more specific all-or-none serial model, in which participants track only one object, and performance is at chance for tracking a second object. Alternatively, in parallel resource theory, an attentional resource is distributed across all tracked objects in parallel; the more objects one tracks, the less resource for each object. We focus on the more specific fixed-capacity parallel model. One can implement this model by forming ~~in which~~ a representation of the stimulus ~~is formed~~ through sampling, and more targets means each target is sampled less. The current study distinguished these two specific models using a dual task with sparse displays to control the contribution of visual crowding. Performance was compared when participants tracked one (single-task) or two (dual-task) targets moving in separate regions of the visual field. The fixed-capacity parallel model predicts a specific dual-task deficit that is smaller in magnitude than that predicted by the all-or-none serial model. Additionally, the all-or-none serial model predicted a negative trial-by-trial correlation between dual-task responses, while the fixed-capacity parallel model predicted no correlation. Results showed a dual-task deficit that is consistent with the all-or-none serial model, but no negative correlation. We discuss alternative models that can account for these results.

Whether driving on a busy street or supervising children on a crowded playground, the ability to track moving objects is important in a dynamic environment. Despite the importance of this task, there are limits to how many objects can be tracked at once. Our ability to track moving objects has often been studied using the multiple object tracking paradigm (Pylyshyn & Storm, 1988). In multiple object tracking, a display is shown with some number of moving objects, and a subset of those objects are marked as targets. Participants track the targets for some length of time, and then are typically asked to either select all target objects, or they are probed with an object and asked whether the probed object is a target or not. Performance has generally been found to decrease with the number of objects tracked (Alvarez & Franconeri, 2007; Pylyshyn & Storm, 1988). The interpretation of this set-size effect is pursued in this article. More specifically, we test two attentional hypotheses that might account for set-size effects in multiple object tracking. But first, we introduce an important non-attentional limit on performance in multiple object tracking.

Crowding Theory

One phenomenon that has been found to influence performance in multiple object tracking is visual crowding. Visual crowding occurs when nearby objects interfere and with one another. Assuming a typical display with fixed width and height, as the number of objects in a display increases, the spacing between objects decreases, leading to a higher likelihood of crowding. One measure of the stimulus conditions under which crowding ~~is most likely to~~ occurs is referred to as Bouma's Law (Bouma, 1970). Under Bouma's Law, each stimulus within a display has a crowding window surrounding it, and any additional stimuli placed within that window result in crowding. The size of the crowding window is roughly equal to half of the object's eccentricity. For example, an object at nine degrees eccentricity has a crowding window surrounding it with a radius of approximately 4.5 degrees of visual angle. This rule captures the well-known result that crowding increases with eccentricity.

Most studies of visual crowding use tasks involving discrimination to show its effects (Ester, Clee, & Awh, 2013; Levi & Carney, 2009; Palomares, Pelli, & Majaj, 2001). Crowding also influences performance in multiple object tracking. In experiments where the spacing between moving objects has been manipulated, performance has been found to be worse when spacing is small versus when it is large (Franconeri, Lin, Pylyshyn, Fisher, & Enns, 2008; Shim, Alvarez, & Jiang, 2008). Thus, it is clear that crowding influences ~~our~~ the ability to track moving objects, and in experiments where object spacing is not controlled, it is difficult to distinguish crowding effects from divided attention effects. In a pure crowding theory, uncrowded displays have no set-size effect on performance in multiple object tracking. Thus, one way to test for attention effects in multiple object tracking is to use sparse, uncrowded displays, and measure whether ~~divided attention effects occur.~~ there are still set-size effects

Serial Switching Theory

The current experiment uses sparse displays to investigate alternative models of attention in multiple object tracking. Specifically, we test two attentional models for set-size effects in tracking: serial switching and parallel resource theory. Under the serial switching hypothesis, participants selectively attend to one object at a time, and must switch attention to track multiple objects. An example of a serial switching model is one in which the locations of objects are recorded and updated over time (Holcombe & Chen, 2013). By this model, when a target object is selectively attended, the location of that target is recorded before attentional selection switches to a different object. When a participant switches their selective attention back to a given target, they return to its most recently recorded location. If the target is still near that location, participants can select it and update their record of that object's ~~spatial position~~ location. However, if the target has moved far away from its most recently recorded location, or if a distractor has moved near to the recorded location, the target is lost. By this hypothesis, set-size effects occur because increasing the number of targets increases the amount of time until selective attention returns to a given object ~~and updates its~~ location. Thus, increasing numbers of targets is associated with worse performance.

An important variable that influences performance in multiple object tracking is speed. As speed increases, performance declines. For the serial switching model, the effect of speed can be understood by the related idea of temporal frequency (Holcombe & Chen, 2013). In this context, temporal frequency is the rate at which objects pass through a given spatial location. For a circular trajectory with a fixed number of objects and a fixed object speed, each point along the trajectory has a frequency at which an object passes through it. More objects along the trajectory results in a higher temporal frequency. Under the serial switching hypothesis, a higher temporal frequency leads to a higher likelihood of the target being lost because there is less time until a distractor occupies the location where the target was most recently selected. Thus, the serial switching hypothesis predicts that performance decreases with increasing temporal frequency.

The serial switching model is difficult to distinguish from models that assume limited-capacity parallel processing. Both models predict set-size effects. To make predictions for serial switching distinct, we study conditions where there is little time to switch attention. Such conditions might result in a specific version of the serial switching hypothesis, called the all-or-none serial model. Under the all-or-none serial model, attentional switching is assumed to not be possible. Instead, participants choose one target to track and stick with it through the duration of the trial without switching attention. Performance is predicted to be at chance for the unattended target. Additionally, this model predicts that there is a negative trial-by-trial correlation between responses in the dual-task condition (Sperling & Melchner, 1978). Such correlations have not been studied in typical MOT experiments (but see Howard & Holcombe, 2008). In other domains, evidence of all-or-none serial switching has been found in tasks where attention is divided across masked words separated in space (White, Palmer, & Boynton, 2018; 2019), visual search tasks that require different stimulus-response mappings (Sperling & Melchner, 1978) and tasks where attention is divided between different features of different objects (Bonnel & Prinzmetal, 1998).

Parallel Resource Theory

The second general hypotheses predicting set-size effects in multiple object tracking is parallel resource theory. Parallel resource theory posits that a limited attentional resource is shared between attended objects, and the more objects that are tracked, the less of the resource is dedicated to each object. One way to implement the idea of a limited attentional resource is to assume that the speed at which objects can be tracked depends on how much of the resource is allocated to each object (Alvarez & Franconeri, 2007). As set size increases, the amount of resources allocated to each target decreases, resulting in worse performance.

To make the predictions of parallel resource theory more concrete, we focus on a specific version called the fixed-capacity, parallel model (Shaw, 1980). Fixed capacity refers to extracting a constant amount of position information from the display per unit time. When estimates of a target object's position become sufficiently noisy, the target is lost. One way to implement this abstract idea is to assume that each target's representation is formed through a process of sampling, and the total number of samples is fixed (the sample size model; Horowitz & Cohen, 2010; Miller & Bonnel, 1994; Smith, Lilburn, Corbett, Sewell, & Kyllingsbæk, 2016).

