

# Motion information is sometimes used as an aid to the visual tracking of objects

Piers D. L. Howe

Psychological Sciences,  
University of Melbourne, Australia



Alex O. Holcombe

Department of Psychology,  
University of Sydney, Australia



In everyday life, observers often need to visually track moving objects. Currently, there is a debate as to whether observers utilize motion information in doing this or whether they rely purely on positional information (e.g., frame-by-frame locations). In our experiments, we had observers keep track of a subset of moving objects. In one condition, the objects moved in straight lines and their future positions were thus predictable. In a second condition, the objects changed directions randomly. Across three experiments, tracking performance was better in the predictable condition, suggesting that observers can use motion to help them track objects, at least when tracking just two. When tracking four objects, performance was not different between the two conditions. We discuss these findings in relation to several theories of object tracking.

Keywords: object tracking, MOT, attention, motion perception, individuating, predictable

Citation: Howe, P. D. L., & Holcombe, A. O. (2012). Motion information is sometimes used as an aid to the visual tracking of objects. *Journal of Vision*, 12(13):10, 1–10, <http://www.journalofvision.org/content/12/13/10>, doi:10.1167/12.13.10.

## Introduction

In everyday life, we often need to use visual information to track moving objects. When driving a car, keeping track of the positions of the other road users reduces the chance of collisions. At a playground, we might need to keep track of our children to make sure that they do not stray onto the road. Tracking is an important part of processing dynamic scenes.

In the laboratory, object tracking can be studied using the Multiple Object Tracking (MOT) paradigm (Pylyshyn & Storm, 1988). Figure 1 shows a typical trial. It starts with the presentation of a number of identical disks, a subset of which are briefly highlighted to indicate that these are the targets to be tracked. After one and a half seconds, the targets become identical to the other disks, and all the disks continue to move about the display in a random fashion. The disks then come to halt, two disks are highlighted, one at a time, and for each highlighted disk, the observer is asked whether it is a target or not. Because all the disks are identical during the movement phase, the observer can answer these questions only if he/she mentally tracked the targets. It is hoped that by understanding how tracking occurs in such simplified settings, we will then be able to understand how it occurs in more realistic situations (Cavanagh & Alvarez, 2005).

An open question is what information observers use to track the targets. According to the *location-only* hypothesis, the observer tracks the targets using only location information. Specifically, the observer remembers the targets' locations and frequently updates this information. Alternatively, according to the *motion* hypothesis, the observer uses motion information, in addition to location information, as an aid to tracking the objects.

Previous studies have shown that in the MOT paradigm, observers can report the movement direction of the targets (Horowitz & Cohen, 2010; Shooner, Tripathy, Bedell, & Ogmen, 2010). For example, Horowitz and Cohen (2010) had their observers view a standard MOT stimulus. When the disks came to a halt, a single target was highlighted and the observer was required to indicate the final direction of motion of that target. Observers were able to answer this question accurately, though precision declined rapidly as the number of targets was increased.

The fact that observers know the direction of motion of the targets is suggestive, but does not show that observers actually use this information to benefit tracking. Keane and Pylyshyn (2006) used the “target recovery” paradigm to address this question more directly. Just before the end of a typical MOT display, the screen would go blank and all the disks would briefly disappear. The disks would then reappear and the observer would be asked to indicate which were the targets.

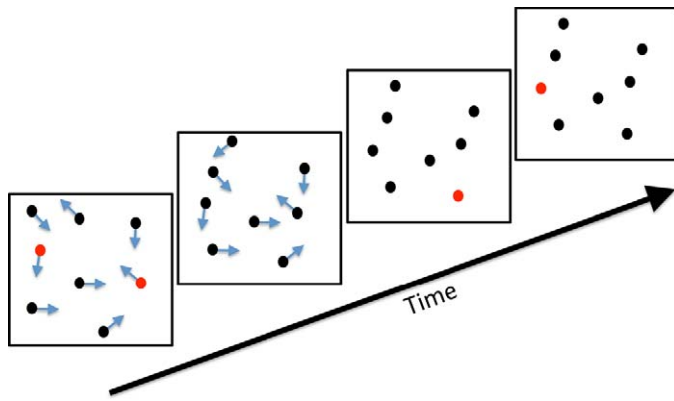


Figure 1. The trial structure used in the experiments. At the start of the trial, a number of the disks turn red for 1.5 s to indicate that these are the targets to be tracked. They then revert to black while continuing to move for 5.5 s, after which they come to a halt and two disks are highlighted in turn. For each highlighted disk, the observer uses the keyboard to indicate whether it is a target or not.

