Cognition 146 (2016) 393–409

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/COGNIT

Position tracking and identity tracking are separate systems: Evidence from eye movements

Lauri Oksama [a,*], Jukka Hyönä [b]

[a] National Defence University, Finland
[b] University of Turku, Finland

ARTICLE INFO

Article history:
Received 28 June 2013
Revised 2 October 2015
Accepted 19 October 2015
Available online 9 November 2015

Keywords:
Multiple identity tracking
Multiple object tracking
Eye movements
Visual attention
Visual scanning

ABSTRACT

How do we track multiple moving objects in our visual environment? Some investigators argue that tracking is based on a parallel mechanism (e.g., Cavanagh & Alvarez, 2005; Pylyshyn, 1989), others argue that tracking contains a serial component (e.g. Holcombe & Chen, 2013; Oksama & Hyönä, 2008). In the present study, we put previous theories into a direct test by registering observers' eye movements when they tracked identical moving targets (the MOT task) or when they tracked distinct object identities (the MIT task). The eye movement technique is a useful tool to study whether overt focal attention is exploited during tracking. We found a qualitative difference between these tasks in terms of eye movements. When the participants tracked only position information (MOT), the observers had a clear preference for keeping their eyes fixed for a rather long time on the same screen position. In contrast, active eye behavior was observed when the observers tracked the identities of moving objects (MIT). The participants updated over four target identities with overt attention shifts. These data suggest that there are two separate systems involved in multiple object tracking. The position tracking system keeps track of the positions of the moving targets in parallel without the need of overt attention shifts in the form of eye movements. On the other hand, the identity tracking system maintains identity–location bindings in a serial fashion by utilizing overt attention shifts.

1. Introduction

Keeping track of multiple moving objects is a central part of our everyday life. For example, a mother may be tracking the whereabouts of her children on a crowded beach, or a car driver approaching a busy intersection is monitoring other vehicles also manoeuvring through the intersection. Moreover, professionals, such as air traffic controllers and fighter pilots, constantly deal with similar dynamic visual environments. However, the demands of different tracking tasks may vary quite notably from each other. Sometimes it is sufficient that we are simply aware of the members of the target set as a whole, for example, when a soccer player is attending to the whereabouts of the opponent team members. Other times it is required that we are aware of the whereabouts of individual members of the target set, for example when a soccer player wishes to pass the ball to his team's top scorer.

In the research literature on tracking of moving objects, the former task bears similarity to the multiple object tracking (MOT) task (Pylyshyn & Storm, 1988), where observers track a set of targets that are visually identical to each other (likened to the tracking of a flock of white sheep). Thus, only location information needs to be encoded, whereas object features are irrelevant to the task (see Fig. 1, top). On the other hand, the latter task is similar to the multiple identity tracking (MIT) task (Horowitz et al., 2007; Oksama & Hyönä, 2004), where distinct objects (likened to individual soccer players) are tracked and where observers need to constantly bind and update identity information with location information (see Fig. 1, bottom). Thus, the MOT task is a position tracking task whereas the MIT task is an identity tracking task by nature.

Several theoretical controversies have emerged concerning the mechanisms of position and identity tracking. Firstly, is tracking achieved by a serial or by a parallel process? Secondly, do position tracking and identity tracking share a common mechanism or are they based on independent mechanisms?

1.1. Is multiple object tracking serial or parallel in nature?

Some investigators argue that tracking is based on a parallel mechanism (Alvarez & Cavanagh, 2005; Cavanagh & Alvarez,

---

* Corresponding author at: National Defence University, P.O. Box 5, FI-04401 Järvenpää, Finland.

E-mail address: loksama@utu.fi (L. Oksama).

http://dx.doi.org/10.1016/j.cognition.2015.10.016
0010-0277/© 2015 The Authors. Published by Elsevier B.V.

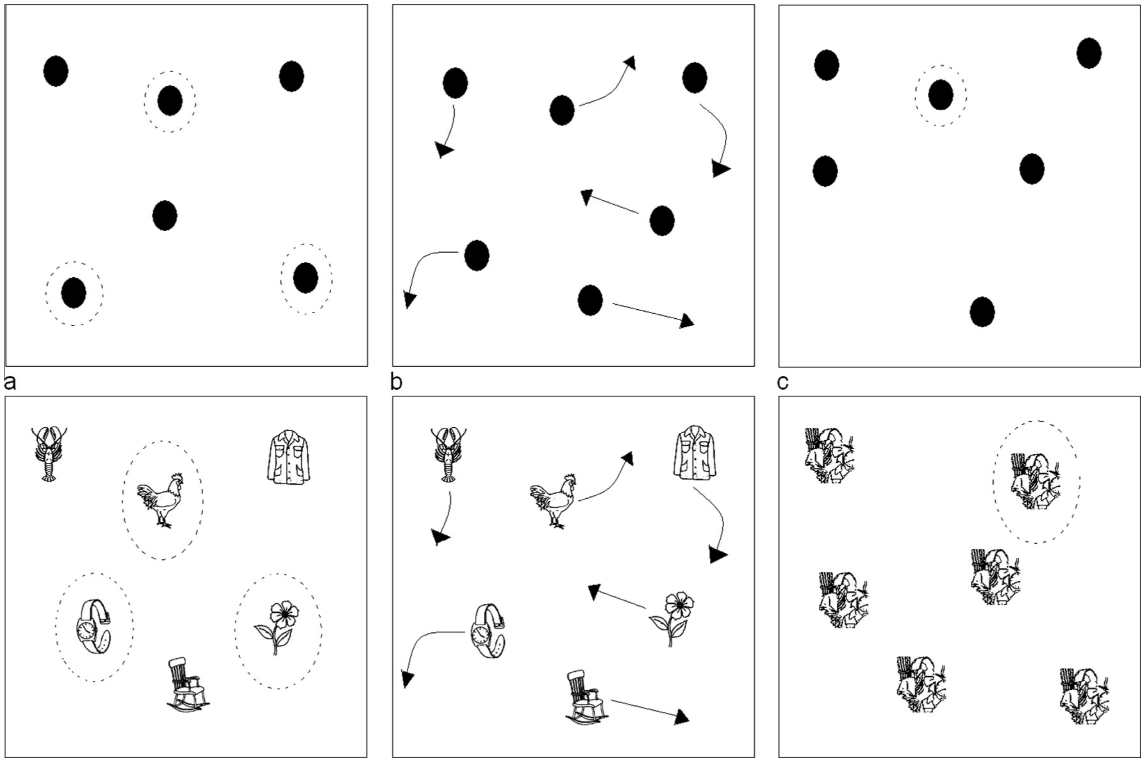

Fig. 1. A schematic depiction of the multiple object tracking task (MOT) with identical objects (top) and the multiple-identity tracking task (MIT, bottom) with distinct objects. Display a: six different objects are presented, and in this trial three of them are designated as targets by flashing a frame around them. Display b: all objects begin to move randomly about the screen. The participant's task is to track the location of the designated targets. Display c: when the motion stops, the participant is asked whether a flashed probe was among the target set flashed at the outset (MOT) or to report the identity of the masked probe (MIT). The target pictures of objects used in MIT were reprinted from Snodgrass and Vanderwart (1980), © 2007 Life Sciences Associates.

2005; Franconeri, Jonathan, & Scimeca, 2010; Howe, Cohen, Pinto, & Horowitz, 2010; Kazanovich & Borisyuk, 2006; Pylyshyn, 1989, 2001), others argue that tracking contains a significant serial component (d'Avossa, Shulman, Snyder, & Corbetta, 2006; Holcombe & Chen, 2013; Oksama & Hyönä, 2004, 2008; Tripathy, Ogmen, & Narasimhan, 2011). Parallel theories are typically based on data collected using the MOT paradigm (but see Howe & Ferguson, 2015), where observers track visually identical targets. According to the FINST theory (Pylyshyn & Storm, 1988), tracking is carried out in parallel for all targets within the capacity limit of about four items. Moreover, the tracking mechanism is assumed to operate pre-attentatively. According to the theory of Cavanagh and Alvarez (2005), tracking requires attention and is based on multiple attentional foci, between which limited attentional resources are allocated. Both versions of the parallel theory assume that serial switching of visual attention between target objects is not needed, either because attention is not needed or because object tracking is parallel in nature. Finally, Alvarez and Franconeri (2007) have proposed a model in which tracking is achieved by a flexibly allocated mental resource; however, they refrain from taking stand whether this resource is serial or parallel in nature.

Oksama and Hyönä (2008) have proposed a serial model especially designed for multiple identity tracking. According to their MOMIT model, observers use only one attentional focus, which needs to be shifted serially from one target to the next. When visual attention is focused on a target, its identity–location binding is created (or updated). In other words, binding identity with location is carried out individually for each target. As other, non-attended targets keep moving, it means that their bindings will be outdated and will not be updated until they are focally attended one at a time. It is further assumed that locations for the tracked targets are temporarily stored in visuo-spatial short-term memory (VSTM). This indexed location information (bound to identities) is then utilized by a mechanism that programs shifts of visual attention between targets. As targets move continuously, location information for all other than the focally attended targets are outdated. The magnitude of this location error is a key factor in predicting tracking accuracy as a function of object speed and target set-size. The size of the location error increases with an increase in target speed and set-size, which results in less efficient switching of attention between targets. Furthermore, it is assumed that serial shifting of attention is controlled partly with the help of this indexed location information stored in VSTM and partly with the help of peripheral vision. According to MOMIT, peripheral vision provides non-indexed (not bound to identities) location information about all moving objects in parallel.

Engineering models describe a generic model for visual sampling in dynamic situations. They are kin to MOMIT, as they are based on serial switching of focal attention. Seminal engineering models of supervisory control provide precise predictions about how often a dynamic display (e.g., flight instruments in a cockpit) should be sampled with focal attention and the eyes (e.g., Carbonell, 1966; Moray, 1984, 1986; Senders, 1964, 1983; Sheridan, 1970; see also Horrey, Wickens, & Consalus, 2006). According to Senders' (1964) model, in order to effectively monitor a dynamic display, it is necessary to sample the display at a rate twice its information bandwidth (bandwidth measured in events/s = Hz). For example, when relevant information in an information channel occurs at a rate of 1 Hz, the optimal observer should visually sample this channel at a rate of 2 Hz. Later Moray (1986) argued that the optimal sampling rate would be equal to the bandwidth. Sampling of dynamic display at this rate is necessary, because the more rapidly the dynamic signal varies, the more quickly it will become impossible to predict its current value on the basis of past observations. In other words, uncertainty about the channel's current state increases as a function of time. Uncertainty may increase either due to endogenous reasons like forgetting or due to exogenous reasons like signal bandwidth. The job of the optimal observer is to reduce uncertainty by switching attention between relevant channels at an optimal rate.

Applied to MOT and MIT, information bandwidth refers to the frequency with which the target objects change their movement direction or their speed. When the targets continue moving in the same direction with the same speed, uncertainty is low and thus the need for sampling the visual environment is also small. In contrast, when the targets change their direction and/or speed, uncertainty increases and so does the need for increasing the visual sampling rate. It should be noted that object speed itself is not relevant, only the possible changes in speed.

Furthermore, Moray (1984) argues that observers are likely to face a speed–accuracy trade-off in monitoring multiple information channels. When the number of channels increases and/or the information bandwidth (events/s) becomes high, the observer either has to delay sampling in order to have enough observation time or s/he must cut short the observation time of a display channel to maintain a desired sampling rate. Such a trade-off will result in performance decrement and in important channel information either being unnoticed or detected too late (e.g., the pilot does not notice a critical value at a flight instrument or it is noticed too late). However, to our knowledge, to date no experiments have been carried out to test the existence of such trade-offs.

As noted above, according to the serial engineering models, the optimal observer aims to minimize uncertainty by periodically sampling the channels. If uncertainty is interpreted as location error, MOMIT and the engineering models are very similar and make similar predictions. Both models assume an increase in the number of targets to also increase the visual sampling rate. Similarly, both models predict variation in speed to influence the sampling rate, albeit for slightly different reasons. As noted above, engineering models do not assume speed as such to affect the sampling rate. However, increase in object speed in MOT and MIT, is associated with more frequent changes in motion direction due to collision prevention.