For multiple stimuli, equal numbers of samples are drawn from each object in parallel. Each object representation can be thought of as having an associated sampling distribution, and the standard deviation of the sampling distribution is smaller with increasing numbers of samples, yielding a more accurate stimulus representation. The more stimuli that are attended, the fewer samples that are drawn from each distribution, which results in a sampling distribution with a larger standard deviation and thus a less accurate representation of the stimulus. Prior work in multiple object tracking has found set-size effects that are consistent with the fixed-capacity parallel model (Horowitz & Cohen, 2010), making it a viable account of resource theory.

Predictions for the all-or-none serial model and fixed-capacity, parallel model can be distinguished using an attentional operating characteristic (AOC; Sperling & Melchner, 1978). AOCs allow one to compare predictions for our two attentional models and predictions of crowding theory for the case of sparse displays with no visual crowding. The AOC method has
static
been commonly used in dual tasks with brief displays to measure divided attention effects. Participants complete either one task (single-task condition), or two tasks simultaneously (dual-task condition). If the two tasks are independent, performance in the dual-task condition for each task is equal to performance in the single-task condition. If the two tasks are dependent in some way, dual-task performance is worse than single-task performance. AOCs have been used in prior work to measure dual-task deficits in multiple object tracking (Alvarez, Horowitz, Arsenio, DiMase, & Wolfe, 2005), however they have not yet been used to distinguish predictions of attentional theories in a task with sparse displays.

The Current Study

The current study used a dual task in which participants tracked either one or two targets that appeared above or below fixation. To differentiate the predictions for crowding theory, discs were widely spaced such that crowding was unlikely (Bouma, 1970). To distinguish the predictions for switching, fast disc motion was used so that attentional switching was unlikely, which leaves one with the all-or-none serial model (Holcombe & Chen, 2013). We also focused on the fixed-capacity parallel model, a special case of parallel resource theory that has been found in prior work to account for set-size effects in multiple object tracking (Horowitz & Cohen, 2010).

The three models described above can be distinguished by measuring the magnitude of the dual-task deficit and the trial-by-trial correlation between accuracy of responses in the dual-task condition. For any given level of single task performance, the all-or-none serial model predicts a specific dual-task deficit and a negative correlation between accuracy for the separate responses in the dual-task condition. The fixed-capacity parallel model predicts a specific dual-task deficit that is smaller in magnitude than that predicted by the all-or-none serial model, and a zero correlation between the two responses in the dual-task condition. For sufficiently sparse displays, crowding theory predicts little or no dual-task deficit, and a zero correlation for accuracy in the dual-task condition. The models are defined more formally in the appendix and the specific predictions are described with the results of the experiment.

Experiment

Method

Participants

There were 11 paid participants. All participants had normal or corrected-to-normal acuity. All gave written and informed consent in accord with the human subjects Institutional Review Board at the University of Washington, in adherence with the Declaration of Helsinki.

To determine the number of participants, we used pilot data from an unpublished pilot study. Participants ($N = 6$) each completed a multiple object tracking task with similar methods. A dual-task deficit of 30% was observed with a standard deviation of 6%, and the correlation between accuracy for each side in the dual-task condition was $r = -.05$ with a standard deviation of .12. Our goal was to distinguish the fixed-capacity, parallel model and the all-or-none serial model. Given single task performance of 80%, we needed to discriminate deficits of 23% and 11%. A power analysis with 80% power suggested a minimum of 4 participants. For the response correlation, we needed to discriminate correlations of -.10 and 0. A power analysis using a one-tailed t-test with 80% power suggested a minimum of 11 participants. Therefore, we used 11 participants.

Apparatus

Displays were presented on a linearized CRT monitor (Sony GDM-FW900) with resolution 1024 by 640 pixels refreshing at 120 Hz. The monitor was viewed from 60 cm and the middle-gray background used in the experiment had a mean luminance of 56 cd/m^2. Stimuli were created with MATLAB (MathWorks) and Psychophysics Toolbox (Brainard, 1997). Gaze position was monitored for all trials using an EyeLink 1000 (SR Research) and the Eyelink toolbox (Cornelissen, Peters, & Palmer, 2002). Trials containing blinks or broken fixations were excluded from analysis. Such excluded trials were infrequent; across participants, blinks occurred on 1.6 ± 0.5% of trials, and broken fixations occurred on 2.8 ± 0.7% of trials.

Stimuli

As illustrated in Figure 1, participants were presented with six black discs that were one degree of visual angle in diameter. Three discs appeared above fixation, and three appeared below fixation. The discs were positioned along invisible circular paths that were centered 6 degrees above and below fixation. The diameter of the circular path was 6 degrees so that the furthest point on the path was 9 degrees above fixation, and the closest point on the path was 3 degrees above fixation. Each disc was equally spaced around the circular trajectory at an angle of 120 degrees around the circle. The linear distance between each disc on a given trajectory was approximately 5.2 degrees in visual angle, which is larger than maximum crowding window as estimated by Bouma's law (Bouma, 1970).

The difficulty of the task was

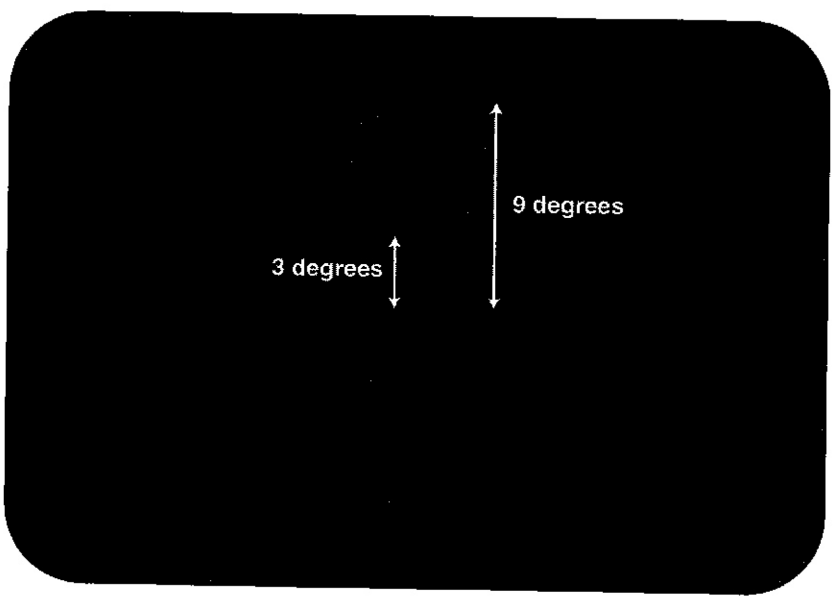

Figure 1. Stimulus object spacing. Discs were positioned along invisible circular trajectories, which are represented here by dashed circles, but were not visible in the experiment. Discs were evenly spaced along the circular trajectory. The circular trajectories were centered at six degrees in visual angle above and below fixation. The furthest points along the trajectory were 9 degrees eccentricity, and the closest points were 3 degrees eccentricity.

manipulated for each participant to maintain average single-task performance between 70-80% correct. Task difficulty was controlled by changing the speed of the disc motion. We varied rotational speeds between 1 and 2.2 rps. The maximum speed was limited to 2.2 rps to maintain the appearance of continuous motion. The average disc speed needed to obtain single-task performance between 70-80% correct was 1.6 rps (range 1.25 to 1.95 rps). This is equivalent to a linear speed of 29.5 degrees per second, or about 6 pixels per frame at 120 Hz.

