Eye movements in repeated multiple object tracking

Jiří Lukavský

Institute of Psychology,
Academy of Sciences of the Czech Republic,
Prague, Czech Republic

Contrary to other tasks (free viewing, recognition, visual search), participants often fail to recognize repetition of trials in multiple object tracking (MOT). This study examines the intra- and interindividual variability of eye movements in repeated MOT trials along with the adherence of eye movements to the previously described strategies. I collected eye movement data from 20 subjects during 64 MOT trials at slow speed (5°/s). Half of the trials were repeated four times, and the remaining trials were unique. I measured the variability of eye-movement patterns during repeated trials using normalized scanpath saliency extended to the temporal domain. People tended to make similar eye movements during repeated presentations (with no or vague feeling of repetition) and the interindividual similarity remained at the same level over time. Several strategies (centroid strategy and its variants) were compared with data and they accounted for 48.8% to 54.3% of eye-movement variability, which was less then variability explained by other peoples' eye movements (68.6%). The results show that the observed intra- and interindividual similarity of eye movements is only partly explained by the current models.

Introduction

Because high-acuity vision is limited to only a small part of retina, humans constantly redirect eyes to objects of their interest. The factors underlying this highly variable behavior have been of great interest to vision scientists for several decades.

In recent years, models based on saliency have become popular, most likely because of the emergence of quantitative models (e.g., Itti & Koch, 2000) that allow comparisons between human performance and predicted models solely based on low-level image features with no or little respect to the content or meaning of the scene. However, as Tatler (2009) noted, there are several problematic aspects of this approach. First, the correlations between prediction maps and human performance are weak (e.g., Einhäuser, Spain, & Perona, 2008; Nyström & Holmqvist, 2008; Tatler, Baddeley, & Gilchrist, 2005). Second, a growing number of studies show that low-level models fail if the task is varied (e.g., Einhäuser, Rutishauser, & Koch, 2008; Foulsham & Underwood, 2008; Rothkopf, Ballard, & Hayhoe, 2007).

The goal of this study is to measure the consistency of eye movements in a simple repeatable task. In eye-tracking studies, subjects are usually asked to simply look at a scene (free viewing), remember a scene (recognition), or search for a specific object (visual search). Typically, performance across various scenes and subjects is evaluated. The problem with these tasks is that subjects cannot perform them meaningfully in a repeated manner because subjects will recognize the scene and look for previously unexplored areas of the scene (in case of free viewing), encode additional details (for later recognition), or find targets more efficiently based on their previous experience. Similarly, Raney and Rayner (1995) found that reading a passage of text for a second time affects eye movements. However, to successfully model eye movements, it is useful to compare within-subject performance for particular stimuli and sort out fixations that are important for the task and those that are less likely to be repeated (potential errors or noise).

Most likely due to difficulties related to finding meaningful tasks, there are few studies of eye movements during repeated presentations of visual stimuli. Studies of contextual cueing (Chun & Jiang, 1998) showed that implicit learning, which manifests as reduced reaction times with more experience, takes place during repeated presentations of visual search trials. The contextual cueing is expected to efficiently guide attentional deployment and thus affect eye movements. Võ and Wolfe (2012) showed that repeated searches for the same object in a scene led to faster responses and the search space substantially decreased. Additionally, they claimed that looking at the target objects and scenes in a preview did not increase the latter search performance. However, Hollingworth

Citation: Lukavský, J. (2013). Eye movements in repeated multiple object tracking. *Journal of Vision*, *13*(7):9, 1–16, http://www.journalofvision.org/content/13/7/9, doi:10.1167/13.7.9.

doi: 10.1167/13.7.9 Received November 19, 2012; published June 17, 2013 ISSN 1534-7362 © 2013 ARVO

(2012) confirmed the preview effect using more sensitive within-subject design.

In visual search or free viewing, subjects scan the scene and may differ not only in the locations on which they fixate but also in the temporal sequence of fixations. It is difficult to determine which differences in fixation order are random and which are influenced by experimental conditions (e.g., change in guidance).

However, in tasks using dynamic stimuli (e.g., free viewing a movie), variations in fixation order are more likely to be significant, because they are attracted to different content. Recently, Dorr, Martinetz, Gegenfurtner, and Barth (2010) compared variability in free viewing in dynamic natural scenes. Their results showed high coherence between scan patterns during repeated presentations in both movie trailers and dynamic natural scenes. The highest coherence was observed in the first presentation of each experimental session; coherence decreased during later presentations throughout the day. Dorr et al. (2010) suggested that the decrease in coherence is caused by rising influence of individual viewing strategies. Additionally, the task is perhaps unclear when subjects watch a movie repeatedly.

This study investigates the similarity of eye movements between trials of a dynamic visual task called multiple object tracking (MOT). This paradigm (Pylyshyn & Storm, 1988) has been used in a number of studies of distributed attention (e.g., Alvarez & Cavanagh, 2005; Horowitz et al., 2007) or visual object qualities (for a review see Scholl, 2009). In MOT tasks, the subject is presented with a set of uniform objects and asked to track a specified subset (one to four objects). After a short period of motion, tracking performance is evaluated. My pilot experiments have shown that subjects do not recognize when an MOT trial is administered repeatedly; this phenomenon makes MOT a good candidate for studying eye movements in repeated visual presentations. Ogawa, Watanabe, and Yagi (2009) investigated changes in performance during repeated MOT tasks and reported 22%–31% recognition rates in their experiments using 15 trial repetitions.

In the majority of MOT studies, subjects are either asked to fixate on the center of the screen and track objects peripherally or they are allowed to move their eyes freely. Recent studies of eye movements during MOT (Fehd & Seiffert, 2008, 2010; Huff, Papenmeier, Jahn, & Hesse, 2010; Zelinsky & Neider, 2008) identified several gaze strategies: target looking and centroid looking. The use of these strategies depends on several factors, especially the number of tracked targets. Zelinsky and Neider (2008) reported that target looking is the main strategy for tracking a single target, while centroid looking is the main strategy for tracking two targets. With three targets, both strategies are used equally, and with four targets, subjects tend to switch between targets rather than looking at the centroid. Fehd and Seiffert (2008) reported a high preference for centroid looking in MOT tasks using three to five targets (41.6%–42.4% of the tracking time) with target looking being used much less frequently (8.1%–10.7%). Fehd and Seiffert (2010) compared the combined center/target-looking strategy to the target-looking strategy and found significantly decreased performance in the latter. They suggested that the target-looking strategy is used mainly to prevent crowding and showed that at the time of center-to-target gaze shift the distractors are closer to targets compared to the target-to-center gaze shifts. Huff et al. (2010) confirmed the preference for centroid strategy in three targets using stricter gaze classification. They found people are faster to make a saccade toward the centroid than targets when the task is interrupted with an abrupt viewpoint change.

Previous studies referred to the term centroid, which is ambiguous. Some studies (Fehd & Seiffert, 2008, 2010) defined the centroid as the center of mass of the object formed by the targets. Zelinsky and Neider (2008) defined the centroid as an averaged spatial position of targets. Fehd and Seiffert (2008) point out that these two methods yield identical results for three targets but differ with higher number of targets. Both methods are methodologically sound. Centroid of the object (here referred as *object centroid*) has been used in previous studies and represents the idea that people might track the targets as a single object. However, there is a problem with centroid calculation if targets form a concave shape—one target is inside of the triangle formed by other three targets—because the object can be defined in three possible ways (see Figure 1). To resolve this ambiguity, I used centroid of convex hull for concave configurations. The averaged spatial position (*target centroid*) is a sound method to calculate center of mass for finite set of points and it also minimizes the sum of squared Euclidean distances between itself and each target.

In a recent review, Schütz, Braun, and Gegenfurtner (2011) suggested a four-layer model for the control of saccadic eye movements based on a similar model of action-perception loops described by Fuster (2004). The proposed four layers are salience, object recognition, value, and plans. For MOT, it is difficult to fit the tracking task to a single layer. Eye movement is influenced by the content of the scene (distribution of the objects); subjects likely try to optimize their viewing position with respect to the spatial resolution of attention (Intriligator & Cavanagh, 2001) and crowding. Eye-movement behavior may be of reflexive oculomotor nature; however, its role in MOT tasks is well defined and very important. Thus, subjects are

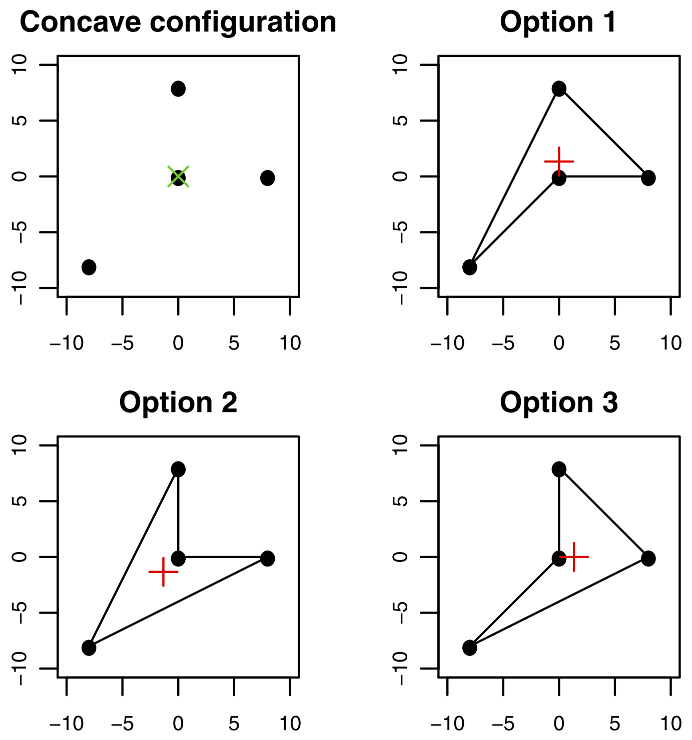

Figure 1. Problem with the definition of the object centroid when targets form a concave configuration. The object shape can be interpreted in three possible ways (Options 1–3), each yielding a different prediction of centroid (red cross). This ambiguity can be resolved by using a centroid of convex hull (green cross) for concave configurations.

likely to control their eye movements and adopt some tracking strategies.

