From Gaussian blobs to naturalistic videos: Comparison of oculomotor behavior across different stimulus complexities

Alexander Goettker*

Abteilung Allgemeine Psychologie, Justus-Liebig University, Gießen, Germany

✉

Ioannis Agtzidis*

Technical University of Munich, Munich, Germany

✉

Doris I. Braun

Abteilung Allgemeine Psychologie, Justus-Liebig University, Gießen, Germany

✉

Michael Dorr

Technical University of Munich, Munich, Germany

✉

Karl R. Gegenfurtner

Abteilung Allgemeine Psychologie, Justus-Liebig University, Gießen, Germany

✉

Research on eye movements has primarily been performed in two distinct ways: (1) under highly controlled conditions using simple stimuli such as dots on a uniform background, or (2) under free-viewing conditions with complex images, real-world movies, or even with observers moving around in the world. Although both approaches offer important insights, the generalizability among eye movement behaviors observed under these different conditions is unclear. Here, we compared eye movement responses to video clips showing moving objects within their natural context with responses to simple Gaussian blobs on a blank screen. Importantly, for both conditions, the targets moved along the same trajectories at the same speed. We measured standard oculometric measures for both stimulus complexities, as well as the effect of the relative angle between saccades and pursuit, and compared them across conditions. In general, eye movement responses were qualitatively similar, especially with respect to pursuit gain. For both types of stimuli, the accuracy of saccades and subsequent pursuit was highest when both eye movements were collinear. We also found interesting differences; for example, latencies of initial saccades to moving Gaussian blob targets were significantly faster compared to saccades to moving objects in video scenes, whereas pursuit accuracy was significantly higher in video scenes. These findings suggest a lower processing demand for simple target conditions during saccade preparation and an advantage for tracking behavior in natural scenes due to higher predictability provided by the context information.

Introduction

Eye movements are an important and integral part of our foveated vision system (for review, see Gegenfurtner, 2016). Why we look where we do has interested scientists for centuries (for review, see Wade, 2010). It is firmly established by now that several factors contribute to determining the next gaze position (for reviews, see Hayhoe, 2017; Schütz, Braun, & Gegenfurtner, 2011; Tatler, Hayhoe, Land, & Ballard, 2011), such as the visual salience of the scene (Itti & Koch, 2000; Kienzle, Franz, Schölkopf, & Wichmann, 2009; Kümmerer, Wallis, & Bethge, 2017), objects in the scene (Einhäuser, Spain, & Perona, 2008), semantic categories and meaning (Henderson & Hayes, 2017; Xu, Jiang, Wang, Kankanhalli, & Zhao, 2014), oculomotor strategies (Tatler & Vincent, 2009), rewards associated with different objects (Navalpakkam, Koch, Rangel, & Perona, 2010; Schütz, Trommershäuser, & Gegenfurtner, 2012), information gain (Geisler & Najemnik, 2005; Najemnik & Geisler, 2008; Peterson & Eckstein, 2012), intended actions (Flanagan & Johansson, 2003; Hayhoe, 2000; Land, Mennie, & Rusted, 1999), and task demands (Gegenfurtner, 2016; Rothkopf, Ballard, & Hayhoe, 2007; Toscani, Valsecchi, & Gegenfurtner, 2013; Yarbus, 1967), as well as individual traits of the observer (Constantino et al., 2017; de Haas, Iakovidis, Schwarzkopf, & Gegenfurtner, 2019; Kennedy, D'Onofrio, Quinn, Bölte, Lichtenstein, & Falck-Ytter, 2017).

However, the complexity of natural scenes together with the large number of factors influencing eye movement behavior make it difficult to gain insight into individual mechanisms of visual processing and their effects on oculomotor control. Therefore, a large part of research on oculomotor behavior has utilized relatively simple and constrained stimuli (e.g., Gaussian blobs, dots on uniform backgrounds) to successfully characterize basic response properties of the oculomotor system. Single-unit recordings, behavioral investigations, computational modeling, and the combination of these three approaches have led to tremendous advances in our understanding of oculomotor control (for reviews, see Lisberger, 2015; Robinson, 1981; Sommer & Wurtz, 2008; Sparks & Mays, 1990). For example, synthetic and controlled stimuli have been used to study the role of stimulus characteristics such as contrast (Doma & Hallett, 1988; Ludwig, Gilchrist, & McSorley, 2004), movement of a stimulus (de Brouwer, Missal, Barnes, & Lefèvre, 2002; Goettker, Braun, & Gegenfurtner, 2019; Schreiber, Missal, & Lefèvre, 2006), or even cognitive and decision-making processes (for reviews, see Glimcher, 2003; Munoz & Everling, 2004). In addition, the controlled settings allow isolating the effect of stimulus characteristics on different oculomotor behaviors, including the effect of target contrast on saccade control (Doma & Hallett, 1988; Ludwig et al., 2004) or its effect on pursuit eye movements (Haegerstrom-Portnoy & Brown, 1979; O'Mullane & Knox, 1999). Although this type of research is valuable for developing an understanding of basic response properties and mechanisms within the oculomotor system, a natural scene and environment will always vary along multiple dimensions with saccadic and pursuit eye movements typically occurring together (Dorr, Martinetz, Gegenfurtner, & Barth, 2010; Orban de Xivry & Lefèvre, 2007; Vig, Dorr, & Barth, 2009). Thus, the strength of experimental control with simple synthetic stimuli that allows us to isolate relevant factors comes at the cost of reduced natural validity.

To combine these potentially complementary lines of research, a logical assumption would be that one can just generalize the findings from simple and synthetic stimuli to understand eye movement behavior in natural scenes; however, this assumption has not been tested well so far. One study by White and colleagues (White, Stritzke, & Gegenfurtner, 2008) even found that there might be potential differences in processing for simple versus complex natural stimuli. White and colleagues demonstrated that saccadic latencies were reduced when stationary Gabor patches of matched levels of visibility were presented within an image of a natural scene, as opposed to a uniform background. However, a systematic comparison of standard oculomotor metrics that indicate the accuracy (saccade position error or pursuit gain) and speed of visual processing (saccade latency) measured under comparable conditions for controlled synthetic lab stimuli and natural content is still missing.

To tackle this, we compared voluntary eye tracking behavior in response to a moving object, which was either a synthetic stimulus (a Gaussian blob on a uniform background) or a naturalistic stimulus (video clips taken in a natural environment from the GazeCom dataset) (Dorr et al., 2010). Importantly, the synthetic and naturalistic stimuli followed identical motion trajectories in both cases, allowing a direct comparison between the two conditions. We specifically chose video scenes of naturally moving stimuli, whose sudden motion onsets and movement behavior have been found to be particularly salient (Dorr et al., 2010; Itti, 2005; Vig et al., 2009). Moving stimuli typically elicit an initial interceptive saccade followed by smooth pursuit, allowing us to study both types of eye movement responses at the same time. Additionally, we tried to look into potential interactions between the saccadic and pursuit eye movements (Goettker et al., 2019; Orban de Xivry & Lefèvre, 2007) by manipulating the relative angle between the two. The goal of our study was to bridge the gap between studying oculomotor behavior with very simple stimuli and in response to naturalistic videos. To do so, we quantitatively compared a range of standard oculomotor measures of eye movement speed and quality for gaze behavior measured with moving spots on a uniform background and the equivalent movement of objects in their natural surrounds.

