Detecting single-target changes in multiple object tracking: The case of peripheral vision

Christian Vater[1] · Ralf Kredel[1] · Ernst-Joachim Hossner[1]

© The Psychonomic Society, Inc. 2016

Abstract In the present study, we investigated whether peripheral vision can be used to monitor multiple moving objects and to detect single-target changes. For this purpose, in Experiment 1, a modified multiple object tracking (MOT) setup with a large projection screen and a constant-position centroid phase had to be checked first. Classical findings regarding the use of a virtual centroid to track multiple objects and the dependency of tracking accuracy on target speed could be successfully replicated. Thereafter, the main experimental variations regarding the manipulation of to-be-detected target changes could be introduced in Experiment 2. In addition to a button press used for the detection task, gaze behavior was assessed using an integrated eyetracking system. The analysis of saccadic reaction times in relation to the motor response showed that peripheral vision is naturally used to detect motion and form changes in MOT, because saccades to the target often occurred after target-change offset. Furthermore, for changes of comparable task difficulties, motion changes are detected better by peripheral vision than are form changes. These findings indicate that the capabilities of the visual system (e.g., visual acuity) affect change detection rates and that covert-attention processes may be affected by vision-related aspects such as spatial uncertainty. Moreover, we argue that a centroid-MOT strategy might reduce saccade-related costs and that eyetracking seems to be generally valuable to test the predictions derived from theories of MOT.

Finally, we propose implications for testing covert attention in applied settings.

Peripheral vision allows us to detect natural objects at large eccentricities at up to 70.5° (Thorpe, Gegenfurtner, Fabre-Thorpe, & Bülthoff, 2001) and to categorize natural scenes at 70° eccentricity with high accuracy (Boucart, Moroni, Thibaut, Szaffarczyk, & Greene, 2013), even under crowded conditions (Li, VanRullen, Koch, & Perona, 2002). Neuroanatomically, the capability to process visual information in the far periphery seems to be controlled by distinct, specialized cortical networks (for a recent review, see Yu, Chaplin, & Rosa, 2015). Furthermore, in comparison to foveal vision, peripheral vision features higher temporal resolution (Hartmann, Lachenmayr, & Brettel, 1979) and superior motion detection (Finlay, 1982; McKee & Nakayama, 1984). Thus, whereas foveal vision with its high spatial resolution is important for extracting detailed information, a crucial role for peripheral vision seems to refer to the processing of changing aspects in the visual environment (for a review, see Strasburger, Rentschler, & Jüttner, 2011).

Hence, a monitoring task on the detection of motion and form changes in the periphery can be regarded as a paradigmatic example of making use of peripheral vision. In an applied sense, those tasks can be found in team sports, where it is often necessary to perceive a number of moving objects across a large functional field of view (Davids, 1984). This applies, for example, in offside decision-making in soccer, where the assistant referee has to monitor the player in ball possession as well as the players at the offside line. In this situation, experts show a gaze strategy in which they anchor their gaze on the

✉ Christian Vater
  Christian.Vater@ispw.unibe.ch

[1] University of Bern, Bremgartenstrasse 145, CH-3012 Bern, Switzerland

offside line while perceiving the player in ball possession peripherally (Catteeuw, Helsen, Gilis, Van Roie, & Wagemans, 2009). Furthermore, Williams and Davids (1998) were able to empirically show that, in a three-versus-three anticipation task in soccer, experts extract a great amount of information regarding players' positions and movements with peripheral vision. The importance of tracking multiple objects in team sports was also underlined by Faubert and Sidebottom (2012), who described the necessity for soccer goalkeepers to simultaneously track the ball, teammates, and opponents while keeping a large and dynamic visual field under surveillance.

In experimental psychology, the demands imposed from ecologically valid situations are generally studied with *multiple object tracking* (MOT), a task introduced by Pylyshyn and Storm (1988) to study visual attention. The task requires tracking a certain number of targets amidst identical-looking distractors for a predefined amount of time. Before the objects begin to move on the computer screen, the targets are briefly highlighted, then all objects start to move in a (quasi)random fashion, and finally, after the targets have stopped, participants are supposed to recall them (see Fig. 1, in which the stimulus materials of the present study are illustrated).

Theoretically, several approaches have been brought forward to explain tracking performance in MOT—namely, (a) the FINST (Fingers of INSTantiation) model, by Pylyshyn and Storm (1988); (b) the grouping model, by Yantis (1992); and (c) the multifocal attention model, by Cavanagh and Alvarez (2005). After brief descriptions of these approaches, we will come back to the question of how the specifics of peripheral vision can be related to these concepts, in particular, and to MOT performance, in general.

(a) In their MOT studies, Pylyshyn and Storm (1988) were mainly interested in how visual attention can be shifted independent of eye movements and whether locations are scanned serially or in parallel. According to their feature-binding-based FINST model, participants track multiple objects in parallel, but each object possesses an individual internal reference at a preattentive stage. The authors describe a serial-tracking algorithm in which target locations (instead of motion vectors) are stored. In a related experiment, motor-response latencies to flashes on targets were measured as a function of the number of targets. Latencies significantly increased from one target to five targets, which could be regarded as an indicator that participants serially tracked the targets (Pylyshyn & Storm, 1988). However, in a second experiment, to test the serial-scanning algorithm, participants had to indicate whether a flash occurred on a target, on a distractor, or at another location. Because the observed performance turned out to be much higher than was algorithmically predicted, Pylyshyn and Storm finally concluded that targets are tracked in parallel rather than serially.

(b) Yantis (1992) alternatively proposed that participants keep track of targets by making use of a virtual polygon formed out of them. Two stages are predicted by this model. The first one, the *group formation* stage, is influenced by Gestalt laws of grouping and can be characterized as preattentive, automatic, and stimulus-driven, and thereby comparable to the model introduced by Pylyshyn and Storm (1988). However, in contrast to the FINST model, the second stage, which concerns *group maintenance*, is characterized as goal-directed, effortful, and attention-demanding. Furthermore, continuous updating of the target representations is assumed and discussed in the context of mental rotations. Yantis (1992) experimentally tested these predictions by applying different Gestalt laws to the MOT task. For example, when either five targets all rotated in the same direction (rigid condition) or three targets rotated in one and two targets in the other direction (nonrigid condition), response accuracies were higher in the rigid condition, presumably because the targets followed the law of common fate. Furthermore, Zhong Ma, Wilson, Liu, and Flombaum (2014) showed that, rather than motion
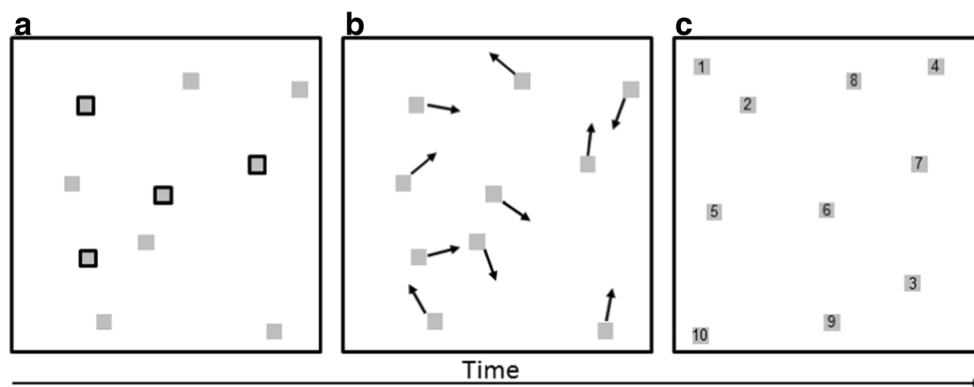

Fig. 1 Experimental paradigm used in Experiments 1 and 2. After the four targets are designated with red frames (**a**), the frames disappear, and all stimuli move in quasirandom fashion for 6 s (**b**). At the end of the motion phase, participants have to recall the targets by naming the respective numbers that are projected onto the objects (**c**)

extrapolation, the recently observed object position seems to be used for tracking. This finding can be interpreted as support for the effortful updating process in the Yantis model because, on the basis of continuous updating, no motion extrapolation would be expected (for an overview of the mixed results from motion extrapolation, however, see Zhong et al., 2014).

(c) Finally, Cavanagh and Alvarez (2005) proposed a multifocal attentional mechanism for MOT that is based on the assumptions that attention is split between the targets and that each selection channel comprises a position tracker and a stream of object features. This model is underpinned, in part, by the results presented by Meyerhoff, Papenmeier, Jahn, and Huff (2013), who were able to show that unexpected changes in targets do affect tracking performance, whereas changes in distractors have no effect even if these changes are in the line of sight. To examine whether motion information is actually used for tracking, Huff and colleagues introduced texture-motion information on the targets that moved in either the same direction as the target or a different direction, thereby not confounding the spatiotemporal information of the objects with their motion information (Huff & Papenmeier, 2013; Meyerhoff, Papenmeier, & Huff, 2013; St Clair, Huff, & Seiffert, 2010). These studies, as well as others, showed that tracking performance declines when the difference between the object direction and the texture direction increases (St Clair et al., 2010), and that the availability and reliability of spatiotemporal and feature information is considered by participants (Papenmeier, Meyerhoff, Jahn, & Huff, 2014). Since both motion and feature information seem to be relevant for tracking performance, these results are perfectly in line with the multifocal attention theory of MOT.