Procedure

The single- and dual-task conditions are shown schematically in Figure 2A. In the single-task condition, participants tracked a single target that appeared above or below fixation. Each trial began with a blank screen for 1.5 seconds, and participants were told that they should use this blank period to blink as much as necessary and then not blink during the moving display. Following the blank period, the cue was shown for 1.5 seconds, during which six discs appeared on the screen, with three above and three below fixation. The single target disc was displayed in red, and all other discs were black. As illustrated in Figure 2B, the cue was also incorporated into the fixation point such that the top half of the stem on the fixation cross appeared in blue when the top half of the display was cued, and the bottom half of the stem on the fixation cross appeared blue when the bottom half of the display was cued.

The target then changed to black to appear identical to the distractors, and each set of three discs

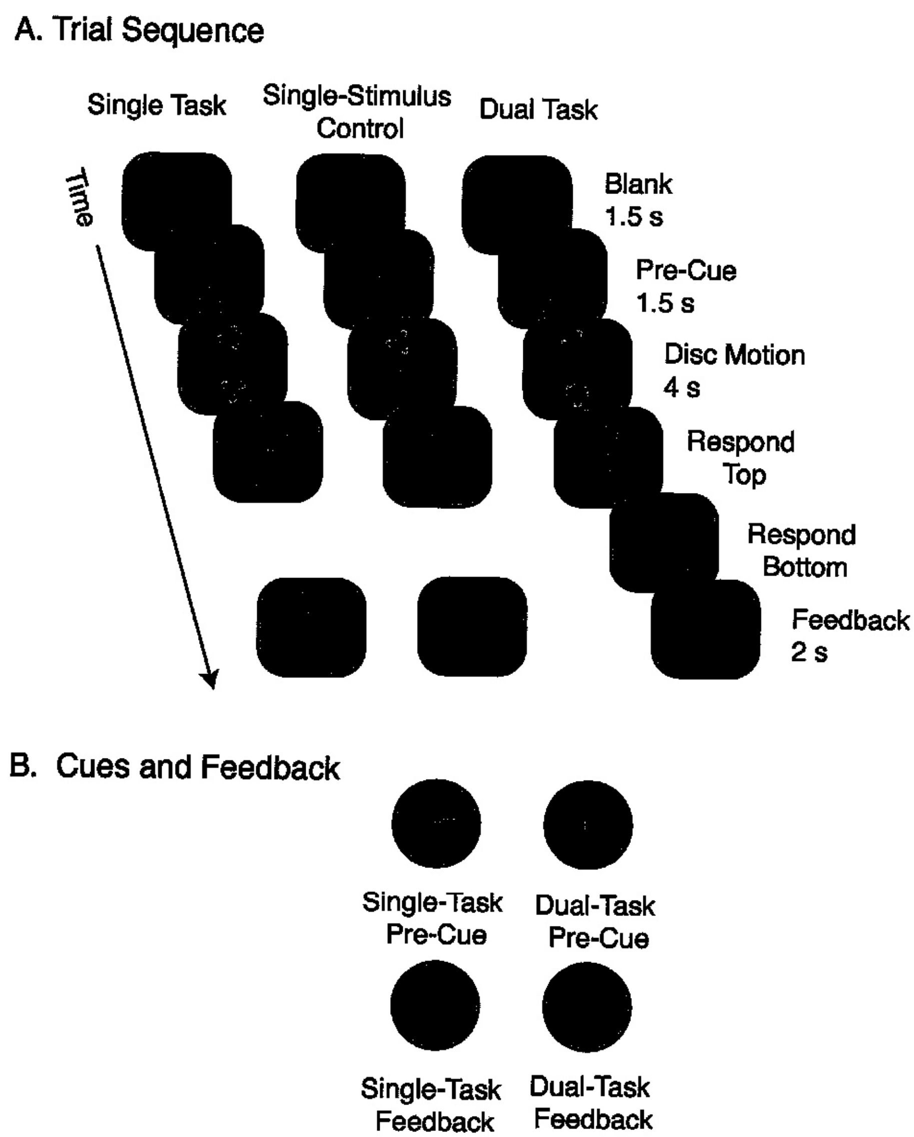

Figure 2. Trial sequence for the single-task, single-stimulus control, and dual-task conditions. In this example of the single-task condition, the top half of the display is being cued and a target is shown above fixation in red. In the dual-task condition, both sides are cued, and two targets are shown in red, one above fixation and one below fixation. The response order in the dual-task condition is counterbalanced. Feedback was incorporated into the cue. Green indicated a correct response, and red indicated an incorrect response.

immediately began moving along an invisible circular trajectory for 4 seconds. During the four seconds of disc motion, each set of discs reversed direction three times, and the timing of those reversals was determined pseudo-randomly and independently for each side of fixation. For one set of discs, an opportunity for reversal occurred every 0.5 seconds, and for the other set of discs, it occurred every 0.6 seconds. This difference in when reversals could occur made it such that the two sets of discs never reversed at exactly the same time. Which side reversed at 0.5 or 0.6 seconds was counterbalanced.

Following the disc motion, participants were prompted to select the target with a mouse-click, and they were given as much time as needed to do so. The response prompt was incorporated into the fixation cross and was identical to the fixation cue shown at the start of the trial. Mouse clicks that did not correspond with any of the three discs on the cued side resulted in a 500 Hz tone being played, after which participants were given another chance to respond. Following response, feedback was shown at fixation for 2 seconds. As illustrated in Figure 2B, feedback was incorporated into the fixation cross in the same manner as the cue, with green indicating a correct response and red indicating an incorrect response.

In the dual-task condition, participants were instructed to track two targets, one above and one below fixation. The trial sequence for the dual-task condition was similar to that for the single-task condition. The key differences were that instead of a single target that appeared in red above or below fixation, there were two targets in red, one above and one below fixation. Additionally, the cue at the start of the experiment indicated that both sides were relevant, thus the full stem of the fixation cross was blue. After the 4 second disc motion, participants were prompted by a response cue to either select the bottom target first and then the top target, or vice versa, and the order of response was counterbalanced. Feedback was shown simultaneously after both responses for 2 seconds, and was incorporated into the fixation cross in the same manner as the cue.

In addition to the single- and dual-task conditions, there was a single-stimulus control condition. For these trials, three discs appeared on the cued side of the display, and the uncued side of the display was blank. The trial sequence, stimulus, and response were otherwise identical to that of the single-stimulus condition. This condition was included as a test of crowding phenomenon.

Prior to the experiment, participants completed 2-3 training sessions, during which they

and an appropriate speed was selected

learned to use the cues ~~and~~ perform the task. Participants then completed 20 experimental sessions, which took between fifteen to twenty hours, and were completed across several weeks. Each session consisted of 8 blocks of 12 trials, making 96 trials per session and 1920 trials per participant. Within a session of 96 trials, there were 24 single-task trials, 24 single-stimulus control trials, and 48 dual-task trials. A mixed design was used such that the three conditions were randomly intermixed throughout a session of 96 trials.