Methods

Selection of baseline trajectories

Our experiments together with the free-viewing analysis are based on the GazeCom dataset (Dorr et al., 2010). This dataset contains high-frequency (250 Hz) gaze data of 54 subjects who were freely viewing 18 high-resolution movies, each 20 seconds in duration, for a total viewing time of approximately 4.5 hours. The movie clips showed a variety of real-world outdoor scenes taken in and around the German city of Lübeck, which typically contained people and animals moving in parks or pedestrian areas, nature sights along waterfronts, or traffic scenes. For the design of our two experimental conditions as summarized in Figure 2, we needed accurate trajectories of single targets shown in the selected videos of the dataset. Contrary to experiments with synthetic stimuli, it is quite challenging to obtain accurate target trajectories in dynamic naturalistic contexts. Even though many automated algorithms exist for motion estimation and optical flow extraction, they would have
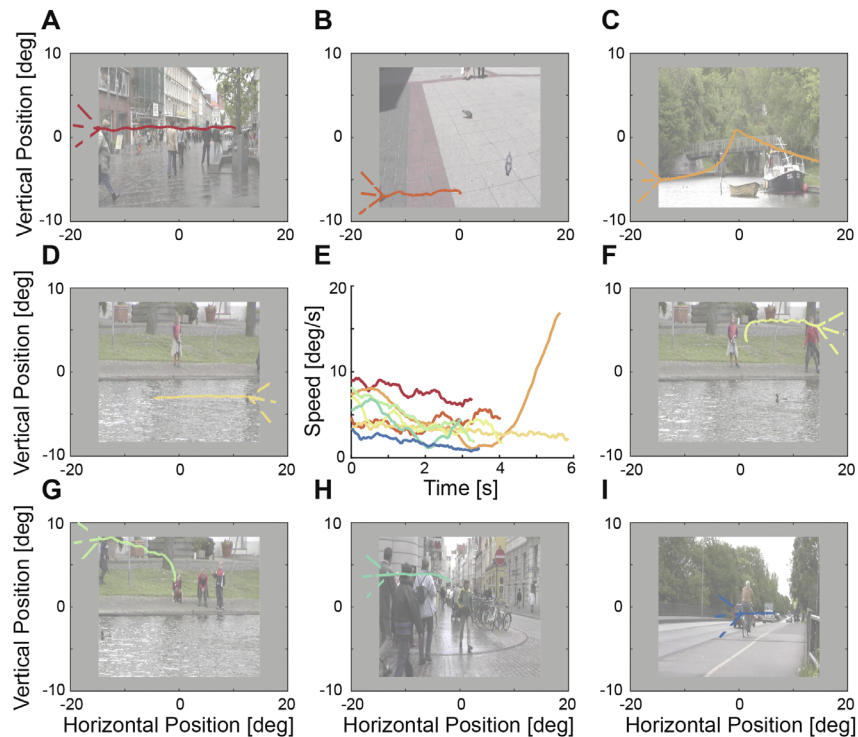
Figure 1. Trajectories and depictions of individual scenes, including still shots from the eight selected videos of the GazeCom dataset. Hand-labeled trajectories of the eight video targets selected for our experiments are shown in different colors. The three dashed horizontal, upward, and downward tilted line segments at the beginning of each trajectory represent the three possible positions of the fixation point at the beginning of each trial at a distance of 5 deg. Note that the contrast of the images was reduced for the overlay of the trajectories. Also shown are the speed profiles of the eight target trajectories as presented in the synthetic and naturalistic conditions. The average velocity of the hand-labeled trajectories for each target have been smoothed by a 100-ms sliding time window. The colors correspond to the target movement shown in the respective videos.

been too noisy for the purpose of our experiments. For this reason, we first selected 45 targets based on the longest duration smooth pursuit sequences of most observers (from 0.8 to 5.9 seconds) as detected by an eye movement classification tool (Agtzidis, Startsev, & Dorr, 2016). We then determined a representative point of each target object (e.g., the nose of a walking person) and manually labeled its position in each video frame.

Baseline trajectories for data collection

For the synthetic and naturalistic experiments, we chose eight targets and used their hand-labeled trajectories from the GazeCom dataset. Our choice was based on the duration of the target movements (>3 seconds) and on a pilot study of ours where we tested whether participants initiated saccades from the initially presented fixation dot to the relevant targets to track their movements. The latter was important to ensure that we could directly compare the oculomotor behavior between stimulus complexities without constraining the task to track a specific object for the naturalistic videos. The movement trajectories of the

eight selected targets and their speed profiles are shown in Figure 1. Target velocities were in general below 10 deg/s, and movements lasted between 3 and 5.9 seconds.

Baseline trajectories for validation

To investigate how well our results generalize to the unconstrained free-viewing condition, we analyzed smooth pursuit responses to all of the 45 hand-labeled targets in the GazeCom dataset with a consistent pursuit response by multiple observers. Automatic labeling of different eye movement types, of smooth pursuit in particular, is still challenging. In order to add ground-truth data to the GazeCom dataset, the automatically labeled eye movement data were visually controlled and manually labeled by two individual annotators and a third conflict-resolving annotator (Startsev, Agtzidis, & Dorr, 2019). This process yielded about 3000 smooth pursuit intervals (comprising about 7% of the total GazeCom viewing time) for the baseline targets.

In our synthetic and naturalistic condition, we varied the relative angle between the initial fixation point and
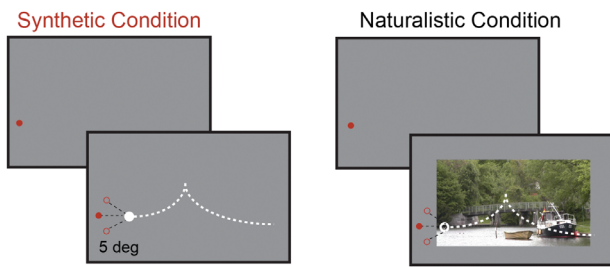
Figure 2. Depiction of the synthetic and the naturalistic conditions. After an initial fixation period, participants saw in the synthetic condition a white Gaussian blob moving on a uniform gray background along the same trajectory as the duck shown in the naturalistic condition in the video clip sequence. In both conditions, the fixation positions and the movement trajectories of the Gaussian blob or the flying duck were identical. The trajectory of the duck corresponds to the orange line in Figure 1C starting from the left side.

the point of the motion onset in the video. This way, the initial saccade and the subsequent smooth pursuit movements were either collinear or at an angle of ±30 deg. From all available data for the hand-labeled trajectories in the GazeCom ground truth, we selected only smooth pursuit intervals that fell within the range of –30 deg to 30 deg relative angle between saccade and pursuit response and obtained 238 such pairs (69 down, 114 collinear, and 55 up in 30 deg-wide bins).

Experimental design

In both conditions, each trial started with the appearance of a fixation cross. The position of the fixation cross was based on the position and the initial direction of the upcoming target trajectory, which we defined by drawing a line between the first position of the target and the position of the target after 250 ms. The position of the fixation cross was either collinear with the motion direction or rotated by 30 deg up- or downward. The moving target always appeared at a distance of 5 deg from the initial fixation cross and moved along the same trajectory as in the natural video clip.

Participants were instructed to look at the fixation cross and to press the space bar on the keyboard to start each trial. No additional instructions were given, and participants were just told to observe the scene (blob) that was presented during each trial. We used the eye position at the moment the space bar was pressed to perform a drift correction at the fixation location. After the key press, a red dot replaced the fixation cross and stayed there for a random duration between 1 and 1.5 seconds (Figure 2). After the dot disappeared, the two conditions differed with respect to the moving target and the background. In the synthetic condition, the

target was a white Gaussian blob ($SD = 0.5$ deg, peak contrast = 50%, mean contrast = 8%) on a uniform gray background. Mean contrast was computed as the average pixel intensity of the object (o) and the mean intensity of its immediate surroundings (s) with the formula: contrast = abs($o – s$)/s. In the synthetic condition, each blob moved along one of the eight hand-labeled target trajectories for the duration of the respective trajectory. In the naturalistic condition, one of the eight selected videos was presented, which contained the moving object (mean contrast = 31%) embedded in its natural context.

Each condition was presented in four blocks of 72 trials each (8 scenes × 3 orientations of the starting position × 3 repetitions). One block lasted approximately 15 minutes. Participants took breaks between blocks and typically performed three blocks per session. Participants randomly started with either the synthetic or naturalistic condition and switched to the respective other condition after completing all blocks for the first one.

Experimental setup

Participants sat at a table facing a 32-inch monitor (Display++ LCD; Cambridge Research Systems, Ltd., Rochester, UK) in a dimly illuminated room. We used a chin and forehead rest to stabilize the participant's head. The eyes of the participants were approximately at the height of the screen center at a distance of 90 cm. We recorded eye movements from the right eye with a desk-mounted eye tracker (EyeLink 1000 Plus, SR Research, Kanata, ON, Canada) with a sampling frequency of 1000 Hz. The experiments were programmed in MATLAB (MathWorks, Natick MA) using the Psychtoolbox (Kleiner, Brainard, & Pelli, 2007). Before each block we used a nine-point calibration to align the gaze data with the screen.

The temporal and spatial resolution of the stimulus, as well as the viewing distance used to collect the GazeCom data (30 Hz, 1280 × 720 pixels, 45 cm), differed from our experimental setup with respect to monitor and viewing distance (120 Hz, 1920 × 1280 pixels, 90 cm). We decided not to resize the scenes or adjust the position of the labeled targets to the new monitor size but instead to present them in the central 1280 × 720 pixels of our larger monitor (Figures 1 and 2). This choice led to differences in the visual field between our experiments (30 deg × 17 deg) and the GazeCom recordings (48 deg × 28 deg), but allowed us to control the initial fixation position of the participants. With the smaller size of the video scenes in the screen center on our monitor we were able to present fixation dots 5 deg from the target starting positions (Figure 2) on the left or right side (Figures 1 and 2).

To account for the different temporal resolution between the original monitor and the movement of synthetic target in our experiments, we resampled the *x* and *y* pixel values from the hand-labeled trajectories to 120 Hz with linear interpolation. In the naturalistic condition, the videos were still presented at 30 Hz by just updating the monitor's content every fourth frame.

Participants

We recorded eye movements of 13 volunteers (mean = 23.5 years old, SD = 3.5 years; 11 female) who were naïve to the purpose of the study and had not seen any of the videos before. Participants were mainly students of Giessen University with normal or corrected-to-normal vision. Before the start of their participation they gave informed consent according to the tenets of the Declaration of Helsinki. All experiments were approved by the local ethics committee (LEK FB06 2017-08). Participants received 8 Euro per hour as monetary compensation.