1.2. Do position tracking and identity tracking share a common mechanism?

Another theoretical controversy in tracking research concerns whether position tracking and identity tracking share a common resource. Pylyshyn (1989) proposed that a preattentive feature-blind position tracking mechanism (FINST) works independently of a feature/identity checking process. The object file theory (Kahneman, Treisman, & Gibbs, 1992), on the other hand, suggests that focusing on an object creates a file that contains location as well as featural or identity information (see also a later model of Pylyshyn, 2004). Thus, featural/identity information comes with location information. More recently, however, Horowitz et al. (2007) have proposed that there may be two separate and independent systems involved in multiple object tracking, one responsible for tracking positional information and one also carrying identity information. Their proposal is based on finding capacity differences between a task only requiring to keep track of target positions versus a task where target identities are also maintained. Horowitz et al. observed a "content deficit": tracking capacity was lower when observers needed to report the location of a particular target rather than the locations of all targets (see also Botterill, Allen, & McGeorge, 2011; Ren, Chen, Liu, & Fu, 2009). Recently, however, Cohen, Pinto, Howe, and Horowitz (2011) and Pinto, Howe, Cohen, and Horowitz (2010) have provided evidence in favor of a unitary tracking model. Cohen et al. demonstrate that observers can trade off performance between location and identity tracking; that is, position and identity tracking may not be entirely dissociable.

It is difficult to differentiate between serial and parallel processing or whether location tracking and identity tracking are carried out by separate systems on the basis of off-line performance accuracy and reaction time measures. In fact, Logan (2002) and Townsend and Wenger (2004a) demonstrated that it is in principle possible to simulate the same behavioral data (e.g., reaction times) using either a serial or parallel model. However, studying tracking in real time via eye movements is likely to be more informative. The eye movement technique is a useful tool to study whether overt focal attention is exploited during tracking. Moment-to-moment allocation of visual attention can be reliably studied by eye movement registration, as shifts of visual attention and eye movements are intimately linked to each other (e.g., Deubel & Schneider, 1996; Hoffman & Subramaniam, 1995; Kowler, Anderson, Dosher, & Blaser, 1995; Shepherd, Findlay, & Hockey, 1986; see also Findlay & Gilchrist, 2003). If position tracking and identity tracking are based on a common mechanism, we expect to find similar eye-movement activity in both tasks. However, if eye-movement activity qualitatively differs between the position tracking and identity tracking tasks, it would be evidence for different mechanisms being responsible for position versus identity tracking.

1.3. Eye-movement studies of MOT

Recent eye movement studies of multiple object tracking provide some support for parallel models (but see below Landry, Sheridan, & Yufik, 2001). Zelinsky and Neider (2008) registered eye movements while observers tracked visually identical sharks moving about on a computer screen. In order to study typical eye movement strategies during MOT, they assigned, using a shortest distance rule, each eye position sample to one of three possible regions: centroid (the geometric center delineated by the moving target objects), target object, or distracter object. With set-size 2, observers looked longest at the centroid, with set-size 3 they looked equally long at the centroid and target, and with set-size 4 they looked longest at the target. Tracking accuracy correlated positively and strongly with fixation time on the centroid but negatively with fixation time on the target, which suggests that the centroid strategy is an optimal eye movement strategy in MOT (see also Fehd & Seiffert, 2010).

Fehd and Seiffert (2008) recorded observers' eye movements when they tracked either one or three identical red dots for 3 s. The analysis of the eye movement data was limited to the trials where all targets were correctly identified. The fixations were grouped either to the centroid or the target or distractor dots using a shortest distance rule, similarly to Zelinsky and Neider (2008). With only one target to be tracked, the eyes fixated almost exclusively on the target. On the other hand, with set-size 3 fixations were predominantly located close to the centroid. In a follow-up experiment, this was also found to be the case for set-size 4 and 5. Fehd and Seiffert (2008, 2010) concluded that observers can accurately maintain a mental representation of object locations without constant foveation. Their data are also consistent with the view that multiple identical items may be tracked in parallel (by perhaps grouping them into single virtual object, see Yantis, 1992) without overtly attending to the individual targets. Finally, Huff, Papenmeier, Jahn, and Hesse (2010) measured eye movements to study tracking of three moving targets and obtained results that were interpreted to be consistent with those of Fehd and Seiffert.

On the other hand, Landry et al. (2001) found evidence for serial processing during tracking. They used a simulated air traffic control task in which participants monitored multiple aircraft for collisions. They found evidence for switching and eye movements between targets. Notice that their air traffic control task resembles very closely the MIT task, as in both tasks distinct objects has to be monitored.

To sum up the above discussion, the available evidence obtained by eye movement studies is consistent with the view that in MOT (identical targets) tracking is achieved using a parallel mechanism without the need for eye movements made on the targets, whereas in MIT the process is based on serial switching (and eye movements).

1.4. Present study

In the present study, we put the above theoretical reasoning into direct test by registering observers' eye movements when they tracked identical (MOT) or distinct (MIT) objects. To examine the likelihood of serial switching of attention between targets, we measured the number of target visits and the number of fixated targets as a function of target set-size and speed. Fixation frequency was used as an additional measure of attentional switching. We also tested whether the average duration of target visits would vary in response to an increase in target set-size and speed. Third, based on eye fixations on targets we calculated the visual sampling rate in Hz and compared it to the bandwidth of the tracking task (i.e., number of events per second, operationalized as changes in movement direction of target objects). Finally, attentional load was measured by pupil size and blink rate; increased load is associated with larger pupil size (e.g., Beatty, 1982; Hyönä, Tommola, & Alaja, 1995) and a smaller number of eyeblinks (e.g., Kramer, 1991; Sirevaag, Kramer, Reisweber, Strayer, & Grenell, 1993; Wilson & Eggemeier, 1991; for a review, see also Irwin & Thomas, 2010). There is also evidence (Thomas & Irwin, 2006; see also Irwin & Thomas, 2010) demonstrating that memory for spatial position as well as for identity–location bindings suffers when an eyeblink occurs.

The present eye movement study differed from the seminal work of Zelinsky and Neider (2008) and Fehd and Seiffert (2008) in several important respects. First, we compared eye behavior during two different tracking tasks (MOT and MIT). Second, we made use of a number of different eye movement measures to tap in more detail into the online tracking performance. Third, unlike Zelinsky and Neider (2008) and Fehd and Seiffert (2008), we did not apply a shortest distance rule; instead fixations were assigned to the different areas of interest only when they actually fell on those regions (for a similar analysis procedure, see also Huff et al., 2010).

1.5. Predictions regarding eye-movements

As argued above, parallel theories predict that the number of eye movements traversing between targets is minimal. Instead, parallel tracking of location information may be optimized by keeping the eyes fixed on a center position delineated by the moving target objects (Fehd & Seiffert, 2008; Zelinsky & Neider, 2008). However, a constant computation and updating of the centroid might be resource demanding, particularly with larger set-sizes (Zelinsky & Neider, 2008), as it takes away limited resources from the actual tracking. Thus, another possibility is to keep the eyes fixed on one location, while the parallel processors (e.g., visual indexes) take care of the tracking of moving objects. Different versions of the parallel theory (preattentative vs. attentative) make the same predictions in terms of eye fixation patterns. However, they differ from each other with respect to the two attentional load measures employed in the present study, pupil size and blink rate. Attentional models (such as the FLEX model by Alvarez & Franconeri, 2007, interpreted as a covert parallel model) predict that an increase in target set-size and speed should result in increased attentional load, which in turn should be reflected in pupil size and blink rate. However, preattentative models, such as FINST (Pylyshyn, 1989), do not make such a prediction. According to single mechanism accounts, the predictions outlined above would hold for both MOT and MIT. However, on the basis of the dual-mechanism account of Horowitz et al. (2007), the above predictions would only hold for MOT.

On the other hand, the serial MOMIT model (Oksama & Hyönä, 2008) makes the following predictions regarding eye movement behavior during MIT. First, number of fixations in general and number of target visits especially increase linearly as a function of set-size within the capacity limit (average capacity of MIT is estimated to be around 3.5–4 items, although there are significant individual differences, see Oksama & Hyönä, 2004). Thus, the increase in fixation frequency is assumed to be monotonic up to four targets, after which it should asymptote. On the other hand, in order to optimize serial updating and tracking, previously created identity–location bindings ought not to become grossly outdated. Thus, it is assumed that the average duration of target visits becomes shorter with increased set-size and speed. However, due to low-level oculomotor constraints, fixation durations cannot decrease ad infinitum but are likely to asymptote to some minimum level. Second, number of fixations made during tracking is also predicted to vary as a function of target speed. When objects move faster, their previously created identity–location bindings become quicker outdated. Thus, the visual sampling rate (i.e., fixation frequency) should increase, but the number of successful target visits should decrease as a function of object speed. On the other hand, average duration of target visits should become shorter with increased object speed as a result of efforts in updating identity–location bindings at a maximal rate. Finally, attentional load, as indexed by pupil size and blink rate, is assumed to increase both as a function of set-size and speed.

The above predictions derived from MOMIT apply to MIT for which the model was built. Although the model is not designed to simulate MOT performance, it is in principle also possible to extend it to MOT. The limited resources FLEX model of Alvarez and Franconeri (2007) interpreted as an overt serial model (with one tracking mechanism, 'flex', in use) would predict a flexible allocation of eye movements in different speed and set-size conditions in both tasks in a similar way as MOMIT. A flexible limited resource system with one 'flex' should increase the number of fixations but decrease the fixation duration on targets as a response to increases in tracking load.

Finally, the engineering models (for a recent review, see Horrey et al., 2006) of visual sampling make similar predictions as MOMIT (Oksama & Hyönä, 2008) regarding the effect of set-size and speed and the trade-off between number of fixations and average fixation duration.

2. Experiment 1: MOT

In the MOT task, the participants tracked 2–5 identical moving objects while their eye movements were registered. Target set-size and speed were manipulated. To briefly recap the previous discussion, parallel models do not stipulate any relationship between eye-movements and tracking. On the other hand, previous eye movement studies of MOT predict that observers keep their eyes fixed in a position, defined by the gravitational center of the to-be-tracked objects, irrespective of set-size and speed (Fehd & Seiffert, 2008, 2010). Moreover, the limited-resource parallel model (e.g., Alvarez & Franconeri, 2007; Cavanagh & Alvarez, 2005) predicts an increase in attentional load in response to increases in set-size and speed, whereas the preattentative parallel model (Pylyshyn, 1989) predicts no effect of attentional load as a function of set-size (up to the capacity limit of 4 items) and speed. On the other hand, serial models (MOMIT and engineering models) predict a linear increase in the number of fixations and number of target visits as a function of set-size and speed. If target locations are serially refreshed, increases in both the target set-size and the target speed would raise the need for more frequent refreshing of target locations. As for MOMIT, this is because an increase in both set-size and speed would render the serially updated location information increasingly outdated. Regarding engineering models, an increase in set-size directly increases uncertainty and sampling rate, while an increase in object speed does so indirectly by increasing changes in movement direction (see above). Finally, as it was possible to estimate the information bandwidth of the tracking task (we used off-line movement trajectories), we were able to test engineering models' predictions concerning the optimal sampling rate. The objects changed their direction and speed on average of 2.75 Hz. The optimal visual sampling rate of targets will then be either 5.5 Hz (Senders, 1964) or 2.75 Hz (Moray, 1986).

2.1. Method

2.1.1. Participants

Eleven participants (psychology students of the University of Turku) were recruited for the experiment (two males, nine females, their mean age was about 21). All participants had normal or corrected-to-normal vision. Previous eye movement studies of MOT used quite comparable sample sizes.

2.1.2. Apparatus

Eye movements were recorded with a desktop mounted Eyelink2000 (SR Research Ltd., Ontario, Canada) system. Sampling frequency was 1000 Hz. The stimuli were presented on a 21″ CRT screen with a screen resolution of 1280 × 1024 pixels and a 100-Hz refresh rate. Participants were seated 57 cm from the screen, and a chin rest was used to stabilize the head. Stimuli were created with the E-prime software (Schneider, Eschman, & Zuccolotto, 2002a, 2002b). The software that generated the motion sequences was written in Visual Basic.

2.1.3. Stimuli

The stimuli consisted of ten identical solid black circles, 2.1 deg in diameter, presented on a white background (25 deg horizontally and 32 deg vertically). Two to five of the total of ten dots was designated as targets by flashing a black frame (2.2 × 2.2 deg) around them. Initial dot positions were generated at random. Dots never overlapped with each other or with the frames of the screen. The motion was created from 167, 233, or 300 static frames by presenting them one after another for 30 ms each. Movement direction for each object was chosen randomly from among the eight compass directions. Each object was assigned a movement duration that was randomly selected from 7 to 37 in 30 ms increments (210–1110 ms), and speed, randomly selected from 1 to 5.7 pixels per frame in the slow condition (the average speed was 2.6 deg/s); from 3 to 12.7 pixels per frame in the medium condition (the average speed was 6.3 deg/s); or from 7 to 18.4 pixels per frame in the fast condition (the average speed was 10.7 deg/s). The range of speeds used here was similar to that of Oksama and Hyönä (2008). The movement duration determined the time for how long the object maintained a certain direction and speed. When the movement duration expired, new random speed, direction and duration values were assigned to the object. Collisions to the other dots and the edges of the display were prevented. If a collision was going to happen, a reverse direction was chosen to these objects. Thus, objects could not intersect each other. This procedure yielded a sequence of frames in which each element moved in a random, independent and continuous way for some period of time (210–1110 ms, or until a collision was about to happen), and then changed direction and speed abruptly and began to move in a new direction. As the coordinates for the moving objects were generated in a random fashion, the movement patterns of the objects covered evenly the entire screen area.