Primary Results

Dual-task deficit

Performance in the single-task condition was 73.5 ± 2.5%, and performance in the dual-task condition was 52.5 ± 2.0% (chance was 33.3%). The difference is a large dual-task deficit of 21.0 ± 1.7%. This difference was statistically significant, $t(10) = 12.39, p < .001, d = 5.22$, 95% CI [17.3, 24.8]. To test how crowding contributes to this dual-task deficit, we compared performance in our single-task and single-stimulus control conditions. The mean difference in performance between these two conditions, which we call the dual-stimulus deficit, was 2.2 ±

0.5%. This difference was statistically significant $t(10) = 4.17$, $p = .002$, $d = 1.26$, 95% CI [1.0, 3.31]. While this difference suggests a crowding effect, it is small compared to the dual-task deficit. The magnitude of the dual-task and dual-stimulus deficits are plotted side-by-side in Figure 3. The y-axis represents the difference in performance between the single-task and dual-task conditions (i.e. the dual-task deficit), and the difference in performance between the single-stimulus control and single-task conditions (i.e. the dual-stimulus deficit). The dual-stimulus deficit is a small fraction of the dual-task deficit. Thus, the crowding effect measured by the dual-stimulus deficits cannot account for the observed dual-task deficit.

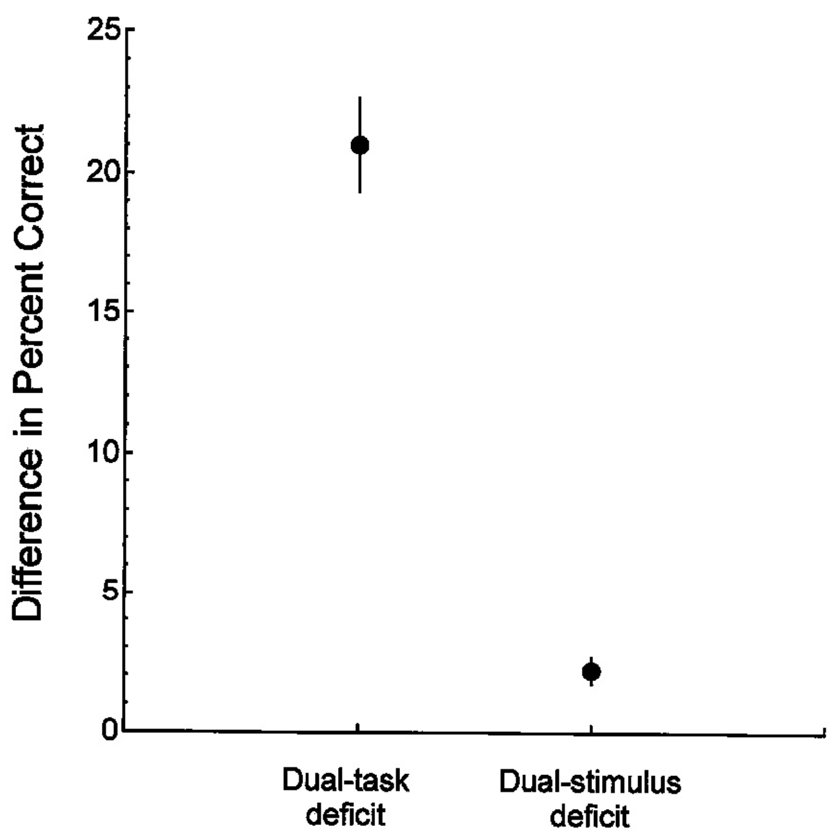

Figure 3. Comparison of the dual-task deficit and the dual-stimulus deficit. The dual-task deficit is the difference in percent correct for the single-task and dual-task conditions. The dual-stimulus deficit is the difference in percent correct for the single-stimulus control condition and the single-task condition. Error bars represent the standard error of the mean.

Figure 4 shows the results compared to model predictions using the attentional operating characteristic (AOC; Sperling & Melchner, 1978). The models are described in the Appendix. The y-axis shows performance when the target is on the top, and the x-axis shows performance when the target is on the bottom. Both axes go from chance performance (approximately 33% correct) to perfect performance (100% correct). The solid lines represent the prediction for crowding theory given a sufficiently sparse display: if our ability to track multiple objects is only limited by perceptual crowding, then there should be no divided attention effect for this sparse display. In this case, accuracy for each of the two targets in the dual-task condition should be equal to that of the single-task condition.

The dashed diagonal line represents the prediction for the all-or-none serial model: if one can track only one target object at a time, and must guess on the location of a second target, there should be a large divided attention effect. The accuracy for the two sides trades off linearly.

The dotted curved line represents a prediction for the fixed-capacity parallel model, where processing for the two sides occurs in parallel, but is fixed in capacity. Assuming signal detection theory and independent samples of the position information, one can calculate the predicted magnitude of the dual-task deficit for this model (for the details of this calculation, see the supplemental materials for White et al., 2018) and the Appendix.

Percent correct for the single-task condition is shown for the top (y-axis) and bottom (x-axis) responses. Single-task performance on the top was 75.3 ± 4.1%, and the bottom was 71.8 ± 2.8% . Dual-task performance is plotted as a point, where the x-value represents dual-task performance for the bottom target, and the y-value represents dual-task performance for the top target. Performance for the top in the dual-task condition was 51.8 ± 3.7%, and performance for the bottom was 53.3 ± 3.2%. dual task ~~deficit~~ performance lies on the diagonal line, which is consistent with the prediction of the all-or-none serial model, and much larger than the prediction of the fixed-capacity, parallel model.

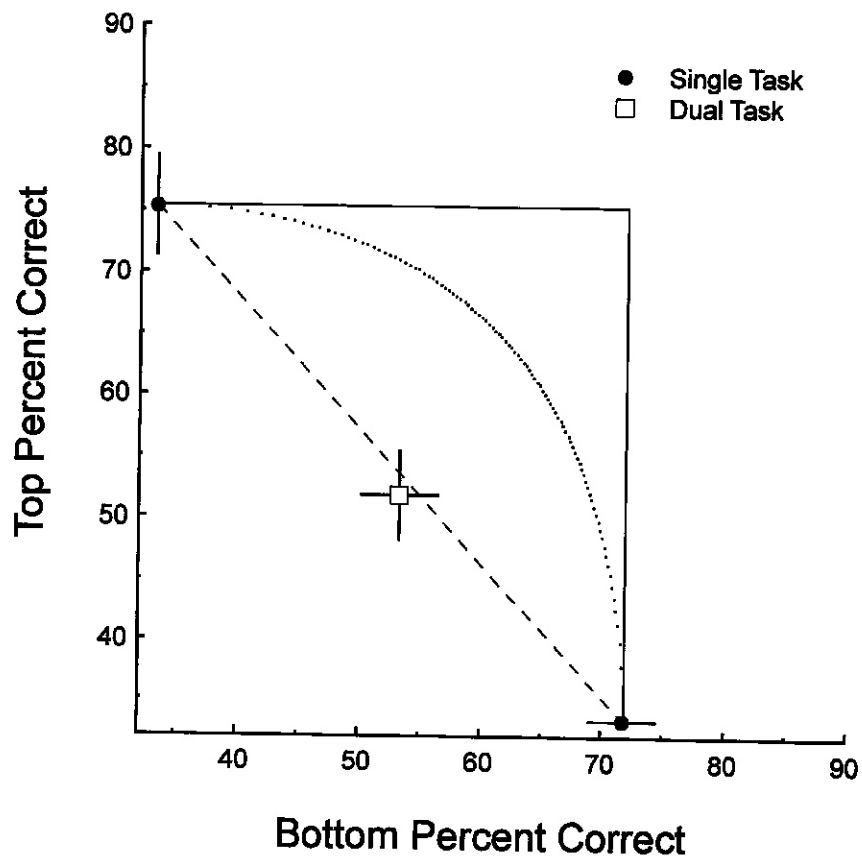

Figure 4. Percent correct for the single- (blue) and dual-task (red) conditions. Error bars represent the standard error of the mean. Model predictions are also shown. The solid lines represent the prediction for crowding theory. The dotted curve represents the prediction for the fixed-capacity parallel model. The dashed diagonal represents the prediction for the all or-none serial model.