Data analysis

For each subject we quantified the characteristics of initial saccades and smooth pursuit and tested for differences with respect to experimental condition and relative angle between saccades and pursuit. We examined these differences for saccade latencies and saccade position errors, as well as for pursuit gain and pursuit directional accuracy. Below, we describe how each measurement and statistic was calculated and mention potential differences or limitations among the different conditions.

For the synthetic and naturalistic conditions, we calculated the saccade latencies as the time difference between the start of the target movement and the onset of the first saccade. For the GazeCom free-viewing validation data such latencies were not defined, because of the start of the target movements happened at different times during the video clips. The saccade position error was defined as the Euclidean distance between the saccade landing position and the labeled target position. We analyzed pursuit movements in the 100-ms time interval starting 50 ms after the saccade completion. This choice excluded post-saccadic oscillations from our calculations and thus returned more robust results while limiting the influence of additional new retinal information after the end of the saccade. Pursuit gain was defined as the average ratio between pursuit and target speeds. Because the targets were not moving linearly we projected the sample-to-sample gaze direction onto the linearly interpolated target direction at each moment in time.

The gain was computed as the average ratio of the projected gaze and target speed during the relevant interval. Pursuit accuracy was defined as the pursuit angular error, which was calculated as the absolute difference between pursuit and target direction, calculated between the first point (50 ms) and the last point (150 ms) of the pursuit interval. We also computed the pursuit precision for each scene, which was defined as the width of a Gaussian distribution fitted to all available segments of the signed pursuit direction error across participants. These included the direction errors measured in the post-saccadic 100-ms interval, but we additionally included segments after this interval if the eye stayed closer than 3 deg to the target and had less than 45 deg of direction error, as here the participants were presumably still tracking the target. We used a sliding window of 100 ms in 10-ms intervals to find these new segments. To test for systematic influences on these statistics we used repeated-measures analysis of variance (ANOVA) with factors relative to angle (down, collinear, up) and stimulus complexity (synthetic vs. natural scene) and Bonferroni-corrected post hoc *t*-tests to compare the individual levels.

In our analysis, we included only trials in which (1) the saccade started from within 2.5 deg of the initial fixation position, and (2) the position error measured between the saccade end position and target position was below 3 deg. The second criterion was introduced to verify that participants were tracking the correct target. This was especially relevant for the naturalistic conditions, as the video scenes could contain multiple potential targets for tracking. Because participants were not instructed to look at a specific object in the video, the exclusion rate for the naturalistic videos was higher, as here participants also could just look at other things shown in the video. Overall, 6077 out of 7488 trials (81%) were included in the analysis. As expected, the amount of excluded trials was much lower in the synthetic condition (390 trials) compared to the naturalistic condition (1021 trials). There were minor differences in the exclusion rates for the eight different videos.

Data and stimuli used for this experiment are available under osf.io/6zc8x.

Results

Here we present measurements of basic saccadic and pursuit parameter across different levels of stimulus complexity across the synthetic and naturalistic condition to bridge the gap to free viewing of natural scenes. If possible, we also present the equivalent statistics of the original GazeCom data set (Dorr et al., 2010).
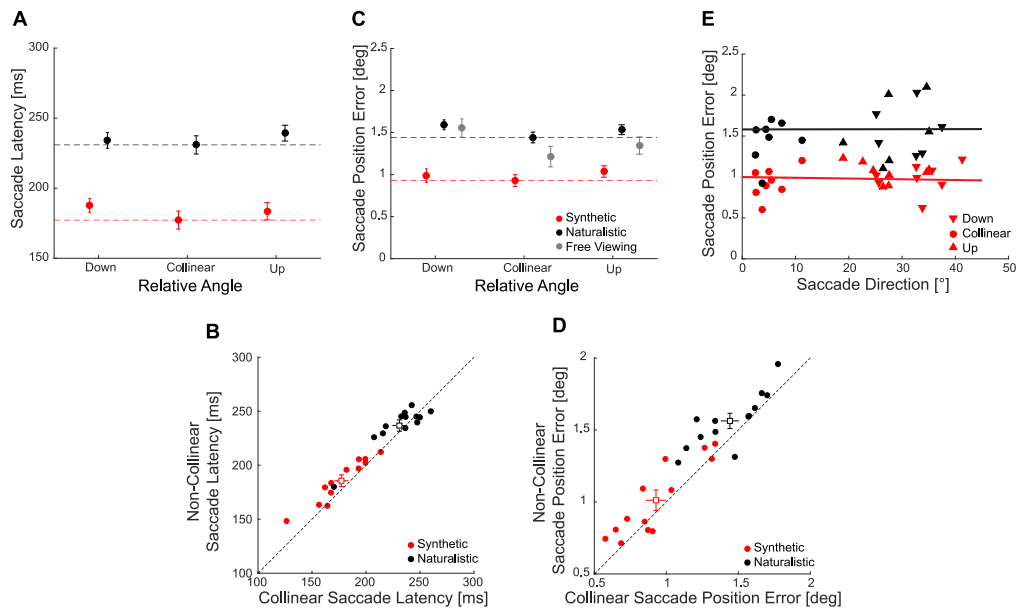
Figure 3. Comparison of saccade latencies and position errors across conditions. (A) Effects of stimulus complexity and relative angle between initial saccades and the upcoming pursuit. The two dashed horizontal lines represent saccadic latency in the collinear case for the synthetic condition (red) or naturalistic condition (black). (B) Saccade latency values of all 12 subjects for collinear movements in the synthetic (red) and the naturalistic (black) condition plotted against the average of the two non-collinear movements. The open squares depict the respective mean. (C) Saccade position errors defined as average Euclidean distance between saccade endpoints and target positions for the synthetic condition (red) and the naturalistic condition (black). For comparison, saccade position errors of saccades of the GazeCom dataset under free viewing conditions are also plotted (light gray). Note that saccade position errors are larger by about 0.5 deg for targets in the natural scenes. (D) Saccade position error of the collinear condition plotted against the non-collinear condition, similarly to C. All error bars depict the standard error of the mean. (E) Saccade position error as a function of the deviation of the saccade direction from the horizontal axis. The three symbols represent the different relative directions between initial saccades and pursuit for the synthetic condition (red) and the naturalistic condition (black). The solid lines represent a linear regression fitted to the data.

Saccadic eye movements

Latencies of initial saccades to the moving blobs or naturalistic targets in video scenes were used as indicators for the processing time needed to program target-directed saccades (Figure 3A). We computed a repeated-measures ANOVA with the saccade target angle (down, collinear, up) and stimulus complexity (synthetic vs. naturalistic) as factors and tested for significant influences on saccade latency. We observed a significant main effect of stimulus complexity, $F(1, 12) = 89.745$, $p < 0.001$, as well as a main effect of relative angle, $F(2, 24) = 6.137$, $p = 0.006$, with no significant interaction between the two. Saccade latencies were shorter by about 50 ms for the synthetic condition compared to the naturalistic condition (183 ms vs. 235 ms on average), suggesting that a single clearly defined target against a uniform background leads to faster saccadic programming. In addition, for both conditions, saccades had shorter latencies when they were collinear (Figure 3B) with the subsequent target movements in comparison to saccades where the target

movements were shifted up- or downward: 177 ms vs. 186 ms for synthetic with $t(12) = 4.21$, $p = 0.001$, and 231 ms vs. 237 ms for naturalistic with $t(12) = 2.06$, $p = 0.06$.

To quantify saccade accuracy we looked at the saccadic position error (Figure 3C), which was measured as the Euclidean distance between the endpoint of the saccade and the position of the target. We calculated the average position errors for each subject depending on the stimulus complexity and the saccade target angle averaged across the eight different natural scenes. As before, we ran a repeated-measures ANOVA with the factors relative saccade target angle and stimulus complexity. We observed a significant main effect of stimulus complexity, $F(1, 12) = 109.262$, $p < 0.001$, as well as relative angle, $F(2, 24) = 5.446$, $p = 0.011$. There was no systematic interaction between the two factors, $F(2, 24) = 2.489$, $p = 0.104$. Saccade position errors were lower in the synthetic condition (average error in synthetic 0.98 deg; average error in naturalistic 1.5 deg). In addition, the variance of the position error was also significantly lower in the
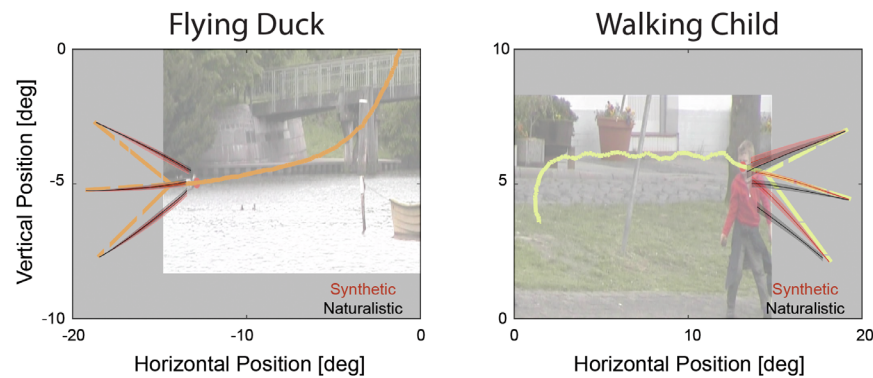
Figure 4. Saccades to moving targets of different sizes. The two images in the background are cutouts of the natural videos, with the flying duck on the left and walking child on the right. Overlaid are the averages of saccade trajectories for the synthetic condition (red) and the naturalistic condition (black) starting from three fixation locations to the moving blob/duck (orange) or to the blob/child (lime). The colors of the target movement trajectories are the same as in Figures 1C and 1F. Note that for the flying duck the black and red curves of the saccade trajectories starting from all three fixation positions lie on top of each other; however, the child saccade trajectories (black) diverge from the synthetic ones (red), and the divergence depends on the fixation position. Shaded areas depict the standard area of the mean of the saccade trajectories.