In the current study, I measured the interindividual and inter-trial similarity of eye movements during MOT. Four different types of comparisons were performed. First, same-subject-same-trial (SSST) comparisons were made to evaluate the similarity between eye movements of the same person in repeated visually identical tasks. Second, same-subject-different-trial (SSDT) comparisons were made to examine the similarity of eye movements of the same person across trials to evaluate potential individual strategies or biases (e.g., looking at the screen center). Third, different-subject-same-trial (DSST) comparisons were made to measure the similarity of eye movements of different people in a visually identical task. Finally, different subject different trial (DSDT) was used as a baseline condition to demonstrate the similarity of eye movements across different people in various trials.

The current study pursued two main goals. First, I wanted to evaluate the role of the task and the role of individual strategies in eye movements during MOT. Second, I wanted to compare the scan patterns with several models to evaluate how much variance can be explained by each model and how much variance remains unexplained when compared to interindividual (DSST) or intra-individual consistency (SSST). If there is a single model that predicts eye movements for all subjects, SSST consistency provides the upper bound for the performance of the model. Additionally, DSST consistency shows how successfully scan patterns in MOT can be predicted based on the scan patterns of other people.

Method

Participants

Twenty students from Charles University participated in this experiment for course credit. All participants had normal or corrected-to-normal vision. The mean age was 22.1 years. All experiments conformed to National and Institutional Guidelines for experiments in human subjects and with the Declaration of Helsinki. Data from 28 participants were originally collected, but data of eight participants were removed (six due to calibration problems or large numbers of blinks and two because of very low tracking accuracy).

Apparatus and stimuli

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

Tracking stimuli comprised of eight mid-gray disks (RGB value: 128 128 128) on a black background (RGB value: 0 0 0). Each disk subtended 1° of visual angle and moved at a fixed speed of 5°/s. The disks were confined to move within an invisible square (20° × 20°) and bounce off the invisible square boundaries and one another with a minimum interobject distance of 0.5°. Besides bouncing the direction of disks was sampled from von Mises (circular Gaussian) distribution with parameter $\kappa = 10$ every 100 ms, creating an impression of Brownian motion (Supplementary Video shows the repeated trials including the recorded gaze positions).

Eye position was recorded at 250 Hz using Eyelink II eye tracker (SR Research, Canada). Drift correction was performed before each trial, and a nine-point calibration was completed after every 16 trials.

Design and procedure

After six training trials, each subject completed 64 experiment trials. In each trial, participants were asked to fixate on a central point (used for drift correction), and then eight objects were presented. Four target objects were highlighted with a light green color for 2 s (RGB value: 0 255 0), and then all objects turned mid-gray and started moving. After 10 s, objects stopped moving and participants selected four tracked objects with the mouse. After they made their choice green word "OK" or the number of errors in red was shown for 500 ms. The participants were instructed to track all four targets carefully and not to deliberately limit their tracking to a subset. They were told that the trial is considered correct only if all targets are successfully identified.

Experimental trials were presented in four blocks; calibration was completed at the beginning of each block. In each block, the odd-numbered trials were unique trials, which were presented only once during the experiment. The even-numbered trials were repeating trials, which were presented once in every block (i.e., four times over the course of the experiment). The trial order used in each condition was random; this semiregular structure was used to ensure that no two repeating trials would be presented consecutively. Thus, eight repeating trials were administered four times each throughout the experiment.

The experiment consisted of 40 trajectories generated in advance. Two protocol versions were administered, which differed in the assignment of unique/repeating tracks—the eight repeating trajectories in Version 2 were unique trajectories in Version 1. Thus the data about sixteen different repeating trajectories were collected. To minimize the chance participants would remember a particular start or end configuration of objects, all prepared trajectories were 12 s long and their 10-s presentation began in a random time between 0 and 2 s, which means that all presentations of the same trajectory across repetitions or participants share a common segment of 8 s (2–10 s).

The experiment lasted approximately 45 min, after which participants were asked about their tracking strategies, asked whether they noticed the repetition, and given a recognition test. The recognition test consisted of eight repeating trajectories, eight unique trajectories, and eight novel trajectories in random order. The participants were asked whether they had seen the presented trial in the experiment (irrespective of number of possible occurrences).

Data analysis

Blink detection

Recorded eye-movement data preceding and following a blink tend to contain artifacts, because the pupil is partially occluded by the eyelid. These artifacts (seen as sudden eye movements downwards and upwards) were partially detected automatically and removed. All data were checked manually to remove remaining blink artifacts. Trials containing more than 10% missing data were discarded. A total of 1,260 trials (98.4%) were included in the analysis (20 trials were discarded because of blinks, missing data, or technical problems).

Data preparation

The collected eye-movement data ranged from −10° to +10° in horizontal and vertical dimensions and from 2 to 10 s in time (time segment shared across all presentations). Because subjects often track objects using smooth pursuit eye movements during MOT, this type of movement is also of interest. Therefore, no fixation detection was performed, and all data samples were included. To facilitate further analyses, the data for each subject and trial were binned in a 3-D spatiotemporal matrix; bin size was 0.25° × 0.25° × 10 ms. The value of each bin represented the number of corresponding eye-movement samples.

To compare the changes in type of eye movements during the experiment, saccades were identified in the recorded data in a two-step procedure. First, the samples with velocity exceeding a high threshold (100°/s) were considered as saccade candidates. Then, the saccade onset and offset were found by adding all adjacent samples exceeding a lower threshold (17°/s).

Normalized scanpath saliency

Scan pattern similarity was measured with normalized scanpath saliency (NSS) similar to the procedure used by Dorr et al. (2010). In this procedure (see Figure 2) the similarity within a group of scan patterns is evaluated using a "leave-one-out" procedure used in machine learning: In each step one scan pattern is compared with the saliency map constructed using all other scan patterns. In other words, we estimate how well one scan pattern can be predicted based on the remaining scan patterns. This process is repeated and the similarity within a group of scan patterns is calculated as mean similarity found in each step.

Specifically, let us suppose that in a given step we compare Scan Pattern A with other patterns B, C, and D. First, the spatiotemporal saliency map was constructed. Eye-movement samples from B to D were transformed into a single spatiotemporal data matrix (see Data preparation) and convolved using a spatiotemporal Gaussian filter ($\sigma_x = \sigma_y = 1.2°$ and $\sigma_t = 26.25$ ms). These parameters were utilized to match the procedure used by Dorr et al. (2010), but to ensure the analysis is not dependent on the particular values,

