Gaze position lagging behind scene content in multiple object tracking: Evidence from forward and backward presentations

Jiří Lukavský¹ · Filip Děchtěrenko^{1,2}

Published online: 26 July 2016
© The Psychonomic Society, Inc. 2016

Abstract In everyday life, people often need to track moving objects. Recently, a topic of discussion has been whether people rely solely on the locations of tracked objects, or take their directions into account in multiple object tracking (MOT). In the current paper, we pose a related question: do people utilise extrapolation in their gaze behaviour, or, in more practical terms, should the mathematical models of gaze behaviour in an MOT task be based on objects' current, past or anticipated positions? We used a data-driven approach with no a priori assumption about the underlying gaze model. We repeatedly presented the same MOT trials forward and backward and collected gaze data. After reversing the data from the backward trials, we gradually tested different time adjustments to find the local maximum of similarity. In a series of four experiments, we showed that the gaze position lagged by approximately 110 ms behind the scene content. We observed the lag in all subjects (Experiment 1). We further experimented to determine whether tracking workload or predictability of movements affect the size of the lag. Low workload led only to a small non-significant shortening of the lag (Experiment 2). Impairing the predictability of objects' trajectories increased the lag (Experiments 3a and 3b). We tested our observations with predictions of a centroid model: we observed a better fit for a model based on the locations of

objects 110 ms earlier. We conclude that mathematical models of gaze behaviour in MOT should account for the lags.

Keywords Eye movements and visual attention · Motion: Integration

We live in a dynamic world in which we must keep track of objects for various reasons. We must avoid other cars when driving, or keep track of children in a crowd when shopping. Tracking objects is a task we practice every day.

In a laboratory, we can study this behaviour with a multiple object tracking (MOT) task (Pylyshyn & Storm, 1988). In the basic variant of this task, a participant sees eight identical discs on a screen. Four discs are targets. They flash, and then all objects start to move randomly. After a few seconds, the objects stop, and the participant is asked to choose the four targets. Scientists use MOT to study distributed attention (e.g., Alvarez & Cavanagh, 2005; Horowitz et al., 2007) or visual object qualities (for a review, see Scholl, 2009). MOT is a convenient task for studying everyday attention because it requires sustained effort. The trial lasts several seconds, and, over that period, the participant has a clear task to track the moving objects.

Because the processing of a visual signal takes some time, it seems beneficial to extrapolate the positions of moving objects. Nijhawan (1994, 2008) suggested that the brain extrapolates the positions of moving objects to counter the processing delay. The supporting evidence is based on mislocalisation errors when an object is flashed in the vicinity of another moving object (flash-lag effect).

The idea of extrapolating visual representation would be useful for MOT. Recent contradictory findings sparked a debate about the role of extrapolation in MOT. Some experiments have indicated that people rely solely on the

✉ Jiří Lukavský
jirilukavsky@gmail.com

¹ Institute of Psychology, The Czech Academy of Sciences, Hybernská 8, 110 00 Prague, Czech Republic

² Department of Software and Computer Science Education, Faculty of Mathematics and Physics, Charles University in Prague, Malostranské nám. 25, 118 00 Prague, Czech Republic

positions of the tracked objects (“location-only” account). The alternative, a “location and direction” account, has demonstrated that people are aware of objects’ motion information and use it in the tracking task. We summarise the arguments of both accounts below. Our main objective is to investigate the extent to which a potential extrapolation or lag is reflected in human eye movements during an MOT task.

One potential argument for motion extrapolation is the ability to track objects behind occlusions. Scholl and Pylyshyn (1999) tested the effect of a brief but complete occlusion of tracked objects on tracking performance. They found that occlusion by a visible or invisible surface (drawn in a background colour) did not impair performance. Conditions with more complicated motion signals, such as sudden jumps (teleporting behind the occluding surface), or imploding and exploding into and out of existence, led to a performance decrease. Unimpaired performance despite occlusion might be regarded as a result of motion extrapolation. However, Franconeri, Pylyshyn and Scholl (2012) showed that participants recover object correspondence after occlusion based on spatiotemporal proximity rather than the direction of movement.

Another source of evidence in this debate comes from querying about the final locations of the tracked targets. Keane and Pylyshyn (2006) interrupted the tracking with a blank screen midway through the trial. The duration of the interruption (150–900 ms) did not affect performance; however, the location after reappearance mattered. The objects could either reappear at the same location (no-move condition), at the extrapolated location given their original direction and velocity (move condition), or at their previous locations (rewind condition, as if they retracted their movement during the blank). People performed best in the no-move and rewind conditions, with worse results in the move condition, arguing against extrapolation.

Fencsik, Klieger and Horowitz (2007) disputed the benefit of the rewind condition. In their subsequent experiments, they demonstrated that participants utilise motion information. In the motion preview condition, the MOT trial was interrupted with a gap, the objects did not move after their reappearance, and participants were asked to respond. In the static preview condition, they presented a static screen with objects before the gap, but the objects moved during the gap, and, thus, the participants were asked to respond about the targets in a slightly jittered layout. The use of motion cues was dependent on the tracking workload. The motion preview yielded no benefit when tracking four targets, but the performance increasingly improved with three or fewer targets. This workload effect is in accord with the following MOT experiment (Horowitz & Cohen, 2010), in which participants were asked directly about the last known direction of a target. The precision of responses declined as the number of targets increased.

Because the objects moved during the blank interval, the importance of motion information was emphasised. However, it is not clear to what extent the motion information is used in a regular MOT task. To address this problem, St.Clair, Huff and Seiffert (2010) modified the MOT paradigm with random-dot textures. The background consisted of a texture, and all objects were covered with a texture, which could move independently of the direction of the moving object. People achieved the best performance when the local textures were static or moving concurrently with the object. Conflicting motion information (orthogonal or reversed) led to impaired tracking. This result supports the use of motion cues in tracking. The finding has been replicated in other studies (Huff & Papenmeier, 2013), and the impairment is selective to the objects with conflicting motion signals (Meyerhoff, Papenmeier, & Huff, 2013).

Similarly to experiments involving interrupted tracking (Keane & Pylyshyn, 2006; Fencsik et al., 2007), other authors asked subjects to recall the exact position of a particular target at the end of the trial. This method yielded mixed results, showing both extrapolation and lags. Iordanescu, Grabowecky and Suzuki (2009) asked people to track multiple objects of different colours. After several seconds, all objects on the screen disappeared, and the participant heard a colour name. The task was to mouse-click the last position of the target with the given colour. Iordanescu et al. (2009) measured the errors of reported locations, and found that the responses were biased towards the anticipated future locations. These results contrasted with previous findings by Howard and Holcombe (2008), who argued for perceptual lags in reported responses about various properties. In particular, they reported position lags of 40–130 ms using a similar paradigm. In their later experiment, they focused solely on reported positions and found lags of 10–70 ms over a range of workload and motion conditions (Howard, Masom, & Holcombe, 2011).