2.1.4. Eye movement analysis

The eye movement data were first parsed into fixations and saccades, after which fixations were assigned to one of five areas of interest (target, distractor, centroid, screen center, or other area). We considered it justified to use eye fixations as the primary data, instead of assigning each eye movement data sample to an area of interest regardless whether it belongs to a fixation or saccade (Fehd & Seiffert, 2008; Zelinsky & Neider, 2008), as the intake of visual information takes place during fixations. Fixation analysis provides a real-time protocol of how observers allocate their visual attention from moment to moment during tracking. Target and distractor objects changed location every 30 ms. Thus, based on their x and y coordinates, each fixation was assigned to an area of interest separately for every 30-ms time slice. As the objects moved during fixations, each fixation was subject to be assigned to more than one area of interest. In order to prevent that from happening, the fixations were assigned to the area of interest within which it was located for more than 50% of its duration. Fixations shorter than 80 ms were excluded. The target area for the moving objects, the centroid, and the screen center was set to 3.4 degrees of visual angle in diameter. It was defined so that different target areas would not overlap with each other. The diameter was only somewhat bigger than the actual diameter of the moving objects.

The centroid was defined as the center of mass of a polygon with the targets as it vertices. In target set-size two, the centroid was the center of the two end points; in set-size three and four, centroid was the center of mass (the centroid of a non-self-intersecting closed polygon). In set-size five, the centroid cannot be defined unambiguously (there could be many different centroids depending on the order of vertices); thus, the results for set-size 5 should be considered with some caution. In set-size five, a geometrical 'average', or the center of two points was determined.

2.1.5. Task

At the beginning of each trial the targets (2–5 objects) were designated by flashing a frame around them. The participants' task was to track the designated targets during the movement phase. At the end of the movement phase, all the objects stopped moving and the probed element was highlighted by flashing a frame around it. Finally, the screen was cleared and a response screen appeared where participants were asked to respond as accurately and quickly as possible whether the probed element was one of the target items. A computer mouse was used to collect the responses (participants were asked to point and click either a framed 'yes' or 'no' picture).

2.1.6. Procedure

A chinrest was used to reduce head movements and to control the viewing distance. Participants were given written instructions prior to the experiment, which outlined the general procedure and explained the trial sequence. They were to note the positions of the flashing targets at the start of each trial and to keep track of them during the movement phase. They were free to move or not to move their eyes during tracking. At the beginning of each block the eye-tracker was calibrated using a 9-point calibration. Drift correction was done after every trial. At the beginning of each trial, the objects were displayed for 1 s. After that, a black frame flashed on and off for ten times (flashing duration was 150 ms; total flashing time was 3000 ms) around the designated targets. The objects then began to move in a random and continuous fashion around the screen. The participants tracked them for 5, 7, or 9 s, after which the movement stopped. This was followed by the flashing of a black frame for five times (flashing duration was 150 ms; total flashing time was 1500 ms) around one of the targets (i.e., the probed item). After a response was given to the probe, the response screen was cleared and an inter-trial screen was presented. The next trial was initiated by the participant pressing the space bar, or after the maximum inter-trial interval (3000 ms) expired. Participants were provided with 10 practice trials; feedback was given after each response during the practice session. Each participant completed two blocks of 72 trials, altogether 144 trials. The order of trials was randomized separately for each participant within each block. The order of blocks was counterbalanced across participants. There was a short rest period between the blocks. The entire session took about 60 min.

2.1.7. Design

There were two manipulated factors in the experiment: number of targets (2–5) and object speed (average speed of 2.6 deg/s, 6.3 deg/s, or 10.7 deg/s; coined ‘slow’, ‘medium’, and ‘fast’). Both factors were within-subject variables. There were 12 trials in each of the 12 conditions.

2.2. Results

2.2.1. Tracking accuracy (error percentage)

The results for the repeated measures analyses of variance (ANOVA) are presented in Table 1. The results showed that tracking performance deteriorated as a function of set-size and speed (see Fig. 2, left panel). The Set-Size × Speed interaction was not significant. Overall tracking accuracies correspond very closely with those obtained by Oksama and Hyönä (2004). We recomputed the analysis for square-root transformed data, but the results remained the same.

2.2.2. Eye-tracking data

The results for the repeated measures analyses of variance (ANOVA) are presented in Table 1.

2.2.2.1. Number of fixations. As evident from Fig. 3 (left panel), the number of fixations made during MOT did not vary as a function of set-size. On the other hand, number of fixations decreased when object speed increased. The significant interaction between set-size and speed reflects the fact that in the slow speed condition the number of fixations increased with set-size, in the medium speed condition it remained constant, while in the high speed condition it decreased slightly.

2.2.2.2. Number of successful target visits. As apparent from Fig. 4 (left panel), the number of visits to target objects remained constant across the different set-sizes, but decreased with increased object speed. The significant Set-Size × Speed interaction is due to the fact that in the slow speed condition the number of target visits increased as a function of set-size, in the medium speed condition remained at a constant level, while in the high speed condition it decreased slightly.

2.2.2.3. Number of updated targets. For this measure, each target visit was regarded as an attempt to update the target location. We then computed how many targets were updated (i.e., visited at least once). As shown in Fig. 5, participants updated on average one and a half target regardless of set-size. The number of updated targets diminished slightly as a function of speed.

2.2.2.4. Average fixation duration. As apparent from Fig. 6 (left panel), average fixation duration did not vary as a function of set-size. It may be noted that the average fixation duration was quite long (around 600 ms). The main effect of target speed approached significance and so did the Set-Size × Speed interaction. There was a tendency for average fixation duration to decrease as a function of set-size in the slow speed condition, to remain constant in the middle speed condition, and to increase in the high speed condition.

2.2.2.5. Pupil size. Pupil size increased both as a function of set-size and speed (Fig. 7, left panel). Their interaction was not significant.

2.2.2.6. Number of blinks. Number of blinks did not vary as a function of set-size but did so as a function of speed (Fig. 8, left panel). Blink rate decreased as the speed increased. The interaction between set-size and speed was non-significant.

2.2.2.7. Visual sampling rate. The participants sampled the visual display on average at the rate of 2.2 Hz; in other words, they made a fixation on average after every 445 ms. Interestingly, however, they made a visit to a target only after every 1853 ms, i.e., at the rate of 0.5 Hz. The optimal rate is estimated to be either 2.75 (Moray, 1986) or 5.5 Hz (Senders, 1964).

2.2.2.8. Average percentage of trial time spent looking at different areas of interest. As apparent from Fig. 9, participants spent most of their trial time (on average 48.2%) looking at some position on the background of the screen that was neither the centroid nor the screen center. They spent much less time looking at the targets (on average 21.0%) and, surprisingly short time looking at the centroid (on average 7.0%), the screen center (on average 3.7%), or distractors (on average 4.0%). What is also apparent from Fig. 9 is that the faster the targets moved, the less time the participants spent looking on the target objects and the more on the background.

Table 1

Results of the repeated measures ANOVA (significant *p* values appear in bold) for the tracking accuracy and the eye measures obtained during the MOT performance (Experiment 1).

	Set-size				Speed				Set-Size × Speed			
	F(3,30)	*p*	MSE	η_p^2	*F*(2,20)	*p*	MSE	η_p^2	*F*(6,60)	*p*	MSE	η_p^2
Accuracy (% error)	4.86	**<.01**	20.90	.33	9.45	**<.01**	50.30	.49	2.22	.13	98.20	.18
Number of fixations	<1	.89	3.24	.02	7.59	**<.05**	13.92	.43	7.29	**<.001**	.99	.42
Number of target visits	<1	.60	.60	.06	19.18	**<.01**	4.51	.66	3.67	**<.01**	.48	.27
Number of updated targets	4.03	.07	.77	.29	21.33	**<.001**	.05	.68	1.89	.10	.04	.16
Pupil size	9.79	**<.01**	23,092	.50	24.75	**<.001**	15,713	.71	1.59	.21	6437	.14
Blinks	1.23	.31	.59	.11	15.02	**<.001**	.09	.60	2.57	.09	.11	.21
Average fixation duration	<1	.58	62,763	.05	4.37	.06	98,266	.30	3.55	.07	57,291	.26

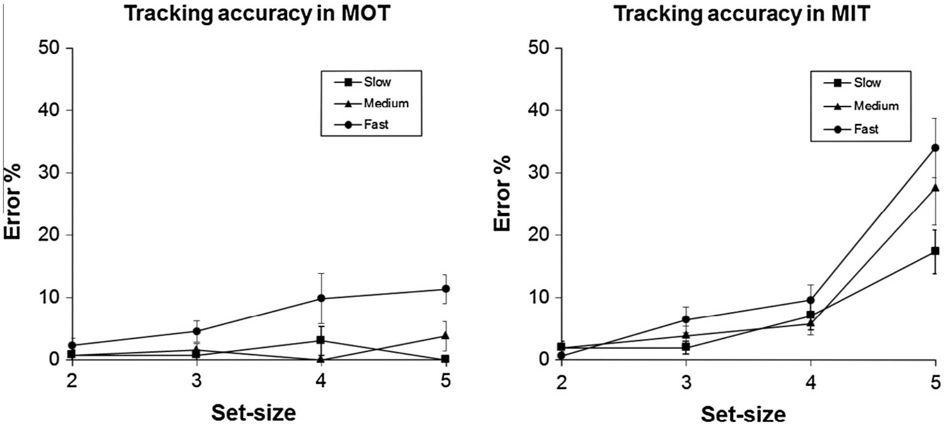

Fig. 2. Error percentage in Experiment 1 for MOT (left panel) and in Experiment 2 for MIT (right panel). Error bars represent the standard error of means (SEMs).

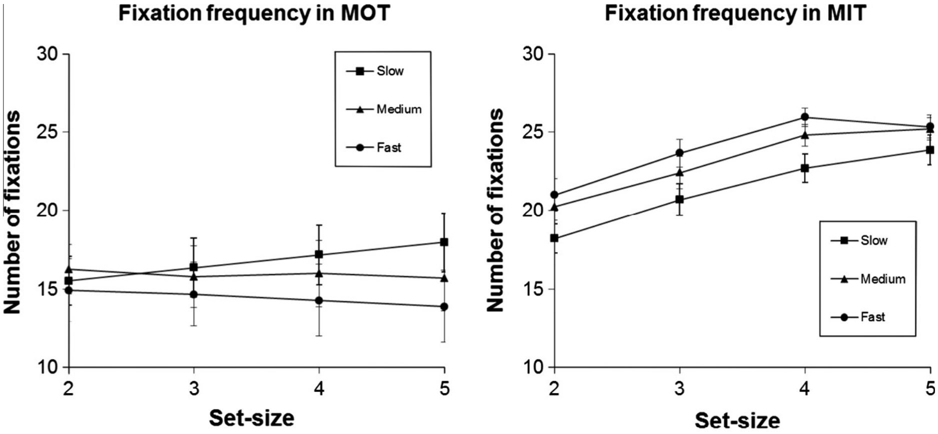

Fig. 3. Number of fixations in Experiment 1 for MOT (left panel) and in Experiment 2 for MIT (right panel). Error bars represent the SEMs.

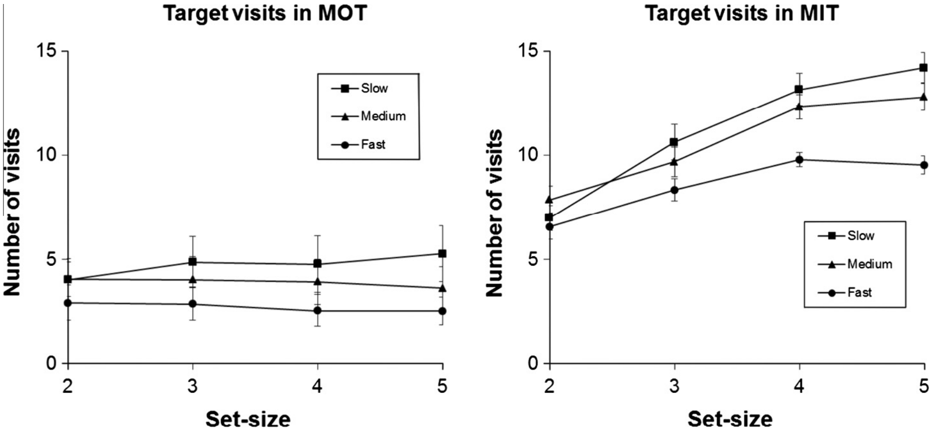

Fig. 4. Number of successful target visits in Experiment 1 for MOT (left panel) and in Experiment 2 for MIT (right panel). Error bars represent the SEMs.