synthetic condition: 0.51 deg for synthetic and 0.62 deg for naturalistic, with $t(12) = 3.67$, $p = 0.003$, indicating more accurate and precise saccades for the synthetic condition. For comparison, we also plot in Figure 3C the saccade position errors for saccadic eye movement data collected in the original GazeCom dataset during free viewing. The overall saccade position error in the GazeCom data is also closer to the position error found for the naturalistic condition. Interestingly, a similar pattern of saccade position errors with regard to the relative angle to the target movement is present (Figure 3D). The collinear saccade-pursuit sequences produced the lowest position errors: synthetic, 0.93 deg for collinear vs. 1.01 deg for non-collinear, with $t(12) = 2.48$, $p = 0.03$; naturalistic, 1.44 deg for collinear vs. 1.56 deg for non-collinear, with $t(12) = 3.28$, $p = 0.006$. Interestingly, this benefit was not due to the overall more horizontal saccades for the collinear conditions, as there was no influence of the absolute saccade direction (Figure 3E).

We observed a significant influence of stimulus complexity on saccade behavior. In the eight naturalistic outdoor scenes, initial saccades to moving objects such as birds or children had significantly longer latencies and larger position errors, although the difference in position error could potentially be due to differences in object size. For a small object, such as a duck flying in the distance, the single point used in the hand labeling may be a sufficiently good representation; however, for larger targets, such as a walking child or a moving car close to the camera, a single point does not always adequately represent the whole target. The issue of object size is evident in the landing locations of initial saccades starting from the three fixation positions, as shown in Figure 4 for two moving target objects of

different sizes: the duck (left) and a child (right). The chosen target position for the flying duck was its beak and for the walking child its nose. For the duck the target location seemed to be well defined for initial saccades starting from all three fixation positions, whereas for the child the landing location of saccades seemed to depend also on their starting position. Although saccades from the higher and central position aimed at the center of the face, saccades starting from the lower fixation point aimed at the middle part of the child's body. Consequently, when only one target location is used to determine the position error for all saccades to a larger object, saccades may show higher position errors because participants simply looked to different parts of the same object depending on the prior fixation location. The situation is even more complicated, because our synthetic target had a fixed contrast, whereas the contrast of targets in natural scenes with respect to their background was variable. Saccade latency and accuracy are well known to depend on contrast (e.g., White, Kerzel, & Gegenfurtner, 2006). In the Effect of scene complexity on saccadic eye movements section (see also Figure 6A), we discuss these issues in more detail.

Pursuit eye movements

After comparing the results for the initial saccadic eye movements, we focus now on the initial pursuit responses analyzed in a time window between 50 and 150 ms after saccade offset. For pursuit gain as well as for pursuit direction errors, we computed a repeated measurement ANOVA with the factors relative angle (down, collinear, up) and stimulus
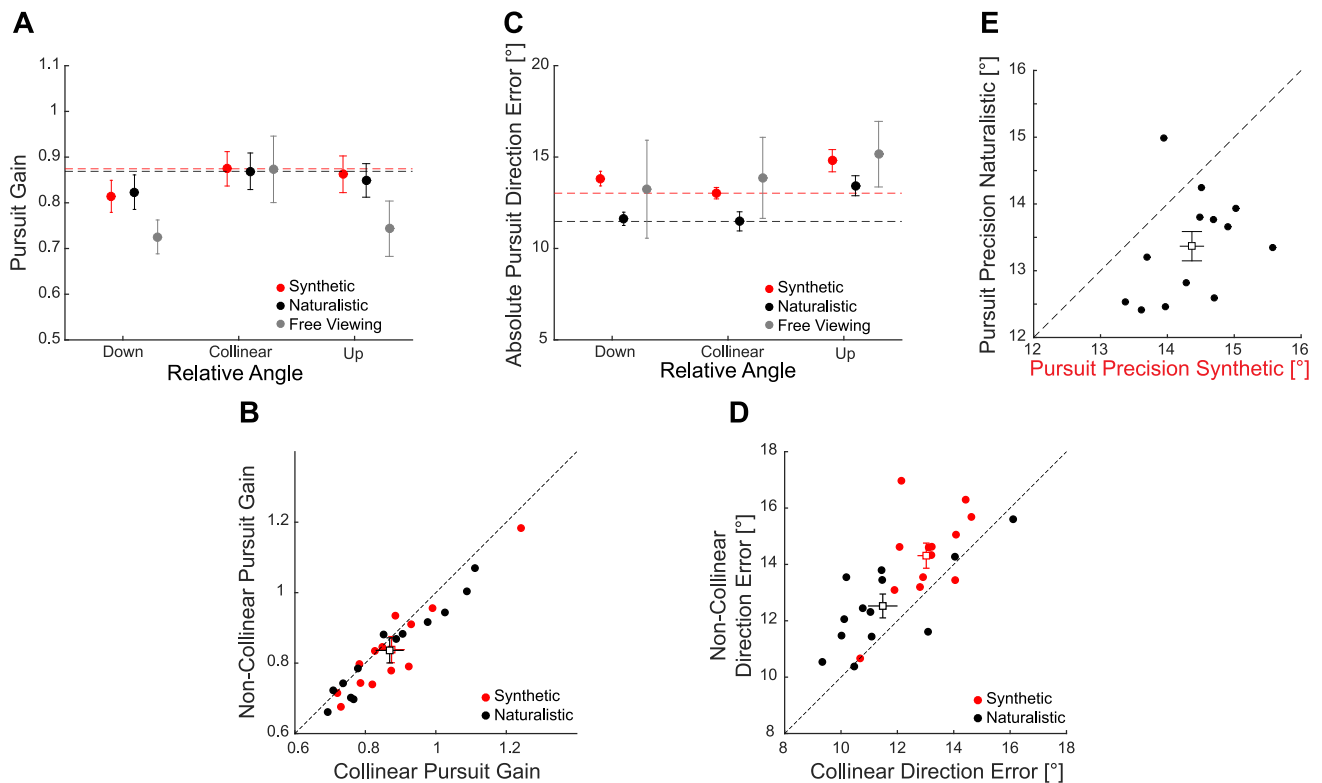
Figure 5. Comparison of pursuit behavior. (A) Improved gain for collinear tracking movements. The average pursuit gain during 50 to 150 ms after the saccade depended on the relative angle between the direction of the saccade and the pursuit movement and on the stimulus complexity. Dashed horizontal lines depict the values for the collinear conditions. (B) Individual values of pursuit gain for the collinear conditions are plotted against the average of the two non-collinear conditions. Colors depict the stimulus complexity, open squares depict the respective mean. (C) The absolute pursuit direction errors depended on the relative angle between the direction saccades and pursuit and on the stimulus complexity; dashed horizontal lines depict the value of the collinear condition. Pursuit direction errors were significantly lower for naturalistic conditions. (D) Comparison of collinear and non-collinear pursuit direction errors. (E) Comparison between the standard deviation of the pursuit direction errors for the synthetic and naturalistic conditions averaged across scenes and relative angles. Note that a lower standard deviation indicates higher precision. Each black dot represents a single subject, the open black square the average. All error bars depict the standard error of the mean.

complexity (synthetic vs. naturalistic). Similar to the results found for saccades, we observed for the pursuit gain a significant main effect of the relative angle, $F(2, 24) = 11.365$, $p < 0.001$, as shown in Figure 5A. Again, there was an improved gain for collinear movements: synthetic, 0.87 for collinear vs. 0.84 for non-collinear, $t(12) = 2.61$, $p = 0.02$; naturalistic, 0.87 for collinear vs. 0.83 for non-collinear, $t(12) = 3.13$, $p = 0.009$ (Figure 5B). However, there was no influence of stimulus complexity, $F(1, 12) = 0.028$, $p = 0.870$.