Alternatively, the motion extrapolation in MOT has been investigated using a probe detection task (Atsma, Koning, & Lier, 2012). The authors presented probes to measure the distribution of attention in the vicinity of a target. They found increased probe detection ahead of the target. Furthermore, they tested how the sensitivity ahead of the target changes when it is about to bounce off an obstacle. The probe detection results argued for linear extrapolation—the performance was better along the direct path than along the reflection path. A bouncing behaviour was anticipated only when the tracking load was low and a single target was tracked.

One problem with the previous studies is that they modify the original MOT paradigm: objects disappear or people receive an additional task to report. Jardine and Seiffert (2011) showed that the orientation indicative of the direction did not improve performance in the traditional MOT task but helped to recover targets after blank periods, when the trials were

interrupted. With interruptions, people can start to utilise surface features (such as colour; Papenmeier, Meyerhoff, Jahn, & Huff, 2014) apart from locations and possible directions.

Howe and Holcombe (2012) retained the traditional MOT design and manipulated the predictability of the movements (the objects moved in straight lines; in the unpredictable condition, their direction was sampled from a uniform distribution every 0.3–0.6 s). They found that the effect of predictability was moderated by tracking workload. Tracking performance was better in the predictable condition when tracking two targets. When tracking four targets, the benefit of predictability disappeared. These results have been replicated when controlling for eye movements (Luu & Howe, 2015).

In the current paper, we pose a related question: how much do people utilise extrapolation in their gaze behaviour, or in more practical terms, when modelling gaze position in MOT task, should we use the present positions of objects or rather their past/anticipated positions?

In MOT studies, participants are either asked to fixate on the centre of the screen, or they are allowed to move their eyes freely. Gaze behaviour in MOT is interesting because understanding eye movement strategies in a situation characterised by distributed attention can be useful for many other everyday tasks. Developing a gaze model for MOT is relatively easier than for other visual tasks: the scene content is simpler, the task is clear throughout the trial, and the dynamic information helps to synchronise gaze among observers (Dorr et al., 2010; Smith & Mital, 2013).

Recent studies on eye movements in MOT (Fehd & Seiffert, 2008, 2010; Huff, Papenmeier, Jahn, & Hesse, 2010; Lukavský, 2013; Zelinsky & Neider, 2008) have identified several gaze strategies: looking at individual targets and looking at their centroid. The preference for each strategy depends on the number of targets. Zelinsky and Neider (2008) reported that people prefer centroid looking when tracking two targets, they use both strategies equally when tracking three, and they switch between targets when tracking four. Fehd and Seiffert (2008) reported a high prevalence of centroid looking (approximately 42 % of the tracking time) in MOT with three to five targets. In other words, people fixate on an empty space for a significant proportion of the time in MOT. Huff et al. (2010) showed that the preference for a centroid strategy increased with object speed. Lukavský (2013) evaluated several models of gaze behaviour in MOT, and found the best results for the centroid strategy or for its variant with weights adjusting for the danger of crowding. People are usually not aware when an identical trial is repeated. Thus, it is possible to compare to what extent people repeat their gaze patterns. Lukavský (2013) found that people tended to make similar eye movements during repeated presentations. Because all tested models performed worse than the predictions based on gaze patterns from previous expositions, the similarity of eye movements is only partly explained by the current models.

One potential difficulty is the proper time reference. Current models operate with the targets'/distractors' locations in the present frame. As previously noted, there are indications that people are sensitive to objects' directions and their future locations (Atsma et al., 2012; Horowitz & Cohen, 2010). Alternatively, other studies have shown that the reported locations lag behind the content of the present frame (Howard & Holcombe, 2008; Howard et al., 2011). Therefore, a future/past frame might be a better reference for the mathematical model.

In the current study, we aimed to find a better time reference for gaze models in MOT. We used a data-driven approach with no a priori assumption about the underlying gaze model (target centroid, crowding reduction, etc.). We extended the method based on comparisons of spatiotemporal saliency maps (Děchtěrenko, Lukavský, & Holmqvist, 2016; Lukavský, 2013). In particular, we presented identical trajectories forward and backward (see Fig. 1). Then, we tested different shift values to find an optimal match. To illustrate, we compared both forward and backward maps without the shift first and obtained a similarity indicator (Pearson correlation). We gradually tested different time adjustments until we reached a local maximum at $\Delta t = 200$ ms. Thus, any given time point T in the forward trial corresponds to the content of the $T - \Delta t$ time point in the backward trial. We can deduce that the gaze position was $\Delta T = \Delta t/2 = 100$ ms ahead of the current content.

Experiment 1

In Experiment 1, we tested the method of measuring lead/lag values with identical MOT trials presented forward and backward. In the traditional MOT paradigm, an experimenter presents a demanding task and compares accuracy levels across conditions. However, the gaze strategy is unclear in the error trials because the participant can track a wrong target or give up tracking some targets. Therefore, we must analyse only trials with correct answers. We chose the MOT variant with a speed of 5°/s, which was still demanding, but the tracking was usually successful.

Method

Participants

Twenty students participated in this experiment for course credit (mean age 21 years, range 19–27 years; three males). All participants had normal or corrected-to-normal vision. All experiments conformed to National and Institutional Guidelines for experiments in human subjects and with the Declaration of Helsinki. Data from 21 subjects were originally collected, but data from one participant were removed due to calibration problems.

Fig. 1 **a–c** Assume there is a single moving object in the scene (*black line*) and the gaze tracking the object is delayed by ΔT (*red line*). We can plot the time and y -coordinates for forward presentation (**a**) and for the

backward presentation (**b**, gaze in *blue*). When we reverse the time coordinate for the backward presentation (**c**), the object positions are aligned, but the gaze positions are misaligned by $2 \times \Delta T$

Apparatus and stimuli

Stimuli were presented on a 19-in CRT monitor with 1024×768 resolution and 85 Hz refresh rate, using a MATLAB script with Psychophysics and Eyelink Toolbox extensions (Brainard, 1997; Cornelissen, M, & Palmer, 2002; Kleiner, Brainard, & Pelli, 2007; Pelli, 1997). Participants viewed the screen from a distance of 50 cm, and their head movements were restrained using a chinrest.

Tracking stimuli comprised eight medium-gray disks on a black background. Each disk subtended 1° of visual angle and moved at a fixed speed of $5^\circ/\text{s}$. The disks were confined to move within an invisible square ($20^\circ \times 20^\circ$) and bounce off the invisible square boundaries and one another with a minimum interobject distance of 0.5° . In addition to bouncing, the direction of disks was sampled from a von Mises (circular Gaussian) distribution with a concentration parameter of $\kappa = 10$ every 100 ms, creating an impression of Brownian motion.