In the three models sketched so far, the theoretical focus was laid on questions regarding visual attention, mostly without applying eyetracking technology to the particular studies. In this vein, Fehd and Seiffert (2008) demonstrated that multiple target objects are tracked among distractors through a center-looking strategy in which participants rely on a virtual centroid, which is defined as the center of mass of the polygon formed by the targets. Furthermore, Fehd and Seiffert showed that this strategy is used independent of speed and object size, and that tracking performance is highest when a center–target switching strategy is used, in which the gaze shifts back and forth between the centroid and the targets. Most interestingly, with regard to the observed gaze behavior, slower stimuli did not lead to more saccades, so a saccade-avoiding strategy was not evident. On the one hand, this result could have been expected, since increased effects of crowding are usually observed at higher stimulus speeds (Franconeri, Lin, Pylyshyn,

Fisher, & Enns, 2008), which would lead to participants needing to saccade more frequently to targets because of potential collisions (Elfanagely, Haladjian, Aks, Kourtev, & Pylyshyn, 2011; Landry, Sheridan, & Yufik, 2001). On the other hand, although saccading helps update the position of single targets (Landry et al., 2001), the costs of eye movements increase with higher stimulus speeds because objects will move farther during the saccade if the stimulus speed is high (Huff, Papenmeier, Jahn, & Hesse, 2010). Additionally, saccadic suppression of information processing (from about 75 ms before saccade onset to 50 ms after saccade offset) makes the updating process even more difficult as object speeds increase (Diamond, Ross, & Morrone, 2000). In this context, the role of "rescue saccades" in MOT has been examined by Zelinsky and Todor (2010), who found that anticipatory saccades are driven by the potential for losing track of a target when it is close to a barrier occluder or a distractor. It could, therefore, be that participants preferably use their peripheral vision to monitor target positions while "anchoring" their gaze on the virtual centroid (Fehd & Seiffert, 2010; Zelinsky & Neider, 2008), and that they initiate saccades just before the distance of a target from other objects becomes too small (Zelinsky & Todor, 2010).

When we relate the attention-related explanations of MOT behavior (see descriptions a–c above) to the just-sketched gaze studies, it seems likely that covert attention is used for tracking, since gaze seems to focus in large part on a virtual centroid. Thus, to examine the actual location of attention, target changes have been included in experimental studies on MOT. In this branch of research, target movements (motion direction) or target properties (color, form) were manipulated to investigate, on the one hand, whether target-motion information is used for tracking, and on the other hand, how attention is distributed between targets and distractors. Regarding the use of motion information, Meyerhoff, Papenmeier, Jahn, and Huff (2013) studied the effect of random motion-direction changes (deviations of up to 60° to the left or right) of targets, distractors, or both. The result, that an *unexpected* change of a target but not of a distractor vector impaired performance, shows that, in line with previous findings (e.g., Fencsik, Klieger, & Horowitz, 2007), target-motion information is crucial for MOT. However, if changes are *expected* and participants have to respond to them, it is possible to figure out where attention is located during MOT. Therefore, Bahrami (2003) introduced color (red, green, blue) and form ("T," "L," "+") changes in his targets and distractors. He found that color and form changes of the target were more often detected than changes of the distractors, again showing that attention is attracted by targets. Moreover, color changes were identified more frequently (~80 %) than form changes (~55 %). In a closely related study, Sears and Pylyshyn (2000) induced either a target or a distractor form change in 50 % of the trials (with the shape of a "seven-segment box figure

eight" becoming either an "E" or an "H"). They showed that target changes were detected much faster than distractor changes and that a higher number of distractors increased response times when the target or distractor changes were evident, underlining the finding that attention is attracted by targets rather than by distractors. This result has also been supported by electrophysiological measures—for example, by Drew, McCollough, Horowitz, and Vogel (2009), who measured event-related potentials (ERPs) while participants had to track two targets amid four moving distractors and four stationary distractors. Task-irrelevant white square probes were presented randomly at the target, on a distractor, in empty space, or on a stationary target. As expected, target probes showed the greatest ERP response, whereas weaker responses were observed for distractors, and the weakest responses for background and stationary targets. Hence, behavioral and electrophysiological results have confirmed that spatial attention is directed more to targets than to other locations.

Summing up, regarding theoretical approaches to explain MOT performance, recent research has supported the multifocal attention theory, since both target motion and feature information are apparently used in the tracking process. Furthermore, the centroid strategy revealed by eyetracking studies suggests that MOT performance is mainly based on covert attention, which, in turn, is particularly allocated to the targets.

As becomes obvious from this summary, despite the fact that eyetracking technology has been applied, the results of MOT studies have mostly been discussed from an attention perspective, and rarely from a vision perspective. This comes as a surprise because many findings could also be explained by the basal properties of the peripheral visual system, in particular by the decreasing visual acuity but increasing contrast sensitivity of the peripheral retina. For example, when Bahrami (2003) reported that target color changes are detected more often than form changes, this could also be traced back to luminance and contrast sensitivity differences of the retina, such that foveal vision might be advantageous for detecting form changes due to its higher spatial resolution (Gralla, 2007). Besides this, the higher temporal resolution (Hartmann et al., 1979) and the superior motion detection capability of the peripheral visual system (Finlay, 1982; McKee & Nakayama, 1984) also predict high detection rates for motion changes in MOT. In this regard, it is really surprising that, to date, no attempt has been made to implement a motion-change detection task to examine the role of peripheral vision in MOT. In such a study it would seem to be crucial to control gaze behavior, since in previous experiments (e.g., Sears & Pylyshyn, 2000) participants were only instructed to track target objects while keeping their eyes on the fixation cross, without confirming this with eyetracking technology. Consequently, up to now, it is completely unknown which natural gaze strategy is executed in MOT to detect target changes in a setting in which no explicit instructions are given

regarding vision. Hence, exactly this empirical question will be addressed in the following experiments.

From an applied perspective, it should be added that these experiments promise to provide helpful advice for the training of gaze strategies—for example, in team-sports situations—as we sketched out at the beginning of this article. In this regard, form changes in MOT can be associated with posture-related changes (e.g., a basketball player indicating with his hands where he wants to receive the ball), and motion changes in MOT with movement-related changes (e.g., detecting an opposing player making a basket). Since the reliability of such recommendations grows with the external validity of the experimental setting, it seems worthwhile to study those changes by displaying the MOT task on a large screen, to overcome the limitation of a restricted field of vision when examining the relevance of peripheral vision. Hence, in a series of two experiments, we first aimed to answer the question of whether classical MOT findings can be replicated under those novel experimental conditions. On the basis of the results of Experiment 1, the main research question was approached in Experiment 2, in which the focus was on the case of peripheral vision in monitoring multiple moving objects while detecting changes in the periphery.

Experiment 1

Experiment 1 was designed as a replication study, in particular for the MOT results reported by Fehd and Seiffert (2008, 2010) on the effects of object speed on response accuracy, on the one hand, and gaze behavior, on the other. To increase the external validity of the study, a large screen was used to display the MOT task. In addition, a novel manipulation detail was introduced in anticipation of the main research question in Experiment 2, regarding form versus motion changes. In this initial experiment, it seemed crucial to manipulate the motion of the targets and distractors in such a way that form- or motion-related changes would occur independently of the (uncontrolled) movements of the centroid, as well as of the (uncontrolled) eccentricities of the to-be-detected event. For this reason, a stationary phase of the centroid was applied in which changes would occur at defined eccentricities (under the assumption that the centroid was actually focused). Despite the larger display and the introduction of a stationary-centroid phase, a successful replication of the findings reported by Fehd and Seiffert (2008, 2010) was expected. Hence, we predicted that response accuracy would decrease as a function of object speed and that participants would prefer a centroid-tracking strategy, independent of speed conditions.

Method

Participants A total of 14 sports science students (seven females and seven males; age: 21.7 ± 1.3 years) participated in

the experiment and received course credits in return. They had self-reported normal or corrected-to-normal vision and were unaware of the research question. The experiment was undertaken in accordance with the Declaration of Helsinki.

Stimuli The motion paths of the objects (ten white squares, 35 × 35 mm, corresponding to 1° × 1° of visual angle) were calculated with MATLAB and then imported to Autodesk 3DS Max (2014) to render single video trials. The stimuli were presented within a quadratic frame (composed of white lines of 25-mm width and measuring 1.40 × 1.40 m, corresponding to 40° × 40° of visual angle) on a black background. All squares appeared in (quasi)randomized starting positions that precluded overlaps with each other. Each trial began with a trial number, followed by the stationary presentation of ten squares, including four targets that were highlighted by red frames (line width of 15 mm; the stimulus and borders together covered an area of 1.7° × 1.7° of visual angle). After 2 s, the target cues disappeared and all stimuli accelerated over 1 s to the final speed of either 6°/s, 9°/s, or 12°/s, which was retained for 4 s. After a subsequent deceleration phase of one further second, all squares stopped, so the overall motion time was 6 s per trial.