For pursuit accuracy, defined as the absolute pursuit direction error, a significant improvement was present when observers followed a moving target in a naturalistic scene (Figure 5C). We found a significant main effect of stimulus complexity, $F(1, 12) = 20.715$, $p < 0.001$, with lower errors when pursuing a moving target in a natural video scene compared to a similarly moving Gaussian blob on a homogeneous background (13.88 deg average error

for synthetic targets; 12.17 deg average error for naturalistic targets). As for the other measures, there was again a significant improvement in accuracy for collinear movements: synthetic, 13.02 deg for collinear vs. 14.31 deg for non-collinear, $t(12) = 3.49$, $p = 0.004$; naturalistic, 11.48 deg for collinear vs. 12.53 deg for non-collinear, $t(12) = 2.87$, $p = 0.014$ (Figure 5D). This benefit for the naturalistic condition was highly consistent across all of our observers and was also present for pursuit precision (see Methods), indicating that the additional richer information provided by the naturalistic objects and their surround in video scenes led to an improved tracking performance: 14.37 deg for synthetic vs. 13.36 deg for naturalistic; $t(12) = 4.311$, $p = 0.001$ (Figure 5E).

A possible explanation for this pursuit benefit could be that participants saw the videos multiple times and were therefore able to learn the movements and improve the tracking responses over time. To test for possible
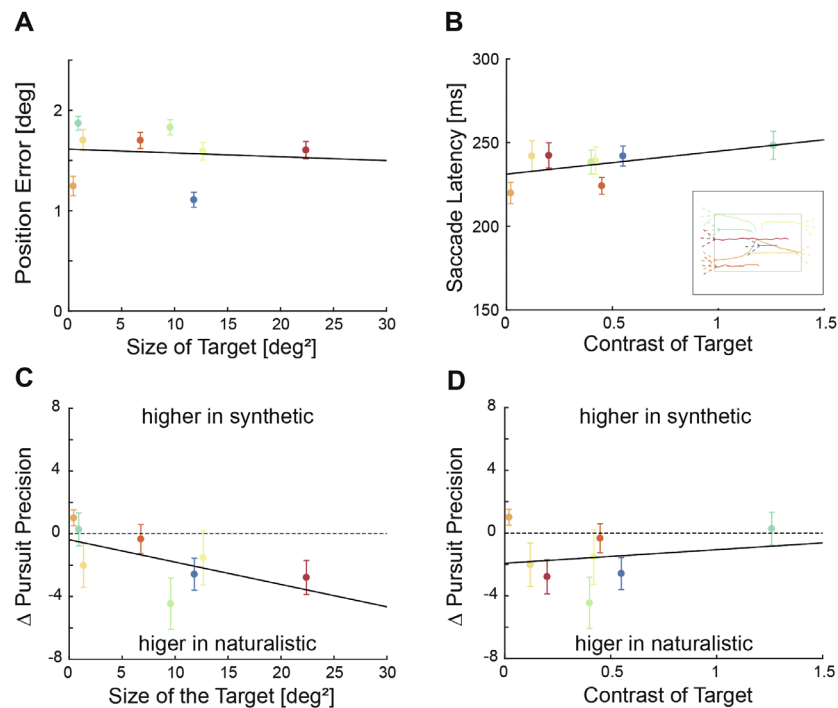
Figure 6. Effects of object size. (A) Saccade position error as a function of target size. The eight targets of the video scenes are depicted in different colors as in Figure 1 (and in the inset of B). The black solid line depicts a linear regression fitted to the data. (B) Saccadic latency as a function of target contrast. (C, D) Relation between the difference in pursuit precision of the synthetic and naturalistic conditions and target size (C) and target contrast (D). Negative values indicate an improved pursuit performance for the objects in the naturalistic conditions. For the eight video objects the same colors are used as in Figure 1. The black solid line depicts a linear regression fitted to the data; the black dashed line depicts no difference between synthetic and naturalistic targets. All data were first averaged across the different relative angles and then across all participants. Error bars depict the standard error of the mean across participants.

effect of repetition and learning we computed pursuit accuracy for early and late trials of seeing each video clip based on a median split. The comparison of early and late trials revealed a similar benefit for pursuit accuracy under naturalistic conditions for both time windows. In comparison to the synthetic condition, there was an improvement of pursuit accuracy of 1.47 deg accuracy for first half and 1.73 deg for second half. The magnitude of the improvement did not differ, $t(12) = 0.40$, $p = 0.70$, suggesting that the benefit for pursuit accuracy was not based on learning effects.

Effect of object size and object contrast

In order to better understand the observed effects in the naturalistic condition, we investigated the relationship between object size and contrast and the measured oculomotor parameter (Figure 6). First, we estimated the size of each target object by manually fitting a mask on a representative frame. The mask was kept constant under the assumption that the size of each target did not change substantially during its presentation. When looking at the saccade results, saccade position errors might become more ambiguous with larger objects; however, there was no significant relationship between position error and object size, $r(8) = -0.11$, $p = 0.80$ (Figure 6A). The position error was also not correlated with the contrast of the target, $r(8) = 0.38$, $p = 0.36$. This suggests that higher position errors in the naturalistic condition were not purely driven by low-level image characteristics. Similarly, there was also no systematic relationship between the object characteristics and the higher saccade latencies in the natural scene: $r(8) = 0.53$, $p = 0.17$ for contrast (Figure 6B); $r(8) = 0.27$, $p = 0.51$ for object size.

For the pursuit results, we correlated the object characteristics with the improvement in pursuit precision (shown in Figure 5E). With our limited dataset of eight objects, we again found no relationship between the low-level object characteristics and pursuit precision: $r(8) = 0.60$, $p = 0.12$ for object size (Figure 6C); $r(8) = 0.18$, $p = 0.66$ for object contrast. Note, however, that we had only a limited number of scenes and therefore cannot exclude potential effects

due to the small number and range of natural video scenes used in our experiment.

Discussion

Most of the eye movement research is performed either in highly controlled lab conditions or with fewer constraints with more natural scenes and environments. This study aimed to bridge the gap synthetic and naturalistic stimuli by investigating whether some widely used statistics for the characterization of eye movement behavior differ for varying stimulus complexities. Our stimuli were either moving objects presented in their typical environment in naturalistic videos, such as a duck flying across a river, or simple synthetic blob targets moving along the same path on a uniform background. For both conditions we computed standard eye movement metrics that reflect processing time and accuracy for oculomotor control. The tracking behavior in both conditions revealed qualitatively comparable results, suggesting that the oculomotor system responded similarly to the movement of synthetic and naturalistic stimuli. For both conditions, we found a significant benefit when saccades and pursuit movements were collinear, which indicated similar interactions between both eye movements irrespective of stimulus conditions. However, we also found systematic differences between the two conditions. Latencies of initial saccades to moving blob targets were significantly lower by about 50 ms, whereas pursuit accuracy for the naturalistic targets was significantly higher. These two results give insight into two important problems that the oculomotor system must solve for successful target tracking. First, time is needed to select and define a target in a complex scene. When only a single target appears on a uniform background, the time required for selection is reduced. Second, the naturalistic context as presented in the videos seems to provide additional information, which can be useful to predict the path and movement behavior of the target to improve pursuit accuracy. Therefore, although tracking eye movements behave generally quite similar for synthetic and naturalistic stimuli, there are some caveats that one needs to keep in mind when comparing results across different stimulus complexities.

Effect of scene complexity on saccadic eye movements

Comparative studies of eye movement responses to targets moving along the same path presented in different contexts are rare. Research of eye movements during real-world scene exploration has mainly focused on fixation locations to understand which and how image properties such as bottom–up or top–down factors determine fixation placements (Baddeley & Tatler, 2006; Dearborn, 1936; Henderson, Nuthmann, & Luke, 2013; Itti & Koch, 2001; Land & Hayhoe, 2001; Torralba, Oliva, Castelhano, & Henderson, 2006). In complex real-world scenes, the detection and selection of a static target object can be potentially difficult if one thinks of the target selection in the context of a race model (Gold & Shadlen, 2007; Tatler et al., 2017). Computational saliency models of gaze behavior (Einhäuser, Rutishauser, & Koch, 2008; Itti & Koch, 2000; Kümmerer et al., 2017; Rai, Gutiérrez, & Le Callet, 2017) predict more variance for naturalistic complex scenes in comparison with clearly defined simple targets in synthetic scenes. An indication for this higher demand with respect to target selection and decision processes is in the present study the overall increase of saccade latency by 50 ms for targets surrounded by natural backgrounds in the video scenes compared to blob targets appearing on the uniform blank screen (Figure 3A). Higher saccadic latencies were also reported by Walshe and Nuthmann (2015) for targets in their uniform condition (203 ms) versus their scene condition (214 ms).

