Extrapolation occurs in multiple object tracking when eye movements are controlled

Tina Luu[1] · Piers D. L. Howe[1]

Abstract There is much debate regarding the types of information observers use to track moving objects. Howe and Holcombe (Journal of Vision *12*(13): 1-10, 2012) recently reported evidence that observers employ extrapolation while tracking. However, their study is potentially confounded because it did not control for eye movements. As eye movements can aid extrapolation, it is unclear whether extrapolation can still occur in multiple object tracking (MOT) when eye movements are eliminated. In the current study, we addressed this question using an eye tracker to ensure that fixation was always maintained on a central fixation point while observers performed a tracking task. In the predictable condition, objects always travelled along linear paths. In the unpredictable condition, objects randomly changed direction every 300–600 ms. If observers employ extrapolation, we would expect performance to be greater in the former condition than in the latter condition. Our results showed that observers did indeed perform better in the predictable condition than in the unpredictable condition, at least when tracking just two objects (Experiments 1, 3, and 4). Extrapolation occurred less when tracking loads increased or when the objects moved more slowly (Experiment 2).

✉ Piers D. L. Howe
pdhowe@unimelb.edu.au

1   School of Psychological Sciences, The University of Melbourne, Parkville, VIC 3010, Australia

Introduction

Our ability to simultaneously track multiple moving objects is critical as it allows us to successfully navigate the dynamic world in which we live. Without this ability, everyday tasks such as crossing the road, driving, or engaging in team sports would not be possible. For example, when crossing a busy street, pedestrians might need to keep track of the positions of oncoming vehicles, cyclists, and/or other pedestrians in order to avoid accidents and injuries. Tracking plays a fundamental role in processing and interpreting dynamic environments.

Tracking has been extensively studied in the laboratory using the multiple object tracking (MOT) paradigm (Pylyshyn & Storm, 1988). In a typical MOT trial (see Fig. 1), observers are presented with a set of identical disks and a subset of these disks is momentarily highlighted to denote that they are the targets to be tracked. The disks then revert to their original color and once again become indistinguishable from the other disks in the display (i.e., the distractors). All disks subsequently move randomly about the display and during this time observers are asked to track the targets. Once the disks come to a halt, observers must identify whether a probed disk is a target or a distractor.

Importantly, the MOT task taps into various properties of real-world visual cognition. Much like the situations that we encounter in our everyday life, be it driving or playing team sports, MOT is an inherently active task that requires the observer to continuously attend to multiple objects over time (Scholl, 2009; Wolfe, Place, & Horowitz, 2007). Because MOT and real-world dynamic tracking both demand sustained attention to multiple objects, it is hoped that researchers will be able to gain a better understanding of how observers track objects in the real world through experiments conducted on MOT in controlled laboratory settings (Cavanagh & Alvarez, 2005).
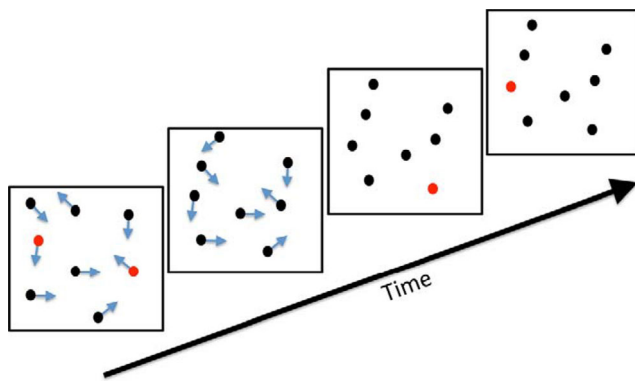
Fig. 1 The structure of the MOT trial used in the present study. The trial begins with either two or four disks turning red to denote that they are the targets. They subsequently revert to black and are free to randomly move about the display. When the disks come to a halt, two disks are highlighted (one after the other). The observer is required to indicate whether each highlighted disk is/is not a target

At present, there is considerable debate regarding how targets are tracked. Proponents of the "no extrapolation hypothesis" argue that observers rely only on location information when tracking the targets (Franconeri, Pylyshyn, & Scholl, 2012; Keane & Pylyshyn, 2006; Vul, Frank, Tenenbaum, & Alvarez, 2009). When targets move from one place to the next, the observer compares the targets' current locations to their last remembered locations. The observer assumes that whichever object is closest to a given target's last remembered location is that target. Conversely, advocates of the "extrapolation hypothesis" claim that observers track targets by using both location information and motion information to extrapolate the future locations of the targets. Targets are identified based on where they are expected to be, not just on where they have been in the past.

To date, studies examining whether motion information is used to track multiple objects have yielded mixed results (Howard, Masom, & Holcombe, 2011; Iordanescu, Grabowecky, & Suzuki, 2009). Keane and Pylyshyn (2006) addressed this question using a "target recovery" paradigm. Unlike the conventional MOT paradigm, a blank screen was briefly introduced at the end of the trial and all the disks disappeared during that period. When the screen was removed, all the disks reappeared and the observers were required to identify the targets. The researchers manipulated the reappearance positions of the disks so that they could either reappear where they had disappeared (i.e., no-move condition) or at a location predicted by their previous movement (i.e., move condition). Tracking accuracy was greater in the no-move condition than in the move condition. Based on this finding, the researchers concluded that only current location information is used during tracking.

Fencsik, Klieger, and Horowitz (2007) argued that while the results of Keane and Pylyshyn (2006) show that observers prefer to use location information over motion information

during tracking, this does not prove that extrapolation cannot be used. It might be that while it is more efficient for observers to utilize location information rather than extrapolation to re-acquire the targets, observers are still able to extrapolate when required to do so (Fencsik et al., 2007). Fencsik et al. (2007) addressed this concern using a slightly modified target recovery paradigm that encouraged observers to employ extrapolation during the blank period. In one condition, the disks continued to move during the blank interval, forcing the observers to extrapolate to anticipate where the targets would reappear. In the other condition, the disks were stationary before the blank interval, thereby making extrapolation impossible. Tracking performance was better in the extrapolation condition than in the static condition for a tracking load of two, but not four, targets. This suggests that observers can use extrapolation to facilitate tracking, but only when tracking two targets. However, the generalizability of this finding is limited because positional information was not available during the blank period in either condition. It might be that in other circumstances whereby positional information is continuously available, observers would not employ extrapolation (Horowitz, Birnkrant, Fencsik, Tran, & Wolfe, 2006).