During the motion time, a repulsion mechanism was used to redirect a square whenever the distance to the surrounding frame or the next square fell below a certain threshold (35 mm, corresponding to 1° of visual angle). To ensure the independence of any findings of centroid-related motion (as well as of the different eccentricities of the target changes in Exp. 2), the targets' paths were manipulated such that the centroid was forced to stay at a constant position for 0.5 s. For this purpose, the path of one distractor was calculated in such a way that it collided with one target at one of three possible points in time (3.0, 3.5, or 4.0 s after motion onset) and that all targets moved after the collision in a concerted way at angles 90° different from each other (e.g., 10°, 100°, 190°, and 280°, in relation to the frame's baseline). Due to the same speeds of the targets, this specification resulted in a stationary position of the centroid of the polygon formed by the four targets. Over the 0.5 s following the critical collision, the five remaining distractors were allowed to move randomly; however, we ensured that no collision of a target with another object or the bordering frame occurred within this time window of 0.5 s. Before and after this interval, all objects were allowed to collide with one another or with the bordering frame without any restrictions.

On the basis of these constraints, 14 mother trials were created, which differed in the positions of the crucial collision, the resulting position of the stationary centroid, and the concerted motion direction of the targets during the stationary-centroid phase. From the mother trials, a total of 126 trials (=14 × 3 × 3) were derived by starting the stationary-centroid phase at three different onsets (either 3.0, 3.5, or 4.0 s after motion onset) and by accelerating the targets and

distractors to three final speeds (either 6°/s, 9°/s, or 12°/s). To balance trials of different speeds within blocks, each block contained five trials of each speed, which were presented in a randomized order (rendering by MAGIX Video Pro X3). Since a single session of 60 min allowed for the presentation of a maximum of nine blocks, and since participants' capability to keep focused on the quite demanding task seemed to be limited, exactly this number of 15-trial blocks was employed. Consequently, in the resulting 135 trials, nine of the 126 trials derived from the mother trials were presented twice (i.e., three per stationary-centroid phase onset and speed).

Apparatus A binocular eyetracking system (EyeSeeCam, 220 Hz) was used to assess the vertical and horizontal rotations of both eyes via infrared reflections from the pupil and the cornea (measurement accuracy: 0.5° of visual angle, with a resolution of 0.01° root-mean square [RMS] within 25° of the participant's field of view). The EyeSeeCam (ESC) was connected to a MacBook Pro via a 20-m fiber-optic Fire Wire link (GOF-Repeater 800, Unibrain), stored in a bum bag. Since the eyetracker was synchronized with a ten-camera VICON-T20 system that tracks retro-reflective markers attached to the ESC, a three-dimensional gaze vector in a laboratory frame of reference could be calculated (Kredel et al., 2011). This gaze vector was updated every 5 ms and allowed for relating the participant's current gaze to the displayed targets or distractors, or to the centroid derived from the current positions of the targets.

The ESC was calibrated at the beginning of each test session. For this purpose, participants had to consecutively fixate five dots of a regular grid with a distance of 8.5° of visual angle between the dots (Kredel et al., 2011). The ESC was recalibrated before each test block whenever the point of gaze deviated more than 0.5° of visual angle from one of the dots in the calibration grid. The stimuli were back-projected (projector: InFocus IN5110) onto a large screen (height 1.87 m, width 3.01 m), where the quadratic frame for the MOT task covered an area of 1.40 × 1.40 m in the middle of the screen. Microsoft Windows Media Player was used to play back the video trials. The gathered data were analyzed using Mathworks MATLAB 2013a. Further statistical analyses were conducted with IBM SPSS Statistics 22.

Procedure Participants were tested individually in a single 1-h session in the institute's sensorimotor laboratory. After participants had read the general information about the study and agreed to participate, the eyetracking system was fitted. Subsequently, they were positioned at 2 m distance from the screen for reading the displayed instructions about the to-be-solved task, which was to recall the targets that were cued at the beginning by naming the respective numbers that were projected onto the targets at the end of each trial. Afterward, the ESC calibration routine was conducted. After each trial,

participants had to recall the targets by naming the respective numbers that were displayed on the ten squares. Participants' verbal decisions were recorded in writing by an experimenter. No augmented feedback on the responses' correctness was given after the trials.

Gaze analysis For the assessment of the location of gaze, a region-of-interest method was applied. For this purpose, 11 regions of interest were defined as circles with a diameter of 5° over each square as well as over the centroid, which was calculated as the center of mass of the polygon formed by the four targets (Fehd & Seiffert, 2010). Whenever the current gaze vector hit a region of interest, the respective frame was counted for the related object. This meant that, in cases of overlapping regions, an allocation to two or even more objects was possible. The number of counted frames per region was then divided by the total number of frames in the trial (i.e., by 1,200 frames for the 6 s of motion time), resulting in percentages of gaze allocation time for each region of interest, a variable that was termed the "gaze overlap time" by Fehd and Seiffert (2010).

Measures The first group of dependent variables referred to the correctness of the 45 responses per speed condition, which can be expressed either as the average number of correctly recalled targets (*n* correct) or as the percentage of trials on which all four targets could be successfully recalled (% correct). Only these trials were regarded as successful and included in the aggregation of gaze-related variables, because in cases in which only a subset of the targets were identified correctly, participants had probably lost one of the targets somewhere over the trial, such that at that point in time their calculation of the virtual centroid would have lost its validity. Thus, the above-defined variable of gaze overlap time was calculated only for trials with perfect target recall (% of total time). Since no specifics of a certain target or a certain distractor were to be expected, gaze overlap times were calculated as the averages of one centroid-related, four target-related, and six distractor-related values, respectively.

All dependent variables were analyzed with repeated measures analyses of variance (ANOVAs) with Speed Condition as the within-subjects factor, in the case of the gaze allocation analysis completed by a second, within-subjects factor, Region of Interest (centroid, target, or distractor). Since the only reason for including the three motion-onset conditions was to prevent participants from learning of the otherwise to-be-anticipated onset time of the change in a trial, and since no theory-related predictions can be derived for this variable, the stationary-centroid phase onset was not included in the ANOVA calculations. Significant interaction effects were further analyzed with planned *t* tests. On the basis of the results reported by Fehd and Seiffert (2010), the alpha level for tests on differences was set a priori to $\alpha = .05$. A-posteriori effect

sizes were computed as partial eta squared (η_p^2), and in cases of nonsignificant tests, the power ($1 - \beta$) was calculated. According to Fehd and Seiffert (2010), the highest response accuracies were expected for the 6°/s speed condition, followed by the 9°/s and 12°/s conditions. Furthermore, gaze overlap was expected to be highest for the centroid, followed by the target, and then the distractor average. Finally, again according to the results obtained by Fehd and Seiffert (2010), the gaze behavior should turn out to be independent of stimulus speed.

Results

As is illustrated in Fig. 2, our results show significant response accuracy differences for the three speed conditions, $F(2, 26) = 62.7$, $p < .01$, $\eta_p^2 = .83$, with the highest accuracies in the 6°/s condition, followed by the 9°/s and 12°/s conditions, and each speed condition differing in terms of response accuracies from each other (all *p*s < .01). The same data pattern is revealed for the average number of correctly identified targets, $F(2, 26) = 54.8$, $p < .01$, $\eta_p^2 = .81$, all *p*s < .01, which means that the slower the objects moved, the more targets were correctly recalled by the participants.

In terms of gaze overlap to the a-priori-defined regions of interest, as is depicted in Fig. 3, a 3 (speed) × 3 (region) ANOVA with repeated measures on both factors revealed a significant main effect for region only, $F(2, 26) = 76.4$, $p < .01$, $\eta_p^2 = .86$. Gaze was allocated more to the centroid than to the targets or the distractors, whereas gaze overlapped more with targets than with distractors, all *p*s < .01. Neither the effect of speed ($p = .90$, $1 - \beta = .93$) nor the interaction of speed and region ($p = .92$, $1 - \beta = .96$) reached significance.

Discussion

In Experiment 1, we aimed to answer the question whether classical MOT findings on gaze patterns, in particular
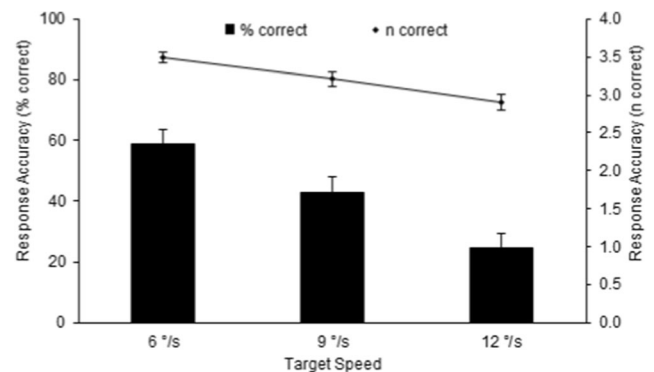