Interestingly, when there is a clearly defined search target and matched contrast across conditions, there might even be a facilitation of saccade processing with rich naturalistic backgrounds (White et al., 2008). White and colleagues observed that, when visibility was equated, latencies were faster by more than 100 ms for natural compared to unstructured backgrounds. This seems to be in contrast to our results but can potentially be explained by some important differences between the two. First, there is a task difference; instead of looking for one specific target as in the experiments of White et al. (2008), participants in our experiment were not instructed to search for a specific object in our experiments but rather to simply observe the scene with no additional constraint. Task demands have been shown to influence processing and gaze behavior, such as when comparing the performance of a real action versus watching a video recording of the same naturalistic input (Foulsham et al., 2011; Martens & Fox, 2007; Walshe & Nuthmann, 2015). Second, and most crucially, in the study of White and colleagues (2008) participants were searching for small, static Gabor targets under varying contrast levels, but our target objects were moving. Motion onsets have a high salience (Itti, 2005; Vig, Dorr, & Barth, 2011), and the observed saccade latencies in our experiment were also much smaller than the ones reported by White and colleagues. We observed average latencies of 183 ms for the synthetic condition and 235 ms for the naturalistic condition, whereas White and colleagues observed saccadic latencies of around 450 ms, which

are drastically different from ours. This indicates clear differences in target visibility, presumably due to the difference in behavior (i.e., movement vs. static and contrast) (Doma & Hallett, 1988; Ludwig et al., 2004; White et al., 2006), thus making a direct comparison between the two studies difficult. Although it is possible to use less complex scenes as a proxy for natural images, there seem to be additional caveats that must be taken into account (Foulsham & Kingstone, 2010).

The faster saccadic latency in the synthetic condition did not come at the cost of a speed–accuracy tradeoff, as the position error for saccades to the synthetic blob targets was also smaller than for the objects in natural scenes (Figure 3) and indicated a more general processing benefit. Still, the position error for objects in our video clips was within roughly 1.5 deg and thus far from being inaccurate. The way we defined the position error of saccades could lead to an overestimation of the position error in the naturalistic video condition, as here only one point of the moving object was labeled as target position and used for the calculation. Because we had not marked a certain location of the moving object as target nor asked participants to look at a certain part of the object, our observers could freely choose their gaze position. Therefore, the landing position in trials in which a participant looked at the relevant object but not on the labeled position (see, for example, Figure 4) could have led to larger position errors of saccades under our naturalistic condition.

The preferred landing position of saccades in the naturalistic condition also seemed to depend on the initial fixation position. We found that observers directed their saccades to different parts of larger objects after fixating higher or lower positions, as found, for example, in the video scene with the walking child (Figure 4). This result seems to be in line with the saccade landing positions of a study in which animals of different sizes were briefly shown in natural scenes. Saccade landing positions revealed a preference for the head of the animal and a secondary preference for the center of gravity (Drewes, Trommershäuser, & Gegenfurtner, 2011). This possibility of having multiple potential targets positions within larger objects is also reflected in less precise saccades measured through an increased variability in position errors (see Results). Interestingly, based on this difference between our synthetic and naturalistic targets one could assume much higher position errors in the naturalistic condition especially for larger objects. However, there was no relationship between the size of the target objects in the naturalistic condition with the magnitude of position errors (Figure 6A). There might simply be idiosyncratic differences in where and what people fixate when looking at certain objects in natural scenes (de Haas et al., 2019).

Effect of scene complexity on pursuit eye movements

For pursuit gain we observed no significant difference between the two experimental conditions, suggesting that for this measurement a generalization between simple lab stimuli and naturalistic videos seems possible. As for the saccadic parameters, we observed the same additional effect of the relative angle between the initial saccade and the pursuit response in both conditions. This suggests a high similarity in the coordination of the two eye movements across stimulus complexities. The lack of a difference in pursuit gain also suggests that the difference in saccade endpoint errors was affected by targeting different parts of the relevant target object. Such behavior led to a higher saccade endpoint error, but the pursuit gain remained comparable as the whole object roughly moved with the same velocity.

However, pursuit behavior can be influenced by background conditions. Studies have shown that pursuit is slower, when a stationary structured background was present, and it is faster when the background was dynamic (Collewijn & Tamminga, 1984; Masson, Proteau, & Mestre, 1995; Niemann & Hoffmann, 1997; Spering & Gegenfurtner, 2008). These differential context effects for stationary or dynamic backgrounds suggest that the pursuit system integrates target and background motion following a vector-averaging algorithm (Groh, Born, & Newsome, 1997; Lisberger & Ferrera, 1997; Spering & Gegenfurtner, 2008).

There are multiple reasons for the lack of a difference in pursuit gain in our study depending on the background.

First, due to the complex background in the naturalistic outdoor scenes we used, there were always multiple potential impacts present at the same time that also differed from scene to scene; for example, static backgrounds were present in street or park scenes. Some backgrounds contained more global movements such as a slowly flowing river. Moving target objects were often accompanied by other moving objects, such as when a person walking along a street was surrounded by other walking people. The different effects of other dynamic objects in the immediate surround of the target object presumably interact and make it difficult to disentangle and assess their potential impact on pursuit gain in general.

Second, the size of the eye movement target was not the same between the two conditions. Recent findings (Heinen, Potapchuk, & Watamaniuk, 2016; Watamaniuk & Heinen, 2015) have shown that larger objects led to an increased pursuit performance and fewer catch-up saccades. In Figure 6C, we have visualized how the sizes of our eight different targets benefited pursuit accuracy. The negative values indicate that target tracking is more accurate for objects in

naturalistic scenes than for the synthetic blob. Although the results of pursuit measurements for only eight different objects in their specific scenes were not statistically significant, the correlation with $r = 0.60$ is quite high, suggesting that with more scenes our results probably could replicate the findings of Heinen and colleagues.

Third, and most important, in our experiment the background shown in the video clips was not meaningless visual noise but rather the natural context for the moving objects. Along those lines, we observed a benefit for the pursuit of naturalistic targets in video scenes with respect to the pursuit direction error and its variability, indicating that the oculomotor system is able to use the additional information embedded in the scene context to improve tracking performance. In line with this result are the findings of a recent study that reported that, during eye tracking of a flying ball, the human oculomotor control was tuned to the realistic properties of the visual scene and was affected when, for example, the gravity was artificially changed (Delle Monache, Lacquaniti, & Bosco, 2019). Prior experience and expectations can also drive anticipatory pursuit (Kowler, 1989); prior information seems to be continuously integrated in a reliability-weighted manner even on a trial-by-trial basis (Darlington, Beck, & Lisberger, 2018; Darlington, Tokiyama, & Lisberger, 2017; Deravet, Blohm, de Xivry, & Lefèvre, 2018). Thus, the results of our study replicate findings suggesting that context information can improve pursuit accuracy and that the visuo-oculomotor system is able to use knowledge of the constraints and typical behavior of moving targets. Knowledge about the continuity and approximate speed of a walking child enables better tracking performance, an advantage that is not present when a moving Gaussian blob appears on a computer monitor.

Saccade–pursuit interactions

For both saccades and subsequent pursuit eye movements we observed a benefit in performance when the two eye movements were collinear: a reduction in saccade position error and increase in pursuit gain. Due to the early interval of 50 to 150 ms after the end of the saccade, the effect on the pursuit gain is unlikely to be based on new retinal information after the saccade. The post-saccadic time interval up to 150 ms is often considered as an open-loop interval (Buonocore, Skinner, & Hafed, 2019; Rasche & Gegenfurtner, 2009; Tavassoli & Ringach, 2009; Wyatt & Pola, 1983), during which, due to processing delays, only very limited new incoming retinal information is affecting the pursuit response. The benefit in pursuit gain could potentially be explained by muscle synergies; for collinear movements the eye can simply keep

moving in the same direction, whereas for the other two conditions the eye might have to decelerate in order to change the movement direction.

These muscle synergies, however, cannot explain why the saccade should be more accurate for collinear movements. This benefit cannot be explained by the orientation of the saccade (Figure 3E), leaving an explanation based on early interactions between the saccadic and pursuit system, as recently shown by Goettker and colleagues (2019). If the last part of the saccade is already affected by the upcoming pursuit response, saccade endpoints will deviate based on the upcoming pursuit direction, which will then increase the position error for the landing position, thus suggesting that saccades and pursuit are based on shared signals (Deravet et al., 2018; Erkelens, 2006; Goettker et al., 2019; Hainque, Apartis, & Daye, 2016; Orban de Xivry & Lefèvre, 2007), which optimize the tracking performance at the transition between saccades and pursuit.

Limitations and future research

Our current study provides a starting point for developing an understanding of the generalizability of results across different levels of stimulus complexity. We are aware of the fact that the two present conditions do not reflect the entire spectrum of possible stimuli, but we considered them to be representative of the two most commonly used conditions in the literature: experiments with simple synthetic stimulus and experiments with naturalistic content. To our knowledge, such a comparison is missing but could help with the interpretation and comparability of results in eye movement research across both scenarios.

For future research, one could further explore the entire spectrum ranging from a simple synthetic stimulus to eye tracking in the natural world. Our synthetic condition was a single simple stimulus on a homogeneous background at one edge of this spectrum, whereas our naturalistic condition with video recordings of natural environments was closer to the real world. Along the entire spectrum of stimulus complexities are many additional possibilities that could provide interesting insights. A version where the target is cropped out of the original video and is presented on a uniform background could be used to bridge the two conditions used in the present study. This would allow better control of differences in target contrast or size; however, the cropped-out target or even only the shape of the target might already provide semantic and contextual information about the object. A condition where participants actually view a natural environment either in virtual reality or in real life might even increase the natural validity further. Also, performing experiments with different tasks for the

observer, such as interacting with the moving object in its natural environment, might change the gaze patterns (Foulsham, Walker, & Kingstone, 2011; Martens & Fox, 2007; Walshe & Nuthmann, 2015). Even if no task is given, there might be implicit assumptions about the goal of the study that can become more pronounced for the more constrained conditions. In our synthetic condition, it is unlikely that the participants fixated on the gray background and they almost always followed the target. This provided us with clear target-directed eye movement responses and a low exclusion rate (see last paragraph of the Experimental design section). In the naturalistic condition, many different objects were shown simultaneously with the target that competed for the participant's attention, and no explicit task was given to the participants. As a result, the natural condition produced a larger variability in the observed gaze patterns.