In one condition, the disks would reappear where they would have been located had they continued to move during the blank interval. Conversely, in another condition, the disks would reappear where they disappeared. Observers were more accurate in the second condition than in the first, indicating that the observers had difficulty using motion information to predict where the disks would reappear. Keane and Pylyshyn (2006) took this as evidence that observers do not use motion information when they track objects.

Fencsik, Klieger, and Horowitz (2007) disputed this conclusion. They pointed out that although the Keane and Pylyshyn (2006) study shows that observers rely more on positional information than on motion information, the study does not prove that observers cannot utilize motion information. To address this issue, Fencsik et al. (2007) also used the target recovery task but encouraged observers to use motion information by including only conditions where the observers were required to extrapolate the disk positions during the blank interval. Their first condition was very similar to that of Keane and Pylyshyn (2006): their disks reappeared where they would have been had they continued to move during the blank interval. Their second condition was identical to their first condition, except that their observers were not shown how the objects moved before the blank interval. Instead they were shown a static display denoting the object locations immediately before the blank interval. Thus, they were provided with the positions of the disks but not the movements of the disks, so could not use motion information to extrapolate the movement of the disks during the blank interval. Accuracy was greater in the first condition than in the second, for a tracking load of two targets but not for four targets,

indicating that observers can use motion information as an aid to extrapolation, at least when tracking only two targets.

A limitation of this finding is that extrapolation was only demonstrated in a situation that encourages observers to use motion information to track objects (Horowitz, Birnkrant, Fencsik, Tran, & Wolfe, 2006). During the blank interval, positional information is not available, so observers can only use motion information. Thus, the finding of Fencsik et al. (2007) that observers can use motion information as an aid to tracking, does not prove the observers actually do use motion information in the standard MOT paradigm where the targets are continuously visible and positional information is thus also continuously available.

To address this concern, StClair, Huff, and Seiffert (2010) utilized a MOT paradigm in which the objects were continuously visible throughout the trial. Each object contained a texture pattern that could move independently of the movement of the object on which it was displayed. It was found that tracking accuracy was worse when the direction of motion of the texture conflicted with the direction of motion of the object on which it was located, presumably because these conflicting motion signals reduced the quality of the motion information available to the observer. This finding is therefore consistent with the hypothesis that observers use motion information to track objects.

In a follow-up experiment, each texture moved in the same direction as the object to which it was attached, but not necessarily at the same speed. Just as conflicting direction information had previously been found to affect tracking accuracy, it was expected that conflicting speed information would also reduce tracking accuracy. Curiously, even when the speed of the texture conflicted with the speed of the object, tracking accuracy was not reduced, which would seem to argue against the motion hypothesis. Consequently, the authors suggested an alternative explanation: when the direction of texture motion conflicted with the direction of object motion, the disks became less visible, which may have reduced the quality of the positional information available to the observer, which itself could account for a reduction in tracking accuracy. Consistent with this hypothesis it was found that when all the objects were made more visible, the previously observed interference effects were substantially reduced (StClair et al., 2010).

An investigation by Vul, Frank, Tenebaum, and Alvarez (2009) avoided the possible visibility confound. In their study the stimuli were highly visible identical black disks, which moved in two dimensions. The speed and direction of the disks varied randomly to an extent determined by an inertia parameter. Thus, by varying the inertia, the motion of the disks could be made more or less predictable. It was found that tracking accuracy

did not vary as a function of predictability, suggesting that observers do not make use of motion information when tracking objects. A possible concern with this paradigm was that the speed of each disk was continuously changing, possibly making it harder to utilize velocity information as a tracking aid. It could be that observers can only utilize velocity information when the speed of the disks is held constant.

A subsequent study also argued that observers use only location information when tracking objects (Franconeri, Pylyshyn, & Scholl, 2012). In one condition, objects traveled behind an occluder and then reappeared either where they would have been expected to reappear based on their direction of motion immediately before the occlusion event, or the objects reappeared in another location that was closer to where they disappeared but inconsistent with their direction of motion. In this way, motion information was set against location information. Performance was greatest when the objects reappeared closest to where they disappeared.

Although this study demonstrated that observers rely more on location information than motion information, it did not rule out the possibility that observers may make use of motion information when it doesn't conflict with location information. In addition, in this study observers always tracked four targets, and it might be that observers can use motion information only when tracking fewer targets (Fencsik et al., 2007).

In summary, it remains unclear whether people actually use motion information as an aid to tracking in the conventional MOT paradigm and, if they do, under what circumstances this occurs. In our study we addressed this question using a MOT paradigm in which all the objects were continuously visible and where the visibility of the objects was the same in all conditions. Unlike the Vul et al. (2009) paradigm, in our paradigm the speed of the disks was constant, which might make the disk motion more predictable, thereby making it more likely that observers would use it as an aid to tracking.