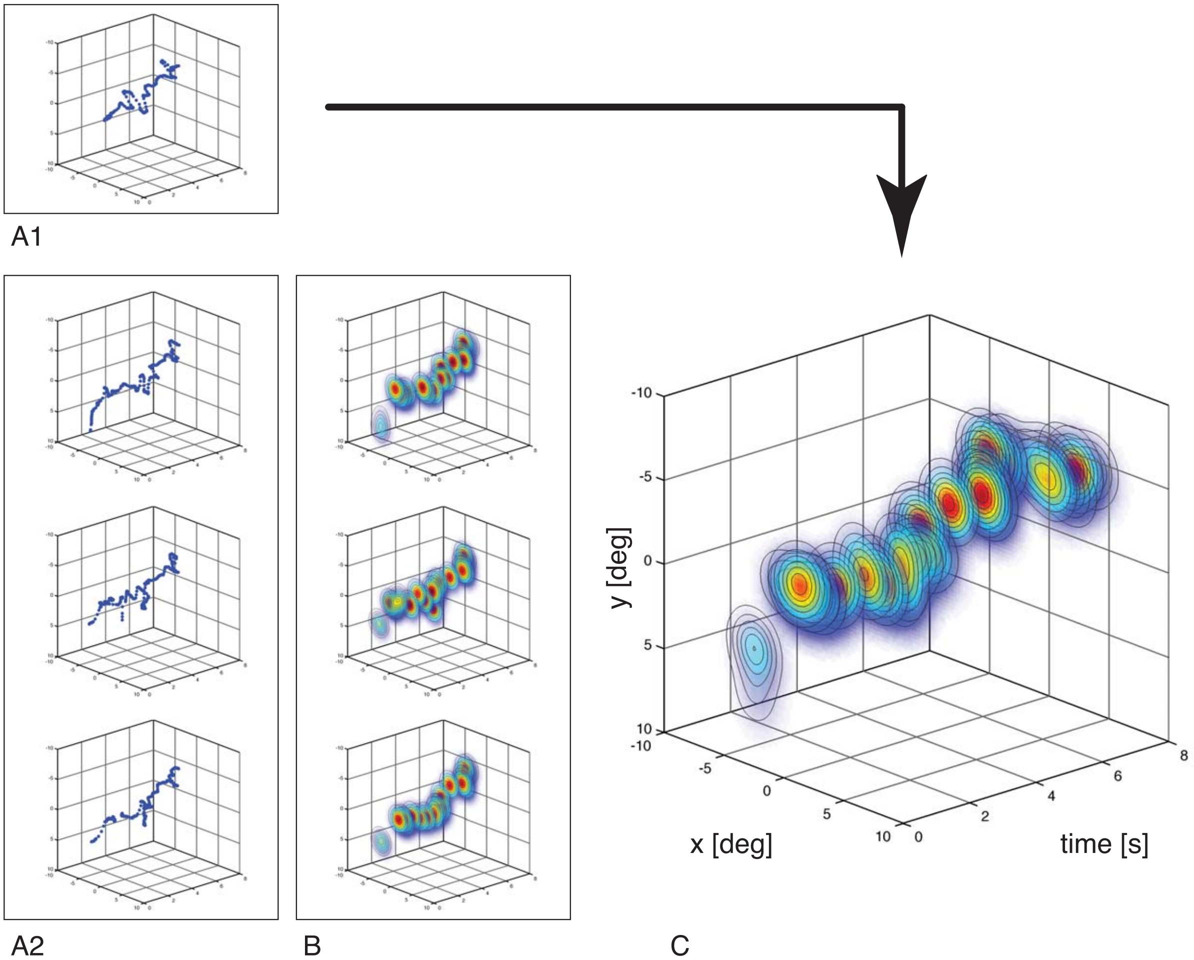

Figure 2. The procedure comparing one target scan pattern (A1) with other three (A2). First, eye-movement samples were binned to a spatiotemporal matrix (bin size 0.25° × 0.25° × 50 ms) and convolved using a three-dimensional Gaussian filter (B). Finally, the spatiotemporal matrices were summed and normalized into a spatiotemporal saliency map (C) and the partial NSS coherence index was calculated as the mean of the values corresponding to the time-places of the target scan pattern (A1). This process was repeated for each of the four scan patterns (the leave-one-out method) and the final NSS coherence index was measured as the mean of the four partial indices.

additional analyses were made using half-sized filter ($\sigma_x = \sigma_y = 0.6°$, $\sigma_t = 13.125$ ms) and double-sized filter ($\sigma_x = \sigma_y = 2.4°$, $\sigma_t = 53.0$ ms), and both led to the same pattern of results.

Second, the obtained spatiotemporal matrix was normalized. The similarity index for a given step was calculated as the mean of the saliency map values, which corresponded to the time-places of the Scan Pattern A. This process was repeated with Scan Pattern B compared to A, C, and D, etc. Finally, the NSS score representing the similarity within the group was calculated as the mean of similarity index found over all steps.

The NSS score is influenced by the number of scan patterns used to construct the saliency map. Using more scan patterns creates a saliency map that is finer with values distributed over larger areas. The peak scores are lower, but in general, a larger proportion of gaze samples to a larger area is considered in concordance with the saliency map. Because the number of possible scan patterns included in the analysis varied (from 4 for SSST, 20 for DSST, to hundreds for DSDT baseline), I performed the analysis first using all possible scan patterns. Later, I sampled only four scan patterns. The first approach yielded mean NSS scores that were on average 0.42 greater (range: 0.28–0.51). In the text to follow, when comparing eye-movement similarity in different comparison perspectives, I report the results obtained via sampling four scan patterns to account for these differences.

Later NSS is used to test the coherence of four repeated trials with a scan pattern based on a strategy, i.e., five scan patterns are compared instead of four. To investigate whether it is possible to compare the values based on four scan patterns (e.g., SSST) and five scan

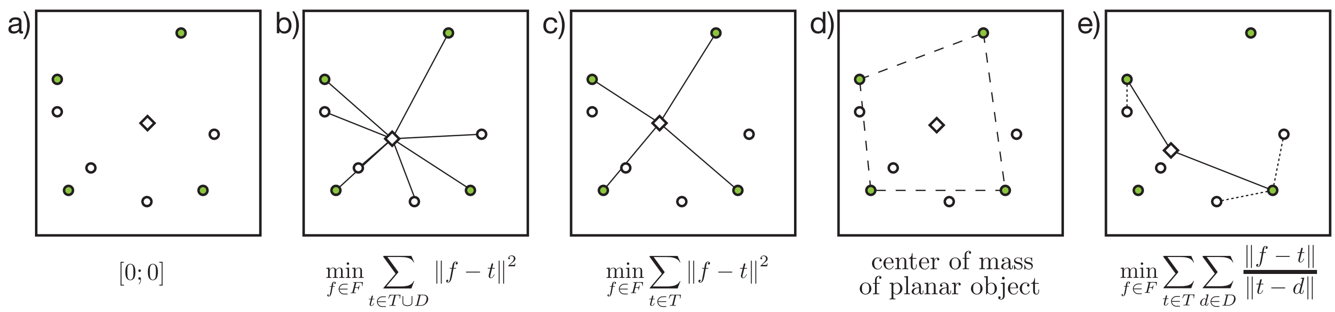
a)
b)
c)
d)
e)
[0; 0]
$\min_{f\in F}\sum_{t\in T\cup D}\|f-t\|^2$
$\min_{f\in F}\sum_{t\in T}\|f-t\|^2$
center of mass
of planar object
$\min_{f\in F}\sum_{t\in T}\sum_{d\in D}\frac{\|f-t\|}{\|t-d\|}$

Figure 3. Strategies used in the analysis. (a) constant strategy—looking at the center of the screen, (b) all-points centroid—minimizing the distance to all objects, (c) target centroid—minimizing the distance to targets, (d) object centroid—center of mass of planar object defined by targets or their convex hull, and (e) anticrowding—minimizing the ratio of the distance of each target to the gaze point and distractors. Targets are shown as green circles, distractors as white circles, and the predicted gaze point as a diamond.

patterns (coherence with a strategy), I compared the NSS values based on sampling four and five patterns. With more patterns involved, NSS values are significantly larger, *diff* = 0.036, $t(1739) = 2.379$, $p = 0.017$, Cohen $d = 0.03 \pm 0.07$, however the effect is very small and thus I consider the NSS values comparable.

For SSST comparisons, the final NSS coherence index was calculated as the mean of four related partial coherence indexes from the leave-one-out procedure described above. For the other comparisons (DSST, SSDT, DSDT), I calculated only the partial NSS coherence index (i.e., how the scan pattern in the particular block can be predicted by scan patterns of other people or scan patterns from other trials) and calculated the mean for all four repetitions of each trial.

Strategies

Five strategies were included in the analysis, and their coherence with the data was evaluated (see Figure 3). The *constant* strategy was a simple strategy of looking at the center of the screen, which may indicate attentional tracking targets without eye movements. This strategy was the only strategy that predicted the same (constant) pattern of eye movements for all trials. In the *all-points-centroid* strategy, observers attempted to minimize the eccentricity of each object falling on the fovea (both targets and distractors). Similarly, in the *target-centroid* strategy, observers attempted to minimize the eccentricity of targets while ignoring distractors. In both of the latter strategies, eccentricity minimization was operationalized as a search for the minimum sum of the squared distances between the gaze point and the object positions and could be easily calculated by averaging the spatial positions of points (similarly to Zelinsky & Neider, 2008). In the *object-centroid* strategy, observers tracked the centroid of the planar object formed by four targets (Fehd & Seiffert, 2008, 2010). When one target was located inside the triangle formed by other three targets, the object centroid was defined as the centroid of the convex hull (i.e., triangle). In the *anticrowding* strategy, observers attempted to minimize the ratio between each target's distance from the gaze point and distance from every distractor. The idea behind this strategy was that the observer would try to minimize the effect of crowding and thus reduce the danger of losing a target. Despite the majority of previous studies have reported the use of centroid strategy, where the positions of distractors are ignored, recent studies show people take crowding into account. Iordanescu, Grabowecky, and Suzuki (2009) found that people are more precise in localizing targets in crowded situations. Similarly, Zelinsky and Todor (2010) reported that people tend to look closer to the targets that are in the proximity of other objects.

Each strategy adds additional information from the scene. The constant strategy uses information about the position of the bounding rectangle with no respect to its content; the all-points-centroid strategy reflects object motion but does not distinguish targets and distractors, which the target-centroid, object-centroid, and anticrowding strategies do. The anticrowding strategy recognizes targets and employs additional expectations about the human ability to distinguish objects in the periphery of the visual field. I expected the lowest performance in constant strategy, followed by all-points-centroid, with all other strategies performing better with no a priori expectations about their order. The aim of this study was to estimate the variance explained by these models compared to the reliability of eye movements in observers, not to differentiate between the models.

Results

Tracking performance

Accuracy was defined as the percentage of trials in which participant correctly identified all four targets (chance = 1.43%). The average percent correct across all participants was 91% (ranging from 77% to 100%). The accuracy did not depend on either time (order of experimental block) or condition (repeating or unique): within-subject analysis of variance (ANOVA) on accuracy showed no significant effects of experimental block, $F(3, 57) = 1.34$, $p = 0.271$, $\eta_G^2 = 0.027$, condition, $F(1, 19) = 1.73$, $p = 0.203$, $\eta_G^2 = 0.009$, or their interaction, $F(3, 57) = 2.16$; $p = 0.103$, $\eta_G^2 = 0.033$.

For the following analysis I used data from the trials where participants correctly identified all targets (1,148 of 1,280 trials). The only exception is SSST comparison, where I included incorrect trials to retain four trials in each NSS comparison (1,260 of 1,280 trials) and later estimated how much it biased the results.

