Acta Psychologica 219 (2021) 103364

Contents lists available at ScienceDirect

Acta Psychologica

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

Replicability and reliability of the background and target velocity effects in smooth pursuit eye movements

Rebekka Schröder, Philine Margarete Baumert, Ulrich Ettinger *

Department of Psychology, University of Bonn, Kaiser-Karl-Ring 9, 53111 Bonn, Germany

ARTICLE INFO

Keywords:
Smooth pursuit eye movements
Reliability
Replicability
Reliability paradox
Individual differences

ABSTRACT

When we follow a slowly moving target with our eyes, we perform smooth pursuit eye movements (SPEM). Previous investigations point to significantly and robustly reduced SPEM performance in the presence of a stationary background and at higher compared to lower target velocities. However, the reliability of these background and target velocity effects has not yet been investigated systematically.

To address this issue, 45 healthy participants (17 m, 28 f) took part in two experimental sessions 7 days apart. In each session, participants were instructed to follow a horizontal SPEM target moving sinusoidally between ±7.89° at three different target velocities, corresponding to frequencies of 0.2, 0.4 and 0.6 Hz. Each target velocity was presented once with and once without a stationary background, resulting in six blocks. The blocks were presented twice per session in order to additionally explore potential task length effects. To assess SPEM performance, velocity gain was calculated as the ratio of eye to target velocity.

In line with previous research, detrimental background and target velocity effects were replicated robustly in both sessions with large effect sizes. Good to excellent test-retest reliabilities were obtained at higher target velocities and in the presence of a stationary background, whereas lower reliabilities occurred with slower targets and in the absence of background stimuli. Target velocity and background effects resulted in largely good to excellent reliabilities.

These findings not only replicated robust experimental effects of background and target velocity at group level, but also revealed that these effects can be translated into reliable individual difference measures.

1. Introduction

When we follow a small, slowly moving target with our eyes, we perform smooth pursuit eye movements (SPEM) in order to hold the image of the target on the fovea (Leigh & Zee, 2015; Lisberger, 2015; Lisberger et al., 1987). SPEM represent a complex sensorimotor behaviour incorporating various perceptual, motor and cognitive processes including attention, prediction and inhibition (Barnes, 2008; Lisberger, 2015). Parameters used to quantify SPEM accuracy include measures such as pursuit velocity gain (ratio of eye velocity to target velocity) or root mean square error (RMSE) as well as more specific measures such as the number of intrusive (e.g. anticipatory saccades) or compensatory saccades (e.g. catch-up saccades) (Barnes, 2008; Lencer & Trillenberg, 2008; Smyrnis, 2008). Velocity gain is considered the primary measure of performance of the smooth pursuit system (Barnes, 2008; Lencer & Trillenberg, 2008; Smyrnis, 2008).

A well-established experimental finding is that SPEM accuracy is reduced in the presence of task irrelevant, visual distractors or structured backgrounds (Barnes & Crombie, 1985; Collewijn & Tamminga, 1984; Hutton et al., 2000; Kaufman & Abel, 1986; Mohrmann & Thier, 1995; Niemann & Hoffmann, 1997; Spering et al., 2006). It has been argued that the optokinetic drive induced by the background slows pursuit eye movements (Barnes, 2008). This process can be influenced by attention directed to enhance target processing and/or reduce background processing (Barnes, 2008), a process related to inhibitory control (Friedman & Miyake, 2004). Importantly, increased deployment of attention may counter, but does not fully abolish the detrimental influence of task irrelevant background stimuli, leading to the replicable observation of this background effect in the literature (Collewijn & Tamminga, 1984; Hutton et al., 2000; Kaufman & Abel, 1986; Mohrmann & Thier, 1995; Niemann & Hoffmann, 1997; Spering et al., 2006).

SPEM is also highly sensitive to target velocity (or target frequency). Horizontal SPEM targets typically either have a constant velocity or follow a sinusoidal velocity pattern, i.e. decelerating towards the

* Corresponding author.
E-mail address: ulrich.ettinger@uni-bonn.de (U. Ettinger).

https://doi.org/10.1016/j.actpsy.2021.103364
Received 6 January 2021; Received in revised form 23 June 2021; Accepted 1 July 2021
Available online 7 July 2021

turnings points and accelerating towards the center of the screen. The latter are described by their peak velocity and/or their temporal frequency given a specific amplitude of target excursion. It has been shown that performance reliably deteriorates with increasing target velocity (or frequency) (Collewijn & Tamminga, 1984; Lisberger et al., 1981). This effect indicates that increasing demands on the system at higher target velocities cannot be fully addressed, leading to the observed performance decreases. The effect might be attributed to eye velocity and/or acceleration saturation at higher target velocities (Buizza & Schmid, 1986).

