

# Disentangling vision and attention in multiple-object tracking: How crowding and collisions affect gaze anchoring and dual-task performance

Christian Vater

Institute of Sport Science,  
University of Bern, Bern, Switzerland

Ralf Kredel

University of Bern, Bern, Switzerland

Ernst-Joachim Hossner

University of Bern, Bern, Switzerland

Previous studies of multiple-object tracking have shown that gaze behavior is affected by target collisions and target–distractor crowding. Therefore, in order to experimentally disentangle this collision-crowding confound, we examined events of target collisions with the bordering frame and crowding with distractors. We hypothesized that collisions are particularly demanding for covert attentional processing, whereas crowding particularly challenges peripheral vision. Results show that gaze is located closer to targets when they are crowded, as would be expected to reduce negative crowding effects by utilizing the higher spatial acuity of foveal vision. However, saccades, which interrupt visual information processing, were instead initiated as a function of target collisions with the bordering frame. Consequently, in a dual-task condition that required the detection of target changes, participants more frequently missed changes if they occurred in time intervals around a collision. Based on these results, superior performance should be expected if foveal gaze is optimally anchored among crowded targets and if potential target changes are monitored with peripheral vision. In addition to the implications for further laboratory research of multiple-object tracking, these findings are relevant to a multitude of tasks that require the monitoring of several targets and the simultaneous detection of certain events in the visual periphery, as it is commonly the case, for instance, in sports.

## Introduction

The ability of humans to visually perceive their environment depends, on one hand, on characteristics of the human visual system, particularly, the high density of cones near the fovea and the high density of

rods in the periphery (Strasburger, Rentschler, & Jüttner, 2011). The resulting high spatial acuity of foveal vision is useful for processing detailed information, while the contrast and motion sensitivity of peripheral vision is useful for processing motion-related changes. On the other hand, human perception is affected by attention (Bruce, Green, & Georgeson, 2003), as studies have shown that both contrast sensitivity and spatial resolution improve if attention is located properly (Carrasco, Ling, & Read, 2004; Gobell & Carrasco, 2005). This interaction between vision and attention is highly relevant in a multitude of real-world settings, especially in sports games. Here, the practical question arises whether foveal vision, with focused (overt) attention to a particular location, should be preferred to peripheral vision, with divided (covert) attention to multiple locations, when monitoring a number of players and the ball for improved decision making by athletes (Williams, Davids & Williams, 1999). To empirically determine how visual and attentional demands affect gaze behavior, investigation in an experimental setting that allows for the controlled and isolated manipulation of both vision and attention is primarily necessary.

An experimental paradigm repeatedly applied to examine perceptual and attentional processes is multiple-object tracking (MOT). The task here is to track a number of targets amid identical-looking distractors and recall the targets at the end of the trial (Pylyshyn & Storm, 1988). A number of underlying tracking mechanisms are discussed in the scientific literature (for a discussion of serial and parallel accounts, see Howe, Cohen, Pinto, & Horowitz, 2010). According to the multifocal-attentional mechanism (Cavanagh & Alvarez, 2005), targets are assumed to be tracked in parallel, meaning that spatial attention is linked to each

Citation: Vater, C., Kredel, R., & Hossner, E.-J. (2017). Disentangling vision and attention in multiple-object tracking: How crowding and collisions affect gaze anchoring and dual-task performance. *Journal of Vision*, 17(5):21, 1–13, doi:10.1167/17.5.21.

doi: 10.1167/17.5.21

Received August 23, 2016; published May 30, 2017

ISSN 1534-7362 Copyright 2017 The Authors



This work is licensed under a Creative Commons Attribution 4.0 International License.

individual target and that target motion information is processed simultaneously (Alvarez & Scholl, 2005; Fencsik, Klieger, & Horowitz, 2007; Howe & Holcombe, 2012; Luu & Howe, 2015; Oksama & Hyönä, 2004).

In MOT, the effective use of spatial attention seems to be challenged by certain target–distractor formations. For example, if distractors approach tracked targets within approximately  $3^\circ$  of visual angle, a crowding effect can be observed (cf. Iordanescu, Grabowecky, & Suzuki, 2009), generally degrading tracking performance in MOT (Alvarez & Franconeri, 2007; Franconeri, Lin, Pylyshyn, Fisher, & Enns, 2008; Shim, Alvarez, & Jiang, 2008; Tombu & Seiffert, 2008). Iordanescu et al. (2009) ascribed this lower tracking performance to higher tracking demands caused by the need to dynamically adjust the distribution of attention to targets that are crowded by distractors. This assumption of heightened inhibition demands due to increased nearby distractors has been tested with probe detections. As predicted, it has been found that distractors close to targets are indeed more inhibited than those further away (Doran & Hoffmann, 2010a; see also Bettencourt & Somers, 2009; Doran & Hoffmann, 2010b; Meyerhoff, Papenmeier, Jahn, & Huff, 2016; Pylyshyn, Haladjian, King, & Reilly, 2008). Nevertheless, the crowding effect on tracking performance can also be explained by the low spatial resolution of peripheral vision (Levi, 2008; Strasburger, 2005; Strasburger et al., 2011), which is predominantly used for target monitoring in MOT tasks (Vater, Kredel, & Hossner, 2016, 2017). To compensate for this low spatial acuity, “rescue” saccades are initiated to clustered target–distractor formations in order to utilize foveal vision (Zelinsky & Todor, 2010; see also Peterson, Kramer, & Irwin, 2004). However, as for crowding, an alternative explanation for these saccades arises in reference to attentional processes. From this perspective, it is of particular interest that the saccades are initiated *before* the separation of a target and a distractor falls below a critical distance, indicating that at the moment of saccade onset, attention must have already been on the critical object formation. The saccades can then be explained by anticipatory attentional shifts to the targets within these crowded target–distractor formations, where crowding causes collisions resulting in directional changes. Thus, a saccade would help to update the objects’ positions after a collision (Fehd & Seiffert, 2010).

In sum, the empirical findings of reduced tracking accuracy from both crowding conditions and the saccades to critical target–distractor formations can be explained twofold, with one assertion for the visual system and one for the attentional system. Therefore, the current study aims to resolve whether gaze location (determined by gaze distances and saccades) and/or spatial attention (determined by target-change detec-

tion rates) are attracted by crowding and/or collisions that induce directional changes in MOT. A *crowding* manipulation was introduced in which three distractors moved at a  $2^\circ$  proximity to three targets without causing a collision between targets or between a target and a distractor. These so-called *group targets* have the individual, near target–distractor pairs moving in similar directions. For the *collision* manipulation, a target was forced to collide with the rectangular frame that surrounded the targets and distractors. This collision with the bordering frame assured that collisions were manipulated independent of crowding, as it would be difficult to discriminate with collisions between targets or between a target and a distractor. With two sources of variance, factors of crowding and collision resulted in a  $2 \times 2$  experimental design. Based on previously reported results on crowding in MOT, reduced tracking accuracies were expected in the crowding condition, which could also be used as a manipulation check. Regarding the two main factors of the experimental design, it was predicted that crowding mainly challenges peripheral vision and consequently leads to a gaze position closer to the crowded set of objects (Strasburger, 2005). In contrast, directional changes of the targets caused by collisions should mainly impose attentional rather than visual demands, because the low spatial acuity of peripheral vision should not be a limiting factor for monitoring motion-direction changes of targets.