Eye position was recorded at 250 Hz using Eyelink II eye tracker (SR Research, Canada). Drift correction was performed before each trial, and a nine-point calibration was completed at the beginning of each block (i.e. after every 15 trials).

Design and procedure

After eight training trials, each subject completed 60 experimental trials. Each trial started with a fixation point, which was used for drift correction and then disappeared. Subsequently, eight objects were presented. Four target objects were highlighted in light green, and after 2 s, all objects turned

medium-gray and commenced moving. The objects stopped moving after 10 s, and participants selected the four tracked targets with the mouse. After they made their choice, feedback was provided: either a green word “OK” or a number of errors in red appeared for 500 ms. The participants were instructed to track all four targets carefully and not to deliberately limit their tracking to a subset. They were told that the trial is considered correct only if all targets are successfully identified.

At the end of the experiment, we asked participants whether they thought some trials had been repeated. We compared the recognition rate with our previous results to check for a comparable level of recognition.

The experimental trials were presented in four blocks separated by eye-tracker calibrations. Each block consisted of five forward trials, five backward trials, and five unique trials. The five forward trials presented five different motion trajectories and were repeated in each block. Each backward trial corresponded to one forward trial, presented the same trajectory in reversed time order and was also repeated in each block. Consequently, in each experiment, we obtained gaze data for five different MOT trajectories: how the gaze data differed across the four repetitions and four backward presentations. The remaining 20 unique trials were used to increase the variability of the trajectories and conceal the repetition. Because the repetition of the exact starting/ending locations might attract attention, all trajectories were 12 s, but only a random 10-s segment of the trajectory was presented in each trial. As a result, all presentations of the same trajectory shared a common segment of 8 s (from 2.0 s to 10.0 s).

Data analysis

Data preparation

Because people often use smooth pursuit to track moving objects, we analysed all data samples with no fixation detection performed. We filtered the data with respect to three criteria. First, we limited our analysis to the trials in which the participant recognised all four targets (91.2 % of trials, range 72.5–100.0 %). Second, we checked the recorded data for blinks and blink artefacts, and trials with more than 10 % of the samples missing were discarded (10.1 %, 0–27.5 % per person). Third, we limited the analysis to the trajectories for which enough data were available. For each subject and trajectory, we collected data from four forward and four backward presentations, and we included only trajectories in which at least two forward and two backward presentations were available. After applying the two criteria, the final sample included 704 trials (88 %, 22–40 trials per person). [Appendix](#) shows a table summarizing the numbers of excluded trials for this and the following experiments.

The raw eye-movement data for each trial ranged from -10° to $+10^\circ$ in the horizontal and vertical dimensions, and from 2 s to 10 s in time (the time segment shared across all presentations, with 4-ms steps). The temporal dimension in the backward map was flipped. To facilitate further computations, the data were binned in a 3D spatiotemporal matrix with a bin size of $0.25^\circ \times 0.25^\circ \times 10$ ms (same as in Lukavský, 2013). The value of each bin represented the number of

corresponding gaze samples. For each trajectory/subject, the data from forward trials and the data from backward trials were pooled together to form two separate spatiotemporal matrices representing gaze samples for forward and backward presentations. Finally, the spatiotemporal matrices were smoothed with a Gaussian filter ($\sigma_x = \sigma_y = 1.2^\circ$ and $\sigma_t = 26.25$ ms) to produce spatiotemporal maps. Smoothing helps to evaluate the proximity of gaze patterns in both space and time. We adopted the particular parameters of the Gaussian filter from a study by Dorr et al. (2010); however, it has been shown that changing the variability parameters to 50 % or 200 % leads to similar results (Lukavský, 2013). The similarity of spatiotemporal gaze maps was evaluated with a Pearson correlation. The alternative analysis using the Normalized Scanpath Saliency for dynamic scenes (Dorr et al., 2010; Lukavský, 2013) led to similar results (not reported here).

Evaluation of lags

The presence and the extent of prediction or lag in eye movements were evaluated for each trajectory/subject combination from the forward and backward spatiotemporal map. The process is illustrated in Fig. 2. First, as previously noted, the temporal dimension in the backward map was flipped (the latest timestamp, 10.0 in the backward map, became the timestamp 2.0, i.e. corresponding to the same object configuration on the display, which was presented at the timestamp 2.0 in the forward map). If no lag/prediction was present, we should see the maximum similarity of the forward and

Fig. 2 **a** All forward repetitions of the same trial were summed and smoothed with a spatiotemporal Gaussian filter (red). The backward repetitions were treated analogously, and the time coordinate was flipped (green). **b** The spatiotemporal maps (here viewed from side)

were aligned to test lead/lag values between ± 250 ms with 10-ms steps. **c** The resulting lead/lag value was the local maximum of all correlation coefficients within the tested interval

backward map with no additional shift. Alternatively, the gaze is based on the object configuration at different times (in past or future). For example, if people move their gaze based on their prediction by 100 ms ahead, their gaze is based on the object configuration that will appear in 100 ms. The object configuration shown at timestamp 4.0 is the basis for their gaze at timestamp 3.9 in the forward map and timestamp 4.1 in the backward map, and we observe the maximum similarity of the forward and the backward map if we shift the backward map by -200 ms (we use negative values for leads and positive for lags).

For each trajectory/subject combination, we searched for the local maximum of the forward/backward map similarity with a maximum shift of ± 500 ms to detect the lead/lag within the range of -250 to $+250$ ms with 10-ms steps. To illustrate, we shifted the backward map by T ms, and expected the same lead/lag value in both maps. Thus, we were actually testing whether the gaze was likely to be $T/2$ ms ahead/behind. The local maximum was found within the range of the interval in the majority of cases (91.5 %), and the non-converging cases were discarded from the analysis.

For each trajectory/subject combination, we obtained a lead/lag value. In this and following experiments, we used mixed linear models to evaluate the statistical significance. The analyses were performed in R (R Core Team, 2014) using the lmer (Bates, Maechler, Bolker, & Walker, 2015) and lmerTest (Kuznetsova, Brockhoff, & Christensen, 2016) packages. In the current analysis, we used a model with an intercept and no further fixed effects. The random effects included individual differences in intercept values, representing the assumption that people differ in their mean lead/lag values. We compared this model with a null model with zero as an intercept.

Results

We found a consistent lag between gaze position and the content of the scene. The search for a local maximum converged in 91.5 % of the cases, and the average lag value was $+114$ ms ($SD = 35$ ms). The lag was present for every participant (see Fig. 3), and the individual means ranged from 52 ms to 173 ms. The model featuring a free intercept represented the data much better than the zero-intercept solution [$\chi^2(1) = 50.04$, $P < .001$]. The estimated lag value based on the mixed linear model was $+111$ ms with the 95 % confidence interval [$+96$; $+126$].

