Investigating attentional theories of multiple object tracking using sparse displays

Miranda Johnson

John Palmer

Geoffrey M. Boynton

Dept of Psychology, U of W

Key words: attention, divided attention, multiple object tracking, crowding

Last Revised 03/31/2025

Intended for AP&P?

JP's Comments 4/2/25

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. Switching theory proposes that attention is switched from one object to the next during tracking. We focus on the more specific all-or-none serial model, which assumes that participants can only track one object, and performance is at chance for tracking a second object. Resource theory proposes an attentional resource that 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, where the a representation of the stimulus is formed through sampling, and more targets means each target is sampled less. The current study distinguishes these two models using a dual-task design 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 all-or-none serial model and fixed-capacity parallel model predict dual-task deficits of differing magnitudes. Additionally, the all-or-none serial model predicts a negative correlation between dual-task responses, while the fixed-capacity parallel model predicts no correlation. Results show 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 is often 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 set 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). This article pursues the interpretation of this set-size effect. More specifically, we test two attentional hypotheses that describe 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 objects presented close in space interfere with one another. Assuming a 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 occur 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 additional stimuli placed within that window result in [illegible] 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 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 task involving discrimination to show its effects (Ester, Clee, & Awh, 2013; Levi & Carney, 2009; Palomares, Pelli, & Majaj, 2001). Visual 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 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. Crowding theory proposes that for uncrowded displays, there is no attentional 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.

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 resource theory. Under the serial switching hypothesis, participants 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 attended, the location of that target is recorded before attentional selection switches to a different object. When a participant switches their 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. 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 selection returns to a given object and updates its location (Holcombe & Chen, 2013). Thus, increasing numbers of targets is associated with worse performance [redacted].

An important variable that influences performance in MOT is speed. As speed increases, performance declines. For the serial switching model, the effect of speed can be understood by the related idea of spatial 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 N 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 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 not 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 correlation between responses in the dual-task condition. This extreme version of serial switching makes distinct predictions for divided attention effects across a variety of tasks. 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, 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).

Resource Theory

The second general hypotheses for why there are set-size effects in multiple object tracking is resource theory. 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 resource theory more concrete, we focus on a specific version of resource theory, 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 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 an obvious version of resource theory to test.

The Current Study

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 us to compare predictions for our two attentional models and predictions of crowding theory, which assumes that in sparse displays with no visual crowding, there are no attentional effects. The AOC method is commonly used in dual tasks 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 uses a dual task where participants tracked either one or two targets that appeared above or below fixation. To make distinct the predictions for crowding theory, discs were widely spaced such that [redacted] crowding was unlikely (Bouma, 1970). To distinguish the predictions for switching, fast disc motion was used so that attentional switching would be unlikely, which leaves one with the all-or-none serial model (Holcombe & Chen, 2013). We also focus on the fixed-capacity parallel model, a resource theory model 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 correlation between accuracy for responses in the dual-task condition. The all-or-none serial model predicts a large dual-task deficit and a negative correlation between accuracy for the top and bottom responses in the dual-task condition. The fixed-capacity parallel model predicts a 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 our sparse displays, crowding theory predicts little or no dual-task deficit, and a zero correlation for accuracy in the dual-task condition.

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. ~~For the dual-task condition~~, we need to discriminate deficits of 23% and 11%. A power analysis with 80% power suggested a minimum of 4 participants. For the response correlation, we need to discriminate correlations of -.10 and 0. A power analysis using a one-tailed test with 80% power suggests a minimum of 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/m2. 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 only 3 ± 2% of trials, and broken fixations occurred on 3 ± 1% 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