In one condition, the objects moved in a very predictable manner, traveling in straight lines except when they bounced off the sides of the display. In the other condition, the disks moved in a less predictable fashion, changing direction randomly every 300–600 ms. From the results of Fencsik et al. (2007), we expected that observers would be able to use motion information when tracking two targets, but less so when tracking four targets. Consistent with this, we found tracking accuracy was higher in the predictable condition than in the unpredictable condition when observers tracked two targets, but not when they tracked four targets. This indicates that observers can use motion information as an aid to tracking, at least when tracking only a small number of targets.

## Experiment 1

The point of this experiment was to investigate whether observers are able to use motion information when tracking objects. Specifically, this experiment tested whether tracking performance is higher when objects move in a predictable fashion as opposed to an unpredictable fashion that would render motion information less useful.

### Method

#### Participants

There were 16 participants. Ages ranged from 18 to 41 years; five were male. A near vision (40 cm) Good-Lite® eye chart was used to verify that all observers had normal or corrected-to-normal visual acuity. In addition, the observers reported not being color blind. The observers provided informed written consent, and the study was approved by the Department Human Ethics Advisory Group in the School of Psychological Sciences at the University of Melbourne. Two participants were excluded from this experiment because a computer error corrupted their data.

#### Apparatus

The participants viewed a 21" CRT monitor at a resolution of 1280 × 1024 pixels with a frame rate of 85 Hz at a distance of 60 cm. Stimuli were presented in MATLAB (Mathworks, Natick, MA) using the Psychophysics Toolbox (Brainard, 1997; Pelli, 1997).

#### Stimuli

We used a 2 × 2 within-subjects factorial design. The two independent variables were type of motion (i.e., predictable or unpredictable motion) and number of targets (i.e., two or four targets). This led to a total of four conditions. In each condition there were always eight black disks on a white background. Each disk subtended 0.75 degrees of visual angle (°). The disks moved freely within the confines of a 15° square with a gray edge. The disks bounced off the inside walls of the square but passed over each other without colliding.

In the predictable motion condition, the disks moved in straight lines except when they encountered the walls of the square. In the unpredictable motion condition, they additionally changed directions randomly every 0.3–0.6 s, according to a uniform probability distribution.

At the start of each trial, the targets were identified by turning red for 1.5 s, after which they reverted to black. For the following 5.5 s, all the disks continued to move around the display. They then came to a halt and



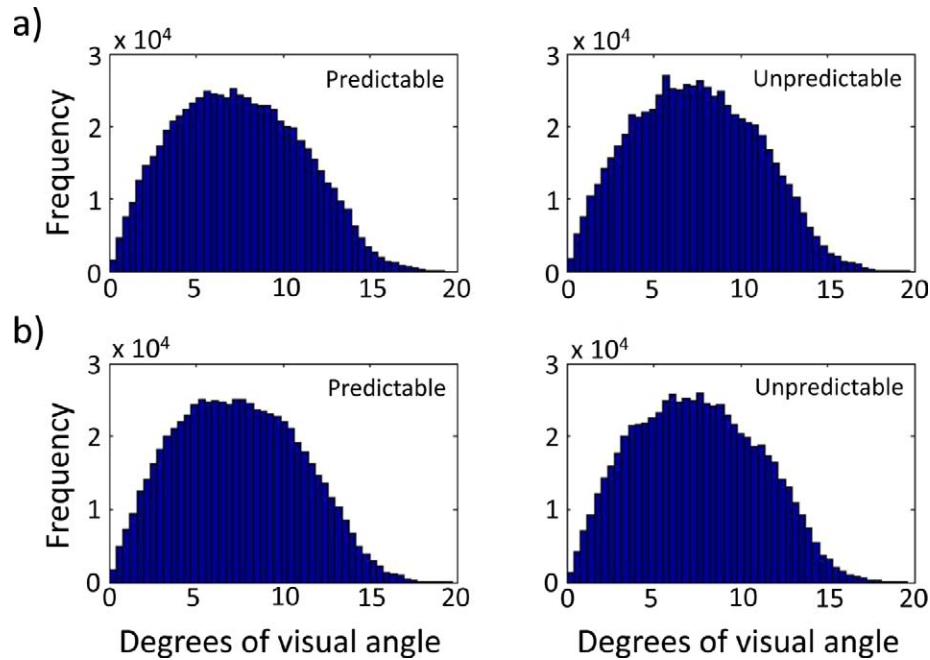


Figure 2. Disk separation histograms for Experiment 1 for (a) the two target conditions and (b) the four target conditions. For a given target number, the distributions are very similar for the predictable motion condition and the unpredictable motion condition.

two disks were highlighted in turn (Figure 1). For each of the two highlighted disks, the observer was required to indicate whether it was a target or a distractor. Regardless of the number of targets in the display, the probability that a given highlighted disk was a target was fixed at 50%. A trial was counted as correct only if both questions were answered correctly, ensuring that chance performance was at 25%, regardless of the number of targets.