2.2.2.9. Fixation heatmap. As apparent from Fig. 9, a substantial number of fixations fell on an area outside any area of interest. In order to get an idea of where the fixations were located during MOT, the distribution of fixation locations was plotted as a fixation heatmap. Fig. 10 shows the statistical differences (*t*-test values based on a multivariate normal distribution) in the fixation distribution for the entire screen area, separately for each pixel. The heatmap depicts how frequently each pixel was fixated during the task performance. As apparent from Fig. 10, the fixation locations during MOT formed an ellipse-shaped area just above the center of the screen.

2.3. Discussion

The results of Experiment 1 provide strong support for parallel models of MOT and evidence against serial models. The clearest supportive evidence comes from the number of fixations, number of target visits and number of updated targets, none of which

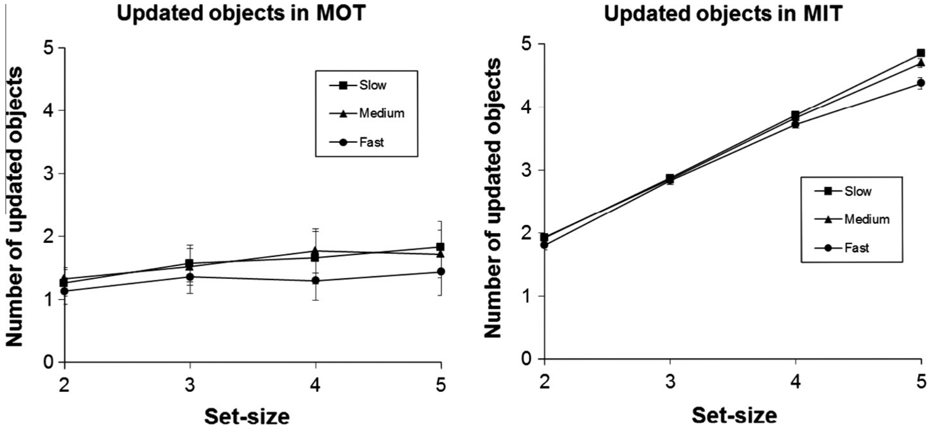

Fig. 5. Number of updated targets in Experiment 1 for MOT (left panel) and Experiment 2 for MIT (right panel). Error bars represent the SEMs.

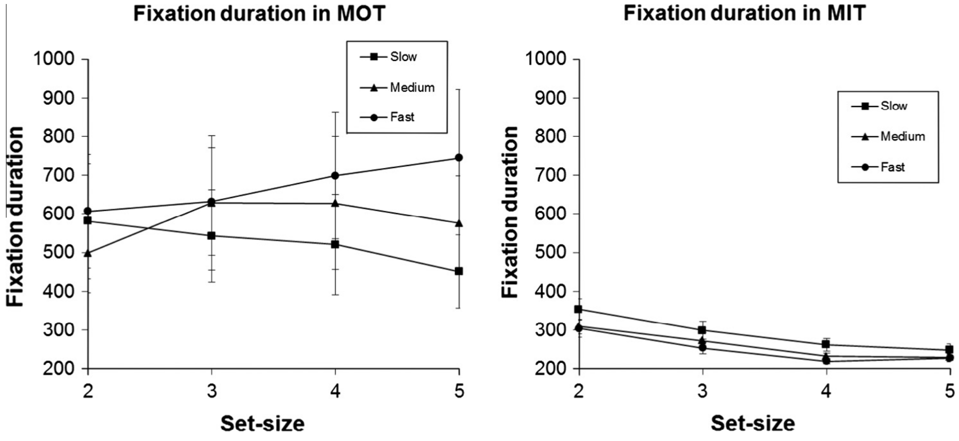

Fig. 6. Average fixation duration in Experiment 1 for MOT (left panel) and Experiment 2 for MIT (right panel). Error bars represent the SEMs.

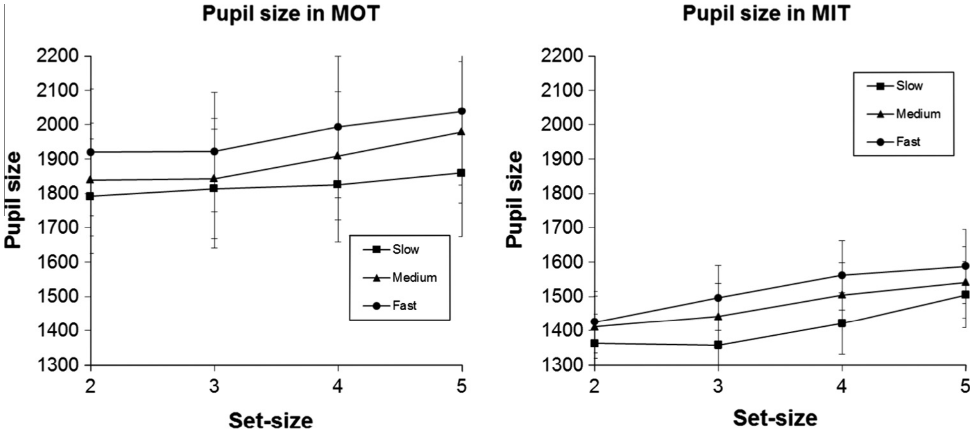

Fig. 7. Pupil size (in pixels) in Experiment 1 for MOT (left panel) and Experiment 2 for MIT (right panel). Error bars represent the SEMs.

was even minimally influenced by target set-size. During MOT, participants appear to fixate at some point on only about one target, regardless of how many targets they track. Thus, these data suggest parallel tracking of visually identical objects in a covert fashion. An increase in pupil size as a function of set-size favor attentative parallel models over preattentative models. Yet, as is apparent from Fig. 7, set-size 2 and 3 did not differ from each other in pupil size. Thus, a preattentative model with a capacity of 3 items may still be defended. Finally, slow object speed appears to encourage participants for serial tracking: In the slow speed condition, number of fixations, target visits and updated targets increased with an increase in set-size. On the other hand, with faster speeds participant resorted to parallel tracking.

The fixation heatmap revealed that the fixations made during MOT formed an ellipse-shape structure located slightly above the screen center (see Fig. 10). This ellipse-shaped area bears similarity to Previc's (1998) model of 3D space. Previc's model proposes several areas responsible for interactions within 3D space. One of

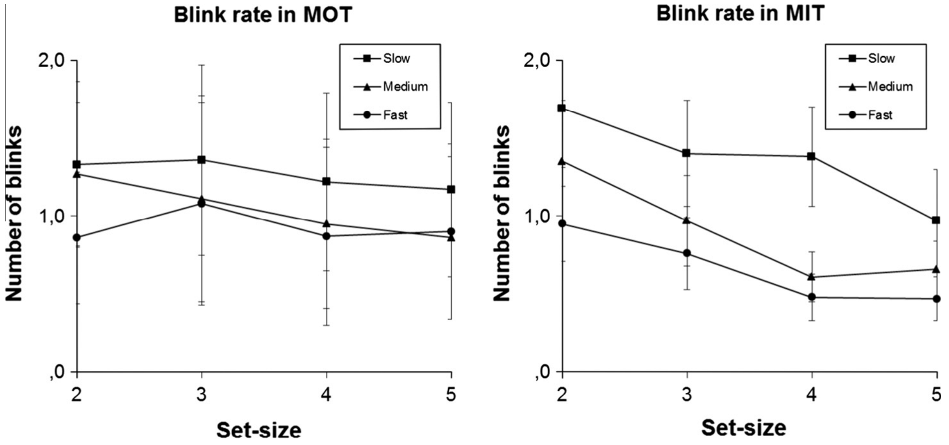

Fig. 8. Number of blinks in Experiment 1 for MOT (left panel) and Experiment 2 for MIT (right panel). Error bars represent the SEMs.

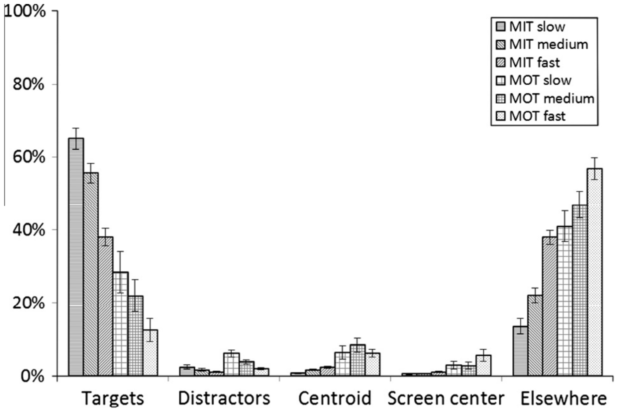

Fig. 9. Percentage of the trial time spent looking at the targets, distractors, centroid of targets, screen center, or elsewhere on the background of the screen in the three speed conditions separately for MOT and MIT. The time is calculated from fixation durations on the aforementioned areas of interest. The rest of the trial time participants spent on saccades, blinks, looking out of the screen (18.7% in MIT and 16.2% in MOT), or fixations lasting less than 80 ms (they were excluded).

those areas, 'focal extrapersonal', is described as 'an (American) football-shaped' region in the front visual space resembling closely the ellipse-shaped fixation pattern found here and is important for visual search and object recognition.

Finally, it is worth noting that the participant's visual sampling rate during MOT differed substantially from those predicted by the engineering models of Moray (1986) Senders (1964). We observed a sampling rate of 0.5 Hz, whereas the optimal rate is estimated by Moray (1986) to be 2.75 Hz and by Senders (1964) to be 5.5 Hz. We return to this issue in the General Discussion.

3. Experiment 2: MIT

In the MIT task, participants tracked 2–5 objects with distinct identities while their eye movements were registered. Analogously to Experiment 1, also target speed was manipulated. The predictions of the parallel and serial models are the same as in MOT. The objects changed their direction and speed during MIT on average of 1.33 Hz. The optimal visual sampling rate of targets will then be either 1.33 Hz (Senders, 1964) or 2.66 Hz (Moray, 1986). The information bandwidth was smaller in MIT than MOT, as MIT included fewer objects, and therefore there were fewer object and edge collisions in MIT than MOT.

3.1. Method

3.1.1. Participants

Thirteen participants (psychology students of the University of Turku) were recruited for the experiment (one male, 12 females, their mean age was about 21). All participants had normal or corrected-to-normal vision.

3.1.2. Apparatus

The same apparatus was used as in Experiment 1.

3.1.3. Stimuli

A set of six stimuli were used. They were vertically oriented line drawings of real objects (flower, coat, lobster, rocking chair, rooster, and watch) also used by Oksama and Hyönä (2008, see Fig. 1, bottom). The pictures were selected from a standardized set of black-and-white line drawings (Snodgrass & Vanderwart, 1980). The stimuli (75 pixels in height and 41–69 pixels in width) were black outline drawings on a white background subtending a visual angle of 1.9 × 1.1–1.8 deg. A subset of two to five objects was designated as targets. After the movement phase, probing was carried out by flashing a black frame (75 × 75 pixels, 2 pixels in width, 1.9 × 1.9 deg) around the target. The frames were not visible during the movement phase. Visual masks (75 × 75 pixels) of variable kind that replaced the pictures at the end of the movement phase were created for the stimuli by copying, rotating and combining parts of the stimulus pictures.

3.1.4. Eye movement analysis

The analysis procedure was identical to that of Experiment 1. Areas that were 3.4 deg in diameter were defined for each moving stimulus, the centroid of the target group and the screen center.