Iordanescu et al. (2009) also investigated whether extrapolation is employed during tracking by examining how targets are recovered after they disappear (see also Howard et al., 2011). The task used in their study differed from the target recovery paradigm used by Keane and Pylyshyn (2006) in that the disks did not reappear after disappearing at the end of the trial. Instead, after all the disks had disappeared, the observers were asked to click on the location of a particular target (e.g., the red one). Having computed the vector between the target's disappearance location and the mouse-click location, the researchers found that observers tended to select locations that matched the direction of the target's trajectory. In other words, they selected locations slightly ahead of where the target disappeared. Furthermore, there was a positive correlation between the degree of displacement and the speed of targets, such that faster target speeds produced larger forward displacements, and vice versa. As such, this study provides evidence that extrapolation is used in tracking.

More recently, an investigation by Franconeri et al. (2012) reported that motion information is not used to recover targets. Instead of having all the objects disappear from the display, individual objects passed behind a vertical occluder whilst they were being tracked. Tracking accuracy was greater when the targets reappeared closest to where they disappeared rather than when they reappeared at the expected location on the other side of the occluder predicted by their motion prior to being occluded. While this finding indicates that observers used the last known positions of the targets more than their extrapolated positions, it does not prove that extrapolation cannot be employed during tracking. Moreover, since four targets were always tracked in that study, it is possible that

reducing the tracking load would allow observers to more readily utilize extrapolation (Fencsik et al., 2007; Iordanescu et al., 2009).

Although a number of the findings from the aforementioned studies provide evidence for extrapolation in object tracking, they do not address whether extrapolation is used to track objects that are continuously visible. Extrapolation may have been used for recovery and not for tracking per se (St. Clair, Huff, & Seiffert, 2010). Evidence of extrapolation following reappearance does not address whether observers actually do extrapolate the future locations of targets during the moment-to-moment tracking of visible objects. It could be that observers only extrapolate when forced to do so because the objects are temporarily not visible.

St. Clair et al. (2010) addressed this concern by using a MOT paradigm in which the targets were always visible. Observers were asked to track a number of disks, each of which contained a texture that could move independently of its motion. Results showed that tracking accuracy declined when the embedded texture moved in the opposite direction to the disk on which it was attached, suggesting that motion information is used to track the disks. A limitation with this study was that it did not control for object visibility. The disks became less visible when the embedded texture moved in the opposite direction to the motion of the disk because this conflicting motion degraded the disks' borders. This reduction in object visibility may have in turn diminished the quality of available positional information and led to the impairment in tracking accuracy.

This visibility confound was avoided by Vul et al. (2009) who in their study presented stimuli that were clearly visible. The ideal observer model proposed by these researchers posits that motion information can be used to predict the future locations of objects, though the extent to which this occurs is determined by an internal observer-specific parameter. By fitting the model to the data obtained from their observers, Vul et al. (2009) could determine the extent to which the observers utilized extrapolation during tracking. Their results indicated that observers do not use extrapolation. However, because the speeds of the disks were constantly changing in their experiment, this may have made it difficult for observers to extrapolate and so could be a potential confound. Keeping the speed of the disks constant would make them more predictable to observers, which would in turn increase the likelihood that observers would use this information during tracking.

Despite a number of studies that have addressed the question, it is still unclear whether observers use extrapolation when tracking continuously visible objects and, if so, under what conditions they occur. Howe and Holcombe (2012) recently conducted an experiment that addressed this question while controlling for the various confounds identified in previous studies. Their study used a MOT paradigm in which the

objects were continuously visible. To address the confounds present in the studies of St. Clair et al. (2010) and Vul et al. (2009), the researchers ensured that the visibility of the objects was the same in all conditions and that the speed of the disks was held constant. Two variables were manipulated: the number of targets to be tracked (two vs. four) and the predictability of object motion (predictable vs. unpredictable).

In the predictable condition, objects always travelled along a linear path, changing direction only when they reached the boundaries of the display. In the unpredictable condition, the disks randomly changed direction every 300–600 ms. When objects move in a predictable manner, the effectiveness of any extrapolation process is maximized (Howard et al., 2011). Conversely, when the same objects move in an unpredictable fashion, extrapolation becomes less helpful. Better performance in the predictable condition would therefore be indicative of observers extrapolating when tracking objects. Across all their experiments, results showed that observers were more accurate in the predictable condition than in the unpredictable condition when tracking two but not four targets. This is consistent with the finding of Fencsik et al. (2007), indicating that observers are able to extrapolate when tracking two targets but are less able to do so when tracking four targets. However, Fencsik et al. (2007) and Howe and Holcombe (2012) did not control for eye movements, and as such their results are potentially confounded because eye movements aid extrapolation (Zhong, Ma, Wilson, Liu, & Flombaum, 2014). In particular, it is unclear whether they would have obtained the same results had they prevented their observers from making eye movements. It is possible that extrapolation only occurs in MOT when observers are free to move their eyes (Zhong et al., 2014).

Although tracking can occur even when observers are required to maintain fixation on a fixation cross throughout the tracking process (Howe, Pinto, & Horowitz, 2010; Intriligator & Cavanagh, 2001), it is becoming increasingly apparent that eye movements can play an important role in MOT under free viewing conditions. When tracking three targets, observers have a tendency to look at the center of the triangle formed by the targets even when none of the targets are located at this position (Fehd & Seiffert, 2008). This tendency is more pronounced when tracking two targets than four targets (Zelinsky & Neider, 2008). This does not occur simply because observers are trying to minimize eye movements but rather this strategy directly benefits tracking performance (Fehd & Seiffert, 2010). When observers are asked to fixate only on the individual targets rather than occasionally also fixating on the center point of the group of targets, their tracking accuracy decreases (Fehd & Seiffert, 2010). This is not to say that observers never need to fixate on the individual targets – they do this periodically, at least in part, to "rescue" targets that are in immediate danger of becoming lost (Zelinsky & Todor, 2010). So while it is clear that the strategy of fixating on the

center point of a group of targets plays an important role in tracking, especially in situations where tracking is particularly difficult such as those containing abrupt viewpoint changes (Huff, Papenmeier, Jahn, & Hesse, 2010), it is not the only factor affecting eye movements (Lukavsky, 2013). In particular, it has recently been suggested that eye movements also play a role in extrapolation (Zhong et al., 2014).