Additionally, although it is desirable to actually record gaze behavior in our natural surround during everyday tasks, one caveat is that, despite improvements in tracking quality, portable video eye trackers still lack the temporal and spatial resolution available in more controlled settings in the lab. Therefore, some interesting measurements, such as the dependence of saccade and pursuit parameters on the relative angle in the early open-loop interval after the saccade, are at the moment not possible. Thus, although we have provided a starting point for comparing stimuli across varying stimulus complexity, we also had to make compromises to obtain the relevant measurements.

Conclusions

One would not expect identical results for tracking naturalistic objects in context and synthetic stimuli without context; however, to compare studies performed in these different contexts it is crucial to understand how large these differences are. While our results do not require a fundamental revision of our understanding of the oculomotor system and we were able to replicate previously known effects (influence of context on pursuit or saccade-pursuit interactions), they do advance the field by bringing together findings for vastly different stimuli.

The results of our study show that, to a large extent, generalization of oculomotor behavior across different stimulus complexities is possible; however, one must keep in mind certain caveats: (1) Different levels of complexity will lead to saccadic latency differences depending on the ease of detecting and identifying the eye movement target. (2) For moving targets, background effects are smaller and saccade latencies are shorter for unstructured backgrounds. (3) For initial saccades to larger and often asymmetrical

moving naturalistic objects, the target location is vaguely defined and influenced by multiple factors, such as object identity and initial fixation location, which makes it more difficult to determine the accuracy of saccadic movements. (4) Pursuit eye movements can become more accurate for larger targets or when context information allows better prediction of the target movement.

Keywords: saccades, pursuit, natural scenes, dynamic targets

References

Agtzidis, I., Startsev, M., & Dorr, M. (2016). In the pursuit of (ground) truth: a hand-labelling tool for eye movements recorded during dynamic scene viewing. In *2016 IEEE Second Workshop on Eye Tracking and Visualization (ETVIS)* (pp. 65–68). Piscataway, NJ: Institute of Electrical and Electronics Engineers.

Baddeley, R. J., & Tatler, B. W. (2006). High frequency edges (but not contrast) predict where we fixate: A Bayesian system identification analysis. *Vision Research, 46*(18), 2824–2833.

Buonocore, A., Skinner, J., & Hafed, Z. M. (2019). Eye position error influence over "open-loop" smooth pursuit initiation. *The Journal of Neuroscience, 39*(14), 2709–2721.

Collewijn, H., & Tamminga, E. P. (1984). Human smooth and saccadic eye movements during

voluntary pursuit of different target motions on different backgrounds. *The Journal of Physiology, 351*(1), 217–250.

Constantino, J. N., Kennon-McGill, S., Weichselbaum, C., Marrus, N., Haider, A., Glowinski, A. L., . . . Jones, W. (2017). Infant viewing of social scenes is under genetic control and is atypical in autism. *Nature, 547*(7663), 340–344.

Darlington, T. R., Beck, J. M., & Lisberger, S. G. (2018). Neural implementation of Bayesian inference in a sensorimotor behavior. *Nature Neuroscience, 21*(10), 1442–1451.

Darlington, T. R., Tokiyama, S., & Lisberger, S. G. (2017). Control of the strength of visual-motor transmission as the mechanism of rapid adaptation of priors for Bayesian inference in smooth pursuit eye movements. *Journal of Neurophysiology, 118*(2), 1173–1189.

de Brouwer, S., Missal, M., Barnes, G., & Lefèvre, P. (2002). Quantitative analysis of catch-up saccades during sustained pursuit. *Journal of Neurophysiology, 87*(4), 1772–1780.

de Haas, B., Iakovidis, A. L., Schwarzkopf, D. S., & Gegenfurtner, K. R. (2019). Individual differences in visual salience vary along semantic dimensions. *Proceedings of the National Academy of Sciences, USA, 116*(24), 11687–11692.

Dearborn, W. F. (1936). How people look at pictures: A study of the psychology of perception in art. *The Elementary School Journal, 37*(1), 66–67.

Delle Monache, S., Lacquaniti, F., & Bosco, G. (2019). Ocular tracking of occluded ballistic trajectories: Effects of visual context and of target law of motion. *Journal of Vision, 19*(4):13, 1–21, https://doi.org/10.1167/19.4.13.

Deravet, N., Blohm, G., de Xivry, J.-J. O., & Lefèvre, P. (2018). Weighted integration of short-term memory and sensory signals in the oculomotor system. *Journal of Vision, 18*(5):16, 1–19, https://doi.org/10.1167/18.5.16.

Doma, H., & Hallett, P. E. (1988). Dependence of saccadic eye-movements on stimulus luminance, and an effect of task. *Vision Research, 28*(8), 915–924.

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

Drewes, J., Trommershäuser, J., & Gegenfurtner, K. R. (2011). Parallel visual search and rapid animal detection in natural scenes. *Journal of Vision, 11*(2):20, 1–21, https://doi.org/10.1167/11.2.20.

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

Einhäuser, W., Spain, M., & Perona, P. (2008). Objects predict fixations better than early saliency. *Journal of Vision, 8*(14):18, 1–26, https://doi.org/10.1167/8.14.18.

Erkelens, C. J. (2006). Coordination of smooth pursuit and saccades. *Vision Research, 46*(1–2), 163–170.

Flanagan, J. R., & Johansson, R. S. (2003). Action plans used in action observation. *Nature, 424*(6950), 769–771.

Foulsham, T., & Kingstone, A. (2010). Asymmetries in the direction of saccades during perception of scenes and fractals: effects of image type and image features. *Vision Research, 50*(8), 779–795.

Foulsham, T., Walker, E., & Kingstone, A. (2011). The where, what and when of gaze allocation in the lab and the natural environment. *Vision Research, 51*(17), 1920–1931.

Gegenfurtner, K. R. (2016). The interaction between vision and eye movements. *Perception, 45*(12), 1333–1357.

Geisler, W. S., & Najemnik, J. (2005). Human and optimal eye movement strategies in visual search. *Journal of Vision, 5*(8), 778–778, https://doi.org/10.1167/5.8.778.

Glimcher, P. W. (2003). The neurobiology of visual-saccadic decision making. *Annual Review of Neuroscience, 26*, 133–179.

Goettker, A., Braun, D. I., & Gegenfurtner, K. R. (2019). Dynamic combination of position and motion information when tracking moving targets. *Journal of Vision, 19*(7):2, 1–22, https://doi.org/10.1167/19.7.2.

Gold, J. I., & Shadlen, M. N. (2007). The neural basis of decision making. *Annual Review of Neuroscience, 30*.

Groh, J. M., Born, R. T., & Newsome, W. T. (1997). How is a sensory map read out? Effects of microstimulation in visual area MT on saccades and smooth pursuit eye movements. *The Journal of Neuroscience, 17*(11), 4312–4330.

Haegerstrom-Portnoy, G., & Brown, B. (1979). Contrast effects on smooth-pursuit eye movement velocity. *Vision Research, 19*(2), 169–174.

Hainque, E., Apartis, E., & Daye, P. M. (2016). Switching between two targets with non-constant velocity profiles reveals shared internal model of target motion. *European Journal of Neuroscience, 44*(8), 2622–2634.

Hayhoe, M. (2000). Vision using routines: A functional account of vision. *Visual Cognition, 7*(1–3), 43–64.

Hayhoe, M. M. (2017). Vision and action. *Annual Review of Vision Science, 3*, 389–413.

Heinen, S. J., Potapchuk, E., & Watamaniuk, S. N. J. (2016). A foveal target increases catch-up saccade frequency during smooth pursuit. *Journal of Neurophysiology, 115*(3), 1220–1227.

Henderson, J. M., & Hayes, T. R. (2017). Meaning-based guidance of attention in scenes as revealed by meaning maps. *Nature Human Behaviour, 1*(10), 743–747.

Henderson, J. M., Nuthmann, A., & Luke, S. G. (2013). Eye movement control during scene viewing: immediate effects of scene luminance on fixation durations. *Journal of Experimental Psychology: Human Perception and Performance, 39*(2), 318–322.

Itti, L. (2005). Quantifying the contribution of low-level saliency to human eye movements in dynamic scenes. *Visual Cognition, 12*(6), 1093–1123.

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

Itti, L., & Koch, C. (2001). Computational modelling of visual attention. *Nature Reviews Neuroscience, 2*(3), 194–203.