The mean initial correlation between the forward and backward presentations was .59 ($SD = 0.14$; range .17 to .85) for the lead/lag of 0 ms. The correlation at the local maximum was on average 9.3 % higher. To obtain baseline data, we also assigned the forward and backward data randomly (data from different subjects and different trials), and the correlation ranged from $-.07$ to $+.57$ (mean

Fig. 3 Individual lead/lag values (means and SEM). The black line represents zero lag between gaze position and content of the scene. The dotted line is the overall mean lag observed in Experiment 1

.11, $SD = 0.13$, median .06). The baseline correlation values were significantly greater than zero [$t(93) = 7.660$, $P < .001$], which is likely the result of a central tendency in participants' gaze.

Several participants reported they noticed repetitions (5 of 21; 24 %). This rate did not differ from that reported in a previous study ($z = 0.787$, $P = .430$), in which we showed that people fail to identify repeated trials despite reporting repetitions (Lukavský, 2013).

Discussion

We found that the gaze position lags behind the scene content by approximately 110 ms. Importantly, the effect was robust in that the lag was present in every subject. Note that we did not make any assumptions about the gaze positions in the MOT task. The employed approach is data-driven, and we did not need to adopt any theory regarding where people look or what strategy they use. Dynamic areas of interest (Papenmeier & Huff, 2010) could be another possible approach, but it would require more assumptions about the classification of gaze (size of AOI, centroid).

These results have a direct application to the modelling of gaze position in MOT (Fehd & Seiffert, 2008, 2010; Lukavský, 2013; Zelinsky & Neider, 2008). Authors typically search for the link between gaze position and currently observed content. However, our results show that the lag can introduce an error of approximately 0.6° in every object (given the speed of $5^\circ/s$). One possible solution to this problem is to include a constant time lag of 110 ms. The lag was different for each subject, but this simple approach would help decrease the average error.

Previous experiments (Horowitz & Cohen, 2010; Atsma et al., 2012; Howe & Holcombe, 2012) showed that sensitivity to objects' motion is dependent on other factors, e.g., tracking workload. In the following experiments, we explored whether the amount of lag in gaze position can be manipulated within the experiment.

Experiment 2: tracking load

In MOT, people are more sensitive to objects' directions when they have more available resources (Horowitz & Cohen, 2010; Atsma et al., 2012; Howe & Holcombe, 2012; Luu & Howe, 2015). With lower tracking workload, they start to use additional features (e.g. background for spatial reference, Jahn, Papenmeier, Meyerhoff, & Huff, 2012). Their performance is less affected by occlusions, and they can afford to make more rescue saccades to recover the targets in moments of uncertainty (Zelinsky & Todor, 2010).

In Experiment 2, we compared the gaze lag in MOT trials with two or four targets tracked. Tracking two targets should demand fewer resources, and the available resources could be used for more effective motion extrapolation. We did not include a more difficult condition because tracking more targets would yield more errors and subsequently fewer trials to analyse.

Method

Participants

Thirty students participated in this experiment for course credit (mean age 23 years, range 19–36 years; 12 males). All participants had normal or corrected-to-normal vision.

Design and procedure

The experiment consisted of eight training trials and 64 experimental trials presented in four blocks and separated by eye-tracker calibrations. Each block consisted of eight forward trials and eight backward trials. In the presented trials, participants were asked to track either two or four targets (low/high workload condition).

Unlike in Experiment 1, we did not use unique trials to conceal repetition because we did not want to increase the length of experiment and account for an additional fatigue effect. The repetition was still concealed using different starting/ending positions.

The procedure was analogous to Experiment 1.

Apparatus and stimuli

The apparatus and stimuli were identical to those used in Experiment 1 except that two targets and six distractors were present in half of the trials.

Data analysis

As previously done, we limited our analysis to the trials in which the participant recognised all assigned targets (two or four). The mean accuracy rate was 92.4 % (range 73.4–100.0 %). We removed a participant with a very low accuracy of 56.2 %. In the next step, we checked the recorded data for blinks and blink artefacts, and trials with more than 10 % of samples missing were discarded (13.8 %, 1.6–37.5 % per person). The data from two additional participants were removed because of the high occurrence of blinks and related artefacts (both had more than 50 % of their data missing). Consequently, the final dataset was based on data from $N = 27$ subjects.

For each subject and trajectory, we collected data from four forward and four backward presentations, and we included only trajectories in which at least two forward and two backward presentations were available. After applying the two criteria, the final sample included 1416 trials (82 %; 27–63 trials per person).

For statistical comparison, we tested a model featuring a fixed effect (tracking workload) against a null model that included an intercept (contrary to Experiment 1). Random effects in both models included individual differences in intercepts and slopes representing the assumption that people differ in their mean lead/lag value and the workload may affect people differently (maximal model; Barr, Levy, Scheepers & Tily, 2013). An analogous approach was used in subsequent experiments.

Results

As in Experiment 1, we found a consistent lag between the gaze position and the content of the scene. The local maximum was found in 94.7 % of cases. In the low workload condition, when people were asked to track only two targets, the lag was slightly lower (mean +93 ms; 95 % CI [+78; +107]) than in the high workload condition (mean +108 ms; 95 % CI [+89; +128]). When tracking only two targets, the lag slightly decreased by 15 ms (95 % CI [−2; +33]). This difference was not significant, and the model featuring a workload condition as a fixed effect did not outperform the null model [$\chi^2(1) = 2.99, P = .083$].

Several participants reported they noticed repetitions (10 of 30; 33 %). This rate did not differ from our previous observations ($z = 0.122, P = .904$) or from Experiment 1 ($z = 0.735, P = .465$). The tracking workload did not affect the accuracy

[paired t -test, $t(26) = 0.0$, $P = 1.0$] or convergence [$t(26) = 1.938$, $P = .064$].

Discussion

In Experiment 2, we replicated the lag of approximately +100 ms that was observed in Experiment 1. However, the critical manipulation did not yield significant results: When people tracked only two objects, the lag was smaller, but the difference was not significant. Tracking two objects is generally easier, but the freed resources did not lead to shortened lags. Nevertheless, the observed difference is in the expected direction.

The non-significant difference may mean that the lag is already very low (floor effect), and it is costly to reduce it further. In addition to available cognitive resources, the ability to extrapolate objects' positions also depends on the predictability of their movement. The objects followed semi-Brownian trajectories, and their randomness was defined by the concentration parameter κ of a von Mises distribution. In Experiments 1 and 2, the parameter was fixed. We decided to manipulate the κ values to introduce either more ballistic movements or more chaotic movements.