For extrapolation to be effective, the observers must first accurately estimate the velocities of the targets. If there is just a single target, the observers can potentially do this by fixating on the target and following it with a smooth eye pursuit (Zhong et al., 2014). By knowing how their eyes are moving, the observers can then estimate the movement of the target. For single targets, observers can indeed accurately extrapolate to where they expect the target to be (Diaz, Cooper, Rothkopf, & Hayhoe, 2013; Hayhoe, McKinney, Chajka, & Pelz, 2012; Land & McLeod, 2000). As the number of targets to be tracked increases, this strategy becomes increasing less effective. This could explain why observers' knowledge of the direction of motion of targets in MOT decreases as the number of targets to be tracked increases (Horowitz & Cohen, 2010; Shooner, Tripathy, Bedell, & Ogmen, 2010). Zhong et al. (2014) have postulated that the only way that observers can extrapolate in MOT is by using eye movements, and this explains why Howe and Holcombe (2012), who in their study enabled observers to freely move their eyes, found evidence for extrapolation when observers tracked two but not four targets.

The purpose of the current investigation was to test the claim that extrapolation in MOT can be achieved only by eye movements. This was done by replicating some of the key experiments of the Howe and Holcombe (2012) study while controlling for eye movements by using an eye tracker to ensure observers maintained fixation on a central fixation cross throughout the tracking task. This also ensured that any eccentricity effects on tracking accuracy (Intriligator & Cavanagh, 2001) would be the same in all conditions and would not vary either with the number of targets or with whether the targets move in a predictable or unpredictable manner.

Experiment 1

The first experiment attempted to replicate Experiment 2 of the Howe and Holcombe (2012) study with the addition of a central fixation cross to control for eye movements. If observers are able to utilize extrapolation, tracking performance should be greater in the predictable condition than in the unpredictable condition because the former condition would render motion information more useful than the latter condition.

Method

Participants

A power analysis run on Experiment 1 of Howe and Holcombe (2012) revealed that for a power level of 0.95 we would need to run 13 subjects. The power analysis was based on the effect between the predictable and unpredictable motion conditions in the two-target case. We decided to run more observers than this to be consistent with the number run in this previous study. A total of 18 (six males, 12 females) undergraduate students from the University of Melbourne aged between 18 and 28 years (M_{age} = 20.8, SD = 2.94) took part in this experiment. Of the 18 participants, two participants were excluded because they performed at ceiling levels (>97 % accuracy in both motion conditions for either the two-target or four-target case) and one participant was excluded because she did not meet the 20/25 visual acuity criterion (i.e., at least 20/25 in either eye). Therefore, the data for the remaining 15 participants were analyzed. All observers that were included in the analysis had normal or corrected-to-normal visual acuity (20/25 or better) as verified using a near vision (40 cm) Good-Lite® eye chart and normal color vision as determined by an Ishihara color blindness test.

Informed written consent was obtained prior to the commencement of the experimental session. The study was approved by the Department Human Ethics Advisory Group in the School of Psychological Sciences at the University of Melbourne.

Apparatus

Stimuli were presented on a 21-in Sony CRT monitor at a resolution of 1280 × 1024 pixels with a refresh rate of 85 Hz at a distance of 60 cm. The experiment was programmed and presented in MATLAB (Mathworks, Natrick, MA, USA) using the Psychophysics Toolbox (Brainard, 1997; Pelli, 1997). A 200-Hz head-fixed ViewPoint EyeTracker® system (Arrington Research, Inc., Scottsdale, AZ, USA) was used to ensure that all participants maintained fixation on a central fixation cross. Any time fixation was broken, which was defined as occurring if the point of fixation left the 1.5° × 1.5° fixation window centered on the fixation cross, the trial was restarted with the positions and motion directions of the objects randomized. The trajectories of the objects were never repeated to prevent the observers from learning them. The number of times a trial was restarted did not vary significantly between conditions, $F(3, 42)$ = 0.982, p = .41, η_p^2 = .07. The participants would therefore not have received significantly more practice with one of the conditions.

Stimuli

The present study employed a 2 × 2 within-subjects factorial design. The independent variables were type of motion (predictable vs. unpredictable) and number of targets (two vs. four). In each of the four conditions, observers were always presented with eight solid black disks (luminance = 1.74 cd/m²) on a white background (luminance = 29.99 cd/m²). A fixation cross (+) subtending 0.95° × 0.95° was presented in the center of the screen. Each disk subtended 0.75° of visual angle. All disks were confined to move within a 15° gray-edged square, bouncing off the inside walls of the square but passing over each other without colliding. In the predictable motion condition, the disks always travelled along a linear path except when the walls of the square were encountered. In the unpredictable motion condition, the disks randomly changed direction every 300–600 s.

Figure 1 illustrates the structure of the MOT trial used in the study. Each trial began with either two or four disks identified as the targets by turning red for 1.5 s. The targets then reverted to black and once again became indistinguishable to all other disks (distractors). The participants were instructed to track the targets while maintaining fixation on a cross at the center of the screen for 5.5 s. When the disks stopped moving, two disks were highlighted, one after the other. Participants were required to indicate whether each highlighted disk was a target or distractor. There was always a 50 % chance that a given probed disk was a target regardless of whether observers initially had to track two or four targets. Since tracking accuracy was defined as the percentage of trials for which observers were able to correctly identify both probed disks at the end of the trial, chance performance was at 25 %.

Procedure

Calibration procedure: Following the completion of ten practice trials, participants performed a calibration procedure which consisted of two 45-trial QUEST staircase routines, one for the two-target predictable condition and the other for the four-target predictable condition (Watson & Pelli, 1983). This procedure determined the speed at which each observer was able to achieve 75 % tracking accuracy in the predictable motion conditions for each target number. The staircase routines were necessary in order to control for individual differences in tracking ability (Oksama & Hyönä, 2004). Equal performance levels in the two-target and four-target conditions enables direct comparisons to be made between the two sets of conditions. Any differences between these two conditions cannot be attributed to differences in tracking performance caused by varying the number of targets.

Main experiment: Using the disk speeds obtained from the calibration process, observers completed in total 120 experimental trials that were presented in a random, interleaved

order. Observers had no prior knowledge of whether the motion for a given trial would be predictable or unpredictable.