Previous research has suggested that the closer a target passes to a distractor, the more likely the two are to be confused (Intriligator & Cavanagh, 2001). It is therefore important that the distribution of spatial separations is similar for both the predictable and unpredictable motion conditions. To verify this, at each screen refresh the distance between all possible disk pairs was calculated. These distances are plotted in histograms in Figure 2. The four histograms are virtually identical, making it unlikely that any difference in performance in the predictable and unpredictable motion conditions can be attributed to differences in the distributions of disk separations.

### Procedure

*Calibration procedure:* The experiment started with 10 practice trials after which a calibration procedure was run. This comprised two 50-trial QUEST staircase routines (Watson & Pelli, 1983), which were used to adjust the disk speed for each observer to ensure that performance did not suffer from either floor or ceiling effects. This was done separately for the two target

conditions and for the four target conditions. This calibration step was necessary because there is a wide variation in tracking ability between observers (Oksama & Hyona, 2004), so we could not use the same disk speed for all observers. These two staircase routines found the speed for each observer and for each target number that yielded an approximate 75% performance level for the predictable motion conditions, i.e., on approximately 75% of the trials, the observer was able to answer *both* end-of-trial questions correctly. This meant that performance for two targets was equal to that for four targets, allowing us to directly compare these two sets of conditions. For a given target number, the speed for the unpredictable condition was set to be the same as the speed for the predictable condition. These speeds are shown in Figure 3.

*Main experiment:* Using these disk speeds, the observers participated in 40 trials for each of the four conditions. Trials were blocked and all the trials from the same condition were run together. For each observer, the four blocks were run in a random order, and before each block the observer was informed whether the trajectories in that block would be predictable or not.

### Results

The results are shown in Figure 4. As can be seen from this figure, the QUEST staircase routine was successful in that performance in the predictable conditions was approximately 75% for both the two

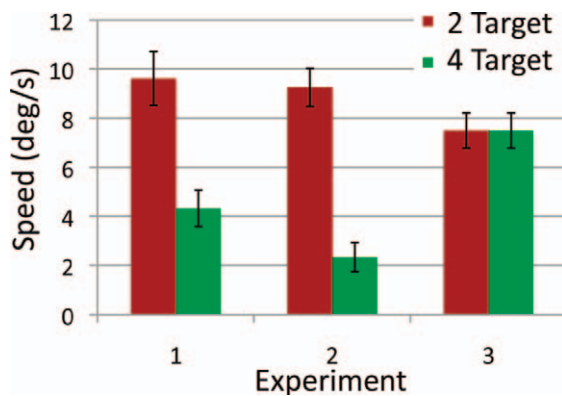


Figure 3. The speed of the disks for the three experiments. For a given number of targets, the same speed was used in the predictable motion condition as in the unpredictable motion condition. In [Experiments 1](#) and [2](#), the disk speed was slower when observers tracked four targets than when they tracked two targets to ensure that tracking performance in both conditions was comparable. In [Experiment 3](#), the same disk speed was used for four targets and two targets. Error bars represent the standard error of the mean.

target condition and the four target condition. A  $2 \times 2$  ANOVA revealed a significant interaction between the type of motion and the number of targets,  $F(1, 13) = 20.6$ ,  $p < 0.001$ . Subsequent  $t$ -tests revealed a significant difference between the predictable and unpredictable motion conditions for the two target case,  $t(13) = 5.37$ ,  $p < 0.001$ , but not for the four target case,  $t(13) = 0.715$ ,  $p = 0.49$ . These results are consistent with our hypothesis that observers can use motion information as an aid to tracking when tracking two targets, but are less able to use motion information as an aid to tracking when tracking four targets.

## Experiment 2

In [Experiment 1](#), the predictable and unpredictable motion conditions were presented in separate blocks. This may have helped observers to adopt the most efficient cognitive strategy for each condition. In [Experiment 2](#), the trials for the two motion conditions were presented in an interleaved and random order so that the participants did not know in advance whether the disks would move in a predictable or unpredictable fashion. Despite this, we found that performance was still higher for predictable motion than for unpredictable motion, at least when the observers were required to track only two targets.