Experiment 3a and 3b: predictability

In Experiments 3a and 3b, we manipulated the predictability of objects' movements within the MOT trial. We decided to perform two separate experiments. Both variants featured the same condition as in the previous experiments (tracking four objects; $\kappa = 10$), which allowed us to compare the lag across experiments. This original condition was contrasted with more predictable trials (Experiment 3a; $\kappa = 20$) or more chaotic trials (Experiment 3b; $\kappa = 5$).

It is difficult to express how the κ values affect the predictability of the movements. We tested several variants and decided to use 50 % and 200 % values, as noted above. To illustrate, Fig. 4 shows how landing positions depend on κ after 1 s of movement.

Fig. 4 The effect of the κ parameter of the von Mises distribution on movement predictability. Each image depicts a simulation of landing positions after 1 s of movement over 1000 repetitions with the given κ

Method

Participants

Twenty-seven and 30 students participated in Experiments 3a and 3b, respectively (Experiment 3a: mean age 22 years, range 19–27 years, seven males; Experiment 3b: mean age 21 years, range 18–27; three males). All participants had normal or corrected-to-normal vision.

Apparatus and stimuli

The apparatus and stimuli were identical to those used in Experiment 1. We manipulated the κ parameter of the von Mises distribution to change the predictability of the objects' movements. In Experiment 3a, we used $\kappa = 20$ (more predictable) and $\kappa = 10$ (more chaotic). In Experiment 3b, we used $\kappa = 10$ (more predictable) and $\kappa = 5$ (more chaotic). Note that the κ value of 10 was used in both experiments as in Experiments 1 and 2, which allowed for a comparison across all experiments.

Design and procedure

Unlike in Experiment 1, we did not use unique trials to conceal the repetition. The experiment consisted of eight training trials and 64 experimental trials, presented in four blocks and separated by eye-tracker calibrations. Each block consisted of eight forward trials and eight backward trials. The predictability of the objects' movements differed in half of the trials (see [Apparatus and stimuli](#)).

The procedure was analogous to that of Experiment 1.

Data analysis

As previously, we limited our analysis to the trials in which the participant recognised all assigned targets. Below, we report the accuracy rates and dropped data for each variant of the experiment.

The direction was sampled every 100 ms, and the object's speed was 5°/s, as in reported experiments. The starting position and the initial direction are indicated by a black dot and arrow, respectively

Experiment 3a (more predictable than baseline). The mean accuracy rate was 88.1 % (range 67.2 % to 100.0 %). We removed data pertaining to three participants with very low accuracies (14 % to 55 %). In the next step, we checked the recorded data for blinks and blink artefacts, and trials with more than 10 % of samples missing were discarded (17.3 %, 3.1 % to 39.1 % per person). One additional participant was removed from the analysis at this step (with 60.9 % trials with missing data). Consequently, the final dataset for Experiment 3a was based on data from $N = 23$ subjects.

For each subject and trajectory, we collected data from four forward and four backward presentations, and we included only trajectories in which at least two forward and two backward presentations were available. After applying the two criteria, the final sample included 1165 trials (79 %; 32 to 62 trials per person).

Experiment 3b (more chaotic than baseline). The mean accuracy rate was 90.2 % (range 70.3 % to 98.4 %). Data from one participant was removed (accuracy 37.5 %). Next, we removed trials with more than 10 % of samples missing (14.2 %, 3.1 % to 32.8 % per person). One additional participant was removed from the analysis at this step (with 46.9 % of trials missing data). Consequently, the final dataset for Experiment 3b was based on data from $N = 28$ subjects. The final sample included 1417 trials (79 %; 34–62 trials per person).

Results

In Experiment 3a, we tested whether the lag between gaze position and the content of the scene varied depending on the predictability of objects' movements. The baseline lag was +95 ms (95 % CI [+80; +110]). In trials with more predictable movements, the lag was shorter (+76 ms with 95 % CI [+59; +94]). However, the lag difference was not significant (difference −19 ms; 95 % CI [−40; +2]). The linear mixed model featuring predictability as a fixed effect did not significantly improve the null model [$\chi^2(1) = 3.04$, $P = .081$]. The local maximum was found in 90.5 % of cases.

In Experiment 3b, the baseline lag was +102 ms (95 % CI [+79; +125]). In trials with more chaotic movements, the lag significantly increased to 134 ms (95 % CI [+118; +150]). The increase by +32 ms (95 % CI [+9; +55]) was statistically significant. The linear mixed model featuring predictability as a fixed effect significantly improved the null model [$\chi^2(1) = 6.57$, $P = .010$]. The local maximum was found in 85.5 % of cases.

In Experiment 3a (more predictable), 12 of 27 participants reported they noticed repetitions (44 %). In the less predictable setting (Experiment 3b), 8 of 30 participants reported repetitions (27 %). These rates did not differ from our previous observations (Experiment 3a: $z = 0.652$, $P = .516$; Experiment 3b: $z = 0.630$, $P = .529$) or from Experiment 1

(Experiment 3a: $z = 1.483$, $P = .139$; Experiment 3b: $z = 0.230$, $P = .818$). The tracking workload did not affect the accuracy [paired t -test, Experiment 3a: $t(22) = 0.694$, $P = .495$; Experiment 3b: $t(27) = 0.584$, $P = .564$] or convergence [Experiment 3a: $t(22) = 0.815$, $P = .424$; Experiment 3b: $t(27) = 0.497$, $P = .623$].

Finally, we pooled the data and compared the baseline lag values across all four experiments (see Fig. 5). We tested a linear mixed model with experiment code as a fixed effect and individual intercepts as random effects. We compared the model to the null model with only random effects. We found no significant effect of experiment [$\chi^2(3) = 1.738$, $P = .629$], which means the observed baseline lag values are comparable across all experiments. In the baseline condition, the gaze position lagged behind the scene content by +103 ms (95 % CI [+94; +112]).

Our results show that gaze models in MOT should incorporate the lag. To support our claim, we compared the gaze data from Experiment 1 with predictions of the centroid strategy. We calculated the predictions for time adjustments of 0 ms, +110 ms (lag) and −110 ms (lead). The predicted values were smoothed with a Gaussian filter and compared with the gaze data using Pearson correlation (analogously to the method described above). With no time adjustment, the mean correlation $r = +.269$ (SD = 0.186). When the predictions were based on positions 110 ms before (lag), the correlation improved [$r = +.271$, SD = 0.185, $t(187) = 2.20$, $P = .029$]. Predictions based on future positions were worse [$r = +.267$, SD = 0.185, $t(187) = 4.09$, $P < .001$].