Results and discussion

The average disk speeds generated by the QUEST staircase routine for all four experiments are shown in Fig. 2. Results for Experiment 1 are shown in Fig. 3. A 2 × 2 repeated measures ANOVA revealed a significant main effect of motion type on tracking accuracy, $F(1,14) = 5.16$, $p = .039$, $\eta_p^2 = .269$, with observers performing better in the predictable motion condition than in the unpredictable motion condition. Accuracy was also significantly greater in the four-target conditions than in the two-target conditions, $F(1,14) = 12.34$, $p = .003$, $\eta_p^2 = .468$, despite the attempt of the QUEST routine to equate performance in the two conditions. There was a significant interaction between the type of motion and the number of targets, $F(1, 14) = 13.92$, $p = .002$, $\eta_p^2 = .499$. Subsequent t-tests revealed a significant difference in tracking accuracy between the predictable and unpredictable motion conditions in the two-target case, $t(14) = 4.33$, $p = .001$, $r^2 = .57$, but not in the four-target case, $t(14) = 0.22$, $p = .83$, $r^2 = .003$.

These results support our hypothesis that observers are able to employ extrapolation when tracking two but not four targets. However, there is a potential confound. In this experiment, the speed at which the disks moved in the four-target conditions was slower than the speed at which they moved in the two-target conditions so as to equate tracking performance in the two sets of conditions. It is possible that the difference between the two-target and four-target conditions was the result of differing
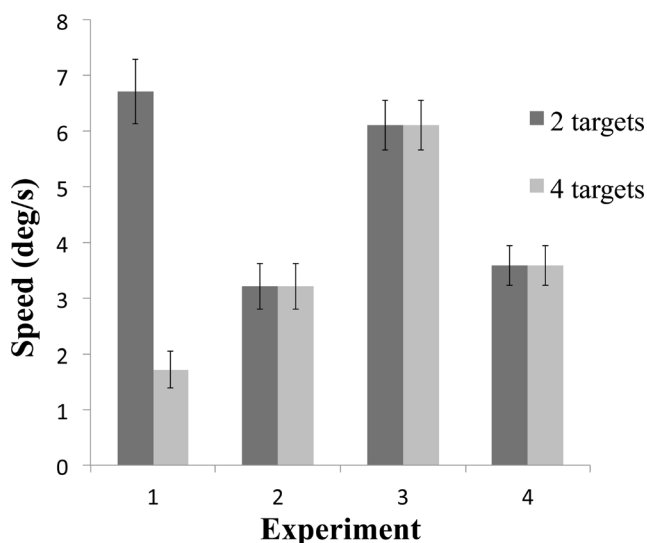
Fig. 2 The average speed of the disks for the four experiments. In Experiment 1, the disk speed for the four-target conditions was slower than the disk speed for the two-target conditions to equate tracking accuracy across these conditions. In Experiments 2–4, the same disk speed was used for all conditions regardless of target number. Error bars represent the standard error of the mean
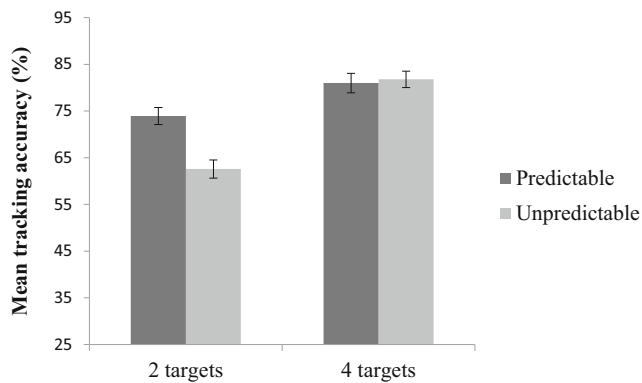
Fig. 3 Tracking performance for Experiment 1 in the predictable and unpredictable motion conditions. Error bars indicate the within-observer standard error of the mean (Morey, 2008)

disk speeds rather than differing target numbers. This issue was addressed in the following experiment.

Experiment 2

In this experiment, we arranged for all conditions to use the same disk speed. This ensures that any observed differences between the conditions are not due to differences in disk speed. This addresses the potential confound in Experiment 1 discussed above.

Method

Participants

Participants were 22 (five males, 17 females) undergraduate students from the University of Melbourne aged between 17 and 27 years (M_{age} = 19.5, SD = 2.48). None of these participants had participated in the previous experiment. Of these 22 participants, the data for seven participants were excluded. One participant was excluded because of ceiling performance (>97 %), one participant was excluded because of floor performance (<25 %), and the remaining five participants were excluded because they did not meet the 20/25 visual acuity criterion. The data for 15 participants were therefore analyzed. All observers included in the analysis had normal or corrected-to-normal visual acuity (20/25 or better) and normal color vision. Informed written consent was obtained from all observers.

Apparatus, stimuli, and procedure

Experiment 2 employed the same apparatus, stimuli, and procedure as that used in Experiment 1 except that the disk speed used was the same for all four conditions. The speed used by each observer was the average of the two-target and four-

target speeds at which they were able to achieve 75 % tracking accuracy in the initial calibration. As before, any time fixation was broken, the trial was abandoned and restarted with new, random positions and speeds. The number of times a trial was restarted did not vary significantly between conditions, $F(3, 42) = 2.62$, $p = .06$, $\eta_p^2 = .16$.

Results and discussion

The results for Experiment 2 are shown in Fig. 4. A 2 × 2 ANOVA revealed no significant interaction between motion type and target number, $F(1, 14) = .04$, $p = .84$, $\eta_p^2 = .003$ and no significant main effect of motion type, $F(1,14) = .02$, $p = .89$, $\eta_p^2 = .001$. The main effect for target number was significant, $F(1,14) = 35.31$, $p < .001$, $\eta_p^2 = .72$, with observers performing slightly better in the two-target conditions than in the four-target conditions. Overall, these results do not support our hypothesis that motion information can be used to extrapolate the future locations of two, but not four targets.

It is possible that our results were caused by our targets moving too slowly. When objects move too slowly, there is not much of an advantage to extrapolating so it could be that the observers in Experiment 2 did not extrapolate under these circumstances. Alternatively, it could be that these observers still extrapolated but the benefits of extrapolation were too small to be observed. Either way, the next logical step would be to repeat the previous experiment but instead use a faster disk speed to see if evidence for extrapolation could then be obtained.