### Participants

There were 15 participants. Their ages ranged from 20 to 47 years, and seven were male. As before, all

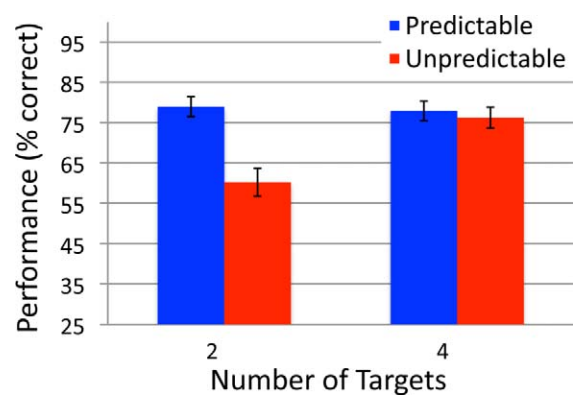


Figure 4. Tracking performance for [Experiment 1](#). Performance was defined as the percentage of trials for which observers identified both probed objects correctly as target or nontarget. Error bars show the within-observer standard error of the mean (Morey, 2008).

observers had normal or corrected-to-normal visual acuity and reported having normal color vision. All observers provided informed written consent and the study was approved by the Department Human Ethics Advisory Group in the School of Psychological Sciences at the University of Melbourne.

### Apparatus, stimuli, and procedure

The apparatus, stimuli, and procedure were identical to that of [Experiment 1](#) except that the trials from the four conditions were randomly interleaved so that the observer would not know in advance for a given trial whether the motion would be predictable or unpredictable.

### Results

The results for this experiment are very similar to those of the last one, as can be seen in [Figure 5](#). Again, a  $2 \times 2$  ANOVA revealed a significant interaction

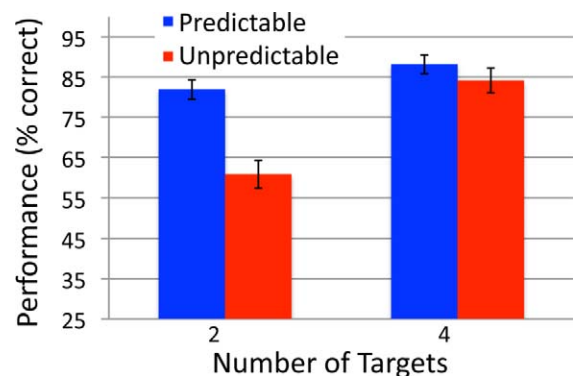


Figure 5. Tracking performance for [Experiment 2](#). Error bars show the within-observer standard error of the mean (Morey, 2008).

between the motion conditions and the target conditions,  $F(1, 14) = 18.6$ ,  $p < 0.001$ . Subsequent  $t$ -tests again revealed a significant difference between the predictable and unpredictable motion conditions for the two target case,  $t(14) = 7.28$ ,  $p < 0.001$ , but not for the four target case,  $t(14) = 1.81$ ,  $p = 0.09$ , again indicating that observers were able to utilize motion information when tracking two targets but less so, if at all, when tracking four targets.

## Experiment 3

In Experiments 1 and 2, the disk speed in the four target conditions was slower than that in the two target conditions. This was done to equate tracking performance in the two sets of conditions. However, because both velocity and direction discrimination do vary with speed (DeBruyn & Orban, 1988), there remains the possibility that the reason for the difference between the two target and four target conditions was due to the differing disk speeds and not due to the differing target numbers. Experiment 3 addressed this issue by using the same disk speed for all conditions. It was still found that tracking accuracy was greater in the predictable condition than in the unpredictable condition for two targets but not for four targets.

## Participants

There were 15 participants, with an age range of 18 to 24 years, five male. As before, all observers had normal or corrected-to-normal visual acuity and reported having normal color vision. All observers provided informed written consent and the study was approved by the Department Human Ethics Advisory Group in the School of Psychological Sciences at the University of Melbourne.

## Apparatus, stimuli, and procedure

The apparatus, stimuli, and procedure were identical to that of Experiment 1 except that the same disk speed was used for all four conditions. For each observer, this was the average of the speeds obtained by the two staircase procedures run in the initial calibration.

## Results

The results for this experiment are very similar to those of Experiment 1, as can be seen in Figure 6. Again, a  $2 \times 2$  ANOVA revealed a significant interaction between the motion conditions and the

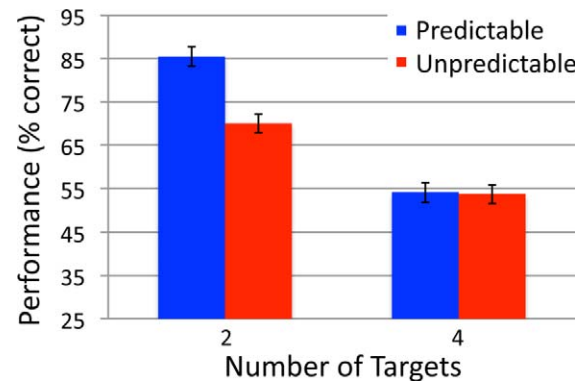


Figure 6. Tracking performance for Experiment 3. Error bars show the within-observer standard error of the mean (Morey, 2008).

target conditions,  $F(1, 14) = 16.1$ ,  $p = 0.001$ . Subsequent within-subject  $t$ -tests revealed a significant difference between the predictable and unpredictable motion conditions for the two target case,  $t(14) = 5.97$ ,  $p < 0.001$ , but not for the four target case,  $t(14) = 0.14$ ,  $p = 0.89$ , showing that observers could utilize motion information when tracking two targets but less so, if at all, when tracking four targets.