Discussion

In Experiments 3a and 3b, we manipulated the predictability of objects' movements. We hypothesised that more predictable movements would allow for easier prediction of the scene content and reduce the lag. Similarly, more chaotic movements would increase the lag because the anticipation of the movement will often be wrong, and people will need to be more reactive. Although the results were in line with the hypothesised directions, only the latter effect was significant.

We suggest that the observed lag is already close to the oculomotor limits.¹ Under the current settings, neither the extra cognitive resources nor an easier anticipation of the movements could significantly reduce the lag. On the other hand, it

¹ We ran a small control experiment in which a single target was being tracked. We considered several movement options: (1) same as the objects in the current experiments, (2) same speed but linear, and (3) same as the centroid in current experiments. We opted for the last condition, which produced slow but irregular movements (both in terms of speed and direction). The centroid is a plausible candidate for the gaze in a MOT task; this control experiment required similar eye movements without the workload of tracking additional objects beyond fovea. In this particular setting, the gaze also lagged behind the object ($N = 7$, $M = 57$ ms, $SD = 25$ ms, median 50 ms, range 25–100 ms).

Fig. 5 Lead/lag times across all four experiments. The *solid line* represents zero lag between the gaze position and content of the scene. The *dotted line* represents the 95 % confidence interval for the baseline

condition (tracking four targets, speed 5°/s, $\kappa = 10$) across all four experiments. *White dots* Baseline condition, *black dots* manipulated conditions

is possible to increase the lag by reducing the predictability of the observed scene. We did not test whether tracking more targets could increase the lag in a similar fashion. Although it is possible to track more than four targets, people vary in their tracking capacity (Oksama & Hyönä, 2004), and the increased difficulty would lead to errors and/or guessing.

General discussion

Following the debate on the use of extrapolation in MOT, we measured the extent to which the motion extrapolation is used in participants' gaze behaviour. We found the gaze to lag behind the scene content in every participant. We experimented to determine whether the lag could be affected by changing various aspects of the task, namely, the tracking workload and predictability. For lower tracking workload and more predictable trajectories, the lag decreased as expected, but the difference was not significant. Only the introduction of less predictable (more Brownian-like) trajectories led to significantly longer lags.

A possible source of contradictory evidence in the “location-only” vs. “location and direction” debate results from changes in the MOT task. In the common variant of the MOT task, targets move around the display, and the participant must denote the targets when they stop. To examine the role of extrapolation, the motion was interrupted (Keane & Pylyshyn, 2006), or a secondary task was introduced. Participants were asked to report targets' directions (Horowitz & Cohen, 2010) or the presence of visual probes (Atsma et al., 2012). The traditional MOT task required participants to select targets from a

visible range of options. Successfully reporting the targets' locations in an empty space (Howard & Holcombe, 2008; Howard et al., 2011; Iordanescu et al., 2009) might require different processes (e.g., visual memory). Similarly, related findings (e.g., more precise localisation of crowded objects) might be a result of an attentional strategy, but other effects (visual memory, clustering) might contribute (Lew & Vul, 2015). The experiments that were closest to the traditional MOT paradigm were based on manipulating the predictability of movements (Howe & Holcombe, 2012; Luu & Howe, 2015).

The observed lag does not rule out the possibility of motion extrapolation. People can use the motion extrapolation to adjust their attention (either to individual objects or to their group as whole) separately from their gaze. Attention and eye movements are usually tightly coupled; they can be separated, but doing so requires notable effort and affects performance (Hoffman & Subramaniam, 1995). Our experiment does not address the extent or lack of extrapolation; we only report that the gaze lags behind scene content, which is consistent with the experiments of Hutton and Tegally (2005), who found that a difficult dual task affects smooth pursuit performance (slower velocity, larger position errors).

The variant of the MOT task used in the presented experiments differed in two potentially important aspects. First, we used repeated presentation to improve the signal-to-noise ratio in the comparison of spatiotemporal maps. We asked the participants whether they thought some trials had been repeated. The subjective recognition rate was similar to that in the previous experiment, in which people failed to identify the repeated trials (Lukavský, 2013).

Ogawa, Watanabe and Yagi (2009) showed people incidentally learn during repeated MOT. This contextual cueing could be also reflected in the level of eye movements (better planning, fewer rescue saccades) and potentially lead to shorter lag times. Previously, we found people perform fewer saccades in later blocks in general, and this trend is stronger in repeated trials. Despite this change in eye movements, the spatiotemporal maps did not start to differ over time (Lukavský, 2013).

Second, the targets' directions were sampled from a von Mises distribution regularly at 10 Hz. We used this method to apply smoother control over the predictability of the trajectories. Consequently, the targets followed more Brownian-like motion than in other studies (e.g., Howe & Holcombe, 2012), in which targets followed a direct path and their direction was sampled from the uniform distribution at irregular intervals (after 0.3–0.6 s). Both approaches may differ in the scope of their predictability: the direction along linear paths may be easier to extrapolate, but the uniform distribution can easily cause the targets to make turnarounds. With a von Mises distribution, the objects follow smooth but not-linear curves, and they rarely return to their previous locations (see Fig. 4). We altered the predictability for all objects in the same manner. It is possible to affect only a subset of objects, and the performance will be proportionally decreased. Even less predictable movements solely in distractors impair performance because the collisions become more difficult to predict (Meyerhoff, Papenmeier, Jahn, & Huff, 2016).

The observed lag of 110 ms was difficult to reduce any further. This finding suggests that its value is already close to a limit. We cannot claim the existence of this limit based on our data; a different experiment featuring more predictability levels would be a stronger argument for the existence of lag lower bound limit. The presumed limit is likely caused by temporal integration: visual features are not sampled immediately but integrated over a time interval. The duration of the interval is not known for position information, but 100 ms is a reasonable estimate based on data from similar features (Snowden & Braddick, 1991; Watamaniuk & Sekuler, 1992). Huff and Papenmeier (2013) showed that the integration of object motion and texture motion occurs at intervals as short as 100 ms in MOT. The temporal integration plays two roles in this task. First, the position information is integrated and, thus, the assumed position would be biased towards past locations. Second, the motion information must be integrated to identify the object's direction if the direction is used for later motion extrapolation.

In smooth pursuit research, a latency of approximately 100 ms is observed before smooth movements start. The latency is believed to be composed of a visuomotor

processing delay (70 ms) and a decision-making delay of 30 ms (Wyatt & Pola, 1987). In our experiments, we did not include a separate condition during tracking of a single object (only results from a small-scale experiment are reported in Footnote 1). We did not include this condition, first, because people are known to be fairly accurate in tracking once the initial delay is over. Second, we do not believe the control task is a good one for MOT. People actually track an empty space (centroid) most of the time in MOT, whereas an object is present at the fovea in single object tracking. The presence of an object may facilitate predictions and catch-up saccades to align with the target (de Brouwer, Missal, & Lefèvre, 2001).