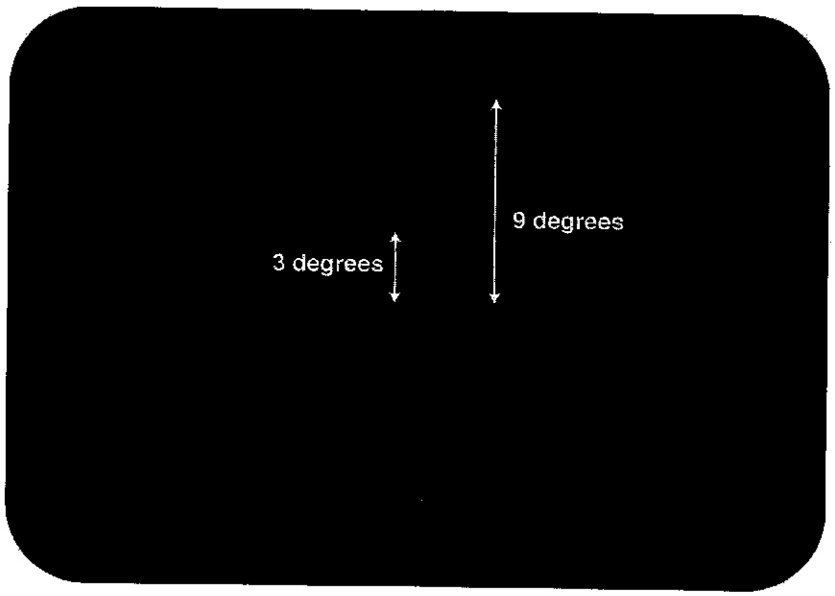

Figure 1. Stimulus object spacing. Discs were positioned along invisible circular trajectories, which are represented here by dashed circles, but were not visible in the experiment. Discs were evenly spaced along the circular trajectory. The circular trajectories were centered at six degrees in visual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

estimated by Bouma's law (Bouma, 1970).

The difficulty of the task was 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 (120 Hz).

Procedure

The single- and dual-task conditions are shown schematically in Figure 2. 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. The cue was 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 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 when those reversals could occur 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. During the 4 seconds of disc motion, participants were instructed not to blink, and trials where participants blinked were not included in the final analysis.

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. 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. (illustrated at the bottom of Figure 2)

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 vertical stem of the fixation cross was blue. After the 4 seconds ~~period~~ of disc motion, participants were prompted by a response cue to either select the bottom target first and then the