To determine the allocation of covert attention, a dual-task approach was applied with MOT as the primary task and a target-stop-detection task as the secondary task (cf. Vater et al., 2017). The stop could occur either at one of the three group targets, manipulated according to the above-sketched  $2 \times 2$  experimental conditions, or at a separate target that was neither part of a crowded configuration nor about to collide with the bordering frame. As covert attention is hypothetically attracted to targets that are about to collide (here, group targets), more target-stop detections should be observed for colliding compared to noncolliding targets (here, separate targets). As observed in previous studies (Vater et al., 2016, 2017), these changes should mainly be detected with peripheral vision, regardless of the manipulation condition.

## Methods

### Participants

Fourteen sports science students (seven women and seven men; aged  $20.4 \pm 1.1$  years) participated in the experiment and received course credit in return. Sample size had been determined a priori on the basis of

previous studies (Fehd & Seiffert, 2008, 2010; Vater et al., 2016, 2017).

Participants had self-reported normal or corrected-to-normal vision and were unaware of the research question. The experiment was conducted in accordance with the Declaration of Helsinki.

## Stimuli

MATLAB (2014a; MathWorks) was used to calculate the motion paths of 10 white squares ( $35 \text{ mm} \times 35 \text{ mm}$  corresponding to  $1^\circ \times 1^\circ$  of visual angle) so that all squares appeared in (quasi-)random starting positions with no overlapping objects (cf. Figure 1). These MOT stimuli were then imported to Autodesk 3ds Max software (Autodesk Inc., Lake Oswego, OR; 2014) to render single video trials. The white squares appeared within a rectangular frame (white line of 25 mm width,  $1.40 \text{ m} \times 1.40 \text{ m}$  corresponding to  $40^\circ \times 40^\circ$  of visual angle) on a black background. For each trial, the trial number was displayed prior to the presentation of the 10 stationary squares. Subsequently, the four targets were highlighted by red frames (line of 15 mm width; frame and stimulus together covering an area of  $1.7^\circ \times 1.7^\circ$  of visual angle). After 2 s, the target-defining cues disappeared, and all stimuli accelerated on straight-line paths for 1 s to reach a final speed of  $6^\circ/\text{s}$ . This final speed was then sustained for 4 s, followed by a subsequent deceleration phase of 1 s, after which all squares stopped. This pattern resulted in a total object-motion duration of 6 s. In the following and final 3 s of each trial, participants were to identify the initially highlighted targets by naming the respective numbers projected onto the now stationary squares.

A repulsion mechanism was used to redirect a square whenever the distance from the rectangular frame or the next square fell below a certain threshold ( $35 \text{ mm}$  corresponding to  $1^\circ$  of visual angle). In each trial, the centroid (i.e., the center of mass of the four targets that is expected to be visually tracked) was forced to stay at a constant position (termed *static-centroid phase*) for a period of 0.5 s, beginning 3.0 s, 3.5 s, or 4.0 s after motion onset. To achieve this, critical collisions between targets and distractors were induced so that the four targets moved in perpendicular directions for at least 0.5 s without any directional changes. Furthermore, at the beginning of the static-centroid phase, three of the targets were located in one quadrant of the display while the remaining target was located in the opposite quadrant, resulting in a configuration of three group targets and one separate target. Before and after the static-centroid phase, collisions were allowed for all objects without any restriction. The to-be-detected target stop, lasting 0.50 s, always

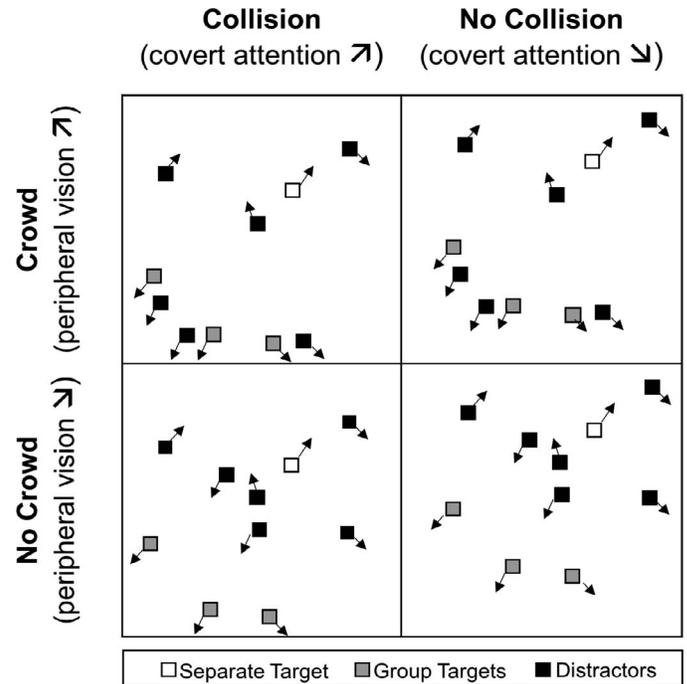


Figure 1. Stimulus configuration at the moment of the target-stop onset for the two crowding and the two collision conditions, with respective hypothetical high ( $\nearrow$ ) or low ( $\searrow$ ) demands on either the peripheral-visual system or covert-attentional processes. The target stop for 0.5 s involves either the separate target (depicted in the top-right quadrant) or one of the group targets (depicted in the bottom-left quadrant).

occurred 0.25 s after the start of the static-centroid phase.

Based on the discussed constraints, five primary trials were created. Each primary trial was then manipulated for the three stop conditions (stop group target vs. stop separate target vs. no stop), two crowding conditions (crowd vs. no crowd), and two collision conditions (collision vs. no collision) to ultimately generate 60 trials. Finally, these 60 trials were threefold rotated by  $90^\circ$  to assure that separate and crowd targets appeared in each of the four quadrants of the display equally often. Each rotation was linked to a new set of numbers finally projected onto the targets to prevent advantageous memory effects. In total, each participant performed 240 test trials. Each of the 20 test blocks grouped 12 test trials in a randomized order and were then rendered with MAGIX Video Pro X3 (MAGIX, Berlin, Germany). Within each block, it was ensured that (a) trials originating from the same primary trial were not presented consecutively, (b) four trials appeared for each stop condition, and (c) the group targets were presented equally often in each of the four quadrants (i.e., three times per block).

## Stimulus manipulations

The crowd versus no-crowd condition and the collision versus no-collision condition are illustrated in more detail in Figure 1. In the crowd condition, one distractor was bound to each of the three group targets over the static centroid phase, maintaining an average distance of  $2^\circ$  (edge-to-edge) from the respective target. In the no-crowd condition, this distance was set to  $20^\circ$  of visual angle. These distances were chosen based on the work of Iordanescu et al. (2009), in which a crowding condition was defined by object separations less than  $3^\circ$  of visual angle.

In order to impose the necessary collisions with the frame, the distance of the group targets to the rectangular frame was manipulated so that collisions would occur at the target-stop time (the time interval in which either the separate target or one of the group targets would stop for 0.5 s). At the onset of the (separate or group) target stop, the distance between the group targets and the frame was set to  $2.2^\circ$  of visual angle. With targets moving at a constant speed, this placement resulted in group target collisions 200 ms after the target-stop onset. In the no-collision condition, the distance of the group targets from the rectangular frame was set to  $4.0^\circ$  such that the collision with the frame occurred 400 ms after target-stop onset. These values were chosen as Atsma, Koning, and van Lier (2012) demonstrated that, in MOT, target-motion trajectories can be extrapolated by about  $3.0^\circ$ . Thus, in the collision condition, it can be expected that the anticipated collision has already attracted covert attention to the group targets at target-stop time since their distance to the frame is less than  $3^\circ$ . This would not be the case in the no-collision condition, as the  $4^\circ$  distance is too far to draw attention at this moment. The  $1.8^\circ$  difference in target location between the two conditions was generated by adding exactly this positional offset to all squares. Generally, it is important to note that the relative distances of the four targets from each other and from the centroid was exactly the same in every variation of the mother trial, as is illustrated in Figure 1, which presents four variants derived from the same primary trial in the same rotation. Thus, over the crucial target-stop time, the four stimulus-manipulation conditions only and selectively differed with regard to the crowding of the group targets and the collisions of group targets with the rectangular frame, anticipated at earlier or later times.