We suggest that models of gaze behaviour in MOT should be based on past objects' positions rather than on the current frame. In addition to workload and predictability, there are likely various other parameters affecting the lag, namely, objects' speeds. We did not manipulate the speed in our experiments. We maintained the speed of 5°/s because previous experiments have shown that people move their eye in this setting (Lukavský, 2013). At higher speeds, people tend to fixate on the centre of the display, and track the targets only with their attention. For modelling in a particular setting, it is either possible to assume the same lag of 110 ms or to run a pilot study and measure a better lag estimate given the exact settings.

Author Note This research was supported by Czech Science Foundation (GA 13-23940S) and RVO 68081740; it is part of research programme of Czech Academy of Sciences Strategy AV21.

Appendix

In the common MOT paradigm, mean accuracy is the key measure. However, gaze strategy can be unclear in error trials, because the participant could track the wrong target or give up tracking some targets. Therefore, we needed to analyse only trials with correct answers. We filtered our data with respect to three criteria:

- (1) We evaluated only correct trials.
- (2) We discarded trials with more than 10 % missing samples (blinks, blink artefacts or looking beyond the tracked area).
- (3) Because we compared groups of scan patterns (four vs. four, initially), we included only trajectories where at least two forward and two backward presentations were available.

Table 1 shown below summarises the subject and trial counts following application of the above-mentioned criteria.

Table 1 Breakdown of subject and trial counts with respect to the exclusion criteria

	Experiment 1	Experiment 2	Experiment 3a (predictable)	Experiment 3b (chaotic)
<i>N</i>	21	30	27	30
Trials per subject	60	64	64	64
Non-unique trials	40	64	64	64
<i>N</i> removed due technical issues	1	0	0	0
<i>N</i> removed due low accuracy	0	1	3	1
Accuracy (mean)	91.2 %	92.4 %	88.1 %	90.2 %
Accuracy (range)	72.5–100.0 %	73.4–100.0 %	67.2–100.0 %	70.3–98.4 %
<i>N</i> removed due to blinks	0	2	1	1
Discarded trials (mean)	10.1 %	13.8 %	17.3 %	14.2 %
Discarded trials (range)	0.0–27.5 %	1.6–37.5 %	3.1–39.1 %	3.1–32.8 %
Analysed trials (<i>N</i>)	704	1416	1165	1417
Analysed trials (%)	88 %	82 %	79 %	79 %
Analysed trials (<i>N</i> per subject)	22–40	27–63	32–62	34–62

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. doi:10.1111/j.1467-9280.2005.01587.x
- Atsma, J., Koning, A., & van Lier, R. (2012). Multiple object tracking: Anticipatory attention doesn't "bounce". *Journal of Vision*, 12(1). doi:10.1167/12.13.1
- Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68(3), 255–278.
- Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48. doi:10.18637/jss.v067.i01
- Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*, 10(4), 433–436. doi:10.1163/156856897X00357
- Core Team, R. (2014). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.
- Cornelissen, F. W., Peters, E. M., & Palmer, J. (2002). The eyelink toolbox: Eye tracking with MATLAB and the psychophysics toolbox. *Behavior Research Methods, Instruments, & Computers*, 34(4), 613–617. doi:10.3758/BF03195489
- de Brouwer, S., Missal, M., & Lefèvre, P. (2001). Role of retinal slip in the prediction of target motion during smooth and saccadic pursuit. *Journal of Neurophysiology*, 86(2), 550–558.
- Děchtěrenko, F., Lukavský, J., & Holmqvist, K. (2016). Flipping the stimulus: Effects on scanpath coherence? *Behavior Research Methods*. doi:10.3758/s13428-016-0708-2
- Dorr, M., Martinetz, T., Gegenfurtner, K. R., & Barth, E. (2010). Variability of eye movements when viewing dynamic natural scenes. *Journal of Vision*, 10(10), 28. doi:10.1167/10.10.28
- Fehd, H. M., & Seiffert, A. E. (2008). Eye movements during multiple object tracking: Where do participants look? *Cognition*, 108(1), 201–209. doi:10.1016/j.cognition.2007.11.008
- Fehd, H. M., & Seiffert, A. E. (2010). Looking at the center of the targets helps multiple object tracking. *Journal of Vision*, 10(19). doi:10.1167/10.4.19
- 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. *Attention, Perception, & Psychophysics*, 69(4), 567–577. doi:10.3758/BF03193914
- 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(4), 691–702. doi:10.3758/s13414-011-0265-9
- Hoffman, J. E., & Subramaniam, B. (1995). The role of visual attention in saccadic eye movements. *Perception & Psychophysics*, 57(6), 787–795. doi:10.3758/BF03206794
- Horowitz, T. S., & Cohen, M. A. (2010). Direction information in multiple object tracking is limited by a graded resource. *Attention, Perception, & Psychophysics*, 72(7), 1765–1775. doi:10.3758/APP.72.7.1765
- Horowitz, T. S., Klieger, S. B., Fencsik, D. E., Yang, K. K., Alvarez, G. A., & Wolfe, J. M. (2007). Tracking unique objects. *Perception & Psychophysics*, 69(2), 172–184.
- Howard, C. J., & Holcombe, A. O. (2008). Tracking the changing features of multiple objects: Progressively poorer perceptual precision and progressively greater perceptual lag. *Vision Research*, 48(9), 1164–1180. doi:10.1016/j.visres.2008.01.023
- Howard, C. J., Masom, D., & Holcombe, A. O. (2011). Position representations lag behind targets in multiple object tracking. *Vision Research*, 51(17), 1907–1919. doi:10.1016/j.visres.2011.07.001
- 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(10). doi:10.1167/12.13.10
- Huff, M., & Papenmeier, F. (2013). It is time to integrate: The temporal dynamics of object motion and texture motion integration in multiple object tracking. *Vision Research*, 76, 25–30. doi:10.1016/j.visres.2012.10.001
- Huff, M., Papenmeier, F., Jahn, G., & Hesse, F. W. (2010). Eye movements across viewpoint changes in multiple object tracking. *Visual Cognition*, 18(9), 1368–1391. doi:10.1080/13506285.2010.495878
- Hutton, S. B., & Tegally, D. (2005). The effects of dividing attention on smooth pursuit eye tracking. *Experimental Brain Research*, 163(3), 306–313. doi:10.1007/s00221-004-2171-z
- Iordanescu, L., Grabowecy, M., & Suzuki, S. (2009). Demand-based dynamic distribution of attention and monitoring of velocities during multiple-object tracking. *Journal of Vision*, 9(1). doi:10.1167/9.4.1
- Jahn, G., Papenmeier, F., Meyerhoff, H. S., & Huff, M. (2012). Spatial reference in multiple object tracking. *Experimental Psychology*, 59(3), 163–173. doi:10.1027/1618-3169/a000139