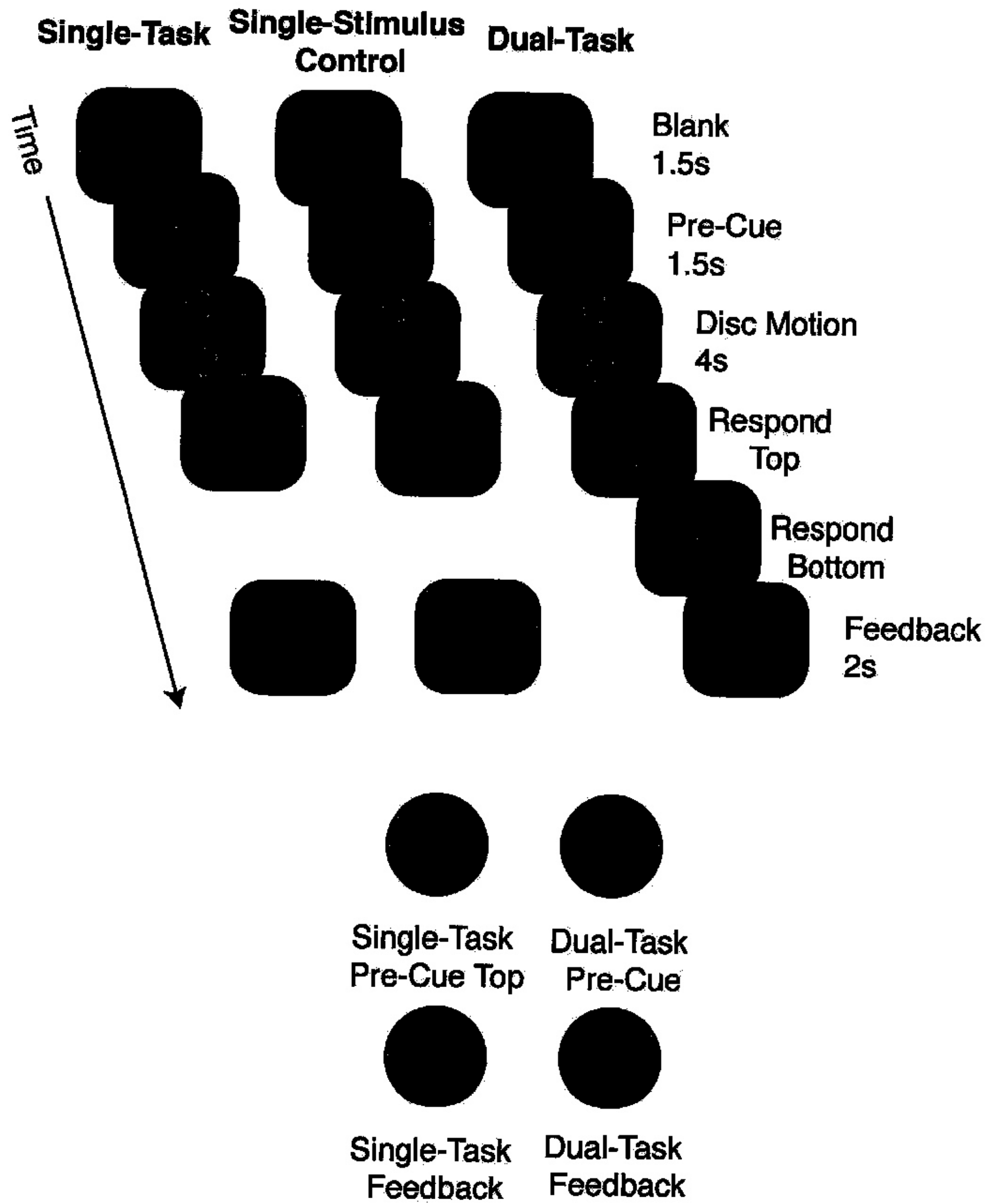

Figure 2. Trial sequence for the single-task, single-stimulus control, and dual-taks conditions. In this example of the single-task condition, the top half of the display is being cued and a target is shown above fixation in red. In the dual-task condition, both sides are cued, and two targets are shown 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.

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 learned to use the cues and perform the task. Participants then completed 20 experimental sessions, which took between fifteen to twenty hours, 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-task control trials, and 38 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 72 ± 2%, and performance in the dual-task condition was 52 ± 2% (chance was 33.3%). The difference is a large dual-task deficit of 20 ± 2%. 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 ± 1%. This difference was