Effect of repetition

After experiment 7 of 20 participants reported they felt that some trials were repeated. In the recognition test participants were presented with 24 trials, which belonged to one of three conditions: (a) repeated trials (seen four times), (b) unique trials (seen once), (c) novel trials (never seen). Participants correctly reported they had seen 47% of repeating trials and 39% of unique trials. However the false alarm rate—reporting novel trials as previously seen—was very high (44%).

Within-subject ANOVA showed that a participant reporting a trial as previously seen was not influenced by the condition (i.e., whether the trial had been presented four times, once or not at all), $F(2, 38) = 0.80$, $p = 0.455$, $\eta_G^2 = 0.020$. There was no difference in recognition performance between participants who reported they noticed some repetitions and those who did not, ANOVA with additional binary factor, reported repetitions, $F(1, 18) = 0.004$, $p = 0.952$, $\eta_G^2 = 0.000$; condition: $F(2, 36) = 0.795$, $p = 0.460$, $\eta_G^2 = 0.021$; no significant interaction, $F(2, 36) = 0.792$, $p = 0.461$, $\eta_G^2 = 0.021$. Both groups did not differ in their tracking performance either, $t(15.01) = 0.382$, $p = 0.708$, $d = 0.17$ ± 0.92.

Eye-movement similarity

The results show that task (trial) played a more dominant role compared to subject and his/her individual strategies (see Figure 4). The NSS coherence for the four possible comparison types was analyzed using repeated measures ANOVA with comparison type as the single categorical predictor and observations matched for subject and trial combinations. Three planned contrasts were performed to evaluate the effect of task/trial (SSST vs. SSDT), the effect of interpersonal differences (SSST vs. DSST), and whether coherence in the same subject across different trials differed from baseline (SSDT vs. DSDT).

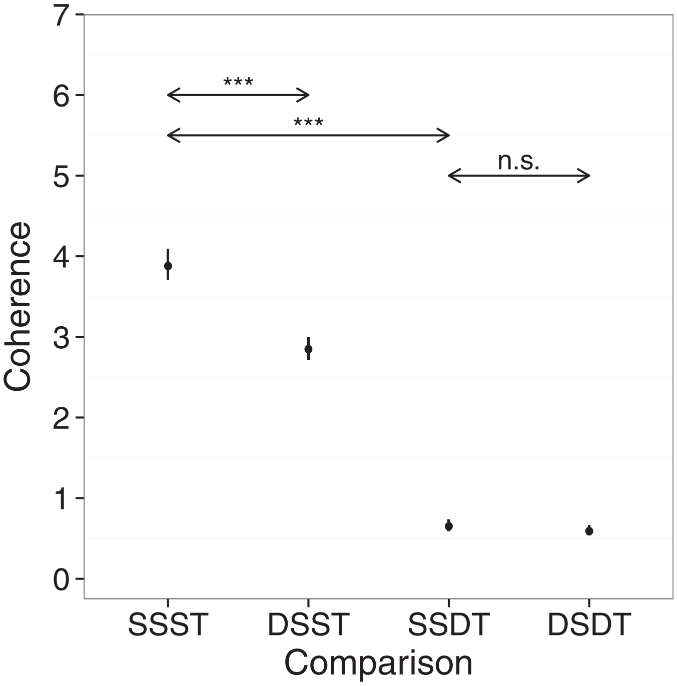

Figure 4. Normalized scanpath saliency scores for different comparisons of eye-movement similarity. SSST = same subject same trial, DSDT = different subject different trial. Error bars show 95% confidence intervals.

The analysis showed that coherence depended on comparison type, $F(3, 477) = 835.15$, $p < 0.001$, $\eta_G^2 = 0.754$. As expected, the greatest amount of coherence was found in intra-individual performance while watching identical trials repeatedly (SSST: $M = 3.89$, $SD = 1.19$). The second greatest amount of coherence was found in different subjects while watching the same trials (DSST: $M = 2.86$, $SD = 0.91$). The difference between coherence in SSST and DSST was significant, $t(159) = 12.96$, $p < 0.001$, Cohen $d = 0.97$ ± 0.23, which means that the interindividual coherence within a trial was significantly lower than the intra-individual coherence (*diff* = 1.03). Coherence in the same subject watching different trials was much lower (SSDT: $M = 0.66$, $SD = 0.47$) and significantly different from coherence in the SSST comparison, *diff* = 3.23, $t(159) = 34.02$, $p < 0.001$, $d = 3.57$ ± 0.35. SSDT coherence did not differ significantly from the baseline, *diff* = 0.06, $t(159) = 1.46$, $p = 0.145$, $d = 0.14$ ± 0.22; baseline DSDT: $M = 0.60$; $SD = 0.42$. The lack of significance may indicate that in the current experimental conditions, subjects actually tracked objects using eye movements; if subjects fixated on the center of the screen and tracked the targets using only their visual attention, eye-movement coherence of unrelated trials would be higher.

I also analyzed the intra-individual variability of the NSS coherence across trials. I limited the scope of the analysis to the SSST and analyzed the differences in two separate one-way ANOVAs. The first ANOVA used subject as a categorical predictor, and the second ANOVA used trial id as a categorical predictor (with 16 levels). I found significant differences in intra-individual coherence across subjects ranging from 1.65 to 5.37, $M = 3.89$, $SD = 0.92$, $F(19, 140) = 9.702$, $p < 0.001$, $\eta^2 = 0.568$. In other words, some subjects were more likely to repeat a pattern of their eye movements than other subjects. The differences in intra-individual coherence between different trials were not significant, $F(15, 144) = 1.061$, $p = 0.398$, $\eta^2 = 0.100$. A significant result would mean that some trials elicit more coherent eye movements than others (given the selection of trials in the present experiment).

I examined whether the eye movements of different people became more similar or diverse over time. The mean NSS value from the DSST comparison for each block ranged from 3.04 (Block 1) to 3.38 (Block 2). To estimate the effect of time I averaged the corresponding NSS values from the first two blocks and last two blocks and compared them with paired *t* test. No significant changes in interindividual eye-movement similarity were observed over time, $t(31) = 0.541$, $p = 0.593$, $d = 0.07 \pm 0.49$.

In order to find out whether the repetition affects the type of eye movements, I analyzed the changes in number of saccades, total time occupied by them, and their median length in both repeating and unique trials. During average trial participants made 9.86 saccades ($SD = 5.58$) with median length 3.24° ($SD = 1.35$), which occupied 330 ms ($SD = 220$) of total 8 s. To estimate the effect of time I grouped the data from the first two blocks and last two blocks and used repeated measures ANOVA with two factors (condition: repeating/unique, and time). I found a significant decrease in number of saccades over time, $F(1, 19) = 5.178$, $p = 0.035$, $\eta_G{}^2 = 0.010$, no effect of condition, $F(1, 19) = 1.689$, $p = 0.209$, $\eta_G{}^2 = 0.001$, and significant interaction, $F(1, 19) = 7.092$, $p = 0.015$, $\eta_G{}^2 = 0.002$, showing the number of saccades decreases faster in unique trials. For the total time occupied by saccades there was a significant interaction, $F(1, 19) = 5.644$, $p = 0.028$, $\eta_G{}^2 = 0.001$, with no effect of time, $F(1, 19) = 3.727$, $p = 0.069$, $\eta_G{}^2 = 0.007$, or condition, $F(1, 19) = 1.336$, $p = 0.262$, $\eta_G{}^2 = 0.001$. There was no significant effect for median saccade length, time: $F(1, 19) = 0.460$, $p = 0.501$, $\eta_G{}^2 = 0.002$; condition: $F(1, 19) = 2.746$, $p = 0.114$, $\eta_G{}^2 = 0.008$; interaction: $F(1, 19) = 1.503$, $p = 0.235$, $\eta_G{}^2 = 0.002$. The results show a decrease in number of saccades during experiment, which was larger in unique trials. This decrease is partly mirrored in less total time occupied by saccades. The length of saccades did not change over time.

In SSST comparison incorrect trials were included to retain the number of trials in each NSS calculation and keep the results comparable. I compared the NSS values for SSST comparisons with and without any error. The majority of comparisons contained no error (122 of 160) and yielded slightly higher NSS values, $M = 3.99$, $SD = 1.18$, with errors: $M = 3.56$, $SD = 1.20$, $t(60.6) = 1.945$, $p = 0.056$, $d = 0.37 \pm 0.37$. Therefore adding incorrect trials lowered NSS coherence in SSST approximately by 0.1.

To provide additional insight into NSS values and their relationship to actual gaze positions, I repeated the analysis using a different approach based on gaze distances. I used the same comparisons as in the previous analysis, but I compared scan patterns using their mutual distances in every frame. For each frame (85 Hz) I calculated all distances between scan patterns involved in the comparison. Because in each frame several gaze samples (up to three) were present, I included only the first sample in the analysis. Then I calculated the mean, median, minimum, and maximum distance in each frame and the median values of these parameters over all frames in each trial. The following results are based on the analysis of mean distances, but all four parameters were highly correlated and yielded similar results.

The analysis based on gaze distances confirmed the results based on NSS (see Figure 5a). In SSST comparison the mean gaze distance was 2.51° ($SD = 1.07$), which was significantly smaller than in DSST comparison, $M = 3.09$, $SD = 0.71$, $t(159) = 12.117$, $p < 0.001$, Cohen $d = 0.93 \pm 0.23$. The mean gaze distances in SSDT comparison were not significantly different from DSDT condition, $M = 5.75$, $SD = 0.95$, $t(159) = 1.324$, $p = 0.188$, $d = 0.12 \pm 0.22$, and significantly larger than in SSST comparison, $t(159) = 32.927$, $p < 0.001$, $d = 3.45 \pm 0.35$.