Fig. 2 Response accuracy as percentages of correct responses (*M*s and *SE*s) and numbers of correctly recalled targets (*M*s and *SE*s) as a function of target speed
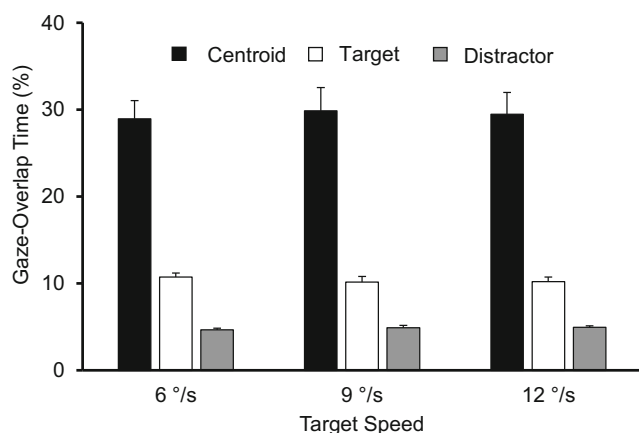
Fig. 3 Gaze-overlap times (*M*s and *SE*s) on the centroid, the target average, and the distractor average as a function of target speed

those reported by Fehd and Seiffert (2010), could be replicated in a novel experimental setting with a large projection screen and the introduction of a stationary-centroid phase. Regarding this replication trial, our results showed that the faster the objects move, the more tracking accuracy was impaired (see Fig. 2). This result is in perfect agreement with previous research (Alvarez & Franconeri, 2007; Fehd & Seiffert, 2010). Relative to the present findings, Fehd and Seiffert (2010) reported higher response accuracies for the 6°/s (90 %) and 12°/s (65 %) conditions, whereas the numbers of correctly tracked targets in our study were comparable to those reported by Alvarez and Franconeri in similar speed conditions (i.e., 7°/s and 14°/s) when the spacing (one stimulus diameter) was also similar (3.4 and 2.4 targets, respectively).

Regarding gaze behavior, Fehd and Seiffert (2010) found that participants were mainly looking at the centroid of the target group, and this was true even at high object speeds. This finding could also be replicated in the present study, since our results showed about 29.5 % gaze overlap with the centroid (Fehd & Seiffert, 2010: 25 %–30 %), 10.5 % with the target average (Fehd & Seiffert, 2010: ~10 %), and 4.8 % with the distractor average (Fehd & Seiffert, 2010: ~5 %; see Fig. 1B in Fehd & Seiffert, 2010). A further analysis of gaze switches revealed that center–target switches (*M* = 5.5, *SD* = 1.9) were more often executed than target–target switches (*M* = 1.3, *SD* = 0.6), which again is in perfect line with the previous results (Fehd & Seiffert, 2010, Exp. 1: 5.9 vs. 1.9). Thus, we could demonstrate again that the centroid seems to have a pronounced value in MOT. From a vision perspective, an obvious explanation for this gaze behavior is that peripheral information about the target positions is used to effectively monitor all targets. On the basis of this finding, the central research question, on the role of peripheral vision in monitoring moving objects with unexpected changes, could reasonably be asked.

Experiment 2

In Experiment 2, the functionality of peripheral vision in tracking multiple objects was under investigation. In particular, we aimed to answer the question of whether participants are able to respond to a target-related change in MOT peripherally—that is, before gaze is on this target. Furthermore, since motion information is used for tracking and because of the high temporal resolution of the peripheral retina, we also predicted that motion changes would be detected better than form changes if both kinds of change detections featured comparable degrees of task difficulty. For analyzing events such as motion or form changes, the evaluated stationary-centroid phase from Experiment 1 would be important, as this manipulation opened the door for precisely controlling the eccentricities of events. For evaluating the dynamics of the visual search behavior that underlies change identification, we analyzed gaze using a saccade-detection algorithm to determine the "saccadic reaction time," which is defined as the time interval between the onset of a change and the onset of the saccade. We expected to find differences in saccadic reaction times between the two manipulation conditions, because if both kinds of change detections are of comparable task difficulty, due to the motion sensitivity of peripheral vision, a motion change should be detected better with peripheral vision than would a form change. More specifically, since peripheral vision is often used to guide saccades (Kowler, 2011) and a rescue saccade is more likely to be used in the form-change condition to perceive the change with overt attention (i.e., while the square is still a diamond), shorter saccadic reaction times would be expected in this case, which would also indicate that participants relied less on peripheral vision in the form-change condition.

In Experiment 2, the motion change was implemented as a sudden target stop from a speed of 6°/s, and the form change as a square-to-diamond change over the stationary-centroid phase. These specifications were made primarily for reasons of ecological validity, which means, in reference to the change detection demands in natural settings. To illustrate this point with the sports examples sketched at the beginning of this article, in a tennis match with an opponent moving close to the court's opposite baseline, this motion change would correspond to an opponent's sudden stop of a—rather slow—running movement at a speed of about 2.5 m/s, whereas the form change would approximately correspond to an opponent's racket movement of 30 cm to the left or right of his or her body. Hence, from an applied perspective, the experimental conditions can definitely claim ecological validity, which is important when it comes to the derivation of recommendations for real-life practices like sports training.

However, beacause of these specifications, and with regard to the above-made predictions on the superior detection rates for motion changes, the only point that could be inferred from

an empirical confirmation would be that motion changes *as they appear in natural settings* can be detected better than form changes *as they appear in natural settings*. From an experimental point of view, this inference satisfies to a limited degree only, since the alternative explanation that superior motion-change detection rates are caused mainly by higher salience cannot be ruled out. The crucial argument would then be that the specified motion change was simply of a minor task difficulty, which could be turned around either by a difficulty increase of the motion-change detection task (e.g., slowing down the target speed from 6°/s to 5°/s) or by a difficulty decrease of the form-change detection task (e.g., if the target were to become a cross of twice the original size). At this point, an objective measure of the respective task difficulties would be desirable, to be able to choose motion and form changes of comparable difficulties. However, due to the dimensional character of the two types of changes, such a simple measure is not available, for a principal reason explained by Rensink (2002) as follows, in his review of change detection research: "Performance depends on the magnitude of the change and there is no simple way to equate the visibility of different kinds of changes" (p. 255).

For this reason, to empirically control for task difficulty differences between the two change conditions, additional data were gathered by putting the two types of changes into foveal vision and collecting response times (since detection rates are generally 100 % for foveal vision). Response times of the same magnitude could then be taken as a hint that the specified changes do *not* differ in task difficulty per se, such that any potentially revealed superiority of motion-change detection rates in peripheral vision should mostly be ascribed to the specifics of the peripheral visual system. In the following experiments, the details of this task difficulty check (TDC) will be reported at the end of each Method subsection, as well as at the beginning of the Results section.

Method

Participants Since performing an MOT task and additionally detecting target changes is a quite complex experimental task, it seemed advisable to conduct the study with participants who were familiar with the basic task of MOT. For this reason, Experiment 2 was run with the participants who had already taken part in Experiment 1 ($N = 14$), accounting the 135 trials of Experiment 1 as a familiarization phase for the increased task demands. Experiment 2 was conducted five weeks after the completion of Experiment 1. For the additional TDC study, 36 volunteers were recruited (12 females, 24 males; age: 29.25 ± 6.65 years). Both experiments were undertaken in accordance with the Declaration of Helsinki.

Stimuli The stimuli were created in the same way as in Experiment 1, but they differed regarding speed, which was set to 6°/s for all trials, since response accuracy was highest at this speed in Experiment 1. Furthermore, motion and form changes of one of the targets were introduced. These changes were evoked in constant relation to the onset of the stationary-centroid phase, which, as in Experiment 1, started either 3.0, 3.5, or 4.0 s after motion onset of the objects. Unknown to the participants, the change always occurred exactly 250 ms after the collision of the target with a distractor, which caused the stationary-centroid phase. At this point in time, depending on the trial's specific manipulation condition, one out of the four targets either stopped for 0.5 s before continuing its motion (motion change) or abruptly became a diamond (i.e., the square was rotated by 45°) for 0.5 s, but without altering the current speed (form change) (see Fig. 4). Furthermore, to ensure a predefined eccentricity of the events, the distance of the event-target was forced to be 15° (if participants, as we would expect from Exp. 1, actually pursued a centroid strategy). To prevent participants from anticipating the changes, a control condition without any target changes was included. Because each control trial matched a change trial with a particular motion onset, despite the absence of any change, the control trials were assigned to a (virtual) stationary-centroid phase onset condition in accordance with this match.

The same stimulus configuration was used in the TDC experiment. This time, however, at the beginning of a single trial, only one target was highlighted with a red frame—namely, the target that would (virtually) change its motion or form over the subsequent motion phase. Participants were instructed to keep this target in foveal vision during the whole trial.

Apparatus In addition to the setup used in Experiment 1, a single-button response device (1000 Hz) was integrated into the VICON data collection system to measure motor-response times.