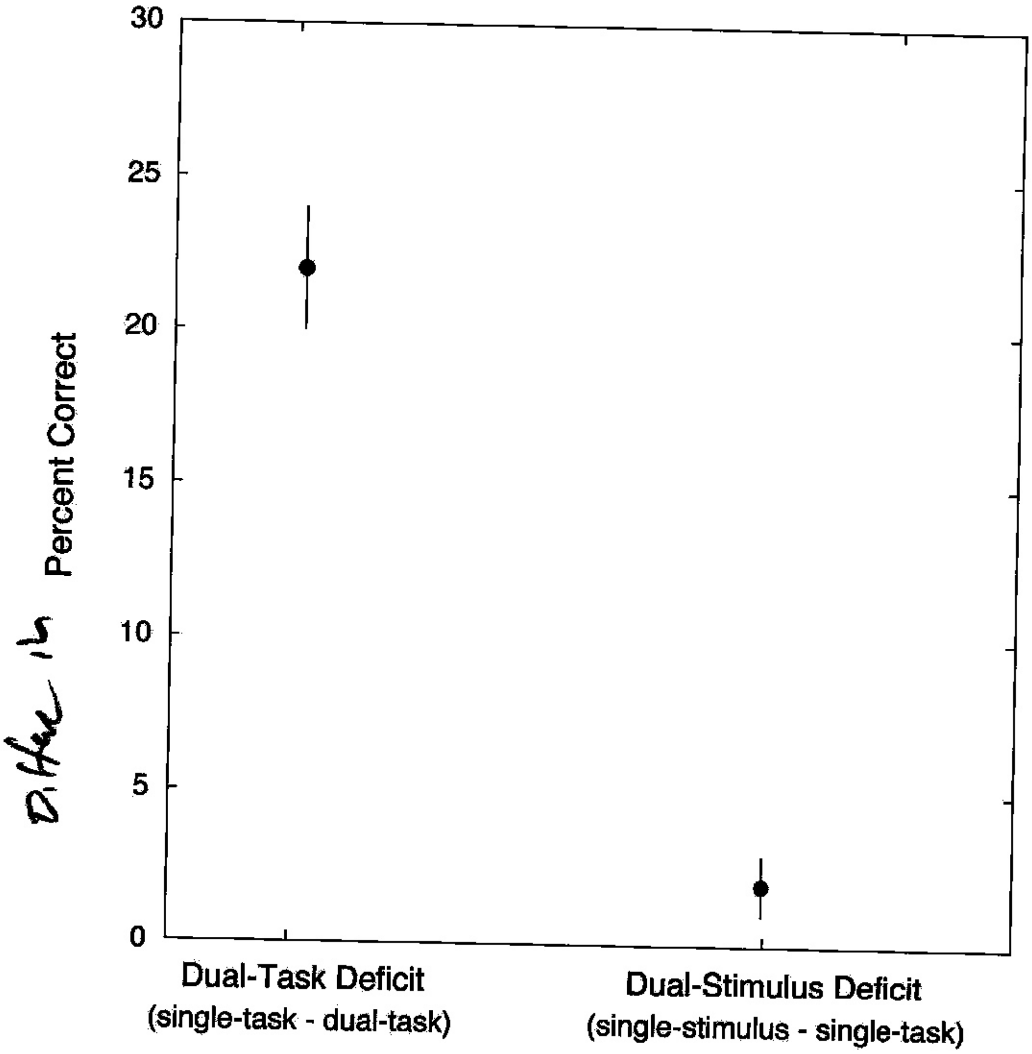

Figure 3. Comparison of the dual-task deficit and the dual-stimulus deficit. The dual-task deficit is the difference in percent correct for the single-task and dual-task conditions. The dual-stimulus deficit is the difference in percent correct for the single-stimulus control condition and the single-task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

(95% CI —, —)

marginally reliable t(5) = 2.47, p = .06. While this difference suggests a small 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 residual crowding effect measured by the dual-stimulus deficits cannot account for the observed dual-task deficit.

Figure 4 shows the results compared to model predictions using the attentional operating characteristic (AOC; Sperling & Melchner, 1978). 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: 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: 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 limited 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.