Experiment 3

This experiment repeated the previous experiment except that we used a faster disk speed. For all conditions, we used the speed that was found in the calibration phase to result in 75 % accuracy in the two-target condition. As shown in Fig. 2, this
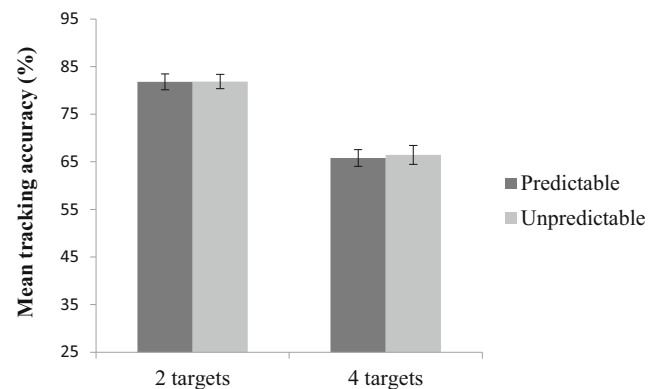
Fig. 4 Tracking performance for Experiment 2 in the predictable and unpredictable motion conditions. Error bars indicate the within-observer standard error of the mean (Morey, 2008)

was a much faster speed than that used in Experiment 2 (6.1°/s vs. 3.2°/s) and comparable to that used in the two-target conditions of Experiment 1 (6.7°/s). As such, we expected to find strong evidence for extrapolation, at least for the two-target conditions.

Method

Participants

Participants were 16 (five males, 11 females) undergraduate students from the University of Melbourne aged between 17 and 34 years (M_{age} = 20.1, SD = 4.23). None of these participants had participated in the previous experiments. All observers had normal or corrected-to-normal visual acuity (20/25 or better) and normal color vision. Data for one participant was excluded because she did not meet the 20/25 visual acuity criterion. Informed written consent was obtained from all observers.

Apparatus, stimuli, and procedure

Experiment 3 employed the same apparatus, stimuli, and procedure as that used in Experiment 1 except that the fast, two-target disk speed (i.e., the speed at which observers could answer both end-of-trial questions correctly on 75 % of trials in the two-target condition) was used in all four conditions. As before, anytime fixation was broken, the trial was abandoned and restarted with new, random positions and speeds. The number of times a trial was restarted did not vary significantly between conditions, $F(3, 42)$ = 1.33, p = .28, η_p^2 = .09.

Results and discussion

The results for Experiment 3 are shown in Fig. 5. A 2 × 2 ANOVA revealed a significant main effect of motion, $F(1,14)$ = 6.13, p = .027, η_p^2 = .305, and target number, $F(1,14)$ = 82.3, p < .001, η_p^2 = .855. Observers performed better in the predictable motion condition than in the unpredictable motion condition. Accuracy was higher in the two-target conditions than in the four-target conditions. Moreover, a significant interaction between the type of motion and the number of targets was found, $F(1, 14)$ = 9.18, p = 0.009, η_p^2 = .396. Subsequent t-tests revealed that there was a significant difference in tracking accuracy between the predictable and unpredictable motion conditions for the two-target conditions, $t(14)$ = 5.22, p < .001, r^2 = .66, but not for the four-target conditions, $t(14)$ = 0.73, p = .48, r^2 = .037. Consistent with our hypothesis, the results of Experiment 3 indicate that observers can extrapolate, at least when tracking just two targets, provided that the targets move sufficiently quickly.

In light of the results of Experiment 3, the results of Experiment 2 seem surprising. The only difference between these
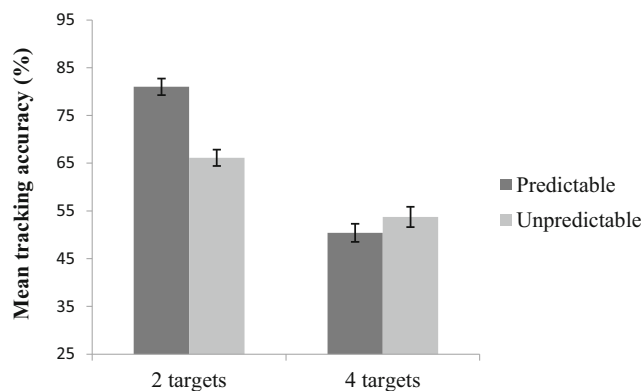
Fig. 5 Tracking performance for Experiment 3 in the predictable and unpredictable motion conditions. Error bars indicate the within-observer standard error of the mean (Morey, 2008)

two experiments was that the disks moved faster in Experiment 3. While this could explain why the difference in accuracies between the predictable and unpredictable conditions would be greater in Experiment 3 than in Experiment 2, it does not explain why there was no difference at all between these accuracies in Experiment 2. Even at the slow speed employed in Experiment 2, we would have expected to find some evidence for extrapolation.

As discussed above, one possibility is that in Experiment 2 the benefits of extrapolation were sufficiently slight that the observers did not extrapolate. It follows that if observers were somehow encouraged to extrapolate, we might still find evidence for extrapolation even at the slow disk speeds utilized in Experiment 2. Experiment 4 tested this hypothesis.

Experiment 4

Experiment 4 repeated Experiment 2 but in a way that encouraged observers to extrapolate. Specifically, Experiment 4 was identical to Experiment 2 except that the trials were presented in a blocked format so that the observer would always know what type of motion would occur in the following trial (i.e., predictable or unpredictable). In this way, we sought to encourage observers to extrapolate while keeping all the parameters of Experiment 2 otherwise identical. This tested the hypothesis that observers could have extrapolated in Experiment 2 had they been sufficiently encouraged to do so.

Method

Participants

Participants were 16 (five males, 11 females) undergraduate students from the University of Melbourne aged between 18 and 25 years (M_{age} = 19.4, SD = 2.00). None of these participants had participated in the previous experiments. All

observers had normal or corrected-to-normal visual acuity (20/25 or better) and normal color vision. Data for one participant was excluded because he did not meet the 20/25 visual acuity criterion. Informed written consent was obtained from all observers.