3.1.5. Task and procedure

At the beginning of each trial, the targets (2–5 objects) were designated by flashing a frame around them. The participants' task was to track the identity of these designated targets during the movement phase. After the movement stopped, all the objects were masked and one of the target objects was probed by flashing a frame around it. Finally, the screen was cleared and a response screen appeared where all the six stimuli present during the tracking phase were arranged into an array of two rows and three columns. A computer mouse was used to collect the responses. A response frame (100 × 100 pixels) surrounded the stimuli. The mouse pointer was initially positioned in the middle of the stimulus array. Participants were asked to select (point and click within a

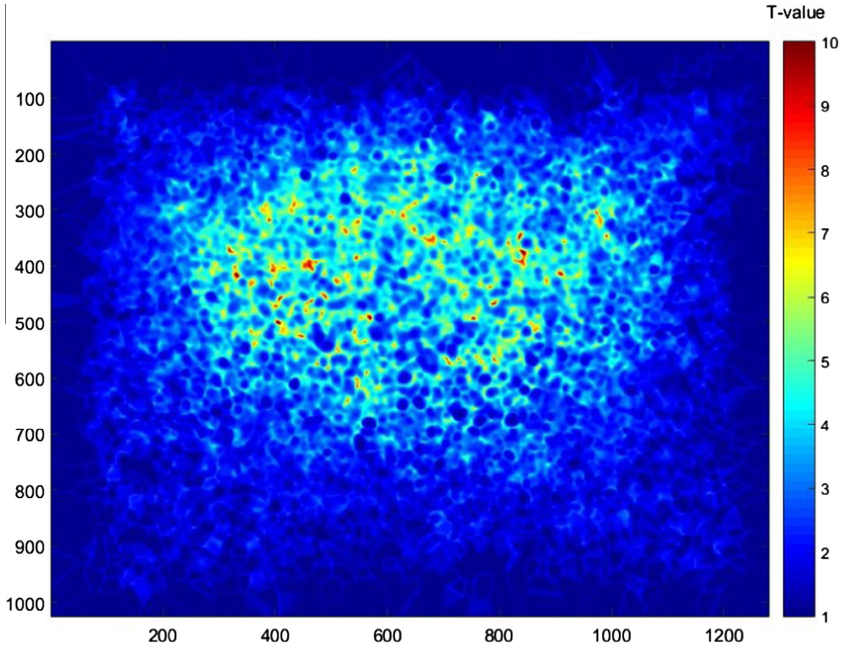

Fig. 10. Statistical differences in the fixation distribution on the entire screen area during MOT. *T*-values over 10 were filtered out (only 0.006% of the total number of pixels was excluded). As the object trajectories were generated randomly, during a trial the positions of the tracked targets were distributed evenly across the screen area (see the Method).

framed picture) the placeholder of the probed target as quickly as possible. They were asked to guess if they did not know the answer. Other aspects of the procedure were identical to that of Experiment 1.

3.1.6. Design

The experimental design was identical to that of Experiment 1.

3.2. Results

3.2.1. Tracking accuracy (error percentage)

The results for the repeated measures ANOVAs are presented in Table 2. The results showed that tracking performance deteriorated as a function of set-size and speed (see Fig. 2, right panel). In addition, their interaction was significant. The interaction reflects the fact that the size of the speed effect increased as set-size increased. Overall tracking accuracies were slightly better than those observed by Oksama and Hyönä (2008), who used more distractors. We recomputed the analysis for square-root transformed data, but the results remained the same.

3.2.2. Eye-tracking data

The results for the repeated measures ANOVAs are presented in Table 2.

3.2.2.1. Number of fixations. The number of fixations made during MIT increased as a function of set-size and speed (Fig. 3, right panel). The Set-Size × Speed interaction was not significant.

3.2.2.2. Number of successful target visits. The number of target visits also increased as a function of set-size, but decreased as a function of speed (Fig. 4, right panel). Their interaction reflects the fact that in the slow and medium speed the number of target visits increased notably as a function of set-size, whereas in the fast speed condition the increase was negligible.

3.2.2.3. Number of updated targets. As is evident from Fig. 5 (right panel), the number of updated targets increased as a function of set-size. Participants were able to update with overt attention shifts the identity–location binding for over four moving targets. Moreover, the number of updated targets decreased as a function of speed (see Fig. 5, right panel). The reliable Set-Size × Speed interaction demonstrates that with slow and medium speed the number of updated targets reached up to almost five targets, while in the high speed condition there is quadratic downward trend in relation to increase in set-size. In the high speed condition participants were able to update with their eyes about four targets.

3.2.2.4. Average fixation duration. Average fixation duration decreased both as a function of set-size and speed (Fig. 6, right

Table 2
Results of the repeated measures ANOVA for the tracking accuracy and the eye measures obtained during the MIT performance (Experiment 2).

	Set-size				Speed				Set-Size × Speed			
	$F(3,36)$	p	MSE	η^2_{p}	$F(2,24)$	p	MSE	η^2_{p}	$F(6,72)$	p	MSE	η^2_{p}
Accuracy (% error)	37.83	**<.001**	299.57	.76	5.33	**<.05**	76.83	.31	3.56	**<.01**	59.27	.23
Number of fixations	37.42	**<.001**	9.14	.76	24.96	**<.001**	3.73	.68	1.85	.18	3.63	.13
Number of visits	63.17	**<.001**	3.31	.84	30.19	**<.001**	3.46	.72	19.67	**<.001**	.57	.62
Number of updated targets	916.11	**<.001**	.06	.99	13.75	**<.001**	.04	.53	7.16	**<.001**	.02	.37
Pupil size	28.43	**<.001**	13,011	.70	45.45	**<.001**	5730	.79	4.56	**<.01**	2091	.28
Blinks	7.76	**<.01**	.73	.39	15.90	**<.01**	.73	.57	2.94	**<.05**	.13	.20
Average fixation duration	21.44	**<.001**	5875	64	12.10	**<.01**	3479	.50	1.37	.27	2005	.10

panel). The decrease levelled off at set-size 4, where the average fixation duration was about 220 ms, which may be close to the lower boundary for a fixation.

3.2.2.5. Pupil size. The pupil dilated both as a function of set-size and speed (Fig. 7, right panel). The reliable interaction is due to the fact that the pupil dilated more robustly as a function of set-size with high than slow speed.

3.2.2.6. Number of blinks. Blink rate decreased both as a function of set-size and speed (Fig. 8, right panel). The interaction reflects the fact that the decrease in blink rate as a function of set-size asymptotes at set-size 4 (approaches zero) in the medium and fast speed conditions, whereas in the slow speed condition it continues to decrease from set-size 4 to set-size 5. The interaction is likely to reflect a floor effect is thus not readily interpretable.

3.2.2.7. Sampling rate. The participants sampled the visual display on average at the rate of 3.3 Hz, i.e., they made a fixation on average after every 306 ms. A target was visited on average after every 690 ms, at the rate of 1.5 Hz, almost exactly (1.33 Hz) as predicted by Moray (1986).

3.2.2.8. Average percentage of trial time spent looking at different areas of interest. As apparent from Fig. 9, participants spent most of the trial time (on average 52.9%) looking at the targets. They spent much less time looking at the background (on average 24.5%) excluding the centroid, screen center and distractors. The time spent looking at the centroid (on average 1.5%), screen center (on average 0.7%), or distractors (on average 1.7%) was negligible.

3.2.2.9. Fixation heatmap. Analogously to Experiment 1, we created a heatmap of fixations made during MIT. As apparent from Fig. 11, the participants' fixations were distributed evenly around the entire screen area, excluding only the area next to the monitor's edges.

3.3. Discussion

The MIT results are perfectly in line with the predictions of serial models. The number of fixations and target visits increased monotonously up to set-size 4, which is shown to be the average tracking capacity in MIT (Oksama & Hyönä, 2004). The data on the number of updated targets showed that participants updated over four targets with overt attention shifts. With respect to the speed manipulation, the results demonstrated an increase in the number of fixations and a decrease in the average fixation duration. That is, participants made more fixations of shorter durations in response to speed increase. The measures reflecting fluctuations in attentional load (pupil size and blink rate) demonstrated a strong involvement of attention in MIT: more demanding conditions were associated with increased attentional load.

As a substantial portion of fixations landed on the targets, it is not surprising that the fixation heatmap (see Fig. 11) covers the entire screen area, where the targets moved about. Finally, the visual sampling rate observed during MIT (1.5 Hz) corresponds with closely with the prediction (1.33 Hz) derived from Moray's (1986) engineering model.

4. Experiment 3: MIT & MOT

Although the MOT results (Experiment 1) differed very clearly from those for MIT (Experiment 2), it may be argued that the differences were due to stimulus differences between the two experiments. In the MOT task filled black circles were used as stimuli, whereas in MIT the stimuli consisted of black and white line drawings. This difference in stimuli may have particularly influenced the pupil size data. There appeared also a peculiarity in the pupil size data in that the overall pupil size appears to be larger in MOT than MIT, which is counterintuitive. However, this may not be real but rather reflect differences in the overall pupil size among the participants who took part in Experiment 1 and 2. To eliminate and control for the differences in stimuli and participants in Experiments 1 and 2, we conducted Experiment 3, where the same participants performed both MOT and MIT with similar stimuli.

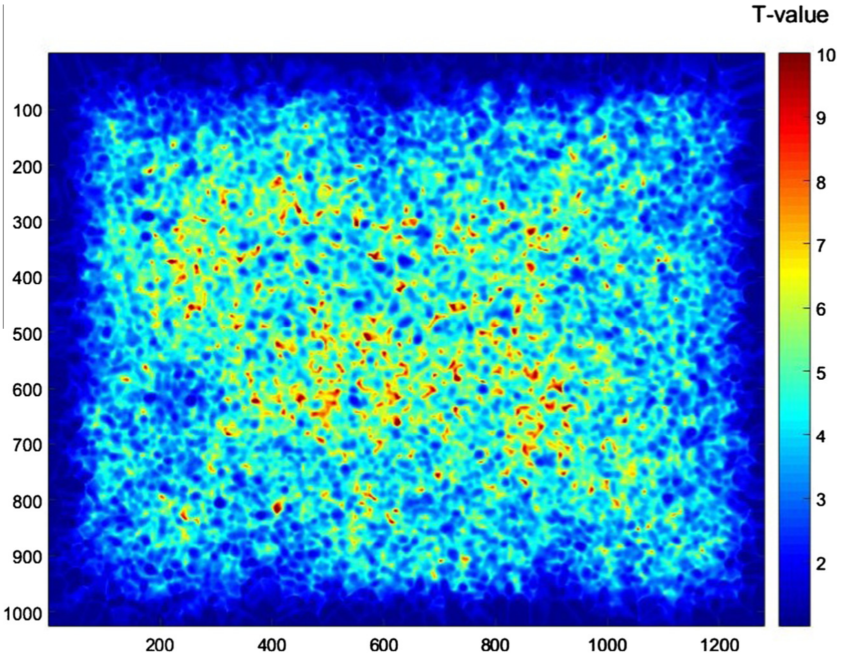

Fig. 11. Statistical differences in the fixation distributions on the screen area during MIT. *T*-values over 10 were filtered out (only 0.06% of the pixels was excluded). As the object trajectories were generated randomly, during a trial the positions of the tracked targets were distributed evenly across the screen area (see the Method).

In Experiment 3 participants tracked identical black-and-white line drawings of lobsters during MOT, while in MIT they tracked four distinct line drawings including the lobster. Similarly to Experiment 1 and 2, the participants' eye movements were registered during task performance. The objects changed their direction and speed at the same rate as in the previous experiments as the same set of trajectories were used (2.75 Hz for MOT and 1.33 Hz for MIT). In order to be able to run the two tasks within a single experimental session, the number of trials employed in Experiment 1 and 2 were distributed between the MOT and MIT tasks. Thus, each task included half the number of trials of Experiment 1 and 2. In order to obtain more reliable condition mean estimates, we excluded from the ANOVA design the speed factor and analyzed the data using a 2 (task: MOT vs MIT) × 4 (set-size: 2–5 targets) within-participants design.

4.1. Method

4.1.1. Participants

Thirteen participants (psychology students of the University of Turku) were recruited for the experiment (three males, ten females, their mean age was about 22). All participants had normal or corrected-to-normal vision. Data from one participant were excluded from the analysis because of poor calibration.

4.1.2. Apparatus

The same apparatus was used as in Experiment 1 and 2.

4.1.3. Stimuli

The stimuli in the MIT condition were the same as in Experiment 2. The stimuli in the MOT condition consisted of six identical line drawings of a lobster, which also appeared as one of the target pictures in MIT. A subset of two to five objects was designated as targets. The MIT procedure was identical to that in Experiment 2 and the MOT procedure was identical to that in Experiment 1. The MOT and MIT trials were presented in separate blocks.

4.1.4. Eye movement analysis

The analysis procedure was identical to that of Experiment 1 and 2.

4.1.5. Design

ANOVAs were computed using two manipulated variables: number of targets (2–5) and task (MOT vs. MIT). Both factors were within-participant variables. There were 18 trials in each of the 8 conditions.

4.2. Results

4.2.1. Tracking accuracy (error percentage)

The results for the repeated measures ANOVAs are presented in Table 3. In tracking accuracy, a significant main effect was found for the number of targets but the main effect of task type was not significant. However, the interaction between set-size and task type was significant. That is, the performance deteriorated steeper for MIT than for MOT as a function of set-size (see Fig. 12). The overall tracking accuracy in MIT corresponds closely with that in Experiment 2. However, the overall tracking accuracy in MOT was slightly lower here than in Experiment 1. We recomputed the analysis for square-root transformed data, but the results remained the same.