In the above analysis, a trial was counted as correct only if the observer was able to correctly identify both of the disks that were highlighted at its end. Alternatively, one could quantify tracking performance by measuring the percentage of the highlighted disks that were correctly identified at the end of the trial. Adopting this new measure did not change our findings: A  $2 \times 2$  ANOVA still revealed a significant interaction between motion conditions and target conditions,  $F(1, 14) = 10.7$ ,  $p = 0.006$ . Subsequent within subject  $t$ -tests revealed a significant difference between the predictable and unpredictable motion conditions for the two target case,  $t(14) = 5.64$ ,  $p < 0.001$ , but not for the four target case,  $t(14) = 0.81$ ,  $p = 0.43$ . Thus, our findings do not depend on using a particular definition for tracking accuracy.

## Discussion

Previous studies have shown that observers are aware of the motion of the targets that they track, but the precision of this knowledge is a steeply decreasing function of the number of targets (Horowitz & Cohen, 2010; Shooner, Tripathy, Bedell, & Ogmen, 2010). Indeed, one study found that observers can use motion information when tracking two targets, but not when tracking four targets (Fencsik et al., 2007). Our results are consistent with this supposition as we also found evidence for use of motion information when tracking two, but not four, targets. The Fencsik et al. (2007) study may have encouraged participants to use motion information

because in their trials, all the objects became invisible for a period of time. During these “blank” intervals, the observer could not use positional information, so was forced to extrapolate the positions of the targets. Our experiments go beyond this study by showing the motion information is used as an aid to tracking even when all the objects are continuously visible.

**Experiment 1** compared tracking ability in a condition where motion information might aid tracking, the predictable motion condition, to a condition where motion information would be less of an aid to tracking, the unpredictable motion condition. Observers tracked either two or four targets. There was a significant interaction between the number of targets and the difference in performance between the two motion conditions. Tracking performance was greater in the predictable motion condition but only when tracking two targets. This interaction is evidence that an observer’s ability to use motion to aid tracking decreases as the number of targets increases. From this data it is unclear whether motion information can be utilized when tracking four targets.

In **Experiment 1**, the conditions were blocked—all the trials from the same condition were presented together. Furthermore, before each block the observer was informed whether motion would be predictable or unpredictable. This was done to maximize the chance of the observers being able to use motion information, at least in those conditions where it would be helpful to do so.

This begs the question of whether the same result would occur if we presented the trials in a nonblocked format. **Experiment 2** investigated this by presenting the trials in a random, interleaved order. The results were very similar to the previous experiment, with performance being greater in the predictable motion condition, but only when observers were tracking two targets. While this result shows that our findings generalize to the nonblocked format, it did not necessarily ensure that observers used the same cognitive strategy in all conditions. It might be that during the first 1.5 s of the each trial, while all the targets were still highlighted, the observer determined whether the motion was predictable or unpredictable and selected a corresponding cognitive strategy.

In the previous two experiments, the speed in the four target conditions was set to be less than that in the two target conditions, so as to equate performance in both sets of conditions. This left open the possibility that the reason why observers were less able to use motion information in the four target conditions was because the objects were moving slower, not because there were more targets. **Experiment 3** addressed this issue by using the same speed in all four conditions. Again it was found that performance was greater in the predictable motion condition than in the unpredictable motion condition, but only when there were two

targets. This indicates that it is the number of targets that determines whether or not observers are able to use motion information as an aid to object tracking.

## Relationship to previous theories

Pylyshyn and Storm (1988) were the first to utilize the MOT paradigm. They investigated whether their data could be explained by a serial model according to which an observer attends to each target in turn. Every time a target is attended, its current location is memorized. When it’s time to reattend a given target, the observer would predict where the target would be using its previously noted location and its previously noted velocity. Consequently, the more predictable the motion, the better the performance of the model.

While this predictive serial model could explain why our observers were sometimes better at tracking targets in the predictable condition than in the unpredictable condition, Pylyshyn and Storm (1988) rejected it as implausible as it would require attention to travel between targets at a speed in excess of 1000°/s. After considering several versions of this serial model, they concluded that MOT is instead achieved by a parallel process in which all targets are attended simultaneously, an inference supported by three subsequent studies (Alvarez & Cavanagh, 2005; Howe, Cohen, Pinto, & Horowitz, 2010; Yantis, 1992).