Figure 5b shows the relationship between mean gaze distance and NSS measures. Despite differences in calculations and the limitations discussed later, both measures show good fit—NSS varied logarithmically with mean gaze distance ($R^2 = 0.885 \pm 0.017$).

Strategies

The four observed scan patterns for each subject and each trial were compared using NSS with idealized scan patterns representing the predicted eye movements for each strategy (constant, all-points centroid, target centroid, object centroid, anticrowding). Table 1 shows distances between predicted gaze positions for each strategy calculated across all frames and all 40 trajectories used in the experiment (both repeating and unique).

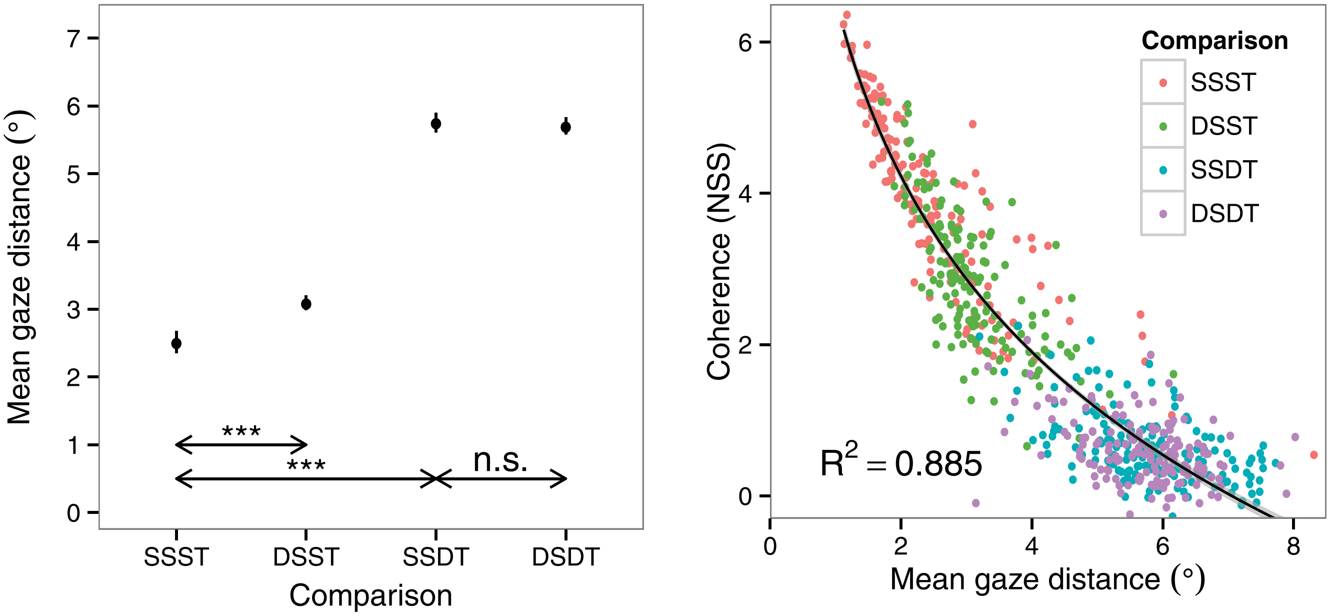

Figure 5. (a) Eye-movement similarity based on mean gaze distance for different comparisons. Error bars show 95% confidence intervals. (b) Normalized scanpath saliency varied logarithmically with mean gaze distance.

In 29% of all frames the targets were in a concave configuration and object centroid was calculated for the convex hull of the targets. As explained before in concave configurations the object shape can be interpreted in three possible ways (see Figure 1), each yielding a different prediction of centroid. The predictions can be far from each other—for each frame I evaluated their spread using maximum distance. The median spread across all frames with concave configurations was 3.58° (the first and third quartiles [2.81; 4.47], maximum 7.63).

Then, I compared the distances between the centroid of the convex hull and each of three possible predictions for every frame. I constructed three types of error estimates by selecting minimum, median, or maximum distance for each frame. For the optimistic estimate (minimum distance) the median error across all frames with a concave configuration was 0.37° (the first and third quartiles [0.18; 0.60], maximum 1.53). The balanced estimate based on the median distance yielded median error 1.18° ([0.92; 1.58], maximum 3.26). For the pessimistic estimate (maximum distance) the median error was 2.35° ([1.78; 2.98], maximum 6.13).

The coherence between the observed data in repeating trials and the predictions based on the selected strategies is shown in Figure 6. NSS results for baseline (DSDT) and intra-individual consistence (SSST) are added for comparison. I used a paired *t* test to compare the fit between different models in each subject and trial type. The reported *p* values have been adjusted for multiple comparisons using the Bonferroni correction (11 tests).

The coherence between the constant strategy and the data was significantly higher than DSDT baseline, $t(159) = 3.906$; $p = 0.001$; Cohen $d = 0.38 \pm 0.22$. As expected, adding extra information to the strategy improved its coherence with the data: the all-points-centroid strategy was better than the constant strategy, $t(159) = 9.809$; $p < 0.001$; $d = 0.86 \pm 0.23$, and all remaining strategies overperformed all-points-centroid model, object centroid: $t(159) = 6.487$; $p < 0.001$; $d = 0.51 \pm 0.22$; target centroid: $t(159) = 8.530$; $p < 0.001$; $d = 0.68 \pm 0.23$; anticrowding: $t(159) = 7.078$; $p < 0.001$; $d = 0.68 \pm 0.23$. However, the difference between these strategies and intra-individual coherence was still very large, object centroid: $t(159) = 14.21$; $p < 0.001$; $d = 1.42 \pm 0.25$; target centroid: $t(159) = 12.68$; $p <$

	All points centroid	Target centroid	Object centroid	Anticrowding
Constant	2.15 [1.37; 2.98]	3.06 [1.84; 4.47]	2.91 [1.90; 4.19]	4.33 [2.83; 5.99]
All-points centroid		2.36 [1.39; 3.22]	2.27 [1.37; 3.08]	3.20 [2.10; 4.46]
Target centroid			0.71 [0.42; 1.05]	2.64 [1.67; 3.95]
Object centroid				3.14 [1.96; 4.60]

Table 1. Median distances between predicted gaze positions for each pair of strategies. The first and third quartiles are shown in brackets.

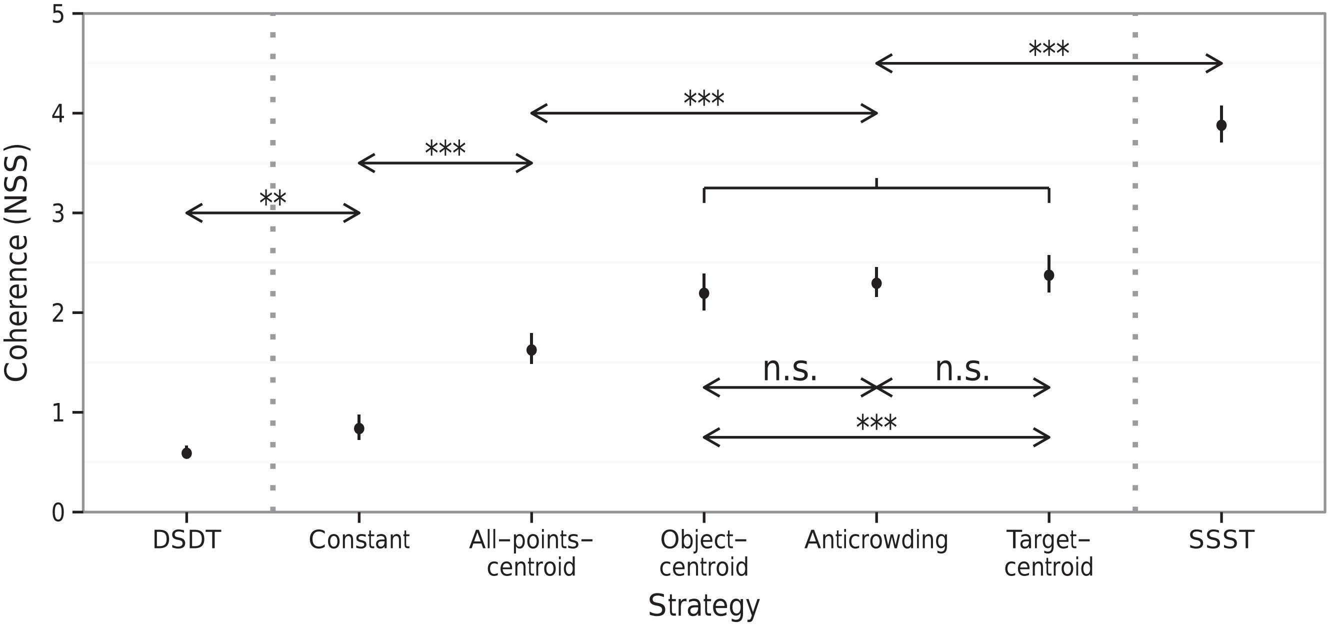

Figure 6. Coherence between selected eye-movement strategies in multiple object tracking and observed data. Coherence measured using NSS, differences tested with Student's *t* test and corrected for multiple comparisons. For clarity, the comparisons with object-centroid, anticrowding, and target-centroid strategies are grouped.