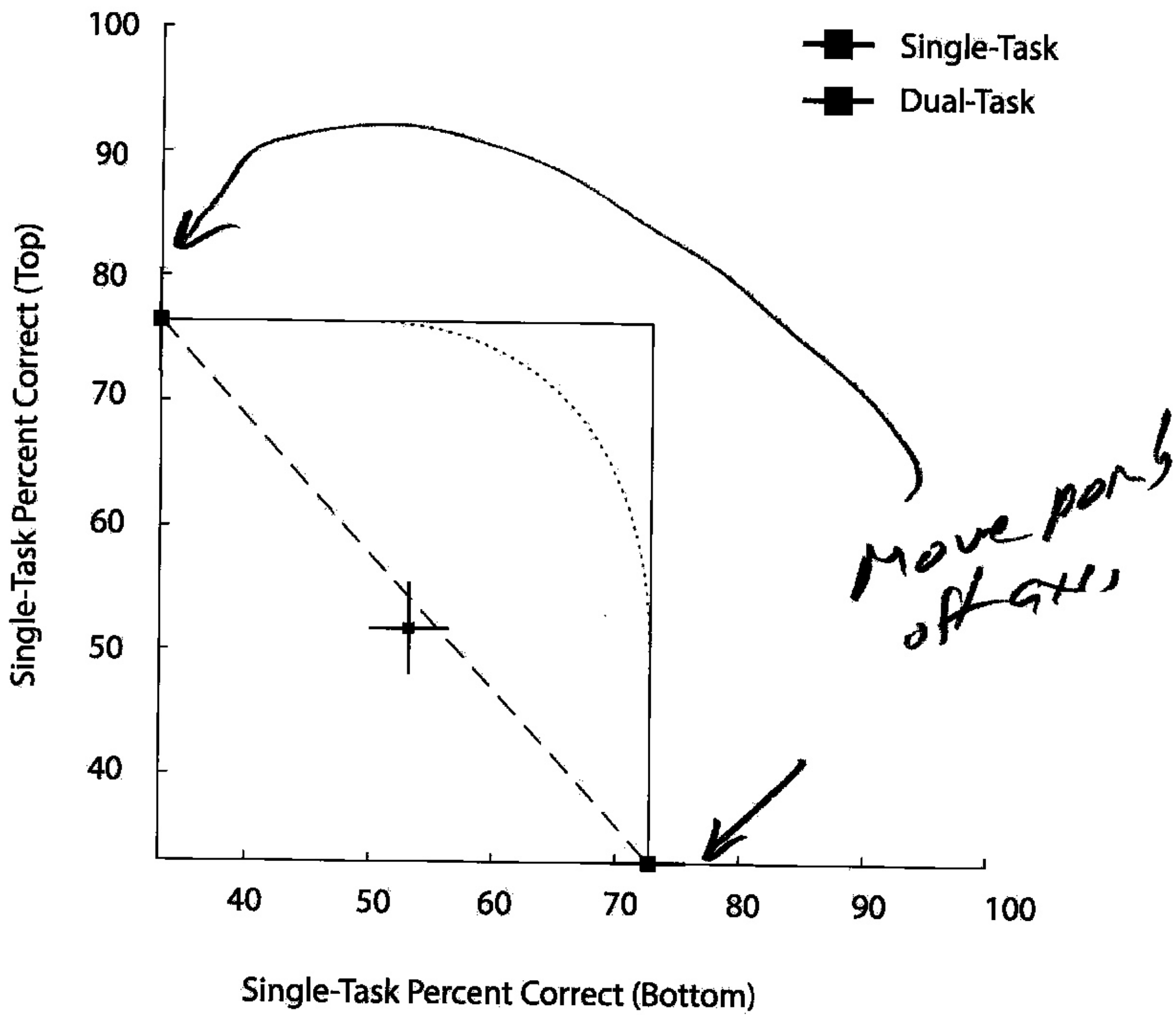

Figure 4. Percent correct for the single- (red) and dual-task (blue) conditions. Error bars represent the standard error of the mean. Model predictions are also shown. The solid lines represent the prediction for crowding theory. The dotted curve represents the prediction for the fixed-capacity parallel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

Percent correct for the single-task condition is shown for the top (y-axis) and bottom (x-axis) responses. Single-task performance was 76 ± 4% for responses on the top, and 73 ± 3% for responses on the bottom. The dual-task deficit is plotted as a point, where the x-value represents dual-task performance for the top target, and the y-value represents dual-task performance for the bottom target. Performance for the top in the dual-task condition was 52 ± 4%, and performance for the bottom was 54 ± 3%. Plotting the dual-task deficit on the AOC reveals that the magnitude of the deficit is consistent with the all-or-none serial prediction, and much larger than predicted by the fixed-capacity parallel model.

Correlation between responses in the dual-task condition

Figure 5 shows the observed 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 dual-task performance is 67% correct for this 3-choice task. This model assumes that participants can only track 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.