Kennedy, D. P., D'Onofrio, B. M., Quinn, P. D., Bölte, S., Lichtenstein, P., & Falck-Ytter, T. (2017). Genetic influence on eye movements to complex scenes at short timescales. *Current Biology, 27*(22), 3554–3560.e3.

Kienzle, W., Franz, M. O., Schölkopf, B., & Wichmann, F. A. (2009). Center-surround patterns emerge as optimal predictors for human saccade targets. *Journal of Vision, 9*(5):, 1–15, https://doi.org/10.1167/9.5.7.

Kleiner, M., Brainard, D., & Pelli, D. (2007). What's new in Psychtoolbox-3? *Perception, 34*(14), 1–16.

Kowler, E. (1989). Cognitive expectations, not habits, control anticipatory smooth oculomotor pursuit. *Vision Research, 29*(9), 1049–1057.

Kümmerer, M., Wallis, T., & Bethge, M. (2017). DeepGaze II: Predicting fixations from deep features over time and tasks. *Journal of Vision, 17*(10), 1147, https://doi.org/10.1167/17.10.1147.

Land, M., Mennie, N., & Rusted, J. (1999). The roles of vision and eye movements in the control of activities of daily living. *Perception, 28*(11), 1311–1328.

Land, M. F., & Hayhoe, M. (2001). In what ways do eye movements contribute to everyday activities? *Vision Research, 41*(25–26), 3559–3565.

Lisberger, S. G. (2015). Visual guidance of smooth pursuit eye movements. *Annual Review of Vision Science, 1*, 447–468.

Lisberger, S. G., & Ferrera, V. P. (1997). Vector averaging for smooth pursuit eye movements initiated by two moving targets in monkeys. *The Journal of Neuroscience, 17*(19), 7490–7502.

Ludwig, C. J. H., Gilchrist, I. D., & McSorley, E. (2004). The influence of spatial frequency and contrast on saccade latencies. *Vision Research, 44*(22), 2597–2604.

Martens, M. H., & Fox, M. (2007). Does road familiarity change eye fixations? A comparison between watching a video and real driving. *Transportation Research Part F: Traffic Psychology and Behaviour, 10*(1), 33–47.

Masson, G., Proteau, L., & Mestre, D. R. (1995). Effects of stationary and moving textured backgrounds on the visuo-oculo-manual tracking in humans. *Vision Research, 35*(6), 837–852.

Munoz, D. P., & Everling, S. (2004). Look away: The anti-saccade task and the voluntary control of eye movement. *Nature Reviews Neuroscience, 5*(3), 218–228.

Najemnik, J., & Geisler, W. S. (2008). Eye movement statistics in humans are consistent with an optimal search strategy. *Journal of Vision, 8*(3):4, 1–14, https://doi.org/10.1167/8.3.4.

Navalpakkam, V., Koch, C., Rangel, A., & Perona, P. (2010). Optimal reward harvesting in complex perceptual environments. *Proceedings of the National Academy of Sciences, USA, 107*(11), 5232–5237.

Niemann, T., & Hoffmann, K. P. (1997). The influence of stationary and moving textured backgrounds on smooth-pursuit initiation and steady state pursuit in humans. *Experimental Brain Research, 115*(3), 531–540.

O'Mullane, G., & Knox, P. C. (1999). Modification of smooth pursuit initiation by target contrast. In *Vision Research, 39*(20), 3459–3464.

Orban de Xivry, J.-J., & Lefèvre, P. (2007). Saccades and pursuit: two outcomes of a single sensorimotor process. In *The Journal of Physiology, 584*(1), 11–23.

Peterson, M. F., & Eckstein, M. P. (2012). Looking just below the eyes is optimal across face recognition tasks. *Proceedings of the National Academy of Sciences, USA, 109*(48), E3314–E3323.

Rai, Y., Gutiérrez, J., & Le Callet, P. (2017). A dataset of head and eye movements for 360 degree images. In *MMSys'17: Proceedings of the 8th ACM on Multimedia Systems Conference*. New York: Association for Computing Machinery.

Rasche, C., & Gegenfurtner, K. R. (2009). Precision of speed discrimination and smooth pursuit eye movements. *Vision Research, 49*(5), 514–523.

Robinson, D. A. (1981). The use of control systems analysis in the neurophysiology of eye movements. *Annual Review of Neuroscience, 4*, 463–503.

Rothkopf, C. A., Ballard, D. H., & Hayhoe, M. M. (2007). Task and context determine where you look. *Journal of Vision, 7*(14):16, 1–2, https://doi.org/10.1167/7.14.16.

Schreiber, C., Missal, M., & Lefèvre, P. (2006). Asynchrony between position and motion signals in the saccadic system. *Journal of Neurophysiology, 95*(2), 960–969.

Schütz, A. C., Braun, D. I., & Gegenfurtner, K. R. (2011). Eye movements and perception: a selective review. *Journal of Vision, 11*(5):9, 1–30, https://doi.org/10.1167/11.5.9.

Schütz, A. C., Trommershäuser, J., & Gegenfurtner, K. R. (2012). Dynamic integration of information about salience and value for saccadic eye movements. *Proceedings of the National Academy of Sciences, USA, 109*(19), 7547–7552.

Sommer, M. A., & Wurtz, R. H. (2008). Brain circuits for the internal monitoring of movements. *Annual Review of Neuroscience, 31*, 317–338.

Sparks, D. L., & Mays, L. E. (1990). Signal transformations required for the generation of saccadic eye movements. *Annual Review of Neuroscience, 13*, 309–336.

Spering, M., & Gegenfurtner, K. R. (2008). Contextual effects on motion perception and smooth pursuit eye movements. *Brain Research, 1225*, 76–85.

Startsev, M., Agtzidis, I., & Dorr, M. (2019). Characterizing and automatically detecting smooth pursuit in a large-scale ground-truth data set of dynamic natural scenes. *Journal of Vision*, *19*(14):10, 1–25, https://doi.org/10.1167/19.14.10.

Tatler, B. W., Brockmole, J. R., & Carpenter, R. H. (2017). LATEST: A model of saccadic decisions in space and time. *Psychological Review, 124*(3), 267.

Tatler, B. W., Hayhoe, M. M., Land, M. F., & Ballard, D. H. (2011). Eye guidance in natural vision: reinterpreting salience. *Journal of Vision, 11*(5):5, 1–23, https://doi.org/10.1167/11.5.5.

Tatler, B. W., & Vincent, B. T. (2009). The prominence of behavioural biases in eye guidance. *Visual Cognition, 17*(6–7), 1029–1054.

Tavassoli, A., & Ringach, D. L. (2009). Dynamics of smooth pursuit maintenance. *Journal of Neurophysiology, 102*(1), 110–118.

Torralba, A., Oliva, A., Castelhano, M. S., & Henderson, J. M. (2006). Contextual guidance of eye movements and attention in real-world scenes: the role of global features in object search. *Psychological Review, 113*(4), 766–786.

Toscani, M., Valsecchi, M., & Gegenfurtner, K. R. (2013). Optimal sampling of visual information for lightness judgments. *Proceedings of the National Academy of Sciences, USA, 110*(27), 11163–11168.

Vig, E., Dorr, M., & Barth, E. (2009). Efficient visual coding and the predictability of eye movements on natural movies. *Spatial Vision, 22*(5), 397–408.

Vig, E., Dorr, M., & Barth, E. (2011). Learned saliency transformations for gaze guidance. In *Proceedings of SPIE 7865, Human Vision and Electronic Imaging XVI*. Bellingham, WA: SPIE.

Wade, N. J. (2010). Pioneers of eye movement research. *I-Perception, 1*(2), 33–68.

Walshe, R. C., & Nuthmann, A. (2015). Mechanisms of saccadic decision making while encoding naturalistic scenes. *Journal of Vision, 15*(5):21, 1–19, https://doi.org/10.1167/15.5.21.

Watamaniuk, S. N. J., & Heinen, S. J. (2015). Allocation of attention during pursuit of large objects is no different than during fixation. *Journal of Vision, 15*(9):9, 1–12, https://doi.org/10.1167/15.9.9.

White, B. J., Kerzel, D., & Gegenfurtner, K. R. (2006). The spatio-temporal tuning of the mechanisms in the control of saccadic eye movements. *Vision Research, 46*(22), 3886–3897.

White, B. J., Stritzke, M., & Gegenfurtner, K. R. (2008). Saccadic facilitation in natural backgrounds. *Current Biology, 18*(2), 124–128.

Wyatt, H. J., & Pola, J. (1983). Smooth pursuit eye movements under open-loop and closed-loop conditions. *Vision Research, 23*(10), 1121–1131.

Xu, J., Jiang, M., Wang, S., Kankanhalli, M. S., & Zhao, Q. (2014). Predicting human gaze beyond pixels. *Journal of Vision, 14*(1):28, 1–20, https://doi.org/10.1167/14.1.28.

Yarbus, A. L. (1967). Eye movements during perception of complex objects. In *Eye movements and Vision* (pp. 171–211). Boston, MA: Springer.