0.001; $d = 1.26 \pm 0.24$; anticrowding: $t(159) = 13.85$; $p < 0.001$; $d = 1.46 \pm 0.25$.

The differences between object centroid, target centroid, and anticrowding strategy were small, object-centroid versus anticrowding: $t(159) = 0.903$; uncorrected $p = 0.368$; $d = 0.09 \pm 0.22$; anticrowding versus target centroid: $t(159) = 0.783$; uncorrected $p = 0.435$; $d = 0.07 \pm 0.22$; object centroid versus target centroid: $t(159) = 6.174$; $p < 0.001$; $d = 0.15 \pm 0.22$. The absolute values of NSS coherence for these three strategies ranged from 2.22 ($SD = 1.19$) for object-centroid strategy to 2.39 ($SD = 1.20$) for target-centroid strategy.

The coherence of a given strategy may also be expressed relative to the variability in the observed data. I rescaled the NSS values to the scale from 0% (DSDT) to 100% (SSST) to calculate how much variance each strategy explained above the baseline condition using the maximum coherence observed in our subjects. The constant and all-centroid strategies explained 7.5% and 31.5% of eye-movement variance, respectively. The more successful strategies, object centroid, anticrowding, and target centroid, explained 48.8%, 51.2%, and 54.3% of the variance, respectively. For comparison, predictions based on other peoples' eye movements (DSST condition) would explain 68.6% of the variance.

In order to test the coherence with strategies over a wider selection of trajectories, I calculated NSS values for each strategy and each individual trial (i.e., not based on four repetitions of the same trial) and included the data from unique trials. Because the repeating trials would be presented four times and overweigh the unique trials, I selected only the repeating trials from the first block, when the participants first encountered them. The final set consisted of 712 correctly answered trials. The observed coherence did not differ in repeating and nonrepeating trials, $t(961.9) = 0.992$, $p = 0.321$, $d = 0.04 \pm 0.09$.

I focused on object-centroid, anticrowding, and target-centroid strategies, which were very similar in the previous analysis; with a wider selection of trials the differences were more apparent but still small (see Figure 7). The coherence of target-centroid strategy was higher than coherence of object-centroid strategy, $t(711) = 13.888$, $p < 0.001$, $d = 0.19 \pm 0.10$, and smaller than anticrowding strategy, $t(711) = 4.705$, $p < 0.001$, $d = 0.19 \pm 0.10$.

Using the same selection of trials, I confirmed the differences between the strategies by comparing the time spent in dynamic areas of interests defined by the strategies, targets, and distractors (Huff et al., 2010; Papenmeier & Huff, 2010). In every frame each predicted position defined a circular area of interest (AOI) with 1° radius. Additional AOIs were similarly defined for each target and distractor. Figure 8 shows proportion of time in each trial spent in each AOI. Participants spent most of the time looking at any of four targets (12.6%, or 3.1% per target), while they spent only 4.0% of time looking at distractors. In accord with the previous results, the anticrowding strategy showed the best fit (12.2%) compared to the target-centroid strategy, 9.0%, $t(711) = 7.268$, $p < 0.001$, $d = 0.34 \pm 0.10$, or object-centroid strategy, 7.7%, $t(711) = 10.135$, $p < 0.001$, $d = 0.50 \pm 0.11$. The

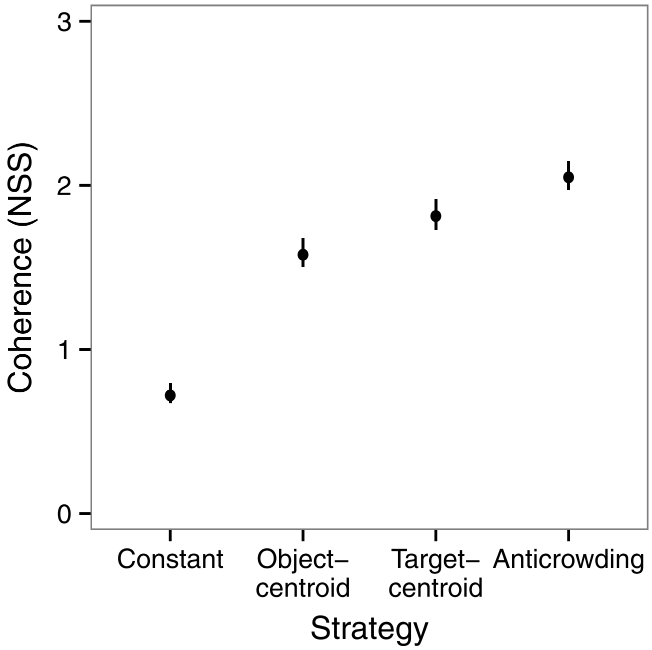

Figure 7. Coherence between selected eye-movement strategies in MOT and observed data based on both repeating and unique trials (individual trials).

difference between the target-centroid and object-centroid strategies was significant, but small, $t(711) = 5.697$, $p < 0.001$, $d = 0.14 \pm 0.10$. The observed proportions are lower than the proportions reported by previous studies, which utilized so-called "shortest distance rule" (Fehd & Seiffert, 2008; Zelinsky & Neider, 2008) but comparable to the values found with dynamic AOI approach (Huff et al., 2010).

When I compared whether people start to differ over time (between first two blocks and last two blocks), I found no significant effect of time, $F(1, 19) = 0.182$, $p = 0.674$, $\eta_G^2 < 0.001$, and no significant Time × Strategy interaction, $F(6,114) = 0.879$, $p = 0.513$, $\eta_G^2 = 0.005$.

Discussion

I compared the similarity of eye movements made during MOT from various perspectives and found a considerable amount of scan pattern similarity when the same MOT task was repeated, both within the same person and across different observers. Intra-individual and interindividual similarity cannot be explained solely by simple strategies derived from the previous studies; this outcome suggests that other aspects of the scene are utilized.

How precisely do people repeat their eye movements? I found that the mean gaze distance over repetitions of the same trial is approximately 2.5° (for illustration, the diameter of each stimulus was 1°). Some part of this variability can be attributed to the measurement error of the eye tracker; however, the results show that people are likely to repeat their eye movements, but the task does not require them to be too accurate in planning their gaze.

The results show a significant relationship between NSS and distances between gaze points in each frame. Both methods are similar as they take the distances into account compared to gaze classification approaches based on AOIs, where the distance is reduced to a

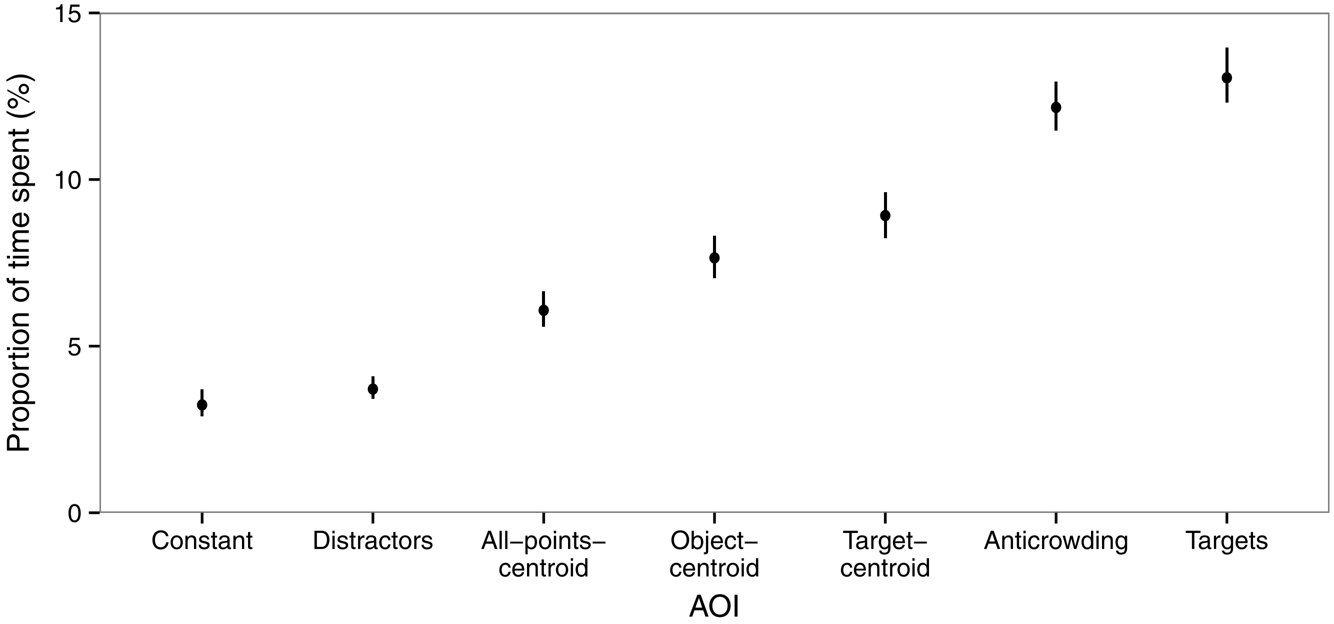

Figure 8. Gaze on AOIs defined by targets, distractors, and predictions of selected strategies. AOIs defined by targets and distractors are four times larger. Error bars show 95% confidence intervals.

binary factor. On general level NSS provides a single number representing the similarity of two or more scan patterns (using "leave-one-out" method). In contrast, the distance approach yields a similarity measure (e.g., mean distance) for each frame, which needs to be later summarized in some manner. Both approaches are partly convertible. The multitude of results can be achieved in NSS by calculating values not for entire scan pattern but for its floating subsets (e.g., from −200 ms to +200 ms for each time step). Due to spatiotemporal filtering NSS method detects similarity not only between gaze samples in the same frame but also over several proximal frames. This feature can be achieved by calculating distances between gaze points for varying lags.