- Jardine, N. L., & Seiffert, A. E. (2011). Tracking objects that move where they are headed. *Attention, Perception, & Psychophysics*, 73, 2168–2179. doi:[10.3758/s13414-011-0169-8](https://doi.org/10.3758/s13414-011-0169-8)
- 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. doi:[10.1016/j.cogpsych.2005.12.001](https://doi.org/10.1016/j.cogpsych.2005.12.001)
- Kleiner, M., Brainard, D. H., & Pelli, D. G. (2007). What's new in Psychtoolbox-3? *Perception 36 ECV Abstract Supplement*.
- Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2016). lmerTest: Tests in linear mixed effects models. R package version 2.0-32. <https://CRAN.R-project.org/package=lmerTest>
- Lew, T. F., & Vul, E. (2015). Ensemble clustering in visual working memory biases location memories and reduces the Weber noise of relative positions. *Journal of Vision*, 15(10). doi:[10.1167/15.4.10](https://doi.org/10.1167/15.4.10)
- Lukavský, J. (2013). Eye movements in repeated multiple object tracking. *Journal of Vision*, 13(7), 1–16. doi:[10.1167/13.7.9](https://doi.org/10.1167/13.7.9)
- Luu, T., & Howe, P. D. L. (2015). Extrapolation occurs in multiple object tracking when eye movements are controlled. *Attention, Perception, & Psychophysics*, 77(6), 1919–1929. doi:[10.3758/s13414-015-0891-8](https://doi.org/10.3758/s13414-015-0891-8)
- Meyerhoff, H. S., Papenmeier, F., Jahn, G., & Huff, M. (2016). Not FLEXible enough: Exploring the temporal dynamics of attentional reallocations with the multiple object tracking paradigm. *Journal of Experimental Psychology: Human Perception and Performance*, 42(6), 776–787. doi:[10.1037/xhp0000187](https://doi.org/10.1037/xhp0000187)
- Meyerhoff, H. S., Papenmeier, F., & Huff, M. (2013). Object-based integration of motion information during attentive tracking. *Perception*, 42(1), 119–121. doi:[10.1068/p7273](https://doi.org/10.1068/p7273)
- Nijhawan, R. (1994). Motion extrapolation in catching. *Nature*, 370(6487), 256–257. doi:[10.1038/370256b0](https://doi.org/10.1038/370256b0)
- Nijhawan, R. (2008). Visual prediction: Psychophysics and neurophysiology of compensation for time delays. *Behavioral and Brain Sciences*, 31(2), 179–239. doi:[10.1017/S0140525X08003804](https://doi.org/10.1017/S0140525X08003804)
- Ogawa, H., Watanabe, K., & Yagi, A. (2009). Contextual cueing in multiple object tracking. *Visual Cognition*, 17(8), 1244–1258. doi:[10.1080/13506280802457176](https://doi.org/10.1080/13506280802457176)
- Oksama, L., & Hyönä, J. (2004). Is multiple object tracking carried out automatically by an early vision mechanism independent of higher-order cognition? An individual difference approach. *Visual Cognition*, 11(5), 631–671. doi:[10.1080/13506280344000473](https://doi.org/10.1080/13506280344000473)
- Papenmeier, F., & Huff, M. (2010). DynAOI: A tool for matching eye-movement data with dynamic areas of interest in animations and movies. *Behavior Research Methods*, 42(1), 179–187. doi:[10.3758/BRM.42.1.179](https://doi.org/10.3758/BRM.42.1.179)
- Papenmeier, F., Meyerhoff, H. S., Jahn, G., & Huff, M. (2014). Tracking by location and features: Object correspondence across spatiotemporal discontinuities during multiple object tracking. *Journal of Experimental Psychology: Human Perception and Performance*, 40(1), 159–171. doi:[10.1037/a0033117](https://doi.org/10.1037/a0033117)
- Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, 10(4), 437–442. doi:[10.1163/156856897X00366](https://doi.org/10.1163/156856897X00366)
- Pylyshyn, Z. W., & Storm, R. W. (1988). Tracking multiple independent targets: Evidence for a parallel tracking mechanism. *Spatial Vision*, 3(3), 1–19. doi:[10.1163/156856888X00122](https://doi.org/10.1163/156856888X00122)
- 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* (pp. 49–77). Boston, MA: MIT Press.
- Scholl, B. J., & Pylyshyn, Z. W. (1999). Tracking multiple items through occlusion: Clues to visual objecthood. *Cognitive Psychology*, 38(2), 259–290. doi:[10.1006/cogp.1998.0698](https://doi.org/10.1006/cogp.1998.0698)
- Smith, T. J., & Mital, P. K. (2013). Attentional synchrony and the influence of viewing task on gaze behavior in static and dynamic scenes. *Journal of Vision*, 13(8), 16. doi:[10.1167/13.8.16](https://doi.org/10.1167/13.8.16)
- Snowden, R. J., & Braddick, O. J. (1991). The temporal integration and resolution of velocity signals. *Vision Research*, 31(5), 907–914. doi:[10.1016/0042-6989\(91\)90156-Y](https://doi.org/10.1016/0042-6989(91)90156-Y)
- St. Clair, R., Huff, M., & Seiffert, A. E. (2010). Conflicting motion information impairs multiple object tracking. *Journal of Vision*, 10(4), 18.1–18.13. doi:[10.1167/10.4.18](https://doi.org/10.1167/10.4.18)
- Watamaniuk, S. N. J., & Sekuler, R. (1992). Temporal and spatial integration in dynamic random-dot stimuli. *Vision Research*, 32(12), 2341–2347. doi:[10.1016/0042-6989\(92\)90097-3](https://doi.org/10.1016/0042-6989(92)90097-3)
- Wyatt, H. J., & Pola, J. (1987). Smooth eye movements with step-ramp stimuli: The influence of attention and stimulus extent. *Vision Research*, 27(9), 1565–1580. doi:[10.1016/0042-6989\(87\)90165-9](https://doi.org/10.1016/0042-6989(87)90165-9)
- Zelinsky, G. J., & Neider, M. B. (2008). An eye movement analysis of multiple object tracking in a realistic environment. *Visual Cognition*, 16(5), 553–566. doi:[10.1080/13506280802000752](https://doi.org/10.1080/13506280802000752)
- Zelinsky, G. J., & Todor, A. (2010). The role of “rescue saccades” in tracking objects through occlusions. *Journal of Vision*, 10(14), 29. doi:[10.1167/10.14.29](https://doi.org/10.1167/10.14.29)