Procedure Experiment 2 was organized in individual single sessions that lasted about 1 h. Participants' task was to press the button as soon as a motion- or form-related target change occurred and to name the number of the changed target at the end of the trial. However, if no change occurred, the task was to recall all four targets. Participants initially practiced each of the three conditions with three trials that were not shown later in the test blocks, resulting in nine practice trials in total. Taking the stimulus constraints into account, a total of 120 trials were derived from 12 mother trials. These trials were quasirandomly ordered in ten blocks of 12 trials each, with the additional specification that, per block, each mother trial appeared exactly once, and each change/control condition and each onset of the stationary-centroid phase occurred exactly four times. As in Experiment 1, verbal decisions were recorded in writing, and no knowledge of the results was given after the test trials.
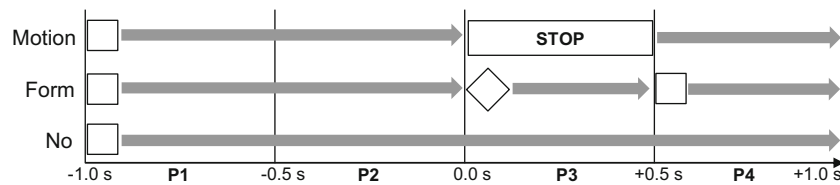
Fig. 4 Temporal properties of the three target-change conditions. In phase P3, the target either stopped (Motion) or became a diamond (Form), or continued to move unchanged (No) for 0.5 s, whereas no differences between the conditions are evident for the remaining phases P1, P2, and P4

In the TDC experiment, participants started with six familiarization trials followed by 30 test trials in two blocks (ten trials each for the form-change, motion-change, and no-change conditions), in which they had to press the button as soon as a motion- or form-related target change occurred. In addition, at the end of each trial, participants had to name the number of the target they had followed with their eyes.

Gaze analysis Since in Experiment 2 the focus was laid on the dynamics of gaze behavior as a function of monitoring object changes with peripheral vision, two kinds of trials had to be excluded from further analysis. First, a trial could not be included if the target change was not correctly signaled by a button press or was not correctly named at the end of the trial's motion phase, because in those trials attention was not on the correct target. Second, trials were excluded from further analysis if, at event start, the distance of the gaze vector to the to-be-changed target was less than 5°—that is, within the range of parafoveal vision (Calvo & Lang, 2005). The reason for this exclusion was that in those trials the participants could have focused on the particular target just by chance, so that the lack of a saccade should be considered as accidental and not be interpreted as resulting from a certain gaze-control strategy. In the Results section, detailed information will be given on the percentage of trials that had to be excluded due to these criteria.

For the remaining trials, two gaze-related groups of variables were aggregated. First, the distances of the gaze vector to the target as well as the centroid were calculated (in degrees) as the average value for the four phases P1–P4 of 0.5-s duration each. As is depicted in Fig. 4, these phases were defined in relation to the change event (in the case of control trials, to the virtual change event), with P1 beginning 1.0 s before, P2 beginning 0.5 s before, and P3 beginning exactly with the target change, and P4 beginning exactly with the change's termination. After the calculation of the mean gaze–target and gaze–centroid distances for the four phases per single trial, the values were averaged for each participant over the valid trials per change/control condition.

To analyze the dynamics of peripheral monitoring of the targets, for the valid motion- and form-change trials, a second group of gaze-related variables referred to the onset and offset of the first saccade to the changed target. For this purpose, saccades were identified by a velocity-based detection

algorithm with adaptive thresholds based on local noise levels (Nyström & Holmqvist, 2010). This analysis resulted in three time intervals (in milliseconds), which (in cases in which the identification was signaled by the participants after the saccade) together comprised the motor-response time indicated by the button press: a *saccadic reaction time* from the onset of the target change to the onset of the saccade, a *saccadic duration* from the onset to the offset of the saccade, and a *fixation duration* from the offset of the saccade to the button press. For each participant, the values per trial were averaged over the valid trials for motion- and form-related changes. It can be argued that the shorter the saccadic reaction time is, the less is change detection based on peripheral vision.

In the TDC experiment, participants were required to use foveal vision for detecting target changes (maximum distance of 3° of visual angle to the target).

Measures Besides the gaze-related variables, response accuracy (as a percentage) was gathered—in the no-change trials, as the percentage of trials on which all four targets could be correctly recalled, and in the motion- and form-change trials, as the percentage of trials on which the target change was correctly signaled by the button press and on which the correct number of the event-target was named at the end of the trial as well. Furthermore, the motor-response time was calculated (in milliseconds) for all correct motion- and form-change trials as the time between target-change onset and the button press, and then averaged over all available trials. Response time calculations in the TDC experiment were conducted in the same manner.

All dependent variables were analyzed with repeated measures ANOVAs, including either all three or just the two target-change conditions as the first variable and, if applicable, an additional gaze- or phase-related variable. Significant main or interaction effects were analyzed further with paired t tests. The alpha level for tests on the differences was set a priori to $\alpha = .05$, a-posteriori effect sizes were computed as η_p^2, and the power $(1 - \beta)$ was calculated in cases of nonsignificant tests.

We expected to find higher response accuracies in the motion-change than in the form-change condition, whereas response accuracies in the no-change condition should not differ from the results obtained in Experiment 1 for the same stimulus speed. Furthermore, we hypothesized that the average gaze distance to the centroid would be smaller than the

distance to the manipulated target in P1 and P2. However, gaze should be closer to the target in the form- and motion-change conditions in P4 (latest), whereas in the no-change condition, the gaze distance should continuously be smaller to the centroid than to the target. For the TDC experiment, we predicted that the participants would *not* be able to respond faster to foveally perceived motion changes than to foveally perceived form changes. Under the assumption that this prediction were to be confirmed, we anticipated superior performance for motion over form changes in the peripheral viewing conditions of the main experiment. With respect to the dependent variables described above, we particularly expected that the motor response would be executed ahead of the saccade onset in more trials for the motion-change than for the form-change condition.

Results

For the TDC experiment, paired t tests were used to test the response time differences for motion- versus form-change detections, and Cohen's d was calculated as the effect size. The data for 35 of the participants could be analyzed. The results not only revealed that motion changes were *not* detected faster than form changes, as had been predicted above, but that form changes were actually detected significantly faster than motion changes when both were in foveal vision (motion, $M = 427.8$ ms, $SE = 72.3$ ms; form, $M = 289.1$ ms, $SE = 48.9$ ms), $t(34) = 11.39$, $p < .01$, $d = 1.7$.

For the main experiment, three participants had to be excluded from the analysis because they were only able to perceive a target change if central gaze was directed by chance at the correct target, leading to less than five trials that fulfilled the above-defined inclusion criteria. In contrast, the remaining 11 participants clearly pursued a centroid-tracking strategy, because at target-change onset, the relevant target was out of parafoveal vision on 83.2 % ($SE = 4.1$ %) of the correctly identified motion-change and 78.0 % ($SE = 6.4$ %) of the correctly identified form-change trials, respectively. When applying the same calculation to the corresponding no-change conditions, based on the same mother trials as the change trials, all four of the targets would be successfully recalled and the relevant target would be out of parafoveal vision at the moment of the (virtual) target change on an average of 79.4 % ($SE = 6.4$ %) of the trials. To provide consistency, the following results refer to the 11 selected participants and to the just-described percentages of valid trials.

As is depicted in Fig. 5, participants were able to detect target-motion changes better than form changes. This was confirmed by the ANOVA based on all trials in which the change could be identified correctly (gray plus black fractions of the bars), $F(1, 10) = 17.2$, $p < .01$, $\eta_p^2 = .63$, as well as when the calculation was restricted to the trials that were valid in that sense that, at target-change onset, the relevant target was out
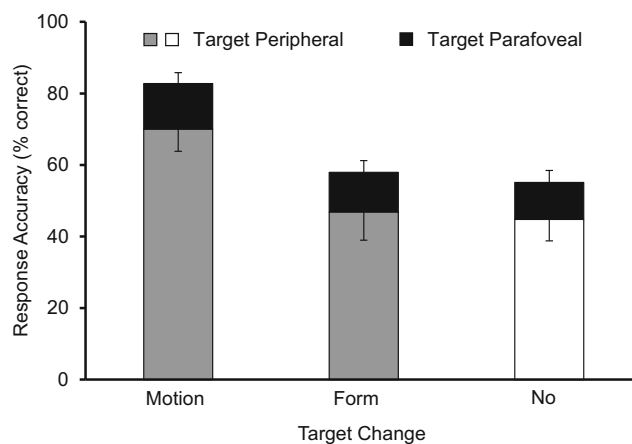

Fig. 5 Response accuracy (*M*s and *SE*s) as a function of target-change condition. In the motion- and form-change conditions, a trial was counted as correct if the button was pressed and the correct target number was named at the end of the trial, whereas in the no-change condition, all four targets had to be recalled correctly. The black fractions of the bars depict the percentages of trials that were excluded from further analyses because, at change onset, the relevant target was already in parafoveal vision

of parafoveal vision (gray fractions of the bars), $F(1, 10) = 10.71$, $p < .01$, $\eta_p^2 = .52$. Furthermore, response accuracy in the no-change condition ($M = 55.2$ %, $SE = 4.5$ %; white plus black fraction of the bar) showed no significant differences from the accuracies obtained in Experiment 1 for the same object speed ($M = 59.6$ %, $SE = 6.1$ %), $t(11) = 0.82$, $p = .43$, $1 - \beta = .74$.