A recent study replicated the background effect, i.e. a general performance decrease in the presence of a stationary background (Meyhöfer et al., 2019). In addition, it was shown that the background effect interacted with the detrimental effects of target velocity on velocity gain, as performance decrements with increasing target velocity were particularly strong in the presence of a stationary background (Meyhöfer et al., 2019). This suggests that the inhibition of the influence of background stimuli and the precise matching of eye to target velocity compete for the same limited cognitive resources or rely on the same underlying system, possibly related to spatial attention or motion processing (Spering & Montagnini, 2011; Van Donkelaar & Drew, 2002).

Whereas the background and target velocity effects on SPEM have been consistently reported in the literature and are thus highly replicable at group level, less is known about their reliability at the level of individual differences. In this context, we refer to replicability as obtaining a similar finding with different random samples that capture the most important facets of the original research (Asendorpf et al., 2013). In contrast, reliability refers to the extent to which a measure consistently ranks individuals (Hedge et al., 2018).

In a widely noted recent methodological analysis, it was argued that robust and replicable experimental task effects at group level cannot necessarily be translated into reliable individual difference measures due to low between-subject variability, a phenomenon referred to as the reliability paradox (Hedge et al., 2018). In this context, Hedge et al. (2018) showed that several classic, attention demanding inhibitory tasks such as the Stroop and Eriksen flanker tasks show relatively poor test-retest reliability despite producing replicable task effects at group level. This phenomenon can be traced back to different objectives in experimental and correlative research fields concerning within- and between-subjects variance maximization.

Previous studies on the reliability of smooth pursuit performance have yielded mostly moderate to good reliability scores (Bargary et al., 2017; Calkins et al., 2003; Ettinger et al., 2003; Roy-Byrne et al., 1995; Versino et al., 1993). Importantly, the only study that distinguished between different target velocities reported better reliability for faster targets (Ettinger et al., 2003), suggesting that reliability might dependent on overall demands on the pursuit system, i.e. task difficulty. Most strikingly, to our knowledge, no previous study has reported the reliability of the background or target velocity effects in SPEM. This is an important gap in the literature, particularly given the observation of strong within-subject variance in these highly replicable group-level effects.

Although SPEM has long been studied in experimental or group designs (Barnes & Asselman, 1991; Haraldsson et al., 2008; Holzman et al., 1973), more recently, there has been an increasing focus on individual differences in SPEM using correlational designs (Bargary et al., 2017; Lenzenweger & O'Driscoll, 2006; Smyrnis et al., 2007). Many of these studies have investigated correlations between SPEM measures and psychosis-spectrum personality traits, based on the continuum hypothesis of individual differences in personality and psychopathology (Ettinger et al., 2014; Haslam et al., 2020). Interestingly, studies on the relationship between SPEM performance and psychosis-related personality traits such as schizotypy typically yield only small correlations (Lenzenweger & O'Driscoll, 2006; Smyrnis et al., 2007), possibly due to low reliability of task performance (as has been argued regarding cognitive tasks; Hedge et al., 2018). Therefore, the current study will focus on exploring the reliability of smooth pursuit in a paradigm with and without a stationary background and at different target velocities in a group of healthy participants, in order to estimate the reliability not only of pursuit performance in general but, specifically, of the target velocity and background effects.

In addition to this important primary objective, this study will address a number of related, secondary questions. As a secondary issue, the role of task duration in measuring reliability will be addressed. In experimental investigations, shorter tasks carry several advantages including economic efficiency, higher acceptability and easier applicability in patient or developmental populations. However, compared to longer tasks they bear the risk of lower reliability (Hedge et al., 2018; Wöstmann et al., 2013). The extent, to which task duration influences the reliability of SPEM, and in particular of the target velocity and background effects, has not yet been characterised. Therefore, this study will also address this issue using a modification of a task that has previously been shown to produce robust background and target velocity effects at group level (Meyhöfer et al., 2019). Specifically, in each session (test, retest), the task is presented twice, in two blocks, and reliability indices are calculated for the short (one block) as well as the long (two blocks) version of the task in order to characterise effects of task duration on reliability.

A further aim of this study is to explore how repeated exposure to the task both across sessions (test, retest) and in each session (first block, second block) influences the magnitude of the background and target velocity effects. While in standard SPEM tasks, in the absence of a background, no or only small effects of repeated exposure on pursuit performance across time points have been observed (Calkins et al., 2003; Ettinger et al., 2003), less is known about the influence of task repetition on the robustness of the background and target velocity effects. Complex interactions between perceptual and motor learning may help to automate behaviour and improve performance over time (Censor et al., 2012; Ostry & Gribble, 2016). By presenting the task in two blocks in each session and twice across two sessions over the course of one week, we aim to investigate whether the background and target velocity effects can be replicated robustly within the same sample and to what extent they are affected by time. If interactions between time or block and background or target velocity are revealed, it is crucial to address this issue in future studies as they can act as confounding factors in longitudinal designs.