4.2.2. Eye-tracking data

The results for the repeated measures ANOVAs are presented in Table 3.

4.2.2.1. Number of fixations. The main effects of set-size, task type and their interaction were all significant (see Table 3). As evident from Fig. 13 (top left), the number of fixations in MIT increased as a function of set-size, but in MOT it remained constant.

4.2.2.2. Number of successful target visits. The main effects of set-size, task type and their interaction were significant. As evident from Fig. 13 (top middle), the number of visits in MIT increased as a function of set-size, but in MOT it remained constant.

4.2.2.3. Number of updated targets. The main effects of set-size, task type and their interaction were significant (Table 3). That is, the number of updated targets in MIT increased as a function of set-size, whereas in MOT it remained constant (see Fig. 13, top right). In MIT, the number of updated targets reached up to over four targets, whereas in MOT participants updated on average one and a half targets regardless of set-size.

4.2.2.4. Average fixation duration. A main effect of set-size was not significant (Table 3). However, the main effect of task type was significant and the Set-Size × Task interaction was marginally significant. That is, average fixation duration decreased in MIT, whereas in MOT it remained constant (see Fig. 13, bottom left). It decreased in MIT with set-size 5 to about 270–280 ms in the medium and fast speed conditions.

4.2.2.5. Pupil size. In pupil size, a significant main effect was found for the number of targets, the task type and the interaction between set-size and task type (Table 3). That is, pupil size increased more steeply in MIT than MOT as a function of set-size (see Fig. 13, bottom middle).

4.2.2.6. Number of blinks. Number of blinks revealed a significant main effect for the number of targets, but the main effect of task type was not significant (Table 3). However, the interaction between set-size and task type was significant. That is, blink rate decreased more steeply for MIT than for MOT as a function of set-size (see Fig. 13, bottom left).

Table 3

Results of the repeated measures ANOVA for the tracking accuracy and the eye measures obtained during the MIT and the MOT performance (Experiment 3).

	Set-size				Task				Set-Size × Task			
	$F(3,36)$	p	MSE	η^2_p	$F(1,12)$	p	MSE	η^2_p	$F(3,36)$	p	MSE	η^2_p
Accuracy (% error)	27.52	**<.001**	2.83	.70	<1	.45	1.57	.05	15.16	**<.001**	1.00	.56
Number of fixations	9.26	**<.001**	2.25	.44	39.21	**<.001**	17.32	.77	25.52	**<.001**	1.77	.68
Number of visits	26.23	**<.001**	1.79	.69	94.15	**<.001**	5.31	.89	64.65	**<.001**	.70	.84
Number of updated targets	127.73	**<.001**	.14	.91	110.72	**<.001**	.47	.90	124.99	**<.001**	.06	.91
Pupil size	40.12	**<.001**	3243.48	.77	6.53	**<.05**	26043.55	.35	22.41	**<.001**	787.69	.65
Blinks	13.99	**<.001**	.08	.54	3.81	.08	.24	.24	4.07	**<.05**	.13	.25
Average fixation duration	1.48	.24	128963.87	.11	9.18	**<.05**	354471.38	.43	3.19	.07	66324.70	.21

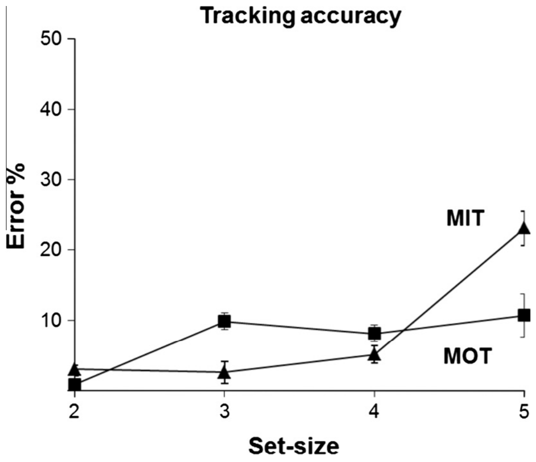

Fig. 12. Error percentages in Experiment 3 for MOT and MIT. Error bars represent the SEMs.

4.2.2.7. Sampling rate. The participants sampled the visual display in MIT on average at the rate of 2.8 Hz (a fixation was made after every 358 ms), whereas in MOT the rate was 2.1 Hz (a fixation was made after every 485 ms). A target was visited in MIT on average after every 902 ms, that is, at the rate of 1.1 Hz, whereas in MOT a target was visited after every 2074 ms, that is, at the rate of 0.5 Hz. The MOT results are practically identical to those in Experiment 1, but in MIT the sampling rate is slightly lower than in Experiment 2.

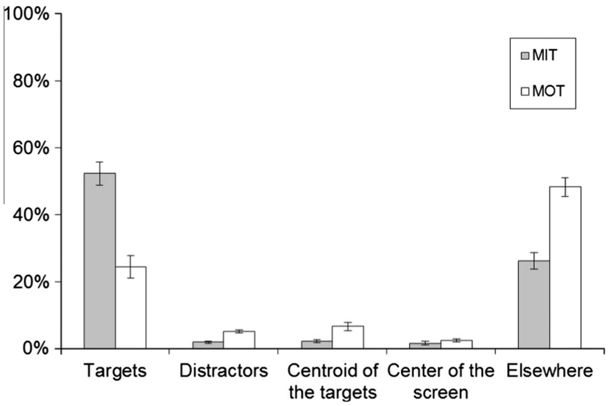

Fig. 14. Percentage of the trial time spent looking at the targets, distractors, centroid of targets, screen center, or elsewhere in the background of the screen in Experiment 3. The time is calculated from fixation durations on the aforementioned areas of interest. The rest of the trial time participants spent on saccades, blinks, looking out of the screen (15.9% in MIT and 13.3% in MOT), or fixations lasting less than 80 ms (they were excluded).

4.2.2.8. Average percentage of trial time spent looking at different areas of interest. As apparent from Fig. 14, participants spent most of the trial time in MIT looking at the targets (52.4%), whereas in MOT they spent much less time on the targets (24.4%). During MOT the participants spent most of their trial time (48.4%) looking at some position in the background of the screen that was neither the centroid nor the screen center, whereas in MIT much less time was spent looking at the background (26.1%). The time spent looking at the centroid (MIT 2.1%, MOT 6.5%), screen center (MIT 1.6%, MOT 2.4%), or distractors (MIT 1.9%, MOT 5.1%) were again surprisingly short. The results are highly similar to those of Experiment 1 and 2.

4.3. Discussion

Experiment 3 differed from Experiment 1 and 2 in two respects. First, the stimuli in Experiment 3 were comparable across the two

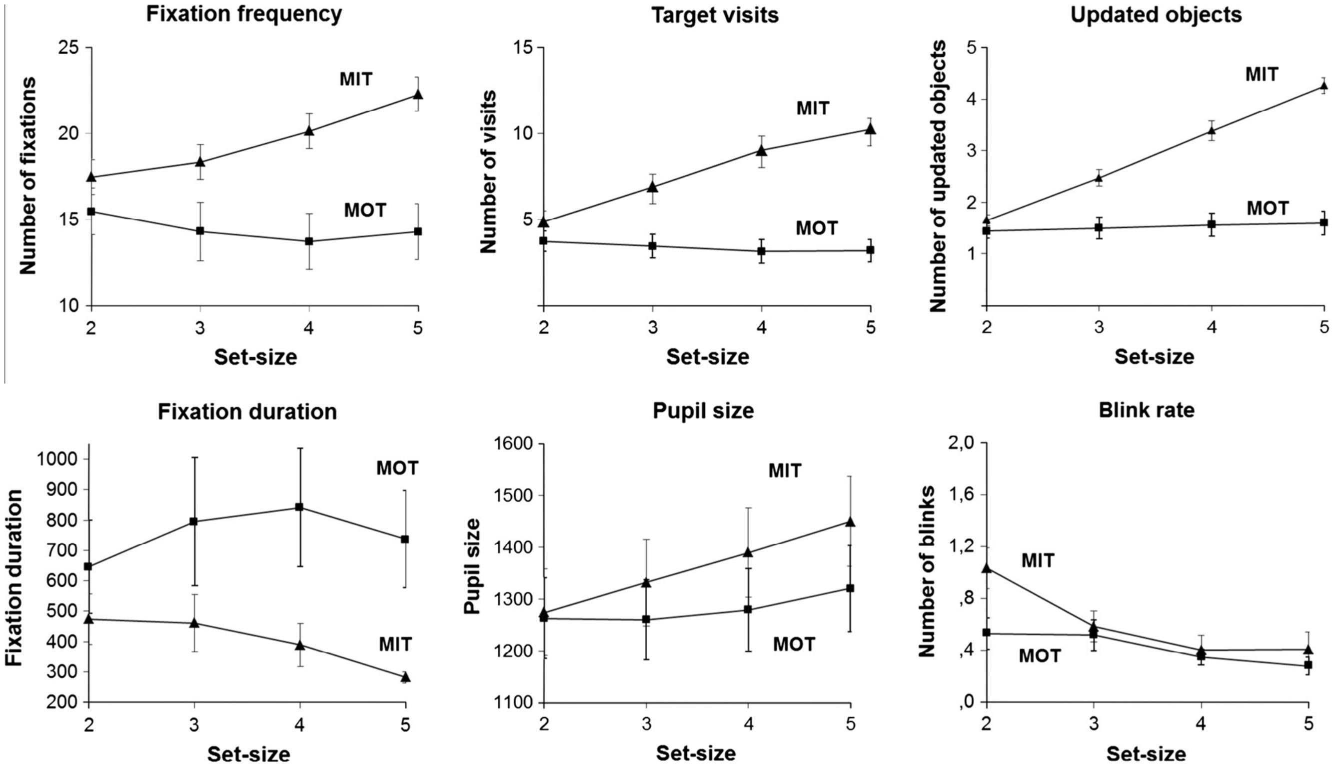

Fig. 13. Number of fixations (top left), number of successful target visits (top middle), number of updated targets (top right), average fixation duration (ms; bottom left), pupil size in pixels (bottom middle) and number of blinks (bottom right) during a trial (average duration was 7 s) in Experiment 3. Error bars represent the SEMs.

tasks in that now also in MOT the stimuli were line drawings (lobsters) instead of filled black circles (Experiment 1). Second, the key comparison between MOT and MIT was now made using a within-participants design. All the key results obtained in Experiment 1 and 2 were replicated in Experiment 3. Most importantly, in MIT a linear increase was observed as a function of set-size in fixation frequency, number of target visits, and number of updated targets, whereas in MOT set-size did not reliably affect these measures indexing sampling of the dynamically changing visual environment. Moreover, pupil size indexing cognitive load increased linearly in MIT as a function of set-size, whereas in MOT the increase was not as robust. Similarly, blink rate decreased more steeply in MIT than in MOT as a function of set-size. Finally, average fixation duration decreased linearly in MIT as a function of set-size, whereas no such trend was visible in MOT. Overall, when performing the MIT task, the participants sampled the dynamic visual display much more frequently than when performing the MOT task.

Experiment 3 clarified one peculiar finding of Experiments 1 and 2. The between-experiment comparison for the pupil size suggested that the pupil size would have been generally larger in MOT (Experiment 1) than in MIT (Experiment 2). We suspected this unexpected result to reflect overall between-participant variation in pupil size unrelated to the performed task. Experiment 3 indicated this claim to be correct. When the same participants performed both tasks, we found generally larger pupil sizes in MIT than MOT, suggesting that MIT is generally a more demanding task to perform than MOT. Alnæs et al. (2014) recently demonstrated that in MOT pupil size faithfully reflects increases in set-size from 2 to 5 targets. Our MOT results are less robust, as there is no difference in pupil size between set-size 2 and 3, and the set-size effect is generally smaller than in MIT. The difference between the present results and those of Alnæs et al. may be ascribed to stimulus differences. In our MOT task, the target stimuli were almost three times bigger than those of Alnæs et al. (a diameter of 1.9 versus 0.7, respectively).

5. General discussion

In the present study, we investigated observers' eye movements when they kept track of the positions of identical moving targets (the MOT task, Experiment 1 and 3) or when they tracked the whereabouts of distinct object identities (the MIT task, Experiment 2 and 3). We found a dramatic difference between these tasks in terms of eye movements. When the participants tracked only position information of the moving objects (the MOT task), the eye-movement measures were not influenced by the set-size manipulation. Moreover, only minimal eye-movement activity was found; rather, the observers had a clear preference for keeping their eyes fixed for a relatively long time in the same position of the screen (typically in the upper center area). In contrast, active eye behavior was observed when the observers tracked position-bound identity information. Now the participants sampled the moving target identities at the same rate as the objects changed their direction (1.3 Hz). Also, the participants updated over four target identities with overt attention shifts (i.e., by making eye fixations on targets), and the eye-movement measures were sensitive to the set-size and speed manipulations.