For the event-related gaze-distance analysis, which is illustrated in Fig. 6, a 3 (Target Change) × 2 (Location) × 4 (Phase) ANOVA with repeated measures on all three factors was conducted. The results showed significant main effects for location, $F(1, 10) = 8.21$, $p = .02$, $\eta_p^2 = .45$, and phase, $F(3, 30) = 30.7$, $p < .01$, $\eta_p^2 = .76$, as well as the following significant two-way interactions: Target Change × Location, $F(2, 20) = 21.3$, $p < .01$, $\eta_p^2 = .68$; Target Change × Phase, $F(6, 60) = 4.9$, $p < .01$, $\eta_p^2 = .33$; and Location × Phase, $F(3, 30) = 42.6$, $p < .01$, $\eta_p^2 = .81$. Furthermore, a significant three-way, Target Change × Location × Phase interaction emerged, $F(6, 60) = 83.8$, $p < .01$, $\eta_p^2 = .89$. Consequently, to further interpret the three-way interaction, paired t tests were used to evaluate the gaze-distance differences to the centroid relative to the manipulated target for all target-change conditions and phases. The results showed that, for all target-change conditions, the gaze vector was closer to the centroid than to the respective targets in the first three phases (all ps ≤ .01), whereas in P4, gaze was closer to the target in both the motion-change condition and the form-change condition (all ps < .01), whereas the opposite was true in the no-change condition ($p < .01$). Furthermore, the gaze vector was closer to the target in P4 of the form-change trials than in P4 of the motion-change trials ($p = .03$).

In Fig. 7, the dynamics of the interplay between peripheral monitoring and saccadic reaction times from the centroid to
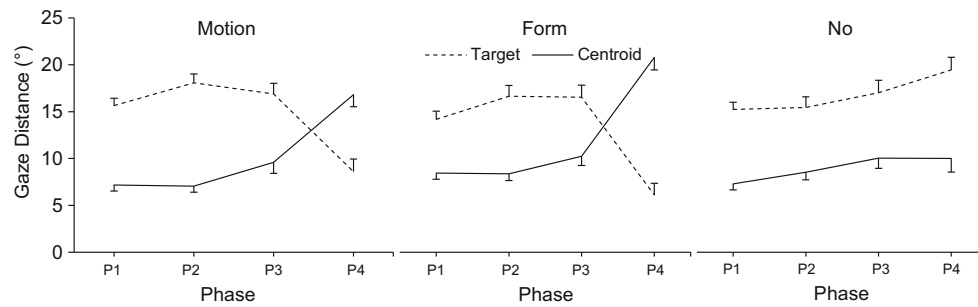
Fig. 6 Gaze distances (*M*s and *SE*s) to the centroid and the changing target for four phases (P1–P4), depicted separately for the three target-change conditions. In the no-change condition, gaze distances were calculated to the target that would change in the matched trials of the other conditions

the changed target are illustrated in more detail. As is depicted in Fig. 7a, the button was pressed before gaze was on the target in a higher percentage of correct trials in the motion-change than in the form-change condition, $F(1, 10) = 8.9$, $p = .01$, $\eta_p^2 = .47$, which means that in those cases, the change was definitely detected by means of peripheral vision. Furthermore, as is shown in Fig. 7b, an ANOVA revealed significant differences in the saccadic reaction times after target-change onset, $F(1, 10) = 6.7$, $p = .03$, $\eta_p^2 = .40$, as well as in the fixation times before the motor response was indicated by the button press, $F(1, 10) = 7.1$, $p = .02$, $\eta_p^2 = .42$, whereas no difference was found for saccade durations, $F(1, 10) = 1.6$, $p = .24$, $1 - \beta = .94$. This means that later saccade onsets to the changed target were observed in the motion-change condition than in the form-change condition, whereas in the form-change condition, targets were fixated longer before the motor response was executed.

Discussion

In Experiment 2, we introduced a novel approach to studying MOT performance, by using an event-driven saccade-based gaze analysis that precisely allowed for examining whether motor responses can be executed when changes in the environment are perceived with peripheral vision only. First of all,

for trials without a change, intra-individual response accuracies did not differ from the values obtained in Experiment 1, which could be seen as an indicator that the additional detection task did not affect tracking performance negatively (as was also suggested by Bahrami, 2003). Furthermore, no learning effects were observed over the five weeks between the two experiments. On this basis, with respect to the main research question, it could be clearly shown that motor responses can be reliably initiated if the decisive stimuli can be perceived by peripheral vision only. Moreover, as we hypothesized, higher response accuracies for the motion-change condition could be revealed than for the form-change condition. For the appropriate interpretation of this finding, the results of the TDC study must be taken into account, since they showed faster responses to form than to motion changes (of exactly the same magnitude as in the main experiment) when the changes were put into foveal vision, such that the superior detection of motion changes in the main experiment can definitely not be ascribed to a mere task difficulty effect, but must be attributed to the specifics of the peripheral visual system.

Having a closer look at the main results, the gaze distance was significantly closer to the centroid than to the targets over the two phases before the event (P1 and P2), and even over the phase of the change (P3), whereas over the last phase (P4), gaze was mostly on the target in change trials, whereas it
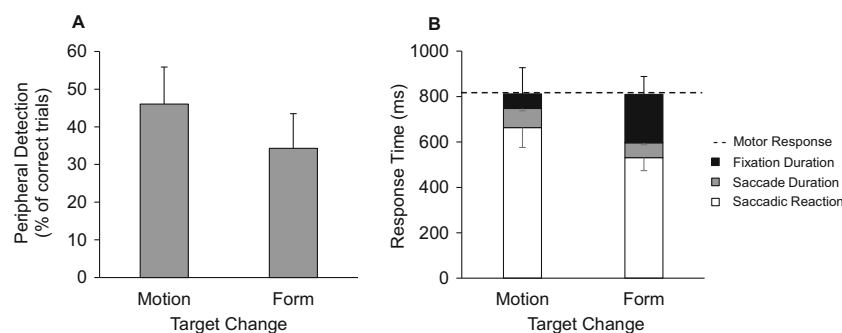

Fig. 7 (a) Peripheral change detection (*M*s and *SE*s) as a function of target change, calculated as the percentages of trials on which the motor response was executed before the gaze vector reached the respective target. (b) Saccadic reaction times, saccade durations, and fixation durations (*M*s and *SE*s) as cumulated responses up to the average motor response time. The "0" value on the vertical axis signifies the onset of the target change

stayed closer to the centroid in no-change trials. At this point, it should be noted that the gaze distances depicted in Fig. 6 resulted from an averaging procedure, which means that a relatively small percentage of trials with no saccades to the target resulted in large gaze–target distances in P4 of the change conditions. The same is true for the gaze–centroid distances in P1–P3 and, in addition, in P4 of the no-change condition, since saccades to targets were occasionally initiated (Elfanagely et al., 2011; Fehd & Seiffert, 2008, 2010), thereby increasing the mean distance to the centroid. For this reason, the respective values ranging between 5° and 10° should, in the end, be understood as the result of focusing on either a target or the centroid in most cases, but not always. Hence, the data should be interpreted in such a way that, in cases of a correct response, participants produced a saccade from the centroid to the respective target whenever a target changed its motion or form.

However, when bearing in mind that the target change over P3 lasted 0.5 s only, and that the change was terminated at the beginning of P4, this finding should not be misinterpreted in such a way that the saccades revealed for P4 were mandatorily needed to detect the change. This conclusion is underpinned by the gaze data depicted in Fig. 7, showing that the mean saccadic reaction times are longer than 0.5 s in both target-change conditions and that, on a considerable percentage of the trials, the saccade was initiated even after the button press—that is, on average, more than 0.8 s after the target change. Hence, in a lot of cases, the decision must have been made before verification by a suitable fixation could have taken place.

When comparing the two target-change conditions in more detail, in line with our predictions, on significantly more correct trials in the motion-change than in the form-change condition, the button press had been executed before the saccade to the changed target was initiated. In both cases, as we already argued above, these changes must have been perceived with peripheral vision, since the gaze position was more than 5° away from the changed target. Furthermore, saccadic reaction times turned out to be longer in the motion-change than in the form-change condition, supporting the assumption that foveal information is not required so much for motion-change as for form-change detection. On the contrary, as we hypothesized, a time of gaze being on the changed target seems to be required in the form-change condition before the button press can be executed (which would also be in line with the results of the TDC experiment, that form changes can be detected faster with foveal vision).

To test whether this saccade was helpful for responding faster, a "saccading strategy" (i.e., gaze is on the target *before* the button is pressed) and a "peripheral strategy" (i.e., gaze is on the target *after* the button is pressed) can be distinguished and checked against each other by post-hoc paired *t* tests for both manipulation conditions and gaze strategies. Eight participants could be integrated into this analysis, because they exhibited correct responses with both gaze strategies in both conditions. The results showed that response times for the two gaze strategies differed in neither the form-change (*p* = .93) nor the motion-change (*p* = .40) condition, indicating that initiating a saccade to the target does *not* help one to respond faster than does using peripheral vision only. Thus, initiating a saccade to the target does *not* lead to faster responses, but it could be interpreted as just being helpful with early tracking of the target, for naming the correct number at the end of the trial.