<u>Correlation between responses in the dual-task condition</u>

Figure 5 shows the observed trial-by-trial correlation along with the predictions of the three hypotheses.

When performance is at chance, all three hypotheses predict a correlation of zero in accuracy between the top and bottom responses. As dual-task performance increases, the all-or-none serial model predicts a negative correlation in the accuracy between top and bottom targets. For example, for dual-task percent correct of 50%, this model predicts a negative correlation of $r = -.10$. By this model, the maximum possible dual-task performance is 67% correct for this 3-choice task. This model assumes that participants can track only one target in the dual-task condition, meaning that a correct response for the top side is associated with an incorrect response for the bottom side, and vice versa.

Tab

Both crowding theory and the fixed-capacity, parallel model predict no correlation in dual-task accuracy across all ranges of dual-task performance. Under crowding theory, responses on each side are independent, and therefore there should not be a correlation. Under the fixed-capacity parallel model, because the limited resource is shared equally between targets in parallel, there should be no correlation.

The correlation between dual-task accuracy for the top and bottom was r = -.05 ± .04, which was not significantly different from zero ($t(10) = 1.14$, p = 0.2808, $d = 0.34$, *95%* CI [-1.4, 0.5]; ~~[illegible]~~. ~~This~~ The confidence interval~~s~~ for this result excluded the r ≅ -.15 predicted by the all-or-none model for the observed dual-task performance.

Secondary Results

We tested for order effects in response accuracy in the dual-task condition by comparing performance for first responses and second responses. The difference in performance for first and second responses in the dual-task condition was -0.8 ± 0.6%, with second responses being slightly better, however this difference was not statistically significant, ($t(10) = 1.27$, $p = .2328$, $d = 0.38$, 95% CI [-2.3, 0.6]). Thus, there is no evidence of memory or response interference for the second response. We also tested for differences in accuracy for the top

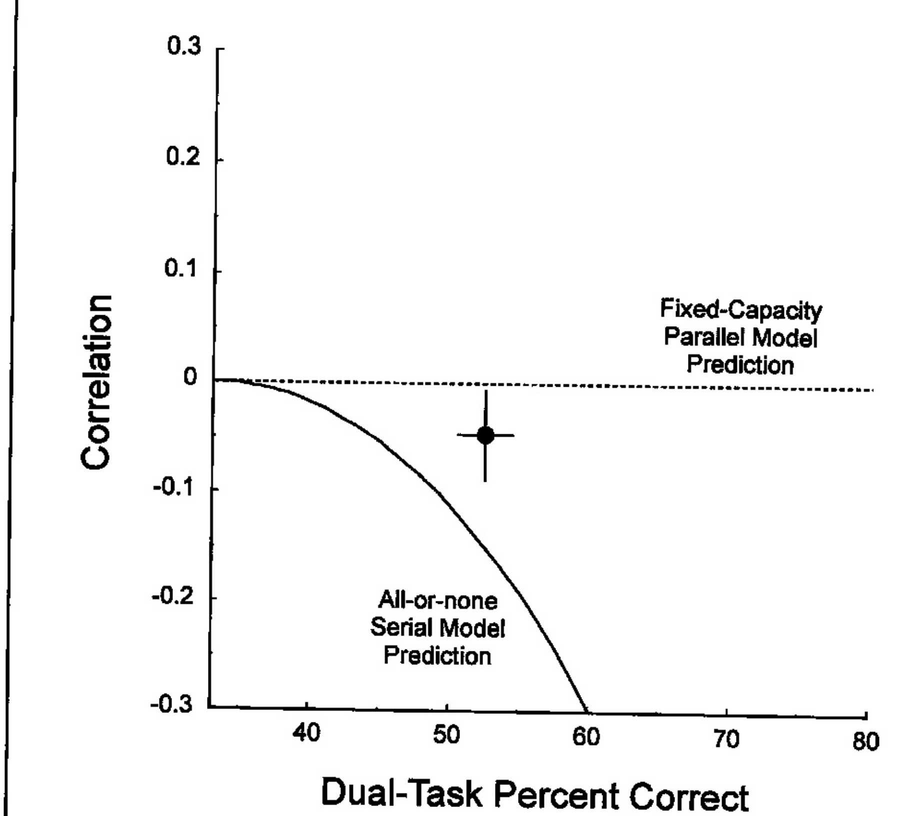

Figure 5. Predicted and observed correlations plotted as a function of dual-task percent correct. The dashed line represents predicted zero correlations under crowding theory and the fixed-capacity parallel model. The curve represents predicted negative correlations under the all-or-none serial model. The closed circle represents the average correlation, and error bars represent the standard error of the mean.

other

and bottom locations (collapsing across all conditions). The difference between the top and bottom location was 1.1 ± 5.1% ($t(10) = 0.21$, $p = .8378$, $d = 0.06$, CI [-10.2, 12.4]). Thus, accuracy was similar for the top and bottom locations.

General Discussion

The results showed a large dual-task deficit, the magnitude of which was consistent with the all-or-none serial model. This observed dual-task deficit allows one to reject the fixed-capacity, parallel model. Additionally, this dual-task deficit with our sparse displays provided evidence against crowding theory, which predicts no set-size effects in multiple object tracking when displays are uncrowded. Our observed dual-task deficit was consistent with those used in prior work that measured AOCs using a dual-task multiple object tracking design (Alvarez et al., 2005). Although the magnitude of the dual-task deficit was consistent with the all-or-none serial model, in the dual-task condition, there was no correlation between accuracy for responses on the top and bottom. This is not consistent with the all-or-none serial model. Thus, none of our three specific models can account for the entire set of results.

Implications for Serial Switching Theory

The key prediction of the all-or-none serial model is that there is a negative correlation between responses in the dual-task condition. The current study is the first to test for negative correlations between dual-task responses using typical multiple object tracking. Large divided attention effects with negative correlations have been found for other tasks using a dual-task design (Bonnel & Prinzmetal, 1998; White et al., 2018, 2019), and these results provide strong evidence consistent with serial processing. In contrast, the results of the current study showed no significant negative correlation between dual-task responses, which is not consistent with the all-or-none serial model.