## Apparatus

A monocular eye-tracking system, EyeSeeCam, 220 Hz (ESC; EyeSeeTech GmbH, Fürstenfeldbruck, Ger-

many) was used to capture the vertical and horizontal rotations of the right eye via infrared reflections from the pupil and the cornea (accuracy:  $0.5^\circ$  of the visual angle; resolution:  $0.01^\circ$  RMS within  $25^\circ$  of the field of view). A MacBook Pro (Apple, Cupertino, CA) was connected to the ESC via a 20 m fiber-optic Firewire link (GOF-Repeater 800; Unibrain, Athens, Greece) to ensure mobility within the lab space. The eye-tracking software running on the MacBook Pro calculated the horizontal and vertical eye-rotation angles, which were streamed in real time over Ethernet to a control PC. Additionally, a 12-camera OptiTrack system (sample rate: 200 Hz; NaturalPoint, Inc., Corvallis, OR) tracked retro-reflective markers attached to the ESC and streams this positional information to the control PC. A custom software application on the control PC was responsible for synchronizing the two data streams, calculating a three-dimensional gaze vector in the laboratory reference frame and providing further functionalities such as trial selection, stimulus presentation, and data logging. In addition, a button press signal (from a Wii remote controller [Nintendo, Kyoto, Japan], connected via Bluetooth) was synchronously received by this custom software application. With this system, the participant's current gaze was related to the displayed objects and the button press in time steps of 5 ms.

For the initial calibration of the ESC at the beginning of the session, the participant consecutively fixated on 5 dots creating a two-dimensional axis with an origin, and with dots separated by the distance of  $8.5^\circ$  of visual angle (Kredel et al., 2011). A recalibration procedure was implemented before test blocks in which the point of gaze deviated more than  $0.5^\circ$  of visual angle from one of the calibration grid dots. A back projection (InFocus IN 5110 projector; InFocus, Portland, OR) onto a large screen (height: 1.87 m; width: 3.01 m) was used to display video stimuli played back on VLC Media Player 2.1.5 software (Softonic, Barcelona, Spain). In this set up, the rectangular frame for the MOT task covered an area of  $1.40\text{ m} \times 1.40\text{ m}$  in the middle of the screen. Participants held the button used for the detection task in their dominant hand. MATLAB (2014a; MathWorks) was used to analyze the gathered data. Further statistical analyses were conducted with IBM SPSS Statistics 23 (IBM Corp., Armonk, NY).

## Procedure

Participants were individually tested in the institute's sensorimotor laboratory in two 1-hr sessions, with Session 2 exactly 7 days after Session 1. Participants first read the general information about the study and the participation agreement and then signed a consent

form before being fitted with the eye-tracking system. Subsequently, participants were positioned at 2.0 m distance from the screen to read the task instructions. The participants' primary task was to recall the four targets cued at the beginning of each trial by naming the respective numbers projected onto the targets at the end of each trial. Their secondary task was to press the button with their dominant hand as fast as possible as soon as they detected a target stop. After providing the instructions, the ESC calibration routine was conducted. At the end of each trial, participants' verbal decisions about the four targets were recorded in writing by an experimenter. No feedback of the responses' correctness was given after the trials.

## Measures

As a manipulation-check and performance measure for the primary monitoring task, tracking accuracy was calculated as the percentage of trials in which all four targets were correctly recalled at the end of a trial. This calculation mainly refers to the question of whether crowding actually impairs tracking behavior (presented below as *tracking accuracy*).

To test the hypothesis that crowding “pulls” the gaze into the direction of the crowded configuration, an algorithm was designed to determine the relative distance from the current point of gaze to the group targets or the separate target. This measure is illustrated in Figure 2. First, for each video frame, the virtual centroid of the four targets was calculated by averaging the targets'  $x$ -coordinates and then  $y$ -coordinates. Next, a straight line was drawn from the separate target to the centroid and one third of this length was extended further from the centroid towards the group targets. By trigonometric reasoning, the resulting endpoint of this line corresponded to the center of mass of the three group targets. Defining this point as the 0% value and the separate target as 100% value of the straight line, the virtual centroid was necessarily located at exactly 25%. In a final step, the point of gaze was perpendicularly projected onto this straight line. The respective percentage value location indicated whether the point of gaze was located closer to the group targets (0%–25%) or to the separate target (25%–100%; presented below as *relative gaze distance*).

To test the hypothesis that target-stop detection accuracy is increased for targets that are about to collide, the percentage of trials with correctly detected group target stops was compared to those with separate target stops (labeled below as *detection accuracy*). In addition, to control for speed–accuracy tradeoffs, the time from target-stop onset to the button press was analyzed for all conditions (labeled below as *response time*).

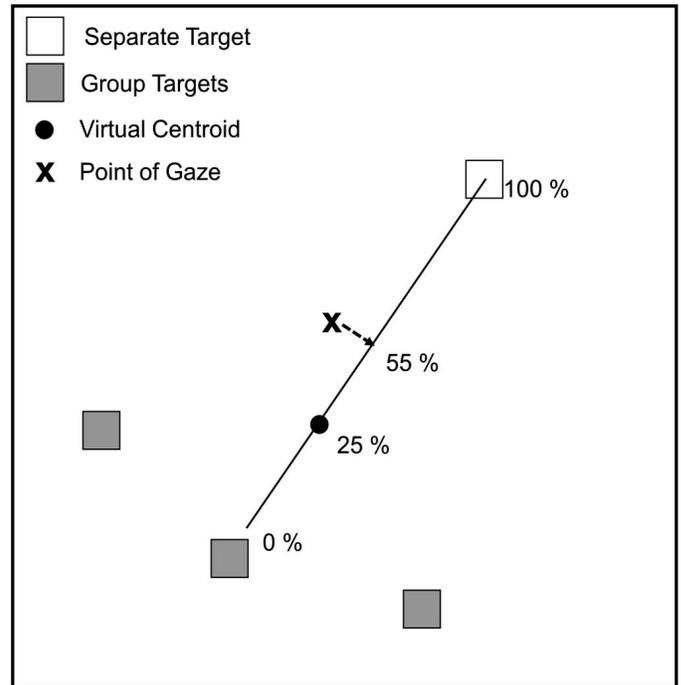


Figure 2. Calculation of the relative gaze distance from the group targets and from the separate target. In the illustrated example, the projection of the current point of gaze onto the line between the group-targets' center of mass and the separate target results in a value of 55%, meaning that the current gaze is closer to the separate target than to the group targets.

To test the hypothesis that target stops are regularly detected with peripheral vision, analysis was additionally directed to the percentage of peripheral detections, defined as the percentage of trials in which a target stop was detected while the target was beyond the range of foveal vision ( $>3^\circ$  of visual angle) before the moment of the button press (labeled below as *peripheral detection*).