General discussion

In the present study, we aimed to investigate whether peripheral vision can be used to monitor multiple moving objects as well as to detect single-target changes. For this purpose, the designated experimental MOT setup with a large projection and a constant-position centroid phase had to be checked first. Therefore, in Experiment 1, our successful replication of the findings regarding the use of a virtual centroid to track multiple objects (cf. Fehd & Seiffert, 2008, 2010) and the dependency of tracking accuracy on target speed (cf. Cavanagh & Alvarez, 2005) were important for conducting the main study. In the subsequent Experiment 2, we could thus clearly show that peripheral vision is used to a great extent to detect changes in MOT. Moreover, for the change conditions, specified on the basis of considerations regarding the ecological validity of the manipulations, we could demonstrate that motion changes can be detected peripherally better than form changes. Finally, an additionally control experiment showed that this superiority must be explained by the particularities of the peripheral visual system, and cannot be attributed to possible differences in task difficulty.

As we argued in the introductory section, findings on MOT are often discussed exclusively from an attention perspective. However, in terms of the gaze-related findings revealed in the experiments at hand, it seems worthwhile to consider classical MOT results also from a vision perspective. In this regard, peripheral vision (i.e., perceiving stimuli with peripheral vision) and spatial covert attention (i.e., allocating attention without eye movements; Strasburger et al., 2011) seem to describe closely related phenomena. Therefore, in the following paragraphs, five reasons will be discussed regarding integrating the properties of the visual system in the context of attention-related findings to understand the role of covert attention in MOT. These reasons refer to (a) the ecological validity of methods to measure covert attention; (b) the detection of feature changes that depend on visual capabilities; (c) the connection of covert attention to a centroid-MOT strategy, and the consequences of saccade-related costs for processes of attentional monitoring; (d) the fact that eyetracking devices

can be generally useful for testing assumptions made in attentional theories of MOT; and (e) spatial uncertainty in the context of covert-attention processes, which could also be related to the voluntary control of saccades. At the end of each paragraph, a follow-up question for future experiments will be derived.

(a) In classical MOT studies, participants were usually instructed to fixate a central point, to examine covert attention in detecting feature changes (e.g., Sears & Pylyshyn, 2000). This approach, however, seems problematic, since, first, the fixation instruction was not checked by the application of gaze registration methods, and second, fixating a central point may not reflect the participant's natural behavior. In contrast, in the present experiment, participants were tested under ecologically valid conditions, left free to move their eyes and their head, because their actual gaze could be controlled by eye- and head-tracking devices. Beyond these advantages regarding ecological validity, event-driven saccade-based gaze analysis offers the opportunity to estimate whether peripheral or foveal vision was used to detect changes, leading to enriched inferences on the role of covert attention in MOT. Thus, in future research, by applying eyetracking methods, covert-attention processes may be examined even in real-life settings in which free gaze behavior is allowed.

(b) The use of covert attention not only depends on psychological but also on visual capabilities, especially when target changes have to be detected in MOT. In former studies (e.g., Bahrami, 2003), participants regularly showed better detection rates for color than for form changes, but visual capabilities were not taken into account in terms of the research questions pursued. Thus, it is not currently known whether the superiority regarding color-change detection actually results from specifics of the attentional process or could be attributed to particularities of the visual system. In the present study, those visual capabilities were taken into account, because finding better detection rates for motion than for form changes was expected, due to the light and motion sensitivity of the peripheral human retina, which comes along with a higher sensitivity to detect motion changes (Finlay, 1982; Goldstein, 2010; McKee & Nakayama, 1984). Indeed, the frequency of detection was found to be higher in the motion-change condition; however, the contrast sensitivity of the peripheral retina obviously allowed for detecting form changes, too. Consequently, the lower detection rates for form changes can be traced back to the lower spatial resolution of the peripheral visual system. Hence, in a follow-up experiment, the low spatial acuity but high motion sensitivity of the peripheral retina should be further investigated by manipulating the eccentricities of the changes. In the case of a visual-system effect on detection rates, the prediction can be derived that, as a function of eccentricity, form-change detection should be impaired to a higher degree than motion-change detection.

(c) Previous research has shown that covert attention is used for tracking purposes (Fehd & Seiffert, 2008, 2010). The gaze-related findings of the present study extend this view by showing that a centroid strategy is preferred not only for tracking, but also for change detection. This gaze strategy is presumably favored because a target-switching strategy with an alternating gaze between targets (Fehd & Seiffert, 2010) would lead to great costs, due to the suppression of information processing before, during, and after a saccade (Diamond, Ross, & Morrone, 2000). As a consequence, the chances to miss the target-feature change would be higher. The participants in our study indeed seemed to avoid costly eye movements by using the virtual centroid strategy: in attentional terms, by monitoring the targets with covert attention and "waiting" for a feature change. With the results at hand, we can show that covert attention is not only used to monitor the distances between targets and barrier occluders (Zelinsky & Todor, 2010), but also to detect target changes. In future experiments, this hypothetical mechanism could be investigated further by manipulating the costs of eye movements. In more detail, due to the suppression of information processing as a consequence of saccades, it would be expected that the more saccades were experimentally induced, the more likely a change would be to remain undetected.

(d) By studying natural gaze behavior, assumptions made by theoretical models on the role of attention in MOT can be assessed more systematically. On the one hand, our findings confirm attentional theories in different respects. In this regard, the observed gaze behavior indicates the value of the center of the polygon for MOT performance (Fehd & Seiffert, 2010), which is in line with the model proposed by Yantis (1992). Furthermore, the higher detection rates in the motion-change condition can be explained by the fact that the anticipated target locations of the polygon (Iordanescu, Grabowecky, & Suzuki, 2009) are no longer perceptible because of a target stop, whereas the polygon still exists in the form-change condition, causing lower detection rates. In this case, presumably because of the high tracking load with four targets (Fencsik et al., 2007; Howe & Holcombe, 2012; Luu & Howe, 2015), no motion extrapolation seems to be used in MOT. Hence, change detection results could be explained better with the multifocal theory and concurrent streams on the targets, where position information was used for tracking because participants were able to detect motion as well as form changes. Thus, a position tracker

and a stream of object features seem to be concurrently used to keep track of the targets.

On the other hand, however, our gaze analyses also pose challenges for attentional MOT theories. For example, in the model proposed by Cavanagh and Alvarez (2005), the relative impacts of the motion tracker and feature stream have to be questioned, since it can be inferred from the results at hand that, when using covert attention to keep track of targets, the position tracker has a more pronounced sensitivity in the concurrent-tracking procedure, because motion changes were detected more often than form changes. This line of thought is supported by the weighting mechanism proposed by Papenmeier et al. (2014), because spatiotemporal information during tracking is reliable until a motion change is induced, but a "false" spatiotemporal reliability is produced in the form-change condition. Hence, the decrease in the distinctiveness of spatiotemporal and feature information could explain the lower detection rates in the form-change condition. Besides the relation of the two streams, the concurrency of attentional streams is challenged by our results, because a number of changes were not detected. Thus, it could be that streams do not permanently exist. Instead, covert attention seems to switch between the targets, leading to a limited amount of time to process the target information, which, in turn, would increase the difficulty of detecting a change and explain the missed changes. Another challenge arises in the context of the FINST model, in which it is assumed that focal attention must be directed to the targets to update their positions (Fehd & Seiffert, 2010; Oksama & Hyönä, 2004). Contrary to this assumption, our research shows that peripheral vision—that is, covert attention—can be used to update target features and detect feature changes, whereas foveal vision—that is, focal attention—seems mainly to be used to keep track of the changed target in order to correctly name its number at the end of the trial. In particular, the assumption of the feature-blind visual index model of Pylyshyn (1989), that feature information cannot be accessed during tracking, is challenged by our data. On the contrary, our results support the findings of Papenmeier et al. (2014), that spatiotemporal *and* feature information can be processed. Consequently, future research on MOT should be directed to empirical testing of the specific predictions of attention-related models, by integrating eyetracking technology. In particular, researchers should test whether the stop of the anticipated target causes its detection, or whether a slowdown would lead to the same effect, to figure out how closely the polygon is linked to the mechanisms of attentional tracking.

(e) Our findings on gaze behavior in MOT also give rise to the speculation that the use of overt attention, measured as saccadic reaction times, is affected by spatial uncertainty as well as by the voluntariness of saccades. First, the ability to use covert attention to detect changes in the environment can be related to spatial uncertainty, because a low contrast in peripheral stimuli results in higher spatial uncertainty (for an overview, see Carrasco, 2011). In the present experiment, form changes presumably caused fewer contrast changes per frame, so that spatial uncertainty can be assumed to be higher in this condition. As a consequence, the early saccade onset in the form-change condition might result from this uncertainty. In contrast, in the motion-change condition, the saccade was initiated later, potentially because of lower spatial uncertainty. Thus, it seems worthwhile to take spatial uncertainty into account in future experiments on covert attention in MOT. Furthermore, we can speculated that the saccades in our study were voluntarily controlled. This speculation is substantiated by the fact that saccades are usually initiated in a bottom-up process, resulting in an involuntary overt-attentional shift as soon as stimuli in the environment change their motion (Drew et al., 2009; Hillstrom & Yantis, 1994). In our case, those shifts of overt attention were only sparsely observed. Hence, the assumption seems to make sense that the saccades to the event-target were voluntarily controlled (Peterson, Kramer, & Irwin, 2004), in order to update the target position because of the higher foveal resolution (Zelinsky & Todor, 2010). Consequently, in a follow-up study, we plan to eliminate the task-dependent evocation of a saccade by pursuing a dual-task approach in which a button has to be pressed as soon as a change is detected and, regardless of a change, all four targets have to be recalled at the end of the trial. If peripheral vision sufficiently allows for change detection, it can be predicted that no, or at least fewer, saccades would be executed, because the main task would be to identify the four targets at the end of each trial. If, by this means, the independence of change detection from subsequent saccades could be confirmed, we also plan to experimentally vary the before-discussed uncertainty of the change event, to investigate whether the degree of uncertainty plays a crucial role in the effectiveness of peripheral monitoring in MOT.