Apparatus, stimuli, and procedure

Experiment 4 employed the same apparatus, stimuli, and procedure as that used in Experiment 2 except that the trials were presented in a blocked format rather than in an interleaved, random order. Because observers are very poor at estimating how predictable object motion is (Vul et al., 2009), we hoped that blocking the trials would allow them to be more aware when the object motion was predictable and when it was not. This in turn was expected to lead to a larger difference between the predictable and unpredictable motion conditions. As before, any time fixation was broken, the trial was abandoned and restarted with new, random positions and speeds. The number of times a trial was restarted did not vary significantly between conditions, $F(3, 42) = 0.56$, $p = .65$, $\eta_p^2 = .04$.

Results and discussion

The results for Experiment 4 are shown in Fig. 6. Similar to the results of Experiment 2, results for this experiment showed no significant interaction between motion type and target number, $F(1, 14) = .53$, $p = .48$, $\eta_p^2 = .037$. As before, a significant main effect for target number was found, $F(1,14) = 26.0, p < .001, \eta_p^2 = .650$, with observers performing better in the two-target conditions than the four-target conditions. Contrary to the findings of Experiment 2, results revealed a significant main effect of motion type, $F(1,14) = 4.69$, $p = .048$, $\eta_p^2 = .251$, with observers performing better in the predictable motion condition than in the unpredictable motion condition. These results confirm the finding of Experiments 1 and 3 that
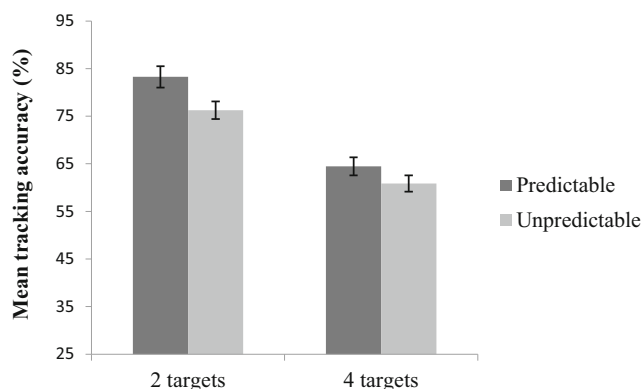
Fig. 6 Tracking performance for Experiment 4 in the predictable and unpredictable motion conditions. Error bars indicate the within-observer standard error of the mean (Morey, 2008)

in some circumstances observers can extrapolate while tracking.

General discussion

The present study investigated whether observers extrapolate object positions when performing MOT. Previous studies have argued that observers do utilize motion information to predict the future locations of objects during tracking (Fencsik et al., 2007; Iordanescu et al., 2009; St. Clair et al., 2010). However, these studies used paradigms that either forced observers to extrapolate because the objects were not continuously visible throughout the tracking task or contained a visibility confound. Thus, none of them were able to conclusively demonstrate that observers actually do extrapolate during traditional MOT. Howe and Holcombe (2012) addressed these concerns by using a task that required the moment-to-moment tracking of highly visible objects and found evidence for the use of extrapolation in traditional MOT. However, their study did not control for eye movements, so is potentially confounded (Zhong et al., 2014). Our study replicated the experiments of Howe and Holcombe (2012) while crucially controlling for eye movements by using an eye tracker to ensure that the observer's fixation was maintained on a central fixation cross throughout each trial. In Experiments 1 and 3, tracking performance was greater when the objects moved predictably than when they moved unpredictably but only when observers were required to track two targets as opposed to four targets. This indicates that observers were able to use motion information to extrapolate the future locations of targets, at least when tracking two targets.

Contrary to these results, we did not find evidence of extrapolation in the two-target conditions in Experiment 2. This may have occurred because in this experiment all the objects moved very slowly, which would decrease the usefulness of extrapolation and instead encourage observers to rely solely on location information. Experiment 4 repeated Experiment 2 but used a block design to ensure that observers would always know before each trial started whether the motion in that trial was predictable or unpredictable. As with Experiments 1 and 3, we found evidence for extrapolation with performance being greater when the objects moved predictably as opposed to when they moved unpredictably.

Taken together, these four experiments reaffirm that observers can extrapolate when tracking two targets and are less able to do so when tracking four targets. Furthermore, evidence for extrapolation is greater when the objects move quicker, though this does not necessarily mean that observers are not extrapolating when the objects are moving more slowly. It could be that when objects move more slowly, the benefits of extrapolation are too small to be reliably detected.

Our findings are in contrast to the predictions of Zhong et al. (2014). Zhong et al. performed a series of computational investigations to determine under what circumstances it would benefit observers to extrapolate given the limitations in the ability of humans to accurately recall the positions and velocities of objects. They concluded that in general it is not worthwhile to extrapolate. They found that it would only be worthwhile to extrapolate in a situation where all the objects move at a constant speed, do not change directions unexpectedly, and the observers are required to track less than four targets. Even in these circumstances, extrapolation resulted in an increase in tracking accuracy by only 2.4 %. This is in contrast to our results where we found that a situation that allows for extrapolation (i.e., the predictable motion condition) can result in an increase in tracking accuracy of approximately 15 % (95 % CI range: 8.8–21) relative to an otherwise identical situation that does not allow for extrapolation (i.e., the unpredictable motion condition), as demonstrated in our Experiment 3.

We expect that the key reason for the discrepancy between our data and their conclusions was that in those situations where we found a benefit for extrapolation, we utilized a faster speed than that which their model assumed. We suspect that the speed assumed by their model was simply too slow for extrapolation to have a large benefit. We found the largest gain due to extrapolation in Experiment 3 where the objects moved at an average speed of 6.1 °/s. Conversely, in their simulations that were most similar to our task (i.e., where the objects moved at a constant speed and did not randomly change directions), they assumed that the objects were moving only at 4.0 °/s. In our Experiment 2, the objects moved at an average speed of 3.2 °/s. For this experiment we found no advantage for extrapolation. This indicates that if the Zhong et al. (2014) simulations were redone using faster object speeds they would likely find a larger benefit for extrapolation.