Finally, an additional variable was calculated to examine the anticipatory saccades that are initiated immediately before a collision of target with the bordering frame (Fehd & Seiffert, 2010). For this purpose, the percentage of trials with saccades to a group target *before* a collision with the frame was computed and contrasted to the percentage of trials with saccades *after* the collision, thereby including all saccades within an interval of  $\pm 200$  ms around the collision (labeled below as *saccades to group targets*). For this computation, saccades were identified by a velocity-based detection algorithm with adaptive thresholds based on local noise levels (Nyström & Holmqvist, 2010). The time of saccade onset was decisive to assign either the value “before” or the value “after” to the respective saccade. The finding that more saccades were initiated before a collision replicates previously reported results on the anticipatory nature

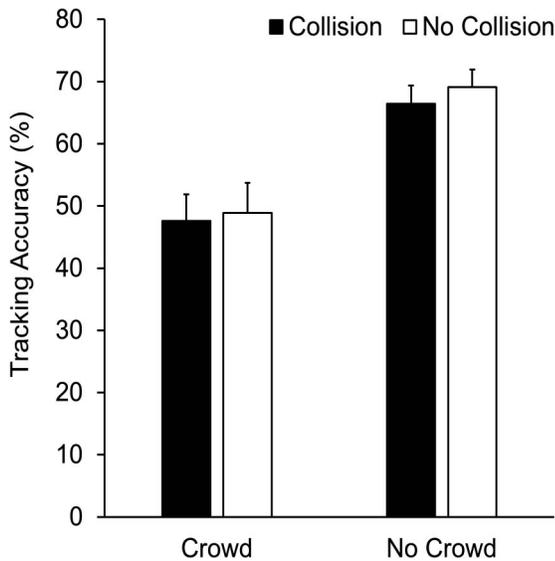


Figure 3. Tracking accuracy ( $M$  and  $SE$ ) in target-stop trials for both crowding and collision conditions.

of saccades (Fehd & Seiffert, 2010; Zelinsky & Todor, 2010). Beyond this replication aspect in terms of the predictors at hand, more anticipatory saccades in the crowd than in the no-crowd condition would indicate the need for foveal vision to separate targets from distractors, while more anticipatory saccades in the collision than in the no-collision condition would indicate that saccades are used to update target positions after a motion-direction change.

Dependent variables were analyzed with Crowding  $\times$  Collision ANOVAs with repeated measures on all factors (tracking accuracy, relative gaze distance). Depending on the variable of interest, either a target factor (separate target vs. group target; detection accuracy, response time, peripheral detection) or a saccade-timing factor (before collision vs. after collision; saccades to group targets) was additionally included. Significant main or interaction effects were further analyzed with paired  $t$  tests, with the  $\alpha$  level set to  $\alpha = 0.05$ , and a posteriori effect sizes were computed as partial eta squares ( $\eta_p^2$ ).

## Results

### Tracking accuracy

In a two-way ANOVA of Crowding  $\times$  Collision (Figure 3), a main effect for crowding was observed,  $F(1, 13) = 103.84$ ,  $p < 0.01$ ,  $\eta_p^2 = 0.89$ , showing that tracking accuracy declines in the crowd in comparison to the no-crowd conditions ( $M = 48.25\%$ ,  $SE = 4.31\%$  vs.  $M = 67.76\%$ ,  $SE = 2.73\%$ ). Neither an effect for

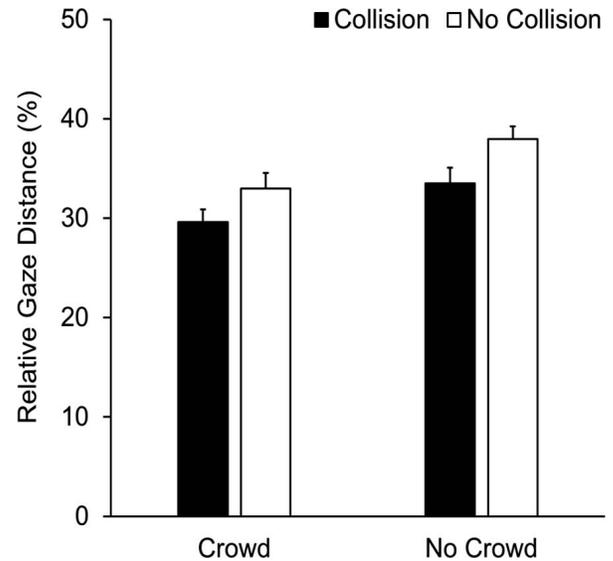


Figure 4. Relative gaze distance ( $M$  and  $SE$ ) with respect to the group targets (0%), the centroid (25%), and the separated target (100%) over the target-stop phase for both crowding and collision conditions.

collision,  $F(1, 13) = 1.36$ ,  $p = 0.27$ ,  $\eta_p^2 = 0.10$ , nor an effect for the interaction between both factors,  $F(1, 13) = 0.15$ ,  $p = 0.71$ ,  $\eta_p^2 = 0.01$ , was revealed.

### Relative gaze distance

Results for the relative distance of the point of gaze from either the separate or the group targets (Figure 4) showed a main effect for crowding,  $F(1, 13) = 92.35$ ,  $p < 0.01$ ,  $\eta_p^2 = 0.88$ , with gaze closer to the group targets in the crowd ( $M = 31.30\%$ ,  $SE = 1.57\%$ ) than in the no-crowd condition ( $M = 35.73\%$ ,  $SE = 1.53\%$ ). A further—slightly smaller—main effect was found for collision,  $F(1, 13) = 33.89$ ,  $p < 0.01$ ,  $\eta_p^2 = 0.72$ , revealing that gaze was closer to the group targets in the collision ( $M = 31.46\%$ ,  $SE = 1.36\%$ ) than in the no-collision conditions ( $M = 35.46\%$ ,  $SE = 1.76\%$ ). The interaction of both factors clearly failed to reach significance,  $F(1, 13) = 0.73$ ,  $p = 0.41$ ,  $\eta_p^2 = 0.05$ .

### Detection accuracy

While indicating detected target stops with a button press, few false alarms (i.e., a button press in absence of a target stop) were observed (4.2% of trials) such that the percentage of hits (i.e., a button press in presence of a target stop) can be designated as a valid measure of detection accuracy. For this variable, the three-way ANOVA Crowding  $\times$  Collision  $\times$  Target (Figure 5) showed a main effect for crowding,  $F(1, 13) = 6.11$ ,  $p = 0.03$ ,  $\eta_p^2 = 0.32$ , indicating that target stops were better