In summary, our experimental approach allows for a more ecologically valid examination of covert attention in MOT. Our empirical results show that the detection of feature changes depends on visual capabilities like spatial resolution and motion sensitivity, that the centroid strategy for MOT seems also to be beneficial for change detection by avoiding saccade-related costs, that feature and spatiotemporal information can be processed with covert attention by concurrent attentional streams, and that spatial uncertainty might affect the voluntary

control of change detection with either overt or covert attention. In any case, the integration of eyetracking methods in the field of attention-related MOT research seems to be worthwhile, since existing findings may appear in a new light, and novel research questions may reasonably be derived. Hence, the follow-up questions that have been suggested above are currently being investigated by our research group.

References

Alvarez, G. A., & Franconeri, S. L. (2007). How many objects can you track? Evidence for a resource-limited attentive tracking mechanism. *Journal of Vision, 7*(13), 14:1–10. doi:10.1167/7.13.14

Bahrami, B. (2003). Object property encoding and change blindness in multiple object tracking. *Visual Cognition, 10,* 949–963.

Boucart, M., Moroni, C., Thibaut, M., Szaffarczyk, S., & Greene, M. (2013). Scene categorization at large visual eccentricities. *Vision Research, 86,* 35–42. doi:10.1016/j.visres.2013.04.006

Calvo, M. G., & Lang, P. J. (2005). Parafoveal semantic processing of emotional visual scenes. *Journal of Experimental Psychology: Human Perception and Performance, 31,* 502–519. doi:10.1037/0096-1523.31.3.502

Carrasco, M. (2011). Visual attention: The past 25 years. *Vision Research, 51,* 1484–1525. doi:10.1016/j.visres.2011.04.012

Catteeuw, P., Helsen, W., Gilis, B., Van Roie, E., & Wagemans, J. (2009). Visual scan patterns and decision-making skills of expert assistant referees in offside situations. *Journal of Sport and Exercise Psychology, 31,* 786–797.

Cavanagh, P., & Alvarez, G. A. (2005). Tracking multiple targets with multifocal attention. *Trends in Cognitive Sciences, 9,* 349–354. doi:10.1016/j.tics.2005.05.009

Davids, K. (1984). The role of peripheral vision in ball games: Some theoretical and practical notions. *Physical Education Review, 7,* 26–40.

Diamond, M. R., Ross, J., & Morrone, M. C. (2000). Extraretinal control of saccadic suppression. *Journal of Neuroscience, 20,* 3449–3455.

Drew, T., McCollough, A. W., Horowitz, T. S., & Vogel, E. K. (2009). Attentional enhancement during multiple-object tracking. *Psychonomic Bulletin & Review, 16,* 411–417. doi:10.3758/PBR.16.3.411

Elfanagely, O., Haladjian, H., Aks, D., Kourtev, H., & Pylyshyn, Z. (2011). Eye-movement dynamics of object-tracking. *Journal of Vision, 11*(11), 280:1–16. doi:10.1167/11.11.280

Faubert, J., & Sidebottom, L. (2012). Perceptual–cognitive training of athletes. *Journal of Clinical Sport Psychology, 6,* 85–102.

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

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

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

Finlay, D. (1982). Motion perception in the peripheral visual field. *Perception, 11,* 457–462.

Franconeri, S. L., Lin, J. Y., Pylyshyn, Z. W., Fisher, B., & Enns, J. T. (2008). Evidence against a speed limit in multiple-object tracking. *Psychonomic Bulletin & Review, 15,* 802–808. doi:10.3758/PBR.15.4.802

Goldstein, E. B. (2010). *Sensation and perception* (8th ed.). Belmont, CA: Wadsworth Cengage Learning.

Gralla, V. (2007). *Peripheres Sehen im Sport—Möglichkeiten und Grenzen dargestellt am Beispiel der synchronoptischen Wahrnehmung* [Peripheral vision in sports—Possibilities and limits depicted with the example of synchronoptic perception]. Dissertation, Ruhr University Bochum. Retrieved from www-brs.ub.ruhr-uni-bochum.de/netahtml/HSS/Diss/GrallaVolker/diss.pdf

Hartmann, E., Lachenmayr, B., & Brettel, H. (1979). The peripheral critical flicker frequency. *Vision Research, 19,* 1019–1023.

Hillstrom, A. P., & Yantis, S. (1994). Visual motion and attentional capture. *Perception & Psychophysics, 55,* 399–411. doi:10.3758/BF03205298

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

Huff, M., & Papenmeier, F. (2013). It is time to integrate: The temporal dynamics of object motion and texture motion integration in multiple object tracking. *Vision Research, 76,* 25–30.

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

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. doi:10.1167/9.4.1

Kowler, E. (2011). Eye movements: The past 25 years. *Vision Research, 51,* 1457–1483. doi:10.1016/j.visres.2010.12.014

Kredel, R., Klostermann, A., Lienhard, O., Koedijker, J., Michel, K., & Hossner, E. J. (2011). Perceptual skill identification in a complex sport setting. *BIO Web of Conferences, 1,* 00051. doi:10.1051/bioconf/20110100051

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

Li, F. F., VanRullen, R., Koch, C., & Perona, P. (2002). Rapid natural scene categorization in the absence of attention. *Proceedings of the National Academy of Sciences, 99,* 9596–9601.

Luu, T., & Howe, P. D. L. (2015). Extrapolation occurs in multiple object tracking when eye movements are controlled. *Attention, Perception, & Psychophysics, 77,* 1919–1929. doi:10.3758/s13414-015-0891-8

McKee, S. P., & Nakayama, K. (1984). The detection of motion in the peripheral visual field. *Vision Research, 24,* 25–32.

Meyerhoff, H. S., Papenmeier, F., Jahn, G., & Huff, M. (2013). A single unexpected change in target—But not distractor motion impairs multiple object tracking. *i-Perception, 4,* 81–83. doi:10.1068/i0567sas

Meyerhoff, H. S., Papenmeier, F., & Huff, M. (2013). Object-based integration of motion information during attentive tracking. *Perception, 42,* 119–121. doi:10.1068/p7273

Nyström, M., & Holmqvist, K. (2010). An adaptive algorithm for fixation, saccade, and glissade detection in eyetracking data. *Behavior Research Methods, 42,* 188–204. doi:10.3758/BRM.42.1.188

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. doi:10.1080/13506280344000473

Papenmeier, F., Meyerhoff, H. S., Jahn, G., & Huff, M. (2014). Tracking by location and features: Object correspondence across spatiotemporal discontinuities during multiple object tracking. *Journal of Experimental Psychology: Human Perception and Performance, 40,* 159–171. doi:10.1037/a0033117

Peterson, M. S., Kramer, A. F., & Irwin, D. E. (2004). Covert shifts of attention precede involuntary eye movements. *Perception & Psychophysics, 66,* 398–405. doi:10.3758/BF03194888

Pylyshyn, Z. (1989). The role of location indexes in spatial perception: A sketch of the FINST spatial-index model. *Cognition, 32,* 65–97. doi: 10.1016/0010-0277(89)90014-0

Pylyshyn, Z. W., & Storm, R. W. (1988). Tracking multiple independent targets: Evidence for a parallel tracking mechanism. *Spatial Vision, 3,* 179–197. doi:10.1163/156856888X00122

Rensink, R. A. (2002). Change detection. *Annual Review of Psychology, 53,* 245–277. doi:10.1146/annurev.psych.53. 100901.135125

Sears, C. R., & Pylyshyn, Z. W. (2000). Multiple object tracking and attentional processing. *Canadian Journal of Experimental Psychology, 54,* 1–14. doi:10.1037/h0087326

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

Strasburger, H., Rentschler, I., & Jüttner, M. (2011). Peripheral vision and pattern recognition: A review. *Journal of Vision, 11*(5), 13:1–82. doi: 10.1167/11.5.13

Thorpe, S. J., Gegenfurtner, K. R., Fabre-Thorpe, M., & Bülthoff, H. H. (2001). Detection of animals in natural images using far peripheral vision. *European Journal of Neuroscience, 14,* 869–876.

Williams, A. M., & Davids, K. (1998). Visual search strategy, selective attention, and expertise in soccer. *Research Quarterly for Exercise and Sport, 69,* 111–128.

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

Yu, H. H., Chaplin, T. A., & Rosa, M. G. P. (2015). Representation of central and peripheral vision in the primate cerebral cortex: Insights from studies of the marmoset brain. *Neuroscience Research, 93,* 47–61.

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

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

Zhong, S.-H., Ma, Z., Wilson, C., Liu, Y., & Flombaum, J. I. (2014). Why do people appear not to extrapolate trajectories during multiple object tracking? A computational investigation. *Journal of Vision, 14*(12), 12:1–30. doi:10.1167/14.12.12