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

Here should be

parallel, no correlation ~~is predicted.~~

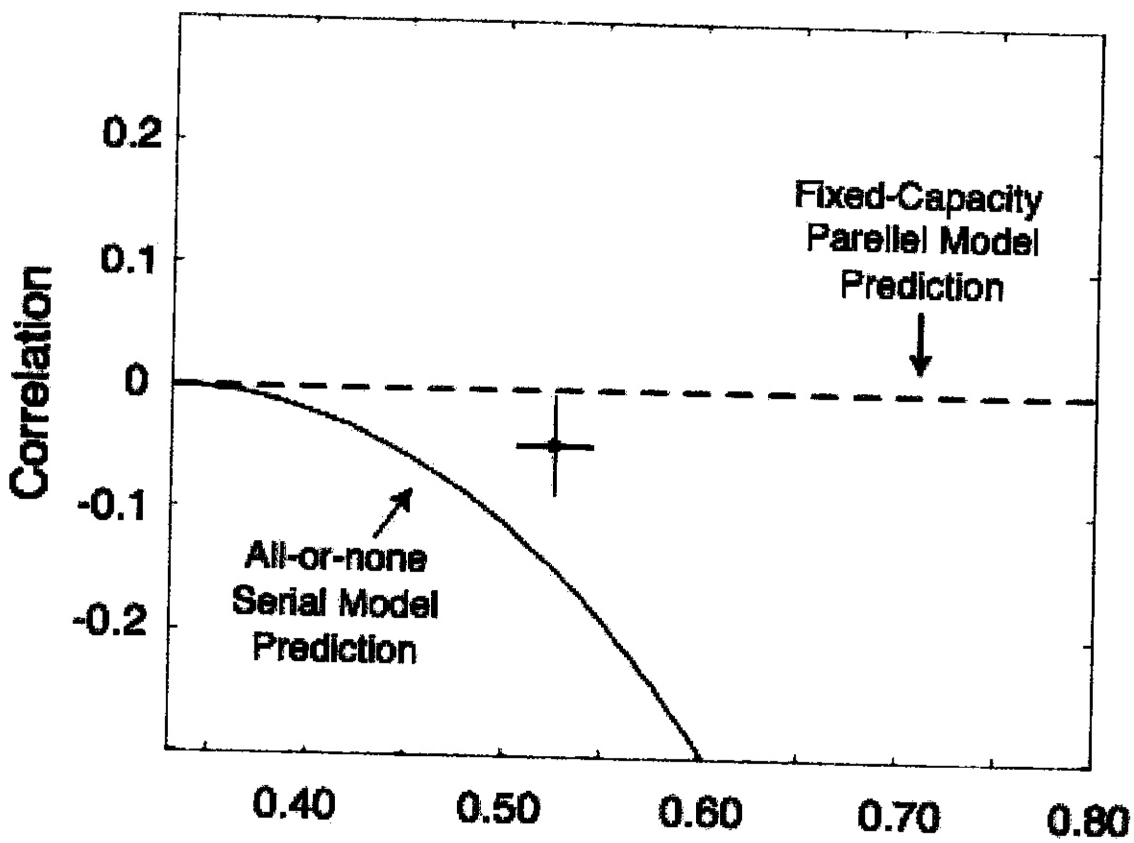

Figure 5. Predicted and observed correlations plotted as a function of dual-task percent correct. The dashed line represents predicted zero correlations under crowding theory and the fixed-capacity parallel model. The curve represents predicted negative correlations under the all-or-none serial model.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

The correlation between dual-task accuracy for the top and bottom was r = -.05 ± .04 (Figure 5). This result was quite different from 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 for our 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 -1 ± 1, with second responses being slightly better, however this difference was not reliable t(5) = -1.94, p = .11. Thus, there is no evidence of memory or response interference for the second response.

General Discussion

Results show a large dual-task deficit, the magnitude of which is consistent with the all-or-none serial model. The observed dual-task deficit allows us to reject the fixed-capacity parallel model. Additionally, this dual-task deficit with our sparse displays provides evidence against crowding theory, which predicts no set-size effects in multiple object tracking when displays are uncrowded. The observed dual-task deficit measured in this study is 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, there was no correlation between accuracy for responses on the top and bottom in the dual-task condition, which is not consistent with the all-or-none serial model. We next put our results in the context of the larger literature, and consider alternative models that can account for this combination of results.

Implications for 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 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., 2019), and these results provide strong evidence consistent with serial processing. Results of the current study showed no negative correlation between dual-task responses, which is not consistent with the serial [redacted] hypothesis.

The all-or-none serial [redacted] hypothesis assumes that 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 & Chen (2013), where performance was worse than a model that assumes participants can only track one object, and must guess on a second object.

Results consistent with the all-or-none serial model have been found in prior work using multiple object tracking. Holcombe & 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. Results showed that speed limits were much lower for tracking two targets versus tracking one. The speed limit for the track-two condition 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 N 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 result in a higher temporal frequency. Additionally, faster disc speeds lead to a higher temporal frequency. Under the 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 target. 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 more general serial switching model. Additionally, speed thresholds were worse than what is predicted by 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. The results of this study can be described by the switching model, but it is unclear whether these results rule out a resource theory-like model.