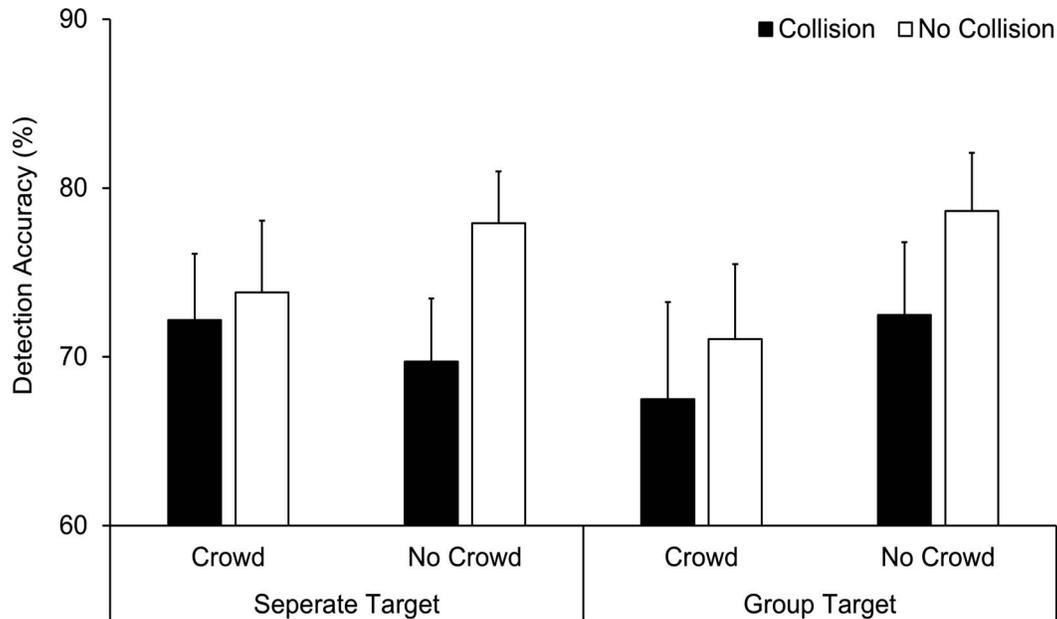


Figure 5. Detection accuracy ( $M$  and  $SE$ ) of separate or group target stops for both crowding and collision conditions.

detected in the no-crowd ( $M = 74.69\%$ ,  $SE = 3.06\%$ ) than in the crowd conditions ( $M = 71.13\%$ ,  $SE = 3.93\%$ ). A further main effect was found for collision,  $F(1, 13) = 16.91$ ,  $p < 0.01$ ,  $\eta_p^2 = 0.57$ , with more detections in the no-collision ( $M = 75.36\%$ ,  $SE = 3.06\%$ ) than in the collision condition ( $M = 70.46\%$ ,  $SE = 3.89\%$ ). In addition to these main effects, an interaction between crowding and target was observed,  $F(1, 13) = 5.93$ ,  $p = 0.05$ ,  $\eta_p^2 = 0.28$ , demonstrating that a group target stop was detected better in the no-crowd ( $M = 75.55\%$ ,  $SE = 3.73\%$ ) than in the crowd condition ( $M = 69.38\%$ ,  $SE = 4.88\%$ ). All other interactions clearly ( $p > 0.51$ ) and the interaction Crowding  $\times$  Collision barely ( $p = 0.07$ ) missed significance.

## Response time

For the time intervals from the onset of a target stop to the initiation of the button press (owing to circumstances, calculated for hits only), a three-way ANOVA Crowding  $\times$  Collision  $\times$  Target (Figure 6) revealed a main effect for target,  $F(1, 13) = 5.36$ ,  $p = 0.04$ ,  $\eta_p^2 = 0.29$ , with faster response times for the separate ( $M = 549$  ms,  $SE = 30$  ms) than for the group targets ( $M = 600$  ms,  $SE = 19$  ms). Furthermore, a main effect was identified for crowding,  $F(1, 13) = 11.95$ ,  $p < 0.01$ ,  $\eta_p^2 = 0.48$ , indicating longer response times in the crowd ( $M = 622$  ms,  $SE = 26$  ms) than in the no-crowd conditions ( $M = 526$  ms,  $SE = 26$  ms). Additionally, the Collision  $\times$  Target interaction showed significant effects,  $F(1, 13) = 8.81$ ,  $p = 0.01$ ,  $\eta_p^2 = 0.40$ . In the collision condition, separate target changes were detected faster ( $M = 515$  ms,  $SE = 34$  ms) than group

target changes ( $M = 637$  ms,  $SE = 20$  ms;  $p < 0.001$ ), while in the no-collision condition, no differences between group and separate target were observed ( $p = 0.57$ ). Additionally, group-target changes were detected faster in the no-collision ( $M = 562$  ms,  $SE = 24$  ms) than in the collision condition ( $M = 637$  ms,  $SE = 20$  ms), while there were no differences for the separate-target changes between collision and no-collision conditions ( $p = 0.15$ ). The interaction Crowding  $\times$  Collision just missed significance ( $p = 0.06$ ), and all other interactions showed no significant effects ( $p > 0.75$ ).

## Peripheral detection

As predicted, target stops were detected by peripheral vision in the vast majority of cases (overall average:  $M = 74.28\%$ ,  $SE = 1.26\%$ ). To examine how this effect was moderated by our experimental conditions, a three-way ANOVA Crowding  $\times$  Collision  $\times$  Target (Figure 7) showed a main effect for the target factor,  $F(1, 13) = 14.92$ ,  $p < 0.01$ ,  $\eta_p^2 = 0.53$ , with more peripheral detections of stops of the separated target ( $M = 88.83\%$ ,  $SE = 1.36\%$ ) than of a group target ( $M = 83.94\%$ ,  $SE = 1.77\%$ ). Secondly, a main effect for collision was found,  $F(1, 13) = 9.52$ ,  $p < 0.01$ ,  $\eta_p^2 = 0.42$ , as more peripheral detections were observed in the collision condition ( $M = 89.33\%$ ,  $SE = 1.30\%$ ) compared to the no-collision condition ( $M = 83.43\%$ ,  $SE = 2.10\%$ ). Besides these main effects, a significant interaction Crowding  $\times$  Target was revealed,  $F(1, 13) = 5.66$ ,  $p = 0.03$ ,  $\eta_p^2 = 0.30$ . In the crowd conditions, peripheral vision was used more often to detect the separate target stops than a group target stops ( $p < 0.01$ ), which was not the case

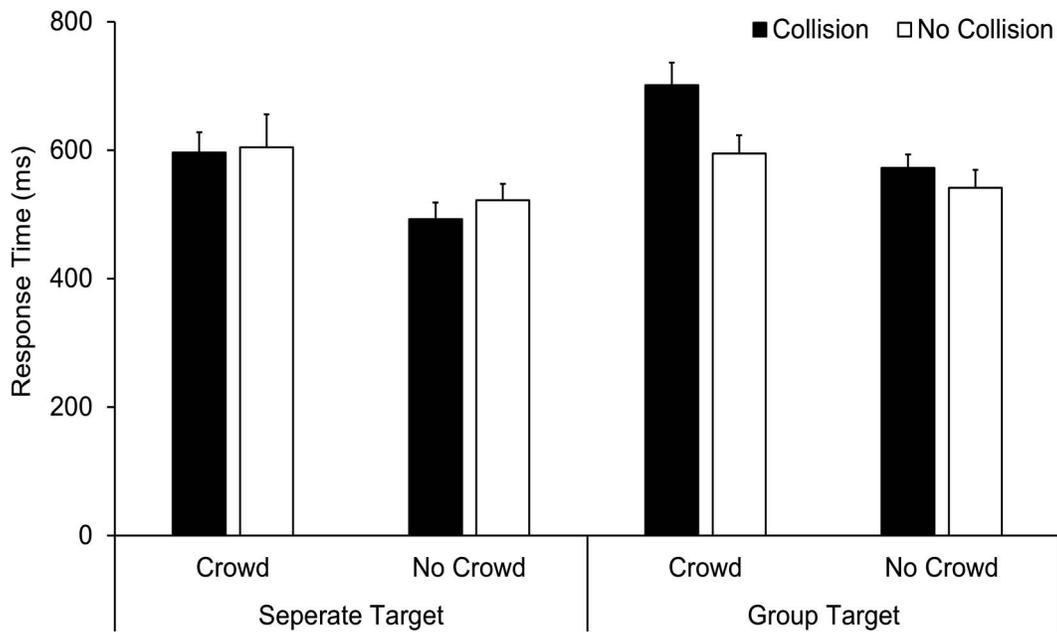


Figure 6. Response times (*M* and *SE*) for correct detections of the separate or a group target stop for both crowding and collision conditions.

in the no-crowd conditions ( $p = 0.61$ ). All other main or interaction effects failed to reach significance (all  $p > 0.15$ ).