There are currently only two models of MOT that explicitly allow for extrapolation. The model by Vul et al. (2009) assumes that either the observers have perfect knowledge of the inertia of the moving object or they otherwise assume that the objects have no inertia, so will move unpredictably. The authors concluded that the second model provided a better description of their behavioral data. In contrast, the model by Zhong et al. (2014) provides a more graded approach, predicting that observers will expect objects to be located at positions that are a weighted average of where the objects were last registered and where one would expect them to appear based on extrapolation. This allows the model to extrapolate conservatively. By varying the extrapolation weighting, this model can account both for situations where observers do not appear to extrapolate (e.g., our Experiment 2) as well as for those situations where observers do extrapolate (e.g., our Experiments 1, 3, and 4). As such, this model could potentially account for our data. However, as currently set up, it accounts for data only in a post-hoc manner, determining the

extrapolation weight so as to provide the best fit. To provide a full account of the data, it would need to predict the extrapolation weight in advance based on the number of targets to be tracked and whether a blocked design was used. Until this is done, the model cannot be described as having true predictive power. For a discussion of other theories of MOT in relation to extrapolation we refer the reader to Howe and Holcombe (2012).

Extrapolation for two versus four targets

Experiments 1 and 3 found that observers are less able to extrapolate when tracking four targets compared to when they track two targets, a finding that is consistent with previous literature showing that extrapolation is more likely under reduced target loads (Fencsik et al., 2007; Howe & Holcombe, 2012). Experiment 4 paints a slightly different picture. While there was a significant effect of motion type in that performance was better for predictable as opposed to unpredictable motion, there was no interaction between target number and motion condition. While it could be argued that this result suggests that extrapolation occurred not only for two targets but also for four targets, that reasoning would require drawing a strong conclusion from a null result – the lack of interaction between target number and motion condition – which is not statistically sound reasoning. Therefore, especially given our previous findings, we instead take Experiment 4 simply as providing further evidence that extrapolation can occur while tracking, leaving unanswered the question of whether this ability depends on the number of targets tracked.

So why does extrapolation generally occur more readily when tracking two targets as opposed to tracking four targets? Previously it was suggested that extrapolation required observers to estimate the velocities of the targets by tracking them with smooth eye pursuit movement. Obviously, the more targets there are, the harder this would be, so one would expect less extrapolation with more targets (Zhong et al., 2014). However, our study has shown that this explanation cannot be correct because extrapolation occurs even when observers do not fixate on the targets.

To extrapolate, one has to remember more information about each target. It is not enough to remember each target's position; one must also remember each target's direction of motion. Having to remember more information about the targets reduces the number of targets that can be tracked. This was most directly illustrated by Pylyshyn (2004). In his study, observers performed a standard MOT task where all the objects were identical during the tracking phase, but each object was assigned a unique identity at the beginning of the trial. It was found that on those trials where the observers were able to keep track of all four targets, they were typically unable to remember information about the individual targets. Specifically, they could not recall each target's unique identity. To be

able to accurately recall the targets' unique identities, observers needed to track fewer targets (Oksama & Hyönä, 2008). Consistent with these findings, it has been found that the precision by which observers are able to indicate the direction of motion of the targets decreases as the number of targets to be tracked increases (Horowitz & Cohen, 2010). We suggest that in our experiments, observers were simply not able to store sufficiently precise information about the motion of the targets to allow for extrapolation when asked to track four targets. They were only able to do this when asked to track two targets, which is why reliable extrapolation was observed only in the two-target cases and not in the four-target cases.

Blocked versus unblocked design

The only difference between Experiments 2 and 4 was that in Experiment 4 the trials were blocked according to whether the motion was predictable or unpredictable. Of the two experiments, it was found that extrapolation occurred only in Experiment 4. At first glance, this result may seem surprising. For the unpredictable trials, unpredictable motion will occur within the first 600 ms of the trial. Thus, one might expect the observers to be able to distinguish between predictable and unpredictable trials within approximately 600 ms. If so, blocking the trials should have little benefit. Why then did blocking provide a clear benefit in Experiment 4?

The clear implication is that observers have difficulty distinguishing between predictable and unpredictable trials just by viewing the trial, so instead have a strong bias to assume that objects always move in an unpredictable fashion. Although this conclusion is counter-intuitive, it is supported by an experiment conducted by Vul et al. (2009), as discussed above. Why would observers assume this? There seems to be at least two reasons for this behavior. One reason is that it is much more costly to assume that motion is predictable when it is not than to assume that motion is unpredictable when it is (Vul et al., 2009). In the first case, the observers are likely to lose the targets whereas in the second case the observers will simply not track the targets as efficiently as they might otherwise do. However, this explanation would seem to be incomplete in that it only argues that observers should make sure that motion really is predictable before assuming that it is predictable, not that observers should never assume that motion is predictable. For this reason, we suggest that observers are also concerned that the type of motion might change during the trial. Thus, although the objects might start off moving in a predictable fashion, their motion might become unpredictable later on in the trial. To be clear, we are not claiming that in our experiments the observers were consciously concerned that halfway through a trial the type of motion (i.e., predictable or unpredictable) might change. Instead we believe from the Vul et al. (2009) study that the default, possibly implicit,

assumption of observers is that motion is unpredictable and to change this assumption requires considerable evidence to the contrary (e.g., blocked trials where all the trials are predictable). We also do not claim that observers necessarily adopt an all or nothing approach. It could be that the expectation of where each target is located is a weighted average of its previously recorded position and its position based on extrapolation (Zhong et al., 2014). The more confident they are that the motion is predictable, the more they may be willing to more heavily weigh this expectation based on extrapolation.

Finally, before we conclude this section we should acknowledge that even in the predictable motion condition some of the motion may have appeared unpredictable to the visual system. In our trials, when the objects reached the boundaries of the displays they would bounce back according to the Newtonian laws of motion. Thus, their motion was entirely predictable. However, this does not mean that the visual system could predict their motion. Although the visual system can predict the trajectory of a single object bouncing off a wall, this does not appear to occur when there are multiple moving objects, as is the case in MOT. Atsma, Koning and van Lier (2012) demonstrated this using a probe technique. In MOT, probes were better detected immediately ahead of the targets, indicating that attention lead the targets. However, when a target approaches a wall, a probe in the location where the target would bounce to was detected worse than a probe in the location where the target would be if the wall did not exist. The authors concluded that anticipatory attention does not bounce. The implication for our study is that even in the predictable motion condition the visual system may regard the motion as still somewhat unpredictable. This may have encouraged the visual system not to employ extrapolation as much as it otherwise would have.