There is no obvious metric for eye-movement similarity; how it should scale under various circumstances and every proposed metric includes some design decisions. Both NSS and distance approach share similar underlying assumptions. If two scan patterns involve only two gaze locations, but they strictly alternate in time, the scan patterns will be considered substantially different. Similarly, if two people follow a single but different target throughout a trial, their strategies will be also considered to be different. However, in the context of MOT task, these assumptions seem plausible.

I did not find changes in eye-movement similarity during repeated presentations of MOT trials. In an experiment with repeated presentations of movie trailers and dynamic natural scenes, Dorr et al. (2010) presented each movie 10 times (five times for two consecutive days). These authors found that coherence was highest in the first presentation and gradually decreased in later presentations within each day. The amount of initial coherence and variability was comparable between the first and second days. They concluded that individual strategies had an increasing influence on coherence with repeated viewings. The presented results suggest that in MOT, eye movements are defined strongly by the demands of the task; subjects do not use individual strategies in later presentations. In contrast, if subjects are instructed to "watch a movie attentively," they may feel the urge to scan the scene for details, which they had not attended in previous presentations. The potential increase of coherence in repeated MOT would suggest that subjects are learning during the course of experiment. Learning can occur either on a skill level (people converge to a similar strategy) or on a trial level (people become more confident in particular situations of crowding; thus, they make fewer rescue saccades). The first type of learning would be evident if adherence of eye movements to a particular strategy increased over time as well as in nonrepeated trials; however, this was not observed in the current study. In general people started to make fewer saccades throughout the experiment.

Previous experiments (Makovski, Vázquez, & Jiang, 2008; Ogawa et al., 2009) showed that people improve in MOT if the assignment of targets and distractors is retained. Although Makovski et al. (2008) reported that the strongest learning effects form within the first few repetitions, their task was much more difficult (similar presentation time, but the object speed was 22.5°/s compared to 5°/s). Ogawa et al. (2009) increased the difficulty by asking subjects to track more objects (5 of 10); they observed that the largest learning effect occurred between the first and second epoch (Repetitions 1–3 and 4–6). Both studies showed that successful recognition of repeated trials did not lead to better performance in MOT tasks. In the current study, 7 of 20 participants reported they noticed that some trials were repeated, which is similar to the proportion reported by Ogawa et al. (2009). This report probably reflects only a general impression from the experiment, because participants were not able to distinguish between repeating and novel trials in a recognition test.

I kept the task relatively easy compared to the majority of MOT studies, because I thereby attempted to encourage subjects to make eye movements. Different experimental conditions (faster object speeds) in which people tend to use less eye activity and follow the objects attentionally only prompt increased interindividual and intra-individual eye-movement similarity. Therefore, I compared the eye-movement patterns with the model predicting constant central fixation for each trial. I found very low similarity (this model explained only 7.5% of the variance within eye movements) and thus, we may conclude that central fixation did not contribute to the observed similarity in the present study. I used an object speed of 5°/s because pilot experiments showed that subjects move their eyes during trials, which may change if objects move with greater speed.

MOT studies that ask subjects to fixate on a central point and track objects without moving their eyes have shown that people can perform an MOT task without eye movements. In the current study, I have shown that when eye movements are allowed, they are characteristic of a trial (i.e., for the objects' positions and trajectories).

In the presented analysis, I compared various eye-movement strategies for MOT with interindividual and intra-individual consistency. It is important to note that both types of results are comparable despite some differences. In the first case, we have a well-defined model and compare its predictions with the observed data; however, there is no clear prediction in case of inter- or intra-individual consistency. Nevertheless, the consistency results are defined as predictions we can make based on knowing three scan patterns (averaged across all four scan patterns using the leave-one-out method). We can imagine that on average any of the observed scan patterns outperforms the predicted strategies in each trial.

Previous studies of eye-movement strategies in MOT (Fehd & Seiffert, 2008, 2010; Zelinsky & Neider, 2008) classified gaze samples based on gaze proximity to targets, distractors, and target centroids. These studies compared the time spent in each region of interest. For four targets, Zelinsky and Neider (2008) reported that participants spent 24% of scanning time near the centroid and 52% of the time on the targets; this ratio varied with the number of targets. Fehd and Seiffert (2008) reported longer times fixated near centroids in their follow-up experiment (42% of the time) compared to the time spent on targets (9% of the time). Huff et al. (2010) used stricter gaze classification using areas of interest of the size equal to the object size. While this classification explained about 35%–47% of gaze, about 10% of gaze was identified as toward centroid. In the presented study I compared the coherence of several strategies with eye-movement data first using NSS and later confirmed the results using the dynamic areas of interest (Huff et al., 2010; Papenmeier & Huff, 2010). Compared to gaze classification approaches, smooth comparisons using NSS allow comparing the strategies to natural variance in eye movements during an MOT task. Considering the observed intra-individual variability it seems beneficial to use an approach that is not dependent on the exact sizes of AOIs.

In the current study I tried to clarify the differences between various definitions of centroid. In previous studies two approaches were employed. First, the centroid was calculated for the solid object defined by the targets (Fehd & Seiffert, 2008, 2010). However the shape of solid object is ambiguous for concave configurations (with more than three targets). To overcome this limitation, I suggested calculating the centroid for a convex hull defined by targets. Second, the centroid can be calculated for a finite set of points/targets (Zelinsky & Neider, 2008), which is also the point with minimum sum of squared Euclidean distances between itself and each target. In the trajectories used in the current experiment the differences between predictions of both approaches were relatively small (median distance 0.71°).

Both centroid models are based solely on the positions of targets. Interestingly, their predictions differed more from the predictions of anticrowding model that demonstrated better coherence with gaze data on a larger selection of different trials (with both unique and repeating trials). The particular definition of the crowding optimization function can be discussed, but the result shows that people take the distractor positions into account.

There are several ways to improve the models of eye movements in MOT. First, while the gaze classification approach reports looking on targets, the employed models rarely predict such events and prefer some central position. Fehd and Seiffert (2010) speculated that participants shift their gaze from the center position to a target when distractors crowd the targets. In an experiment using occluding objects in an MOT task, Zelinsky and Todor (2010) found that subjects make fewer rescue saccades (saccades towards targets that are in danger of being lost) as the distance to the next closest object increases. In the current experiment, eye movements were a mixture of smooth pursuit and saccadic movements. Saccades disrupt visual processing; observers likely fail to detect changes in the scene during saccades, and their perception of space and time is affected (Morrone, Ross, & Burr, 2005; Ross, Morrone, Goldberg, & Burr, 2001). Advanced models could account for the costs of saccades, e.g., predicting that observers make saccades towards targets (or other more distant locations within the display) only when the danger of losing track of them while at the current gaze spot is greater than the danger related to making the saccade.

Second, the strategies proposed in the current study are static; they make predictions based only on the objects' positions in the current frame. Do people predict the positions of tracked objects? Previous studies in which objects disappear during MOT (Fencsik, Klieger, & Horowitz, 2007; Keane & Pylyshyn, 2006) have showed that participants perform better if objects reappear at their last known positions rather than positions based on their last known direction. However, participants process the directions of objects and report the last known directions for targets more accurately than for distractors (Horowitz & Cohen, 2010). When there are fewer targets, participants extrapolate the motion of tracked objects (Fencsik et al., 2007). If there is a smaller capacity for using directional information, participants may use directional cues in a flexible manner. Atsma, Koning, and van Lier (2012) showed that in MOT participants are more sensitive to probes presented in the direction of movement, but they did not anticipate the bouncing of objects unless the attentional load was low. Recent studies (Huff & Papenmeier, 2013; St.Clair, Huff, & Seiffert, 2010) confirmed the importance of directional information in MOT using objects with dynamic textures; the performance decreased when the texture motion did not correspond to the direction of objects. Additionally, participants may preferentially use directional cues as a heuristic to track objects for which less accurate information is available (i.e., they are further away from the gaze point or near other objects).

It is possible that eye-movement strategies are more effective when derived using object positions from some earlier moment. When participants are asked to report the last position of the tracked objects, both motion extrapolation (Iordanescu et al., 2009) and a perceptual lag have been reported (Howard, Masom, & Holcombe, 2011). More sophisticated models may account for perceptual lags caused by the integration of temporal information and some level of motion extrapolation.

Conclusion

Contrary to other tasks (free viewing, recognition, visual search), participants often fail to recognize the repetition of the trials in MOT. This phenomenon allows us to estimate the natural intra-individual and interindividual variability of eye movements made during this task. Many studies have shown that people can perform the task without using eye movements; here, I show that when eye movements are allowed, they are characteristic of a trial. The intra-individual variability provides an important upper bound when evaluating any model that describes eye-movement strategies in an MOT task. I compared the adherence of the observed data to several simple models and found the highest adherence in anticrowding strategy closely followed by target-centroid and object-centroid strategies (these strategies accounted for 48.8% to 54.3% of eye-movement variability, when compared to the empirical baseline and intra-individual variability). Predictions based on other peoples' eye movements (DSST condition) would explain 68.6% of the variance. I conclude that improvement is needed and more sophisticated models are necessary to describe the eye movements used in MOT. Furthermore, intra-individual variability should be used as a benchmark, whenever possible, when evaluating models for visual tasks.