### Saccades to group targets

A three-way ANOVA Crowding  $\times$  Collision  $\times$  Timing (Figure 8) revealed a huge main effect for timing,  $F(1, 13) = 244.45$ ,  $p < 0.01$ ,  $\eta_p^2 = 0.95$ , showing that saccades to group targets were more often initiated

before ( $M = 64.19\%$ ,  $SE = 3.82\%$ ) rather than after ( $M = 9.30\%$ ,  $SE = 1.43\%$ ) the collision of a group target with the rectangular frame. Furthermore, a Collision  $\times$  Timing interaction was found,  $F(1, 13) = 5.48$ ,  $p = 0.03$ ,  $\eta_p^2 = 0.31$ , indicating that before the collision, more saccades were initiated to group targets in the collision compared to the no-collision condition ( $p = 0.01$ ), where the collision was rather delayed to the target stop. No differences were detected between the collision conditions after the collision ( $p = 0.18$ ). Neither an effect for crowding,  $F(1, 13) = 0.02$ ,  $p = 0.90$ ,  $\eta_p^2 <$

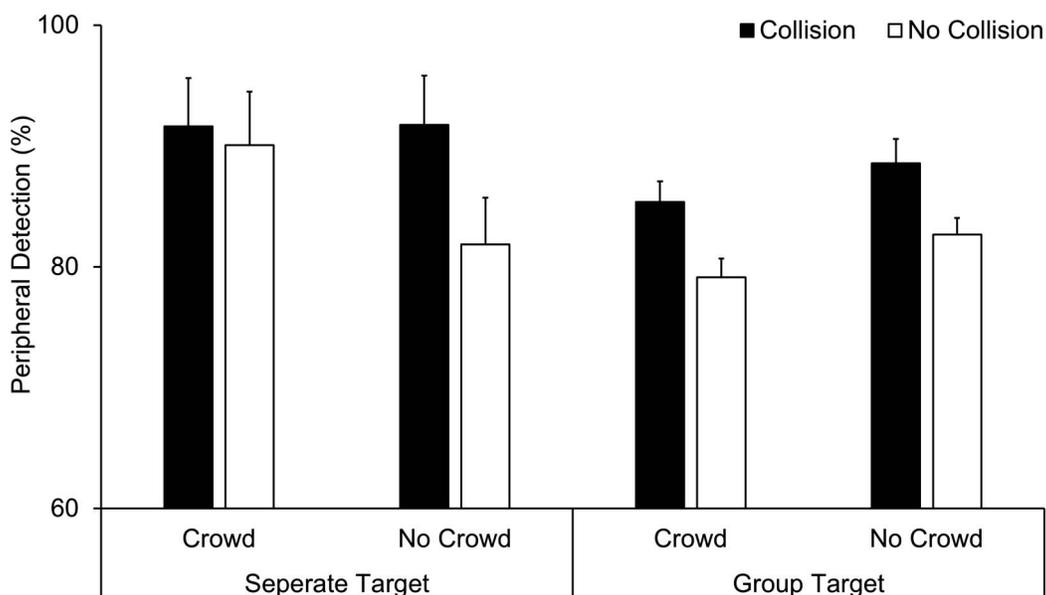


Figure 7. Peripheral detection (*M* and *SE*) of separate or group target stops for both crowding and collision conditions.

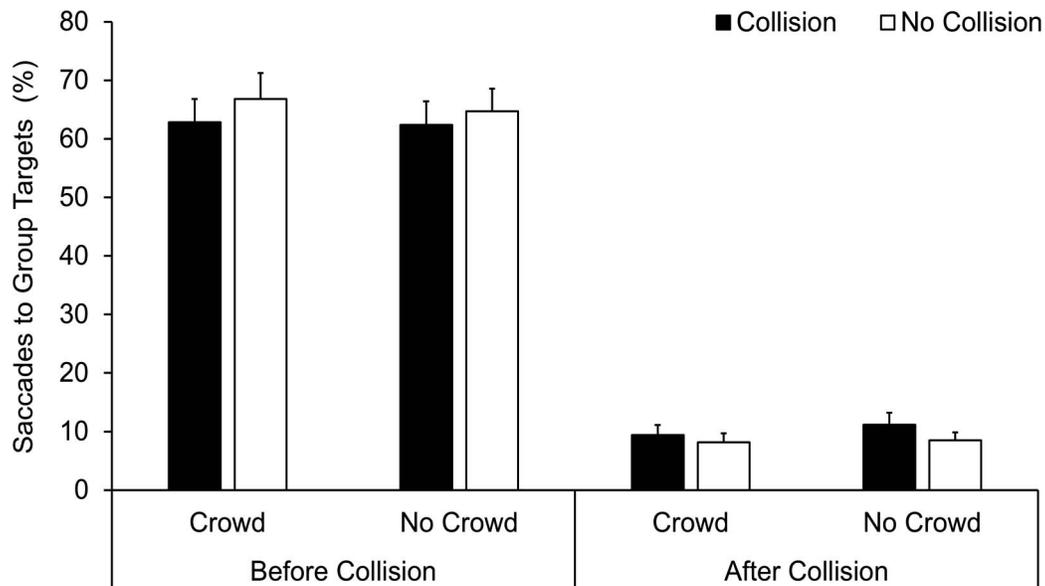


Figure 8. Saccades to group targets as percentage of trials ( $M$  and  $SE$ ) with saccades initiated to group targets either 200 ms before or 200 ms after a collision with the bordering frame for both crowding and collision conditions.

0.01, nor for collision,  $F(1, 13) = 0.26$ ,  $p = 0.62$ ,  $\eta_p^2 = 0.02$ , were observed. All other interactions likewise failed to reach significance (all  $ps > 0.18$ ).

## Discussion

The current study aimed to determine the effects of target–distractor crowding and target collisions on tracking accuracy, gaze behavior, and target-change detection accuracy in MOT. While in previous studies collisions and crowding were inevitably inseparable, we experimentally isolated a crowding from a collision condition by introducing collisions with the rectangular frame instead of collisions with another target or distractor. With this approach, we were able to identify how visual (crowding) and/or attentional (collisions) aspects might differentially affect tracking accuracy, gaze behavior and change-detection accuracy.

Results verified a successful crowding manipulation, with a tracking accuracy decrease of approximately 20% when targets were crowded by distractors (cf. Alvarez & Franconeri, 2007; Franconeri et al., 2008; Shim et al., 2008; Tombu & Seiffert, 2008). Furthermore, our prediction was confirmed that gaze location, measured by relative gaze distance, is “pulled” by crowding. However, countering our expectation, target collisions comparably showed a similar effect. Our second prediction claimed that target stops are better detected if targets are about to collide with the rectangular frame (i.e., with group-target stops  $>$  separate-target stops; collision  $>$  no collision), because attention should be attracted to collisions. This

prediction was not confirmed, as detection accuracy was neither better for the group targets than for the separate target, nor better in the collision than in the no-collision conditions. Indeed, the comparison of the collision and no-collision conditions revealed an effect in the opposite direction, rather with higher detection rates in the no-collision than in the collision conditions. That this effect is not due to a mere speed-accuracy tradeoff is substantiated by the additionally measured response times, as longer detection times were generally observed for conditions with lower detection rates. However, as expected and regardless of the manipulation condition, participants utilized a peripheral detection of target stops in the vast majority of cases. Finally, regarding saccades to group targets, it was additionally shown that anticipatory saccades are rather caused by collision than by crowded conditions and that these saccades are mainly initiated before rather than after the moment of collision with the rectangular frame.