5.1. Eye-movement strategies in MOT and MIT

Fehd and Seiffert (2008, 2010) have provided evidence in support of a centroid-target-switching strategy in MOT, in which participants alternate between looking at the centroid (the geometrical center of the moving targets) and the targets. According to Fehd and Seiffert (2010), observers spend about 20% of the trial time employing this strategy. They also found that participants viewed the centroid more often than the targets or distractors. Consistent with this, Huff et al. (2010) found that centroid looking became more frequent with increasing object speed. However, in their study only about 10% of the gaze time was identified as centroid looking.

The present MOT results are inconsistent with the centroid-target switching strategy. Our observers did not spend much time on the centroid (only 7% of the total viewing time). This compares favorably with the study of Zelinsky and Neider (2008) who also failed to find support for the centroid strategy for tracking four targets (see also Doran, Hoffman, & Scholl, 2009). Our finding is also similar to that of Huff et al. (2010), who found observers to spend only little viewing time on the centroid. However, unlike us they interpret their data to be supportive of the centroid looking strategy. Similarly to the present study, Huff et al. defined the areas of interest dynamically, which is a more stringent system than the shortest distance rule used in other previous studies. When applying the shortest distance rule, every fixation is assigned to one of the chosen interest areas (e.g., centroid, target, or distracter) depending what is closest to the fixation. However, when dynamic interest areas are used, the fixation has to actually land on one of the chosen interest areas, otherwise it is coded as a miss (Huff et al.) or be positioned on the background (the present study). Understandably, the latter procedure yields significantly lower estimates for the time spent on looking at the centroid than the shortest distance rule. Yet, we consider it difficult to interpret these data to suggest the centroid looking strategy to be most prevalent in MOT. In the present study, the observers spent much more time on the targets (a bit over 20%) and, particularly on an unspecified area of the screen (48%) that was neither the centroid nor the screen center. The fixation heatmaps showed this area to form an ellipse-shaped area positioned slightly up from the screen center. Moreover, observers updated only about one target with their eyes during the MOT task. That is, they did not switch their gaze between targets. Fehd and Seiffert (2010) admit that the centroid-target switching strategy does not necessarily reflect participants' natural behavior during tracking. When they instructed the participants to use a centroid-looking, target-looking or free-looking strategy, both the centroid-looking and the target-looking strategy was observed to impair the tracking performance relative to the free-looking condition (the centroid-looking strategy was better than the target-looking strategy).

Another possible eye-movement strategy in MOT is the look-at-one-target-strategy (see Zelinsky & Neider, 2008), in which observers track one target with their gaze, while keeping the other targets in their peripheral vision. Zelinsky and Neider did not find support for this strategy when multiple identical targets were tracked. Interestingly, however, our results somewhat resemble the look-at-one-target-strategy in that the observers updated with their eyes only one target. Yet, they spent ample time looking at the upper center area of the screen. One possibility is that observers choose one target among the target set to be overtly tracked by the eyes. They would fixate that target but more often the blank space in the vicinity of that target. The rest of the targets would be tracked peripherally in a covert fashion. Such a strategy seems both psychologically and computationally more plausible than the continuous computation of the centroid dynamically delineated by the moving targets, which appears computationally quite demanding (cf. a similar kind argument put forth by Fehd & Seiffert, 2010). A non-precise look-at-one-target-strategy is not computationally difficult and may thus serve as a readily available heuristic to be used in MOT.

There is only one prior study that has examined eye behavior during MIT. In a simulated air traffic control task, Landry et al.

(2001) found evidence for attentional switching between targets. The present MIT results are consistent with those of Landry et al. They both speak very clearly in favor of an eye movement strategy where the eyes are sent from one target to the next. Participants spend clearly most of their tracking time on the targets.

5.2. Trade-off between sampling rate and observation time may be a key to capacity limitation in MIT

Moray (1984) argues that observers trade off between sampling rate and observation time (i.e., the time needed to process the visual information at hand) when several channels of dynamic information have to be monitored. Moray speculates that this trade-off can be solved either by shortening the observation time (and increasing the sampling rate) or by delaying the sampling rate (and keeping the observation time constant). Studying multiple identity tracking via eye movements and manipulating the tracking load provides a direct test of this hypothesis, as was done in the present study.

Our results suggest that in MIT participants use both strategies. With smaller set-sizes (2–4) in Experiment 2 participants increased sampling rate and decreased observation time in response to an increase in the tracking load. The strategy proved successful: the identity tracking accuracy was over 90% for 2–4 targets. However, in set-size 4 the observation time and sampling rate levelled off: fixation duration asymptoted at 220 ms and the visual sampling rate (fixations per second) reached its maximum value of 3.7 Hz (where a fixation was made on average after every 270 ms). With set-size 5 the observers changed their strategy: they no longer decreased fixation durations or increased fixation frequency. As a consequence, the tracking accuracy for set-size 5 dropped dramatically (from 90% to 65%). This drop was less dramatic in Experiment 3. Moreover, in Experiment 3 we observed a linear decrease in average fixation duration with no asymptote reached.

The minimum fixation duration observed in Experiment 2 for MIT (about 220 ms) is comparable to average fixation duration on objects during the recognition of single objects: Leek, Patterson, Paul, Rafal, and Cristino (2012) observed the fixation duration to be 218 ms in such a task. In visual search, also mimicking multiple object tracking, average fixation duration varies between 180 ms and 275 ms, depending on the complexity of the search array (Rayner, 2009). The fixation duration of 220 ms is located toward the lower boundary for what is obtained in visual search. Thus, we argue that fixation duration of 220 ms may approximate the minimum time required to update an identity–location binding with overt attention; hence, a serial updating mechanism cannot increase its sampling rate below this limit. In Experiment 3, this lower boundary was not reached; the shortest average fixation duration with set-size 5 was 280 ms. In Experiment 3, the MIT task lasted for only half the time than in Experiment 2. Thus, it is possible that the participants in Experiment 3 did not have enough exposure to the task to obtain sufficient practice in order to reach the asymptote in fixation duration. Hence, we consider the data of Experiment 2 to be more informative than the data of Experiment 3 with respect to the trade-off notion discussed above. If so, the asymptote observed in fixation duration in Experiment 2 may be a root cause for the capacity limit of three or four objects in identity tracking, as within the speed range of Experiment 2 the serial mechanism cannot update and track more than three or four moving targets. Notice also that the shortest fixation times observed in Experiment 2 and 3 for MIT are very close to the 'binding refresh time' estimate obtained by mathematical simulations of the MOMIT model (the best fitting refresh time was about 240 ms, see Oksama & Hyönä, 2008).

Interestingly, in MOT fixation frequency and duration were not influenced by target set-size. Clearly, the MOT performance is not related to overt gaze behavior that would be reflected as fluctuations in fixation duration or sampling rate. Therefore, the capacity limitations in MOT must come from a different source. However, after 25 years of MOT research, the nature of the capacity limitations in MOT is still under discussion. Several candidates have been proposed and debated, such as fixed-limit architecture (Cavanagh & Alvarez, 2005; Pylyshyn, 1989), flexible resources (Alvarez & Franconeri, 2007), spatial interference (Franconeri et al., 2010; but see Holcombe & Chen, 2012), and temporal frequency (Holcombe & Chen, 2013).

5.3. Is tracking parallel or serial?

The present eye movement results suggest an "ecumenical" solution to the recent theoretical controversy about whether tracking is achieved by a serial or parallel processing mechanism. The present study demonstrates that the answer depends on the type of tracking task. The pattern of results strongly points to two separate systems with different processing qualities: position tracking in the MOT task is achieved by a covert parallel system, whereas identity tracking in the MIT task is achieved by an overt serial system. The position tracking system is surprisingly effective: it can keep track of positions of multiple moving objects without directing the eyes toward the targets. The identity tracking, on the other hand, is based on more laborious visual scanning of the to-be-tracked items.

The present results rule out a unitary *parallel* tracking mechanism (such as the object file theory of Kahneman et al., 1992), as the eye-movement data during MIT speak against such mechanism. A unitary *serial* tracking mechanism for MIT and MOT seems also unlikely in the light of the present data. It is unlikely because, to be consistent with the present data, we should propose a covert serial mechanism for MOT but an overt serial mechanism for MIT. We cannot think of a reason why a unitary serial system would use overt attention in one task and covert attention in another task. Thus, it seems more coherent, plausible and economical to assume a tracking system with two components: an effective parallel position tracking component and a serial, identity tracking component. Note, however, that the above theorizing is not in line with some recent evidence that is in favor of serial processing in MOT (Holcombe & Chen, 2013; Tripathy et al., 2011).

Howe and Ferguson (2015) recently presented evidence supporting the idea that in some instances MIT can be carried out in parallel. They applied the simultaneous-sequential paradigm (Eriksen & Spencer, 1969; Shiffrin & Gardner, 1972) and systems factorial technology (Townsend & Nozawa, 1995; Townsend & Wenger, 2004b) to MIT tasks and found evidence in favor of parallel processing. While their results are interesting, it seems that the employed techniques changed the nature of the MIT tasks in a serious way. For instance, eye movements were prevented, identities were not visible or simultaneous dual-tasks were introduced. These differences may have encouraged their participants to try to use more parallel strategies during MIT. Nevertheless, more research is needed on this interesting and theoretically relevant topic. Clearly, the present eye movement results are inconsistent with the idea of parallel processing in MIT.

5.4. Is tracking achieved by a unitary or two independent tracking systems?

Are the position and identity tracking systems truly separate and independent of each other? Our results suggest an affirmative answer to this question. They must be independent, because the covert position tracking system cannot yield identity information.

Otherwise, the observers would not have needed to make eye movements to successfully perform the MIT task. Yet, eye fixations on targets are needed to keep the target identities separate from each other and dynamically update identity–location bindings for each target. Thus, in addition to position tracking, another system is needed which dynamically recognizes the target identities with overt attention shifts (eye fixations) as the objects move about in the visual environment and which attaches position information to them.

On the other hand, there are strong reasons to believe that the two systems work in close co-operation in MIT. The idea is that the parallel position tracking system provides location information to the serial identity tracking system. Position information for the moving objects located in the peripheral vision is required to program serial attention shifts between the moving targets (cf. the assumptions of the MOMIT model, Oksama & Hyönä, 2008). It is assumed that only the position tracking system is in operation during MOT, whereas both systems are activate during MIT. The finding that people are able to report more locations than identities (Horowitz et al., 2007) support this claim: the effective position tracking system provides location information also for those targets that cannot be bound to identities.

Recently, however, Cohen et al. (2011) have provided evidence, which may be taken as evidence in favor of a unitary tracking model. They demonstrated that observers can trade off performance between location and identity tracking during a high tracking load (fast speed or large set-size). They also found some evidence for flexible and strategic attention allocation between the identity and location information. Clearly, this trade-off is not consistent with the idea of two fully independent systems, as independent systems not demonstrate mutual trade-offs. Yet, these data do not necessarily point to a unitary tracking system, as it possible that the two systems share a common component/resource but differ in other important respects. The common resource between MIT and MOT may be a general attentive or executive resource, maybe related to sustained attention (see Oksama & Hyönä, 2004). Our pupil data (see also Alnæs et al., 2014) indicate that both MIT and MOT are attention-demanding (thus ruling out preattentive models, such as FINST of Pylyshyn, 1989), so it is feasible to expect some kind of trade off in difficult tracking conditions (Cohen et al. did not find trade off in the low tracking load conditions). If position tracking becomes more difficult and demands more attention, then attention resources may be withdrawn from identity tracking and the MIT performance would suffer (this may also work the other way around, but to our knowledge this have not been observed).

Moreover, Papenmeier, Meyerhoff, Jahn, and Huff (2013) presented evidence supporting the view that the mechanism behind location tracking can utilize feature information in solving the correspondence problem during motion, especially when spatiotemporal information is not reliable. In particular, they claim that feature information can be processed automatically. These data go clearly against the idea of feature-blind position tracking mechanism (cf. theories of Pylyshyn, 1989, 2001; and Kahneman et al., 1992). These results also lend some support to the unitary view on tracking in which a single (parallel) mechanism can utilize both location and feature information. However, a single automatic feature/location tracking mechanism is not consistent with the present data or other results attesting effortful maintenance of feature-location bindings (Oksama & Hyönä, 2008; Pinto et al., 2010; Saiki, 2003a, 2003b). In fact, the idea that the location tracking mechanism can utilize also feature information does not necessarily go against the view of two separate tracking systems. The position tracking system may utilize both spatiotemporal and feature information as its input, but it may yield only unbounded and unindexed location information as its output. On the other hand, a separate identity tracking mechanism may utilize both feature and location information as its input but yield identity–location-bindings as its output. What is critical is not the input but the output of the system. Moreover, it may be that the correspondence problem and the maintenance of what–where-bindings are not dealt with by the same process or mechanism. The correspondence problem may related to the position tracking mechanism and the maintenance of what–where-binding to the identity tracking mechanism.