By the all-or-none serial model, participants do not switch attention, but instead track only one target through the duration of the trial. One possibility for the lack of a negative correlation is that participants tried to switch attention between targets even though it was advantageous to track only one target. If participants attempt to switch attention in conditions where switching is difficult (e.g. when fast speeds result in short temporal frequencies), dual-task performance decreases over time, but no negative correlation is predicted because both sides get a chance to be tracked. Such a scenario predicts results consistent with those found in the current study. This scenario is also consistent with the findings of Holcombe and Chen (2013), where performance was worse than a model that assumes participants can only track one object, and must guess on a second object.

Some prior multiple object tracking work has found results consistent with the all-or-none serial model. Holcombe and Chen (2012) tasked participants with tracking one or two targets that moved along a circular trajectory that was centered at fixation. Tracking speeds ranged from 0.7 to 1.9 rps, and psychometric functions were fitted to data for each participant. Speed limits (the speed at which participants were 68% correct for tracking one or two targets) were estimated for each participant. The observed speed limits were much lower for tracking two targets compared to tracking one. They calculated that the observed speed limit for tracking two targets was consistent with the all-or-none serial model.

In a follow-up study, Holcombe and Chen (2013) tasked participants with tracking one to three targets among distractors. The main manipulation was of temporal frequency, the rate at which objects passed through a given spatial location. For a circular trajectory with a fixed number of objects and a fixed object speed, each point along the trajectory has a frequency at which an object passes that point. More objects along the trajectory result in a higher temporal frequency. Faster disc speeds also lead to a higher temporal frequency. Under the general serial switching hypothesis, a higher temporal frequency leads to a higher likelihood of the target being lost because there is less time until a distractor occupies the location where the target was most recently selected. Temporal frequency was manipulated by varying the number of objects on a given trajectory. On each trial, there were either 3, 6, 9, or 12 objects per trajectory. Speed thresholds, the speed at which performance fell midway between ceiling and chance, were measured for each of the tracking conditions.

The results of Holcombe and Chen (2013) showed that speed thresholds decreased with increasing numbers of targets. Speed thresholds also decreased with increasing numbers of objects on each trajectory. Importantly, when speed thresholds were converted to temporal frequencies, there was no difference in temporal frequencies for the six, nine, and twelve object conditions. These results indicate that it is not speed that led to a difference in performance across object conditions, but temporal frequency. These set-size and temporal frequency effects are consistent with the general serial switching model. Additionally, speed thresholds were worse than the prediction of an all-or-none-type model that assumes participants can only track one target and must guess on the location of a second target. The authors proposed that performance can be worse than the all-or-none serial prediction if participants attempt to switch attention and track multiple targets, rather than giving up on switching and tracking a single target while guessing on additional targets. In summary, the results of this study and our own study can be described by the general serial switching model, but it is unclear whether these results rule out parallel resource theory.

Implications for Parallel Resource Theory

Other prior research in multiple object tracking has attributed set-size effects and speed limits in multiple object tracking to an attentional resource that is shared across targets in parallel (Alvarez & Franconeri, 2007; Chen, Howe, & Holcombe, 2013; Holcombe & Chen, 2012). It has been proposed that fast object speeds exhaust attentional resources, making it more difficult for participants to track targets. Additionally, with increasing set size, there is a decrease in the amount of the attentional resource given to each target. Thus, participants require a slower speed to track larger numbers of targets.

Although prior work has found results that are consistent with a limited attentional resource, discussions of parallel resource theory are often vague about the specific characteristics of the resource and the mechanisms that determine performance in multiple object tracking. To make the predictions of resource theory concrete, we focused on a specific version of parallel resource theory, the fixed-capacity, parallel model. This model can be conceptualized using the sample size model (Bonnel & Miller, 1994; Smith et al., 2018), where a fixed number of samples is shared between targets, and the resolution of each target's stimulus representation decreases with decreasing numbers of samples. Set-size effects that are consistent with the sample size model have been found in prior work using a multiple object tracking task in which participants reported the direction of motion for target objects (Horowitz & Cohen, 2010). However, the dual-task deficit observed in the current study was larger than what is predicted by models that assume continuous, parallel sampling of target position such as the sample size model.

Dynamic Models

An interesting alternative to the static models developed here are dynamic models. These can be developed for either serial or parallel accounts. Here we illustrate the idea with a particularly simple parallel dynamic model. Assume that tracking depends on maintaining position information that can be lost with a given probability for each unit of time. Specifically, assume that maintaining position information depends upon repeated discrete, parallel sampling, described further in the Appendix. Like the fixed-capacity parallel model, a sampled discrete parallel model assumes that samples are shared between objects. However, sampled information is lost over time in a discrete manner, meaning that at a given time, information about targets is either maintained or completely lost. The loss of information compounds error possibilities over time and predicts that dual-task performance decreases faster over time than single task performance. For the 4 second trials used in this experiment, the model can predict a dual-task deficit that is of a similar magnitude as (or even larger than) that predicted by the all-or-none serial model. Critically, it also predicts a zero correlation. Thus, this version of parallel resource theory is consistent with the current results.

Implications for Crowding and Spatial Interference Theory

Although our sparse displays rule out crowding theory, they do not rule out the more general spatial interference theory (Franconeri, Jonathan, & Scimeca, 2010). In spatial interference theory, attentional selection of multiple targets is influenced by spatial interactions specifically between targets (Shim et al., 2008). These interactions can occur at spatial distances larger than those associated with crowding. One way that this idea can be conceptualized is to assume that locations selected in space have a suppressive surround similar to the center-surround receptive fields found in brain areas associated with visual processing. Because object processing is assumed to occur in higher-level visual processing areas, the size of the suppressive surround is thought to span the full visual field in a manner similar to receptive fields for these brain areas. When there is only a single target, it can be selected and tracked. However, when there are multiple targets, there is competition between targets that result in interactions between the selective region for one target and suppressive surround for the other target.

The influence of target-target interactions on multiple object tracking performance can be measured by varying the spatial distance between targets. Shim et al. (2008) conducted a series of experiments to test the influence of target-target spacing on tracking performance. In the first experiment, they used translational disc motion and varied target-target and target-distractor spacing. Target-target spacing ranged from 0.45 to 2.91 degrees of visual angle, and target-distractor spacing ranged from 0.5 to 3 degrees of visual angle. Both target-target and target-distractor spacing influenced performance, with larger spacing being associated with better performance.

In a second experiment, a quadrant design was used to control target-target spacing. Participants tracked either one or two targets. In the two-target condition, the targets appeared either in the same quadrant, or in different quadrants. Performance was better for trials with targets in different quadrants compared to those in the same quadrant, indicating that greater target-to-target distances were associated with better performance. In a third experiment, they used a circular display that was divided into 8 sections, and again, performance was better when the two targets were presented in separate sections versus when they were presented in the same section. Additionally, in trials where targets were presented in separate sections, larger distances between sections was associated with better performance.