In summary, along with the successful replications of decremented MOT-performance due to crowding and participants’ capacity to peripherally detect target stops, our main prediction that gaze is “pulled” by crowding was confirmed. However, the unexpected findings that (a) gaze distance is also reduced by collisions and that (b) detection accuracy is negatively affected by both crowding and collisions calls for further explanation.

(a) Regarding relative gaze distance, the current results extend previous findings by ascribing crowding as primarily a “vision problem,” as participants’ gaze positions were closer to the group of targets in the crowd than in the no-crowd conditions. This effect can

be attributed to the spatial acuity of peripheral vision, which is too low to separate targets from distractors. Consequently, it is advantageous to locate one's gaze closer to crowded targets, thereby allowing for better discrimination of targets and distractors (Levi, 2008; Strasburger, 2005; Strasburger et al., 2011). Besides this predicted crowding effect, however, a nonpredicted effect of collisions indicated closer gaze distances to the group targets in the collision than in the no-collision conditions (see Figure 4). An explanation of this finding might be that, in the collision conditions, closer gaze distances result from saccades, since, as empirically proven, saccades are frequently initiated due to an anticipated collision (see Figure 8). Consequently, gaze anchoring should be more reliably ascribed to either crowding or collisions when successfully controlling for saccades. Hence, if crowding rather than collisions was the decisive cause of closer gaze anchoring (as was originally hypothesized), a post hoc analysis of relative gaze distances for trials *without* saccades should reveal shorter gaze distances from the group targets in the crowding but not in the collision conditions. To test this prediction, relative gaze distances were compared again, but this time solely based on no-saccade trials. The respective ANOVA indeed shows a main effect for crowding,  $F(1, 13) = 13.78$ ,  $p < 0.01$ ,  $\eta_p^2 = 0.52$ , with gaze closer to the group targets in the crowd ( $M = 30.21\%$ ,  $SE = 1.85\%$ ) than in the no-crowd conditions ( $M = 34.57\%$ ,  $SE = 1.68\%$ ;  $p < 0.01$ ). For collision conditions, however, no difference between conditions (collision:  $M = 33.50\%$ ,  $SE = 2.19\%$ ; no collision:  $M = 31.29\%$ ,  $SE = 1.37\%$ ;  $p = 0.16$ ) and no significant interaction ( $p = 0.13$ ) were found. Thus, it can be concluded that gaze is anchored closer to the group targets because the targets are crowded by distractors and not because of an anticipated collision of a target with the rectangular frame. In turn, the main effect for collisions (depicted in Figure 4) can be classified as a mere by-product of saccades to colliding targets before the collision. Hence, it seems reasonable to conclude that the functionality of a gaze-anchoring strategy closer to the crowded targets is rooted in the higher spatial resolution near the fovea to reduce the negative effects of crowding (Strasburger, 2005).

(b) Regarding the detection accuracy of target stops, the predicted higher detection rates for targets that are about to collide was not found for the group targets compared to the separate target nor, within the group targets for the collision compared to the no-collision conditions. Rather than a general enhancement, the results show a notable reduction in detection accuracies in the collision (and crowding) conditions compared to the no-collision (and no-crowding) conditions. For an explanation of this unexpected finding, one might again refer to the results revealed for saccades to group targets (see Figure 8). Since saccades were regularly

initiated *before* the collision and thus resulted in temporal overlaps with the target stop, the accompanying interruption of information processing could be the reason for decreased detection rates in the collision condition (Diamond, Ross, & Morrone, 2000). To further test whether worsened target-stop detections were actually caused by interfering saccades, a post hoc analysis for detection accuracy was conducted with saccade as an additional factor, contrasting trials with saccades immediately before the collision with trials without such a saccade. The respective ANOVA Crowding  $\times$  Collision  $\times$  Saccade shows a significant interaction of collision and saccade,  $F(1, 13) = 5.35$ ,  $p = 0.04$ ,  $\eta_p^2 = 0.29$ , indicating that collisions negatively affect change-detection rates if saccades are executed just before the collision (no collision:  $M = 76.00\%$ ,  $SE = 2.80\%$ ; collision:  $M = 68.71\%$ ,  $SE = 3.95\%$ ), but not if saccades are absent (no collision:  $M = 74.32\%$ ,  $SE = 4.15\%$ ; collision:  $M = 73.50\%$ ,  $SE = 4.03\%$ ). Thus, the possibility for detecting target changes is decreased by saccades, likely due to the disruption of the continuous flow of information processing. In the absence of saccades, however, detection rates did not vary between the collision and no-collision conditions, which indicates that in both cases attention was successfully distributed to the targets. This attention allocation seems to be partially impaired by crowding, since detection accuracy decreased in the crowding condition (see Figure 5). This decrement of accuracy could be explained by the allocation of more attentional resources to the monitoring of difficult target–distractor configurations (Franconeri et al., 2008; Shim et al., 2008), leaving fewer attentional resources available for the detection task.

For the two unexpected findings discussed above, the explanations point to a general dysfunctionality of saccades in crowding and collision conditions. Based on our results presented here, we can clearly deduce that it is *not* crowding that induces saccades before a collision (see also Fehd & Seiffert, 2010; Zelinsky & Todor, 2010). Rather, saccades seem to be initiated in order to update target positions with expected motion-direction changes caused by collisions. Further, these results indicate that target–motion-direction changes do not fall into the same category of detection strategy as target stops. The current study illustrated this separation as foveal vision was used to perceive motion-direction changes, while peripheral vision was used to detect target stops. After a target stop, however, the target continued its trajectory from before the stop, which may explain why peripheral vision is sufficient here. Hence, as information from peripheral vision alone seems sufficient to perceive directional changes (especially under no-crowd conditions), future research should address whether saccades are compulsory to update target positions after collisions in MOT.

In this context, it should further be examined whether (“rescue”) saccades might be the cause of reduced tracking accuracies in crowded conditions. If this is the case, gaze anchoring on a distance-, difficulty-, and event-optimized “pivot point” (Ripoll, Kerlirzin, Stein, & Reine, 1995) could provide an optimal strategy when facing a dual-task situation that requires both the tracking of multiple targets and the detection of target changes.

Beyond the field of MOT research, the results of the current study are relevant to a number of real-world settings, specifically for those in which the main task requires high visual and attentional demands for multiple object monitoring and event detection in the environment. Looking to the complex world of sports as an example, specific predictions can be derived from the above conclusions for the optimization of athletes’ sensorimotor behaviors. In team sports, for instance, it can be predicted that peripheral vision would be beneficial for monitoring teammates and opponents. However, in crowded situations with many players grouped at a similar location, the ability to track single players is expected to be impaired. Thus, to effectively continue tracking in those situations, foveal gaze should be directed closer to the crowd of players to increase spatial acuity and thus, reduce the negative crowding effect. Too many saccades during the monitoring process, however, seem to inevitably incur the cost of an increased risk of missing relevant events. Due to the interrupted information processing induced by saccadic suppression, decision-making performance during this time interval would be impaired, especially if players must initiate critical actions. The practical implication rather favors peripheral vision to process event-related changes.