Pylyshyn and Storm (1988) suggested that targets are tracked by visual indices known as FINSTs, an acronym that stands for Fingers of INSTantiation (see also Pylyshyn, 1989). These mental pointers operate in parallel and indicate the locations of the targets continuously. Although the initialization of the FINSTs may require attention (Pylyshyn, 2003), after initialization they operate pre-attentively. Consequently, each FINST is assumed to have access only to its target’s location information and is explicitly assumed not to be able to use motion information to help track the target (Keane & Pylyshyn, 2006). As such, it is unclear how FINST theory could explain our finding that observers can sometimes use motion as an aid to tracking.

As an alternative to FINST theory, Alvarez and Franconeri (2007) proposed FLEX theory. FLEX indices are similar to FINSTs in that they continuously indicate the locations of the targets. Unlike FINSTs, FLEXs are attentive, so they can have access to motion information, which, in principle, they could use to help track objects. However, this aspect of FLEX theory has not been developed and there is not yet an explicit mechanism to allow FLEXs to utilize motion information.

Taking a different tack, Kazanovich and Borisyuk (2006) proposed a connectionist model of object tracking. Their model posits that targets are tracked via an oscillatory neural network. However, their model



does not utilize motion information, so it cannot account for our data.

As mentioned earlier, Franconeri and colleagues (2012) suggested that tracking errors are due to spatial interactions between disks. For each screen refresh rate, all the disk-disk distances were measured. Figure 2 shows the resultant histograms for Experiment 1, which are essentially the same as those for Experiments 2 and 3. The distribution of interdisk distances was very similar for the predictable and unpredictable motion conditions, regardless of whether the observer was tracking two or four targets. It is therefore unclear how spatial interaction theory could explain why performance was not the same for the predictable and unpredictable motion conditions in the two targets case.

Vul et al. (2009) proposed a computational model of object tracking. They considered each MOT trial to be comprised of discrete time steps, with the positions of the targets at time step  $t$  predicted by their positions and velocities at time step  $t - 1$ . Although they concluded that in their experiment, observers did not utilize velocity information in tracking objects, they acknowledged that in other circumstances, observers might do so. Their model is unique in that it is the only computational model that formalizes how observers could utilize motion information. As such, it could, in principle, explain our findings.

The final theory of MOT that we will consider is by Tripathy, Ogmen, and Narasimhan (2011). The authors argue that in MOT, each object will leave a trace in iconic memory as it moves, similar to the contrail that a jet leaves behind it in the sky. They suggest that observers use these iconic memory traces to help track the targets. Specifically, they argue that observers track the targets by attending to each in turn in a serial manner. When it is time to reattend a given target, the observer uses the target's trace in iconic memory to determine where the target has moved to. Providing each target is reattended before its trace fades, this method will allow the targets to be recovered with perfect accuracy, thereby avoiding the objections Pylyshyn and Storm (1988) raised against serial models. Presumably, these iconic memory traces will be as robust in the predictable motion condition as in the unpredictable motion condition. Consequently, it is not clear how this model could explain why tracking performance was not equal in the predictable and unpredictable conditions.

## Conclusion

In this paper we have presented evidence that in some circumstances, observers can use motion information as an aid to object tracking. Specifically, we have shown that tracking performance is higher when

objects move predictably than when they move unpredictably. However, the benefits of predictable motion decrease as the observer tracks more targets, indicating that observers are less able to utilize motion information as tracking load increases.

This last finding is particularly interesting as it could not be predicted by an ideal observer model. According to such a model, the observer would utilize all the information available in order to maximize tracking performance. Thus, the model would predict that observers would utilize motion information just as readily when tracking four targets as when tracking two targets.

These findings also seem to be inconsistent with most theories of object tracking but could, in principle, be explained by the predictive serial model of Pylyshyn and Storm (1988) and the parallel model of Vul et al. (2009), although both theories would need to be elaborated before they could be applied to our data. Both models suggest that observers use the motion of the targets to predict where they will be located a short time later and these predictions should aid the tracking of the targets. Neither theory explains why observers are more likely to do this when tracking two targets as opposed to four targets.