The long-range spatial interactions described by spatial interference theory can account for dual-task deficits for targets that are widely spaced. Therefore, such a model cannot be ruled out using sparse displays such as those in the current study. One possible way to investigate such a model is to manipulate the perceptual organization of the stimulus. This can be done using grouping. Grouping occurs when individual objects are made to appear as though they are components of a larger perceptual object. Grouping has been found to influence performance in multiple object tracking (Erlikhman, Keane, Mettler, Horowitz, & Kellman, 2014; Keane, Mettler, Tsoi, & Kellman, 2011; Yantis, 1992). If the selection mechanism described by spatial interference theory is object-based, grouping two targets together would allow both targets to fall within the selective region and neither target would fall within the suppressive surround, thus reducing the amount of competition between them. Therefore, the dual-task deficits should be smaller for targets that are grouped versus those that are not grouped.

Conclusion

The current study investigated two broad hypotheses for set-size effects in multiple object tracking: serial switching and parallel resource theory. We focused on specific versions of each hypothesis: the all-or-none serial model and the fixed-capacity, parallel model. Sparse displays were used to minimize contributions from visual crowding. Results of the current study showed large dual-task deficits for a tracking task in which participants tracked either one or two targets. In addition, the correlation between responses was near zero. These results were not consistent with either the all-or-none serial model or the fixed-capacity parallel model. Instead, these results can be accounted for by other models that fall within in the more general categories of serial switching and parallel resource theory.

Cathleen Acknowledgments. We thank Steven Franconeri, Alex Holcombe, Todd Horowitz, and [redacted] M. Moore for critical comments and helpful suggestions.

Funding. This research was partly supported by grants to G. M. Boynton from the National Institute of Health EY12925.

Declarations

Conflicts of Interest. None.

Ethics Approval. The experiments were approved by the University of Washington Institutional Review Board.

Consent to Participate. All participants gave informed consent.

Consent for Publication. All participants gave consent for publication.

Availability of Data and Materials. The data are available in a repository of the Open Science Framework: osf.io/5efcr/

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*(4), 1–10. https://doi.org/10.1167/7.13.14.

Alvarez, G. A., Horowitz, T. S., Arsenio, H. C., DiMase, J. S., & Wolfe, J. M. (2005). Do Multielement Visual Tracking and Visual Search Draw Continuously on the Same Visual Attention Resources ? *Journal of Experimental Psychology: Human Perception and Performance, 31*(4), 643–667. https://doi.org/10.1037/0096-1523.31.4.643

Bonnel, A., & Prinzmetal, W. (1998). Dividing attention between the color and the shape of objects. *Perception & Psychophysics, 60*(1), 113–124.

Bouma, H. (1970). Interaction effects in parafoveal letter recognition. *Nature, 226*(5241).

Chen, W. Y., Howe, P. D., & Holcombe, A. O. (2013). Resource demands of object tracking and differential allocation of the resource. *Attention, Perception, and Psychophysics, 75*(4), 710–725. https://doi.org/10.3758/s13414-013-0425-1

Cornelissen, F. W., Peters, E. M., & Palmer, J. (2002). The Eyelink Toolbox : Eye tracking with MATLAB and the Psychophysics Toolbox. *Behavioral Research Methods, Instruments, & Computers, 34*(4), 613–617.

Erlikhman, G., Keane, B. P., Mettler, E., Horowitz, T. S., & Kellman, P. J. (2014). Automatic feature-based grouping during multiple object tracking. *Journal of Experimental Psychology: Human Perception and Performance, 39*(6), 1625–1637. https://doi.org/10.1037/a0031750

Franconeri, S. L., Lin, J. Y., Pylyshyn, Z. W., Fisher, B., & Enns, J. T. (2008). Evidence against a speed limit in multiple-object tracking. *Psychonomic Bulletin and Review, 15*(4), 802–

808. https://doi.org/10.3758/PBR.15.4.802

Holcombe, A. O., & Chen, W. (2012). Exhausting attentional tracking resources with a single fast-moving object. *Cognition, 123*(2), 218–228. https://doi.org/10.1016/j.cognition.2011.10.003

Holcombe, A. O., & Chen, W. (2013). Splitting attention reduces temporal resolution from 7 Hz for tracking one object to 3 Hz when tracking three. *Journal of Vision, 13*, 1–19. https://doi.org/10.1167/13.1.12.

Horowitz, T. S., & Cohen, M. A. (2010). Direction information in multiple object tracking. *Attention, Perception, & Psychophysics, 72*(7), 1765–1775. https://doi.org/10.3758/APP

Howard, C. J., & Holcombe, A. O. (2008). Tracking the changing features of multiple objects: Progressively poorer perceptual precision and progressively greater perceptual lag. *Vision Research, 48*(9), 1164–1180. https://doi.org/10.1016/j.visres.2008.01.023

Howe, P. D. L., Cohen, M. A., Pinto, Y., & Horowitz, T. S. (2010). Distinguishing between parallel and serial accounts of multiple object tracking. *Journal of Vision, 10*(8), 1–13. https://doi.org/10.1167/10.8.11.

Keane, B. P., Mettler, E., Tsoi, V., & Kellman, P. J. (2011). Attentional Signatures of Perception : Multiple Object Tracking Reveals the Automaticity of Contour Interpolation. *Journal of Experimental Psychology: Human Perception and Performance, 37*(3), 685–698. https://doi.org/10.1037/a0020674

Miller, J., & Bonnel, A. (1994). Switching or sharing in dual-task line-length discrimination ? *Perception & Psychophysics, 56*(4), 431–446.

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

Shaw, M. L. (1980). Identifying Attentional and Decision-Making Components in Information Processing. *Attention and Performance*, *8*, 277–296.

Shiffrin, R. M., & Gardner, G. T. (1972). Visual processing capacity and attentional control. *Journal of Experimental Psychology*, *93*(1), 72–82. https://doi.org/10.1037/h0032453

Shim, W. M., Alvarez, G. A., & Jiang, Y. V. (2008). Spatial separation between targets constrains maintenance of attention on multiple objects. *Psychonomic Bulletin and Review*, *15*(2), 390–397. https://doi.org/10.3758/PBR.15.2.390

Smith, P. L., Lilburn, S. D., Corbett, E. A., Sewell, D. K., & Kyllingsbæk, S. (2016). The attention-weighted sample-size model of visual short-term memory : Attention capture predicts resource allocation and memory load. *Cognitive Psychology*, *89*, 71–105. https://doi.org/10.1016/j.cogpsych.2016.07.002

Sperling, G., & Melchner, M. J. (1978). The Attention Operating Characteristic : Examples from Visual Search. *Science*, *202*(4365), 315–318.

White, A. L., Palmer, J., & Boynton, G. M. (2018). Evidence of serial processing in visual word recognition. Psychological Science, 29, 1062-1071.

White, A. L., Palmer, J., & Boynton, G. M. (2019). Visual word recognition : Evidence for a serial bottleneck in lexical access.

Yantis, S. (1992). Multielement Visual Tracking : Attention and Perceptual Organization. *Cognitive Psychology*, *24*, 295–340.

Appendix

In this appendix, we describe three models. The first two are static models adopted from the prior literature on dual tasks with brief static displays. The third is a dynamic model adopted from the memory literature to illustrate an alternative approach to multiple object tracking that lasts many seconds. All of the models are defined for the specific task used in this article. Specifically, two sets of stimuli are presented in a circular configuration either above or below fixation. Each set contains n equally spaced stimuli (here n=3) with one target and the rest distractors. In the single task, only one set of stimuli is relevant and has a target (either above or below fixation). In the dual task, both sets of stimuli are relevant and each has a target.