Together, our results extend previous findings on crowding and collisions in MOT in several respects. It was shown that (a) gaze is located closer to a set of targets if they are crowded, presumably to reduce the negative effects of crowding by exploiting the higher spatial resolution near the fovea; (b) saccades are initiated as a consequence of collisions in order to update target positions after a motion-direction change, rather than as a consequence of local crowding between the colliding target and a distractor or another target; (c) peripheral vision is naturally used to detect target changes in both crowding or collisions conditions, emphasizing the general functionality of peripheral vision in MOT, and that (d) target changes are more frequently missed in collision than in crowding conditions, presumably because of the interrupted information flow caused by saccades. For MOT research, these results should be taken as first steps to disentangle the existing findings on crowding and collisions. For applied fields like sports, predictions on the functionality of peripheral vision can be derived

from the reported findings and applied for sport-specific empirical tests in future research.

*Keywords:* peripheral vision, covert attention, saccades, gaze behavior, gaze anchoring, change detection

## Acknowledgments

Commercial relationships: none.

Corresponding author: Christian Vater.

Email: Christian.Vater@ispw.unibe.ch.

Address: University of Bern, Bern, Switzerland.

## 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. [PubMed] [Article]
- Alvarez, G. A., & Scholl, B. J. (2005). How does attention select and track spatially extended objects? New effects of attentional concentration and amplification. *Journal of Experimental Psychology: General*, 134, 461–476.
- Atsma, J., Koning, A., & van Lier, R. (2012). Multiple object tracking: Anticipatory attention doesn’t “bounce.” *Journal of Vision*, 12(13):1, 1–11, doi:10.1167/12.13.1. [PubMed] [Article]
- Bettencourt, K. C., & Somers, D. C. (2009). Effects of target enhancement and distractor suppression on multiple object tracking capacity. *Journal of Vision*, 9(7):9, 1–11, doi:10.1167/9.7.9. [PubMed] [Article]
- Bruce, V., Green, P. A., & Georgeson, M. A. (2003). *Visual perception: Physiology, psychology and ecology* (4th ed.). Hove, UK: Psychology Press.
- Carrasco, M., Ling, S., & Read, S. (2004). Attention alters appearance. *Nature Neuroscience*, 7, 308–313.
- Cavanagh, P., & Alvarez, G. A. (2005). Tracking multiple targets with multifocal attention. *Trends in Cognitive Sciences*, 9, 349–354.
- Diamond, M. R., Ross, J., & Morrone, M. C. (2000). Extraretinal control of saccadic suppression. *Journal of Neuroscience*, 20, 3449–3455.
- Doran, M. M., & Hoffman, J. E. (2010a). Target enhancement and distractor suppression in multiple object tracking. In J. Brooks, A. Belopolsky, M. Matsukura, & M. Palomares, Object Perception,

- Attention, and Memory (OPAM) 2009 Conference Report 17th Annual Meeting, Boston, MA, USA. *Visual Cognition*, 18, 126–129.
- Doran, M. M., & Hoffman, J. E. (2010b). The role of visual attention in multiple object tracking: Evidence from ERPs. *Attention, Perception, & Psychophysics*, 72, 33–52.
- Fehd, H. M., & Seiffert, A. E. (2008). Eye movements during multiple object tracking: Where do participants look? *Cognition*, 108, 201–209.
- Fehd, H. M., & Hoffman, J. E. (2010). Looking at the center of the targets helps multiple object tracking. *Journal of Vision*, 10(4):19, 1–13. [PubMed] [Article]
- 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.
- Franconeri, S. L., Lin, J., Pylyshyn, Z., Fisher, B., & Enns, J. (2008). Evidence against a speed limit in multiple object tracking. *Psychonomic Bulletin & Review*, 15, 802–808.
- Gobell, J., & Carrasco (2005). Attention alters the appearance of spatial frequency and gap size. *Psychological Science*, 16, 644–651.
- Howe, P. D., Cohen, M. A., Pinto, Y., & Horowitz, T. S. (2010). Distinguishing between parallel and serial accounts of multiple object tracking. *Journal of Vision*, 10(8):11, 1–13, doi:10.1167/10.8.11. [PubMed] [Article]
- 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. [PubMed] [Article]
- 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. [PubMed] [Article]
- Kredel, R., Klosterman, 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, <http://dx.doi.org/10.1051/bioconf/20110100051>.
- Levi, D. M. (2008). Crowding—An essential bottleneck for object recognition: A mini-review. *Vision Research*, 48, 635–654.
- Luu, T., & Howe, P. D. (2015). Extrapolation occurs in multiple object tracking when eye movements are controlled. *Attention, Perception, & Psychophysics*, 77, 1919–1929.
- Meyerhoff, H. S., Papenmeier, F., Jahn, G., & Huff, M. (2016). Not FLEXible enough: Exploring the temporal dynamics of attentional reallocations with the multiple object tracking paradigm. *Journal of Experimental Psychology: Human Perception & Performance*, 42, 776–787.
- Nyström, M., & Holmqvist, K. (2010). An adaptive algorithm for fixation, saccade, and glissade detection in eye tracking data. *Behavior Research Methods*, 42, 188–204.
- 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.
- Peterson, M. S., Kramer, A. F., & Irwin, D. E. (2004). Covert shifts of attention precede involuntary eye movements. *Perception & Psychophysics*, 66, 398–405.
- Pylyshyn, Z. W., Haladjian, H. H., King, C. E., & Reilly, J. E. (2008). Selective nontarget inhibition in multiple object tracking (MOT). *Visual Cognition*, 16, 1011–1021.
- Pylyshyn, Z. W., & Storm, R. W. (1988). Tracking multiple independent targets: Evidence for a parallel tracking mechanism. *Spatial Vision*, 3, 179–197.
- Ripoll, H., Kerlirzin, Y., Stein, J. F., & Reine, B. (1995). Analysis of information processing, decision making, and visual strategies in complex problem solving sport situations. *Human Movement Science*, 14, 325–349.
- Shim, W. M., Alvarez, G. A., & Jiang, Y. V. (2008). Spatial separation between targets constrains maintenance of attention on multiple objects. *Psychonomic Bulletin & Review*, 15, 390–397.
- Strasburger, H. (2005). Unfocused spatial attention underlies the crowding effect in indirect form vision. *Journal of Vision*, 5(11):8, 1024–1037, doi:10.1167/5.11.8. [PubMed] [Article]
- 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. [PubMed] [Article]
- Tombu, M., & Seiffert, A. E. (2008). Attentional costs in multiple-object tracking. *Cognition*, 108, 1–25.
- Vater, C., Kredel, R., & Hossner, E.-J. (2016). Detecting single-target changes in multiple object tracking: The case of peripheral vision. *Attention, Perception, & Psychophysics*, 78, 1004–1019.
- Vater, C., Kredel, R., & Hossner, E.-J. (2017). Detecting target changes in multiple object

tracking with peripheral vision: More pronounced eccentricity effects for changes in form than in motion. *Journal of Experimental Psychology: Human Perception & Performance*, *43*, 903–913.

Williams, A. M., Davids, K., & Williams, J. G. (1999).

*Visual perception and action in sport*. London, UK: Taylor & Francis.

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. [PubMed] [Article]