Conclusion

The present study has provided evidence that observers are able to employ extrapolation in MOT. In particular, we have found that tracking accuracy was greater when objects move predictably, allowing for extrapolation to occur, compared to when they move unpredictably, which would reduce the usefulness of extrapolation. The advantage of predictable motion was evident when observers tracked two objects but diminished when observers tracked four objects or when the objects moved sufficiently slowly, demonstrating that observers were less able to employ extrapolation as tracking loads increased and as object speeds decreased. Crucially, our results cannot be attributed to eye movements since we used an eye tracker to ensure that observers maintained fixation on a central fixation cross in all trials. While we do not claim that observers always extrapolate when tracking objects (or at least do not always gain a measurable advantage from doing so), it is clear that in

some circumstances they do extrapolate and the benefits from extrapolation can be quite large, resulting in an increase in tracking accuracy of approximately 15 %. Future models of object tracking would do well to consider extrapolation.

References

Atsma, J., Koning, A., & van Lier, R. (2012). Multiple object tracking: anticipatory attention doesn't "bounce". *Journal of Vision, 12*(13): 1, 1–11.

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial vision, 10,* 433–436.

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

Diaz, G., Cooper, J., Rothkopf, C., & Hayhoe, M. (2013). Saccades to future ball location reveal memory-based prediction in a virtual–reality interception task. *Journal of Vision, 13*(1): 20, 1–14.

Fehd, H. M., & Seiffert, A. E. (2008). Eye movements during multiple object tracking: Where do participants look? *Cognition, 108,* 201–209.

Fehd, H. M., & Seiffert, A. E. (2010). Looking at the center of the targets helps multiple object tracking. *Journal of Vision, 10*(4): 19, 1–13.

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

Franconeri, S. L., Pylyshyn, Z. W., & Scholl, B. J. (2012). A simple proximity heuristic allows tracking of multiple objects through occlusion. *Attention, Perception, & Psychophysics, 74,* 691–702.

Hayhoe, M. M., McKinney, T., Chajka, K., & Pelz, J. B. (2012). Predictive eye movements in natural vision. *Experimental Brain Research, 217,* 125–136.

Horowitz, T. S., Birnkrant, R. S., Fencsik, D. E., Tran, L., & Wolfe, J. M. (2006). How do we track invisible objects? *Psychonomic Bulletin & Review, 13,* 516–523.

Horowitz, T. S., & Cohen, M. A. (2010). Direction information in multiple object tracking is limited by a graded resource. *Attention, Perception, & Psychophysics, 72,* 1765–1775.

Howard, C. J., Masom, D., & Holcombe, A. O. (2011). Position representations lag behind targets in multiple object tracking. *Vision Research, 51,* 1907–1919.

Howe, P. D., Pinto, Y., & Horowitz, T. S. (2010). The coordinate systems used in visual tracking. *Vision Research, 50,* 2375–2380.

Howe, P. D., & Holcombe, A. O. (2012). Motion information is sometimes used as an aid to the visual tracking of objects. *Journal of vision, 12*(13), 1–10.

Huff, M., Papenmeier, F., Jahn, G., & Hesse, F. (2010). Eye movements across viewpoint changes in multiple object tracking. *Visual Cognition, 18*(9), 1368–1391. doi:10.1080/13506285.2010.495878

Intriligator, J., & Cavanagh, P. (2001). The spatial resolution of visual attention. *Cognitive Psychology, 43,* 171–216.

Iordanescu, L., Grabowecky, M., & Suzuki, S. (2009). Demand-based dynamic distribution of attention and monitoring of velocities during multiple-object tracking. *Journal of Vision, 9*(4):1, 1–12.

Keane, B. P., & Pylyshyn, Z. W. (2006). Is motion extrapolation employed in multiple object tracking? Tracking as a low-level, non-predictive function. *Cognitive psychology, 52,* 346–368.

Land, M. F., & McLeod, P. (2000). From eye movements to actions: How batsmen hit the ball. *Nature Neuroscience, 3,* 1340–1345.

Lukavsky, J. (2013). Eye movements in repeated multiple object tracking. *Journal of Vision, 13*(7):9, 1–16.

Morey, R. D. (2008). Confidence intervals from normalized data: A correction to Cousineau (2005). *Tutorial in Quantitative Methods for Psychology, 4,* 61–64.

Oksama, L., & Hyönä, J. (2004). Is multiple object tracking carried out automatically by an early vision mechanisms independent of higher-order cognition? An individual difference approach. *Visual Cognition, 11*(5), 631–671.

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

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial vision, 10,* 437–442.

Pylyshyn, Z. (2004). Some puzzling findings in multiple object tracking: I. Tracking without keeping track of object identities. *Visual cognition, 11,* 801–822.

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

Scholl, B. J. (2009). What have we learned about attention from multiple object tracking (and vice versa? In D. Dedrick & L. Trick (Eds.), *Computation, cognition, and Pylyshyn*. Cambridge, MA: MIT Press.

Shooner, C., Tripathy, S. P., Bedell, H. E., & Ogmen, H. (2010). High-capacity, transient retention of direction-of-motion information for multiple moving objects. *Journal of Vision, 10*(6): 8, 1–20.

St. Clair, R., Huff, M., & Seiffert, A. E. (2010). Conflicting motion information impairs multiple object tracking. *Journal of Vision, 10*(4): 18, 1–13.

Vul, E., Frank, M. C., Tenebbaum, J. B., & Alvarez, G. (2009). Explaining human multiple object tracking as resource-constrained approximate inference in a dynamic probabilistic model. *Advances in Neural Information Processing Systems, 22,* 1–9.

Watson, A. B., & Pelli, D. G. (1983). QUEST: A Bayesian adaptive psychometric method. *Perception & psychophysics, 33,* 113–120.

Wolfe, J. M., Place, S. S., & Horowitz, T. S. (2007). Multiple object juggling: Changing what is tracked during extended multiple object tracking. *Psychonomic bulletin & review, 14,* 344–349.

Zelinsky, G. J., & Neider, M. B. (2008). An eye movement analysis of multiple object tracking in a realistic environment. *Visual Cognition, 16,* 553–566.

Zelinsky, G. J., & Todor, A. (2010). The role of "rescue saccades" in tracking objects through occlusions. *Journal of Vision, 10*(14): 29, 1–13.

Zhong, S-H., Ma, Z., Wilson, C., Liu, Y., & Flombaum, J. I. (2014). Why do people appear not to extrapolate trajectories during multiple object tracking? A computational investigation. *Journal of Vision, 14*(12): 12, 1–30.