As an alternative to using motion to predict the future positions of the targets, it could be that observers use motion information only to disambiguate a target from a distractor once they have become confused. We shall call this the error recovery hypothesis. This hypothesis was first considered by Pylyshyn and Storm (1988), but then dismissed because on its own, it could not fully account for their data. However, this does not mean that error recovery plays no role in object tracking, and indeed it could help explain the finding of StClair et al. (2010) that observers use direction information but not speed information as an aid to object tracking. Whereas the predictive hypothesis considered by Pylyshyn and Storm (1988) required the observers to use both direction information and speed information to predict the future locations of the targets, the error recovery hypothesis allows for the possibility that only direction information might be used as a target can often be distinguished from a distractor based solely on its direction of motion. Future studies will need to determine whether the predictive hypothesis or the error recovery hypothesis provides a better account of how observers are able to use motion information as an aid to object tracking.

## Acknowledgments

PH received funding from the University of Melbourne. AH received funding from an ARC Future



Fellowship and Discovery Project (DP110100432). We thank Charlotte Hudson for assistance in the data collection.

Commercial relationships: none.

Corresponding author: Piers D. L. Howe.

Email: pdhowe@unimelb.edu.au.

Address: Psychological Sciences, University of Melbourne, Australia.

## References

- Alvarez, G. A., & Cavanagh, P. (2005). Independent resources for attentional tracking in the left and right visual hemifields. *Psychological Science*, *16*(8), 637–643.
- Alvarez, G. A., & Franconeri, S. L. (2007). How many objects can you track? Evidence for a resource-limited attentive tracking mechanism. *Journal of Vision*, *7*(13):14, 1–10, <http://www.journalofvision.org/content/7/13/14>, doi:10.1167/7.13.14. [PubMed] [Article]
- Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, *10*(4), 433–436.
- Cavanagh, P., & Alvarez, G. A. (2005). Tracking multiple targets with multifocal attention. *Trends in Cognitive Sciences*, *9*(7), 349–354.
- DeBruyn, B., & Orban, G. (1988). Human velocity and direction discrimination measured with random dot patterns. *Vision Research*, *28*, 1323–1335.
- 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 and Psychophysics*, *69*(4), 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 and Psychophysics*, *74*(4), 691–702.
- Horowitz, T. S., Birnkrant, R. S., Fencsik, D. E., Tran, L., & Wolfe, J. M. (2006). How do we track invisible objects? *Psychonomic Bulletin and Review*, *13*(3), 516–523.
- Horowitz, T. S., & Cohen, M. A. (2010). Direction information in multiple object tracking is limited by a graded resource. *Attention, Perception and Psychophysics*, *72*(7), 1765–1775.
- Howe, P. D., Cohen, M. A., Pinto, Y., & Horowitz, T. S. (2010). Distinguishing between parallel and serial accounts of multiple object tracking. *Journal of Vision*, *10*(8):11, 1–13, <http://www.journalofvision.org/content/10/8/11>, doi:10.1167/10.8.11. [PubMed] [Article]
- Intriligator, J., & Cavanagh, P. (2001). The spatial resolution of visual attention. *Cognitive Psychology*, *43*(3), 171–216.
- Kazanovich, Y., & Borisyuk, R. (2006). An oscillatory neural model of multiple object tracking. *Neural Computation*, *18*, 1413–1440.
- 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*(4), 346–368.
- Morey, R. D. (2008). Confidence intervals from normalized data: A correction to Cousineau (2005). *Tutorial in Quantitative Methods for Psychology*, *4*(2), 61–64.
- Oksama, L., & Hyona, 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*, 631–671.
- Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, *10*(4), 437–442.
- Pylyshyn, Z. W. (1989). The role of location indexes in spatial perception: A sketch of the FINST spatial-index model. *Cognition*, *32*(1), 65–97.
- Pylyshyn, Z. (2003). *Seeing and visualizing*. Cambridge, MA: MIT Press.
- Pylyshyn, Z. W., & Storm, R. W. (1988). Tracking multiple independent targets: Evidence for a parallel tracking mechanism. *Spatial Vision*, *3*(3), 179–197.
- 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, <http://www.journalofvision.org/content/10/6/8>, doi:10.1167/10.6.8. [PubMed] [Article]
- StClair, R., Huff, M., & Seiffert, A. E. (2010). Conflicting motion information impairs multiple object tracking. *Journal of Vision*, *10*(4):18, 1–13, <http://www.journalofvision.org/content/10/4/18>, doi:10.1167/10.4.18. [PubMed] [Article]
- Tripathy, S. P., Ogmen, H., & Narasimhan, S. (2011). Multiple-object tracking: A serial attentional process. In C. Mole, D. Smithies, & W. Wu (Eds.), *Attention: Philosophical and Psychological Essays* (pp. 117–144). Oxford, UK: Oxford University Press.
- Vul, E., Frank, M. C., Tenenbaum, 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 and Psychophysics*, 33(2), 113–120.

Yantis, S. (1992). Multielement visual tracking: Attention and perceptual organization. *Cognitive Psychology*, 24(3), 295–340.