Static models

All-or-none serial model

In serial models, one can attend and process only a single stimulus at a time. For typical serial switching models, one can switch attention from stimulus to stimulus. But there is a special case in which there is no switching. Instead, one attends to only one stimulus throughout the tracking task. In such an *all-or-none serial model*, no information is obtained about the other target or the distractor stimuli. If tested on the unattended target one must guess.

To formalize this model, denote the probability of error for an attended stimulus by q_s. Then for the single task condition S, the probability correct is simply

$$p_S = 1-q_s.$$

For the dual task condition D, label the two tasks by A and B. For this case, we need to add an attention parameter a for the fraction of trials for which task A is attended. This parameter is used to sweep out the AOC curve. For task A, the dual task performance is

$$p_{DA} = a(1-q_s) + (1-a)(1/n).$$

And performance for task B is

$$p_{DB} = (1-a)(1-q_s) + a\,(1/n).$$

For equally attended tasks, these equations simplify to

$$p_D = (1-q_s)/2 + (1/n)/2.$$

In summary, q_s is the free parameter, a is used to sweep out the AOC function, and n is determined by the experiment. To provide a numerical example, if q_s=.2, a=.5, and n=3, then p_S = .8 and p_D = .567.

Fixed-Capacity, Continuous Parallel Model

The parallel model considered here processes the separate targets independently and in parallel. For our continuous version, tracking is based on a continuous noisy estimate of the position of the target. When that estimate of position becomes sufficiently noisy, tracking is lost and a distractor is selected.

For this fixed-capacity model, a constant amount of information is maintained about multiple targets. Thus with two targets, half as much information can be maintained about each target. This can be implemented for a continuous representation by the sample size model (Shaw, 1980). The idea is that there are a fixed number of independent samples of position information and more samples yield a better estimate. With multiple stimuli, the samples must be distributed across the stimuli.

To be specific, assume the target's position in a single task is represented by the one-dimensional random variable U. For simplicity, let it be normally distributed in units of position around the circle of possible stimuli with mean zero and standard deviation s for the single task. Strictly speaking, this application needs a circular distribution but the predicted difference in performance is trivial. Further assume the tracking becomes inaccurate when the perceived position is greater than d or less than minus d. This value is assumed to be half the distance between the target and the distractors on either side. With these assumptions, the probability correct for the single task S is

$$p_S = Prob(|U| < d).$$

Denoting the cumulative distribution of the random variable U by F with standard deviation s, this can be rewritten as

$$p_S = F(d) - F(-d).$$

For the dual task D, the standard deviation of the representation of position depends on the fraction of trials each task is attended. Let a be the fraction of trials when task A is attended. Using the sample size model, the number of samples are proportional to the attention parameter a. By this model, the standard deviation of the random variable for position is $s/\sqrt{a}$. With these assumptions, the probability correct in the dual task condition for task A is

$$p_{DA} = F(d\sqrt{a}) - F(-d\sqrt{a})$$

and for task B is

$$p_{DB} = F(d\sqrt{1-a}) - F(-d\sqrt{1-a}).$$

Assuming equal attention to the two task, this becomes

$$p_D = F(d\sqrt{1/2}) - F(-d\sqrt{1/2}).$$

In summary, s is the free parameter, a is used to sweep out the AOC function, d and n are determined by the experiment. To provide a numerical example, if s=46.8°, a=.5, d = 60°and n=3, then p_S = .800 and p_D = .635.

Discrete Parallel Model

Next we introduce a parallel dynamic model. One could construct a similar serial model or a continuous model but we do not pursue them here. This model processes the separate targets independently and in parallel. Unlike the continuous model, tracking is either maintained or lost completely on each time step. There is no partial information. The dynamic part of the model is to update tracking over time expressed in time steps $i = 0, \ldots, Infinity$. For each time step, there is a probability of losing the target. For the single task condition S, this probability is given by q_s Over each time step, i, the probability of successfully tracking the target is given by $V_S(i)$:

$$V_S(i+1) = (1-q_s)\ V_S(i)$$

where $V_S(0) = 1$. Combining this probability with guessing when a target is not tracked successfully yields a probability correct for the single task at time i:

$$p_S(i) = V_S(i) + (1-V_S(i))/n.$$

For the dual task condition D, we implement fixed capacity using a variation on the sample size model but now applying it to a discrete representation (for a related model see the appendix in Popovkina, et al, 2021). The idea is that there is a sampling process that can be applied to either single stimulus or distributed across multiple stimuli. For each sample, there is a chance of losing track denoted q_1. For multiple samples of the same stimulus, one loses track only if it is lost for all of the samples. So for r of m independent samples, the probability of losing track $q(r/m)$ is

$$q(r/m) = q_1^r.$$

Defining the fraction of samples directed to task A by $a=r/m$, we can rewrite this equation as

$$q(a) = q_1^{am}. \qquad (1)$$

Using this equation, for a single task condition with all of the samples allocated to one target we have

$$q_s = q_1^m.$$

Expressed in terms of q_1,

$q_1 = q_s^{1/m}$

This can be substituted into Equation 1 to yield

$q(a) = (q_s^{1/m})^{am}$,

which simplifies to

$q(a) = q_s^a$.

This formulation eliminates the need to specify the number of samples m or the probability of an error for single sample q_1. Only q_s is needed and it becomes the free parameter in this model.

For the dual task condition with task A, the probability of maintaining tracking is

$V_{DA}(i+1) = (1-q_s^a)V_{DA}(i)$

and for task B is

$V_{DB}(i+1) = (1-q_s^{(1-a)})V_{DB}(i)$.

Similar to before, let $V_{DA}(0) = V_{DB}(0) = 1$. Adding in guessing, the probability of a correct response is

$p_{DA}(i) = V_{DA}(i) + (1-V_{DA}(i))/n$, and

$p_{DB}(i) = V_{DB}(i) + (1-V_{DB}(i))/n$.

Insert 35A

For equal attention (a=.5), ~~these equations become~~ we have

$\cancel{V_D(i+1) = (1-q_s^{1/2})V_D(i)}$, ~~and~~

$p_D(i) = \cancel{V_D(i) + (1-V_D(i))/n}$ $[(n-1)(1-q_s^{1/2})^i+1]/n$

For continuous time, these equations become exponential decay functions.

For this model, the AOC changes shape with time. It starts with a small dual-task deficit and a convex AOC, passes through a linear AOC, and with enough time, performance in the dual task approaches chance faster than the single task resulting in a concave AOC. The much worse performance for the dual task compared to the single task is due to compounding the unequal error probabilities over time.

35A

With some algebra, these iterative equations can be simplified to:

$$P_S(i) = \left[(n-1)(1-q_s)^i + 1\right]/n$$

$$P_{DA}(i) = \left[(n-1)(1-q_s^a)^i + 1\right]/n$$

$$P_{DB}(i) = \left[(n-1)(1-q_s'^a)^i + 1\right]/n.$$

In summary, q_s is the free parameter, a is used to sweep out the AOC function, i is used to represent time, and n is determined by the experiment. To provide a numerical example, if q_s =.163, a=.5, and n=3, then at time 2, p_S = .800 and p_D = .570 and at time 6, p_S = .563 and p_D = .363.