Implications for Resource Theory

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 is proposed that fast object speeds exhaust attentional resources, making it more difficult for participants to track targets. Additionally, with increasing set size, the amount of the attentional resource given to each target decreases, meaning that participants require slower speeds to track larger numbers of targets.

Results consistent with resource theory have been found in experiments using the simultaneous-sequential manipulation (Shiffrin & Gardner, 1972). Howe et al. (2010) used the simultaneous-sequential manipulation in a multiple object tracking task where target objects either move simultaneously during a single time interval, or sequentially over multiple time intervals. A model assuming unlimited attentional resources would predict no advantage for sequentially-presented stimuli over simultaneously-presented stimuli. However, a model assuming limited, parallel attentional resources would predict a sequential advantage because each time interval contains fewer moving targets to be tracked, and thus more resources are given to each target per time step. In one experiment included in this study, participants were more accurate in tracking targets in the sequential condition than the simultaneous condition, which is consistent with resource theory.

Although prior work has found results that are consistent with a limited attentional resource, discussions of resource theory are often vague about the specific characteristics of the resource and the mechanisms through which performance in multiple object tracking is inferred. To make the predictions of resource theory concrete, we focused on a specific version of a 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 drawn from the stimulus and shared between targets, and the resolution of each target's stimulus representation decreases with increasing numbers of targets. Set-size effects that are consistent with the sample size model have been found in prior work using a multiple object tracking task where participants reported the direction of motion for target objects (Horowitz & Cohen, 2010). However, the dual-task deficit observed in the current study is larger than what is predicted by models that assume continuous, parallel sampling of target position. Our results can instead be predicted by a model that assumes repeated discrete, parallel sampling.

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 information about targets is either maintained at a given time, or completely lost. The loss of information compounds error possibilities over time and predicts that dual-task performance decreases over time. 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 resource theory is consistent with the current results.

Implications for Crowding and Spatial Interference Theory

Although our sparse displays allow us to rule out crowding theory, we cannot rule out the more general spatial interference theory (Franconeri, Jonathan, and Scimeca, 2010). Spatial interference theory proposes that attentional selection of multiple targets is influenced by spatial interactions 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 occurs 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 is able to be selected and tracked. However, when there are multiple targets, there is competition between targets that result in interactions between selective regions and suppressive surrounds for each target.

The influence of target-target interactions on multiple object tracking performance can be measured by varying the spatial distance between targets and testing whether performance changes. Shim et al. (2008) conducted a series of experiments to test the influence of target-target spacing on tracking performance. In the first experiment, both target-target and target-distractor spacing were varied, and translational disc motion was used. 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 (Shim et al., 2008). 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 where targets were presented in different quadrants compared to those where they were presented in the same quadrant, indicating that greater target-to-target distances were associated with better performance. A third experiment 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 predict dual-task deficits for targets that are widely spaced, therefore such a model cannot be ruled out using the sparse displays such as those in the current study. One way to rule out such a model is to manipulate the perceptual organization of the stimulus 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. ~~It would be therefore be predicted that~~ Thus, the dual-task deficits would be smaller for targets that are grouped versus those that are not grouped.

Conclusion

The current study ~~tested~~ investigated two broad hypotheses for set-size effects in multiple object tracking: serial switching and resource theory. We focused on specific versions of each hypothesis: the all-or-none serial model and the fixed-capacity parallel model. Sparse displays

In addition, the correlation between regions was near zero.

were used to control contributions from visual crowding. Results of the current study show large dual-task deficits for a tracking task where participants track either one or two targets. ~~Our~~ These results are not consistent with either the all-or-none serial model or the fixed-capacity parallel model. Instead ~~However~~, these results can be accounted for by other models that fall within in the more general categories of serial switching and resource theory.

Add author notes and declarations (see journal info)

In particular, thank Horowitz, Holcombe and Franconeri, all of whom made suggestions at your poster.