Keywords: eye movements, attention, spatial vision, active vision, multiple object tracking

Acknowledgments

This research was supported by the Czech Science Foundation (GPP407/10/P607) and RVO 68081740. I would like to thank two anonymous reviewers for helpful comments on the manuscript.

Commercial relationships: none.

Corresponding author: Jiří Lukavský.
Email: lukavsky@praha.psu.cas.cz.
Address: Institute of Psychology, Academy of Sciences of the Czech Republic, Prague, Czech Republic.

References

Alvarez, G. A., & Cavanagh, P. (2005). Independent resources for attentional tracking in the left and right visual hemifields. *Psychological Science, 16*(8), 637–643. doi:10.1111/j.1467-9280.2005.01587.x.

Atsma, J., Koning, A., & van Lier, R. (2012). Multiple object tracking: Anticipatory attention doesn't "bounce." *Journal of Vision, 12*(13):1, 1–11, http://www.journalofvision.org/content/12/13/1, doi:10.1167/12.13.1. [PubMed] [Article]

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision, 10*(4), 433–436.

Chun, M. M., & Jiang, Y. (1998). Contextual cueing: Implicit learning and memory of visual context guides spatial attention. *Cognitive Psychology, 36*(1), 28–71. doi:10.1006/cogp.1998.0681.

Cornelissen, F. W., Peters, E. M., & Palmer, J. (2002). The Eyelink Toolbox: Eye tracking with MATLAB and the Psychophysics Toolbox. *Behavior Research Methods, Instruments, & Computers, 34*(4), 613–617.

Dorr, M., Martinetz, T., Gegenfurtner, K. R., & Barth, E. (2010). Variability of eye movements when viewing dynamic natural scenes. *Journal of Vision, 10*(10):28, 1–17, http://www.journalofvision.org/content/10/10/28, doi:10.1167/10.10.28. [PubMed] [Article]

Einhäuser, W., Rutishauser, U., & Koch, C. (2008). Task-demands can immediately reverse the effects of sensory-driven saliency in complex visual stimuli. *Journal of Vision, 8*(2):2, 1–19, http://www.journalofvision.org/content/8/2/2, doi:10.1167/8.2.2. [PubMed] [Article]

Einhäuser, W., Spain, M., & Perona, P. (2008). Objects predict fixations better than early saliency. *Journal of Vision, 8*(14):18, 1–26, http://www.journalofvision.org/content/8/14/18, doi:10.1167/8.14.18. [PubMed] [Article]

Fehd, H. M., & Seiffert, A. E. (2008). Eye movements during multiple object tracking: Where do participants look? *Cognition, 108*(1), 201–209. doi:10.1016/j.cognition.2007.11.008.

Fehd, H. M., & Seiffert, A. E. (2010). Looking at the center of the targets helps multiple object tracking. *Journal of Vision*, *10*(4):19, 1–13, http://www.journalofvision.org/content/10.4.19, doi:10.1167/10.4.19. [PubMed] [Article]

Fencsik, D., Klieger, S., & Horowitz, T. (2007). The role of location and motion information in the tracking and recovery of moving objects. *Attention, Perception, & Psychophysics*, *69*(4), 567–577.

Foulsham, T., & Underwood, G. (2008). What can saliency models predict about eye movements? Spatial and sequential aspects of fixations during encoding and recognition. *Journal of Vision*, *8*(2):6, 1–17, http://www.journalofvision.org/content/8/2/6, doi:10.1167/8.2.6. [PubMed] [Article]

Fuster, J. M. (2004). Upper processing stages of the perception-action cycle. *Trends in Cognitive Sciences, 8*(4), 143–145.

Hollingworth, A. (2012). Task specificity and the influence of memory on visual search: Comment on Võ and Wolfe (2012). *Journal of Experimental Psychology: Human Perception and Performance, 38,* 1596–1603. doi:10.1037/a0030237.

Horowitz, T. S., & Cohen, M. A. (2010). Direction information in multiple object tracking is limited by a graded resource. *Attention, Perception, & Psychophysics, 72*(7), 1765–1775. doi:10.3758/APP.

Horowitz, T. S., Klieger, S. B., Fencsik, D. E., Yang, K. K., Alvarez, G. A., & Wolfe, J. M. (2007). Tracking unique objects. *Perception & Psychophysics, 69*(2), 172–184. doi:10.3758/BF03193740.

Howard, C. J., Masom, D., & Holcombe, A. O. (2011). Position representations lag behind targets in multiple object tracking. *Vision Research, 51*(17), 1907–1919. doi:10.1016/j.visres.2011.07.001.

Huff, M., & Papenmeier, F. (2013). It's time to integrate: The temporal dynamics of object motion and texture motion integration in multiple object tracking. *Vision Research, 76,* 25–30. doi:10.1016/j.visres.2012.10.001.

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

Intriligator, J., & Cavanagh, P. (2001). The spatial resolution of visual attention. *Cognitive Psychology, 43*(3), 171–216. doi:10.1006/cogp.2001.0755.

Iordanescu, L., Grabowecky, M., & Suzuki, S. (2009). Demand-based dynamic distribution of attention and monitoring of velocities during multiple-object tracking. *Journal of Vision*, *9*(4):1, 1–12, http://www.journalofvision.org/content/9/4/1, doi:10.1167/9.4.1. [PubMed] [Article]

Itti, L., & Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research, 40*(10-12), 1489–1506.

Keane, B. P., & Pylyshyn, Z. W. (2006). Is motion extrapolation employed in multiple object tracking? Tracking as a low-level, non-predictive function. *Cognitive Psychology, 52*(4), 346–368. doi:10.1016/j.cogpsych.2005.12.001.

Kleiner, M., Brainard, D. H., & Pelli, D. G. (2007). What's new in Psychtoolbox-3? *Perception, 36*(ECVP Abstract Supplement), http://www.perceptionweb.com/abstract.cgi?id=v070821.

Makovski, T., Vázquez, G. A., & Jiang, Y. V. (2008). Visual learning in multiple-object tracking. *PLoS ONE, 3*(5), e2228. doi:10.1371/journal.pone.0002228.

Morrone, M. C., Ross, J., & Burr, D. (2005). Saccadic eye movements cause compression of time as well as space. *Nature Neuroscience, 8*(7), 950–954. doi:10.1038/nn1488.

Nyström, M., & Holmqvist, K. (2008). Semantic override of low-level features in image viewing–both initially and overall. *Journal of Eye Movement Research*, *2*(2), 1–11.

Ogawa, H., Watanabe, K., & Yagi, A. (2009). Contextual cueing in multiple object tracking. *Visual Cognition, 17*(8), 1244–1258. doi:10.1080/13506280802457176.

Papenmeier, F., & Huff, M. (2010). DynAOI: A tool for matching eye-movement data with dynamic areas of interest in animations and movies. *Behavior Research Methods, 42*(1), 179–187. doi:10.3758/BRM.42.1.179.

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision, 10*(4), 437–442.

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

Raney, G. E., & Rayner, K. (1995). Word frequency effects and eye movements during two readings of a text. *Canadian Journal of Experimental Psychology, 49*(2), 151–172.

Ross, J., Morrone, M. C., Goldberg, M. E., & Burr, D. C. (2001). Changes in visual perception at the time of saccades. *Trends in Neurosciences, 24*(2), 113–121. doi:10.1016/S0166-2236(00)01685-4.

Rothkopf, C., Ballard, D., & Hayhoe, M. (2007). Task and context determine where you look. *Journal of Vision*, *7*(14):16, 1–20, http://www.journalofvision.org/7/14/16, doi:10.1167/7.14.16. [PubMed] [Article]

Scholl, B. J. (2009). What have we learned about attention from multiple-object tracking (and vice versa)?. In D. Dedrick & L. Trick (Eds.), *Computation, Cognition, and Pylyshyn* (pp. 49–77). Cambridge, MA: MIT Press.

Schütz, A. C., Braun, D. I., & Gegenfurtner, K. R. (2011). Eye movements and perception: A selective review. *Journal of Vision*, *11*(5):9, 1–30, http://www.journalofvision.org/content/11/5/9, doi:10.1167/11.5.9. [PubMed] [Article]

St.Clair, R., Huff, M., & Seiffert, A. E. (2010). Conflicting motion information impairs multiple object tracking. *Journal of Vision*, *10*(4):18, 1–13, http://www.journalofvision.org/content/10/4/18, doi:10.1167/10.4.18. [PubMed] [Article]

Tatler, B. W. (2009). Current understanding of eye guidance. *Visual Cognition, 17*(6-7), 777–789. doi:10.1080/13506280902869213.

Tatler, B. W., Baddeley, R. J., & Gilchrist, I. D. (2005). Visual correlates of fixation selection: Effects of scale and time. *Vision Research, 45*(5), 643–659. doi:10.1016/j.visres.2004.09.017.

Võ, M. L.-H., & Wolfe, J. M. (2012). When does repeated search in scenes involve memory? Looking at versus looking for objects in scenes. *Journal of Experimental Psychology: Human Perception and Performance*, *38*(1), 23–41. doi:10.1037/a0024147.

Zelinsky, G. J., & Neider, M. B. (2008). An eye movement analysis of multiple object tracking in a realistic environment. *Visual Cognition*, *16*(5), 553–566. doi:10.1080/13506280802000752.

Zelinsky, G. J., & Todor, A. (2010). The role of "rescue saccades" in tracking objects through occlusions. *Journal of Vision*, *10*(14):29, 1–13, http://www.journalofvision.org/content/10/14/29, doi:10.1167/10.14.29. [PubMed] [Article]