Finally, the postulation of two separate tracking mechanisms fits nicely with the idea proposed by Leibowitz and Post (1982), Previc (1998) and Horrey et al. (2006). They distinguish between the focal and ambient visual systems. The focal system, operating mainly with visual information available in the foveal vision, is specialized in tasks requiring high visual acuity, such as visual search and object recognition. In contrast, the ambient visual system is involved in the maintenance of spatial orientation and postural control in locomotion. The focal system is tightly linked with eye fixations, whereas the ambient system is not. The two systems are probably in some way connected to each other; ambient vision provides position information to the focal system. According to Previc (1998), ambient vision serves as the bedrock for focal vision. A similar distinction has been made between two cortical pathways, the "what" and "where" systems (Milner & Goodale, 1993; Ungerleider & Mishkin, 1982). The "what" system is responsible for visual object recognition, while the "where" system encodes object positions in space. As we have argued above, the present eye movement data are perfectly in line with this theorizing.

6. Conclusions

The present eye movement study suggests that there are two separate subsystems involved in multiple object tracking. The ambient position tracking system keeps track of the positions of the moving targets in parallel without the need of overt attention shifts in the form of eye movements. On the other hand, the focal system recognizes individual object identities and binds them to their locations one by one by using overt visual scanning. We argue that in MOT observers employ the ambient position tracking system, whereas in MIT both the position and focal identity tracking systems are in operation. Hence, in future studies (e.g., in engineering psychology, traffic psychology or neuropsychology) the MOT task can be used as a tool to measure the workings of the ambient vision and the MIT task as a tool to measure the workings of both the ambient and focal vision.

Acknowledgements

This work was supported by the Academy of Finland. We thank Sanni Aalto and Marjaana Melentjeff for their help in collecting the data, Satu Arola for her help in the data analysis and Lauri Nummenmaa for his help in heatmap programming.

References

Alnæs, D., Sneve, M. H., Espeseth, T., Endestad, T., van de Pavert, S. H. P., & Laeng, B. (2014). Pupil size signals mental effort deployed during multiple object tracking and predicts brain activity in the dorsal attention network and the locus coeruleus. *Journal of Vision, 14*(4), 1–20. http://dx.doi.org/10.1167/14.4.1, pii: 1.

Alvarez, G. A., & Cavanagh, P. (2005). Independent resources for attentional tracking in the left and right visual hemifields. *Psychological Science, 16*, 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*, 1–10.

Beatty, J. (1982). Task-evoked pupillary responses, processing load, and the structure of processing resources. *Psychological Bulletin, 91*, 276–292.

Botterill, K., Allen, R., & McGeorge, P. (2011). Multiple-object tracking. *Experimental Psychology, 58*, 196–200.

Carbonell, J. R. (1966). A queuing model of many-instrument visual sampling. *IEEE Transactions on Human Factors in Electronics, HFE-7*, 157–164.

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

Cohen, M. A., Pinto, Y., Howe, P. D., & Horowitz, T. S. (2011). The what–where trade-off in multiple-identity tracking. *Attention, Perception, & Psychophysics, 73*, 1422–1434.

d'Avossa, G., Shulman, G. L., Snyder, A. Z., & Corbetta, M. (2006). Attentional selection of moving objects by a serial process. *Vision Research, 46*, 3403–3412.

Deubel, H., & Schneider, W. X. (1996). Saccade target selection and object recognition: Evidence for a common attentional mechanism. *Vision Research, 36*, 1827–1837.

Doran, M. M., Hoffman, J. E., & Scholl, B. J. (2009). The role of eye fixations in concentration and amplification effects during multiple object tracking. *Visual Cognition, 17*(4), 574–597.

Eriksen, C. W., & Spencer, T. (1969). Rate of information processing in visual perception: Some results and methodological considerations. *Journal of Experimental Psychology, 79*, 1–16.

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

Fehd, H. M., & Seiffert, A. E. (2010). Looking at the center of the targets helps multiple object tracking. *Journal of Vision, 10*(4). http://dx.doi.org/10.1167/10.4.19.

Findlay, J. M., & Gilchrist, I. D. (2003). *Active vision: The psychology of looking and seeing*. Oxford: Oxford University Press.

Franconeri, S. L., Jonathan, S. V., & Scimeca, J. M. (2010). Tracking multiple objects is limited only by object spacing, not by speed, time, or capacity. *Psychological Science, 21*, 920–925.

Hoffman, J. E., & Subramaniam, B. (1995). The role of visual attention in saccadic eye movements. *Perception & Psychophysics, 57*, 787–795.

Holcombe, A. O., & Chen, W. Y. (2012). Exhausting attentional tracking resources with a single fast-moving object. *Cognition, 123*, 218–228.

Holcombe, A. O., & Chen, W. Y. (2013). Splitting attention reduces temporal resolution from 7 Hz for tracking one object to <3 Hz when tracking three. *Journal of Vision, 13*, 1–19.

Horowitz, T., Klieger, S., Fencsik, D. E., Yang, K. K., Alvarez, G. A., & Wolfe, J. M. (2007). Tracking unique objects. *Perception & Psychophysics, 69*, 172–184.

Horrey, W. J., Wickens, C. D., & Consalus, K. P. (2006). Modeling drivers' visual attention allocation while interacting with in-vehicle technologies. *Journal of Experimental Psychology: Applied, 12*, 67–78.

Howe, P. D., & Ferguson, A. (2015). The identity–location binding problem. *Cognitive Science, 39*, 1622–1645.

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

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

Hyönä, J., Tommola, J., & Alaja, A. M. (1995). Pupil dilation as a measure of processing load in simultaneous interpretation and other language tasks. *Quarterly Journal of Experimental Psychology, 48*, 598–612.

Irwin, D. E., & Thomas, L. E. (2010). Eyeblinks and cognition. In V. Coltheart (Ed.), *Tutorials in visual cognition* (pp. 121–141). New York, NY: Psychology Press.

Kahneman, D., Treisman, A., & Gibbs, B. J. (1992). The reviewing of object files: Object-specific integration of information. *Cognitive Psychology, 24*, 175–219.

Kazanovich, Y. B., & Borisyuk, R. M. (2006). An oscillatory neural model of multiple object tracking. *Neural Computation, 18*, 1413–1440.

Kowler, E., Anderson, E., Dosher, B., & Blaser, E. (1995). The role of attention in the programming of saccades. *Vision Research, 35*, 1897–1916.

Kramer, A. F. (1991). Physiological metrics of mental workload: A review of recent progress. In D. Damos (Ed.), *Multiple task performance* (pp. 279–328). London: Taylor & Francis.

Landry, S., Sheridan, T., & Yufik, Y. (2001). A methodology for studying cognitive groupings in a target-tracking task. *IEEE Transactions on Intelligent Transportation Systems, 2*, 92–100.

Leek, E. C., Patterson, C., Paul, M. A., Rafal, R., & Cristino, F. (2012). Eye movements during object recognition in visual agnosia. *Neuropsychologia, 50*, 2142–2153.

Leibowitz, H. W., & Post, R. B. (1982). The two modes of processing concept and some implications. In J. Beck (Ed.), *Organization and representation in perception* (pp. 343–363). Hillsdale, NJ: Erlbaum.

Logan, D. G. (2002). Parallel and serial processing. In H. Pashler & J. Wixted (Eds.), *Stevens' handbook of experimental psychology. Methodology in experimental psychology* (pp. 271–300). NJ: John Wiley & Sons.

Milner, A. D., & Goodale, M. A. (1993). *The visual brain in action*. New York: Oxford University Press.

Moray, N. (1986). Monitoring behavior and supervisory control. In K. R. Boff, L. Kaufman, & J. P. Thomas (Eds.). *Handbook of perception and human performance* (Vol. 2, pp. 40.1–40.51). New York: Wiley.

Moray, N. (1984). Attention to dynamic visual displays in man–machine systems. In R. Parasuraman & D. Davies (Eds.), *Varieties of attention* (pp. 485–513). Orlando: Academic Press.

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*, 631–671.

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

Papenmeier, F., Meyerhoff, H. S., Jahn, G., & Huff, M. (2013). Tracking by location and features: Object correspondence across spatiotemporal discontinuities during multiple object tracking. *Journal of Experimental Psychology: Human Perception and Performance, 40*, 159–171.

Pinto, Y., Howe, P. D., Cohen, M. A., & Horowitz, T. S. (2010). The more often you see an object, the easier it becomes to track it. *Journal of Vision, 10*, 1–15.

Previc, F. H. (1998). The neuropsychology of 3-D space. *Psychological Bulletin, 124*, 123–164.

Pylyshyn, Z. W. (1989). The role of location indexes in spatial perception: A sketch of the FINST spatial-index model. *Cognition, 32*, 65–97.

Pylyshyn, Z. W. (2001). Visual indexes, preconceptual objects, and situated vision. *Cognition, 80*, 127–158.

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

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

Rayner, K. (2009). Eye movements and attention in reading, scene perception, and visual search. *Quarterly Journal of Experimental Psychology, 62*, 1475–1506.

Ren, D., Chen, W., Liu, C. H., & Fu, X. (2009). Identity processing in multiple-face tracking. *Journal of Vision, 9*(5). http://dx.doi.org/10.1167/9.5.18.

Saiki, J. (2003a). Feature binding in object-file representations of multiple moving items. *Journal of Vision, 3*, 6–21.

Saiki, J. (2003b). Spatiotemporal characteristics of dynamic feature binding in visual working memory. *Vision Research, 43*, 2107–2123.

Schneider, W., Eschman, A., & Zuccolotto, A. (2002a). *E-prime user's guide*. Pittsburgh: Psychology Software Tools Inc.

Schneider, W., Eschman, A., & Zuccolotto, A. (2002b). *E-prime reference guide*. Pittsburgh: Psychology Software Tools Inc.

Senders, J. (1964). The human operator as a monitor and controller of multidegree of freedom systems. *IEEE Transactions on Human Factors in Electronics, HFE-5*, 2–6.

Senders, J. W. (1983). *Visual sampling processes*. Hillsdale, NJ: Erlbaum.

Shepherd, M., Findlay, J. M., & Hockey, R. J. (1986). The relationship between eye movements and spatial attention. *Quarterly Journal of Experimental Psychology, 38*, 475–491.

Sheridan, T. (1970). On how often the supervisor should sample. *IEEE Transactions on Systems Science and Cybernetics, SSC-6*, 140–145.

Shiffrin, R. M., & Gardner, G. T. (1972). Visual processing capacity and attentional control. *Journal of Experimental Psychology, 93*(1), 72–83.

Sirevaag, E. J., Kramer, A. F., Reisweber, C. D. W. M., Strayer, D. L., & Grenell, J. F. (1993). Assessment of pilot performance and mental workload in rotary wing aircraft. *Ergonomics, 36*, 1121–1140.

Snodgrass, J. G., & Vanderwart, M. (1980). A standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity and visual complexity. *Journal of Experimental Psychology: Human Learning and Memory, 6*, 174–215.

Thomas, L. E., & Irwin, D. E. (2006). Voluntary eyeblinks disrupt iconic memory. *Perception & Psychophysics, 68*, 475–488.

Townsend, J. T., & Nozawa, G. (1995). Spatio-temporal properties of elementary perception: An investigation of parallel, serial, and coactive theories. *Journal of Mathematical Psychology, 39*(4), 321–359.

Townsend, J. T., & Wenger, M. J. (2004a). The serial-parallel dilemma: A case study in a linkage of theory and method. *Psychonomic Bulletin & Review, 11*, 391–418.

Townsend, J. T., & Wenger, M. J. (2004b). A theory of interactive parallel processing: New capacity measures and predictions for a response time inequality series. *Psychological Review, 111*(4), 1003.

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 & psychological essays* (pp. 117–144). Oxford: Oxford University Press.

Ungerleider, L. G., & Mishkin, M. (1982). Two cortical visual systems. In D. J. Ingle, M. A. Goodale, & R. J. W. Mansfield (Eds.), *Analysis of visual behavior* (pp. 549–587). Cambridge, MA: MIT Press.

Wilson, G., & Eggemeier, T. (1991). Psychophysiological assessment of workload in multi-task environments. In D. Damos (Ed.), *Multiple task performance* (pp. 329–360). London: Taylor & Francis.

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

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