In summary, this study had six specific aims: (1) to replicate earlier findings of reduced smooth pursuit performance in the presence of a stationary background and at higher target velocities, (2) to explore the test-retest and split-half reliability of smooth pursuit performance at different target velocities with and without a stationary background, (3) to specifically explore the reliability of the background and target velocity effects, (4) to investigate whether reliability increases when using a longer version of the task, (5) to explore the effects of repeated task blocks on performance, and finally, (6) to explore how repeated exposure to the task across sessions influences the magnitude of the background and target velocity effects.

2. Materials and methods

2.1. Power

To detect a significant intraclass correlation ICC of 0.40 (the threshold for fair reliability according to Cicchetti, 1994) compared to no correlation at all with 0.80 power, a sample size of least 39.5 participants is necessary (Walter et al., 1998). With at least 43.5 participants it is possible to detect a difference between fair (0.40) and good (0.60) reliabilities with 0.80 power (Walter et al., 1998). Concerning the background and target velocity effects, at least 26 participants were needed to detect a large effect ($\eta_p^2 = 0.26$) with at least 0.80 power (Faul et al., 2007). Thus, the minimum sample size we aimed for was 44 participants.

2.2. Participants

Participants were recruited via advertisements on the campus of the University of Bonn, circular emails and social media. Exclusion criteria were current diagnosis of physical, psychiatric or neurological condition and current consumption of prescription or over-the-counter medication (except for oral contraceptives in women, nutritional supplements or thyroid drugs). Smokers were asked to abstain from smoking for at least 2 h prior to the sessions. Participants were included if they were healthy university students aged 18–35 years. All participants had normal or corrected-to-normal vision. The study procedures were approved by the research ethics committee of the Department of Psychology at the University of Bonn. Participants provided written, informed consent and received course credits for participation.

2.3. Study design and procedure

Upon recruitment, participants were asked to confirm inclusion and exclusion criteria and to fill in a short online questionnaire to assess demographic information (age, sex, current occupation, years of education, and handedness, assessed via the Edinburgh Handedness Inventory (Oldfield, 1971)). If suitable, they were invited to take part in two sessions in the eye-tracking laboratory. The assessments were carried out at the same time of the day (± 2 h) in two sessions (T1, T2) approximately one week apart (7 days, ± 2 days).

In each session, the task described in 2.4 was carried out two times (block A vs. block B) with a short break in between. In total, one session took no longer than 15 min.

2.4. Task

The task was the same as the one used by Meyhöfer et al. (2019) and is available here (https://osf.io/qbtcf). It was written using Experiment Builder (SR Research Ltd., Ontario, Canada, version 1.10) and presented on the inner 1680 × 1050 px of a flat-screen BenQ monitor (screen dimensions 42.9 × 22.2 cm; resolution 1920 × 1080 pixels; refresh rate 144 Hz). The task was presented on a black background (RGB = 0, 0, 0) in a block design in randomized order. The target was a grey circle (RGB = 128, 128, 128; diameter = 15px/0.27°, stroke width = 5px/0.09°) moving horizontally between ±432px (7.89°) across the screen in a sinusoidal waveform at three different target velocities (corresponding to frequencies of 0.2 Hz, 0.4 Hz, 0.6 Hz), always starting from the central position (0°, 0°). The sinusoidal pattern indicates that target velocity constantly changed over time, accelerating towards the center of the screen, and decelerating towards the turning points. Peak and average velocities were 9.91°/s and 6.31°/s for the 0.2 Hz target, 19.83°/s and 12.62°/s for the 0.4 Hz target and 29.74°/s and 18.94°/s for the 0.6 Hz target. Target trajectories for the three target velocity conditions are depicted in Fig. 1 along with exemplary eye position data from one participant.

Each target velocity condition was presented once on a blank and once on a structured stationary background (Fig. 2), resulting in a total of 6 blocks, each lasting 30 s. The stationary background consisted of a six-by-six grid of white circles (RGB = 255, 255, 255; diameter = 15px/0.27°, stroke width = 5px/0.09°) symmetrically distributed along the horizontal and vertical plane of the screen (corner coordinates in pixels: 408, 310; 408, 740; 1272, 310; 1272, 740).

Between blocks, a fixation circle was presented. Participants rested their head on a chin-rest and were instructed to follow the target as accurately as possible with their eyes while keeping their head still. At

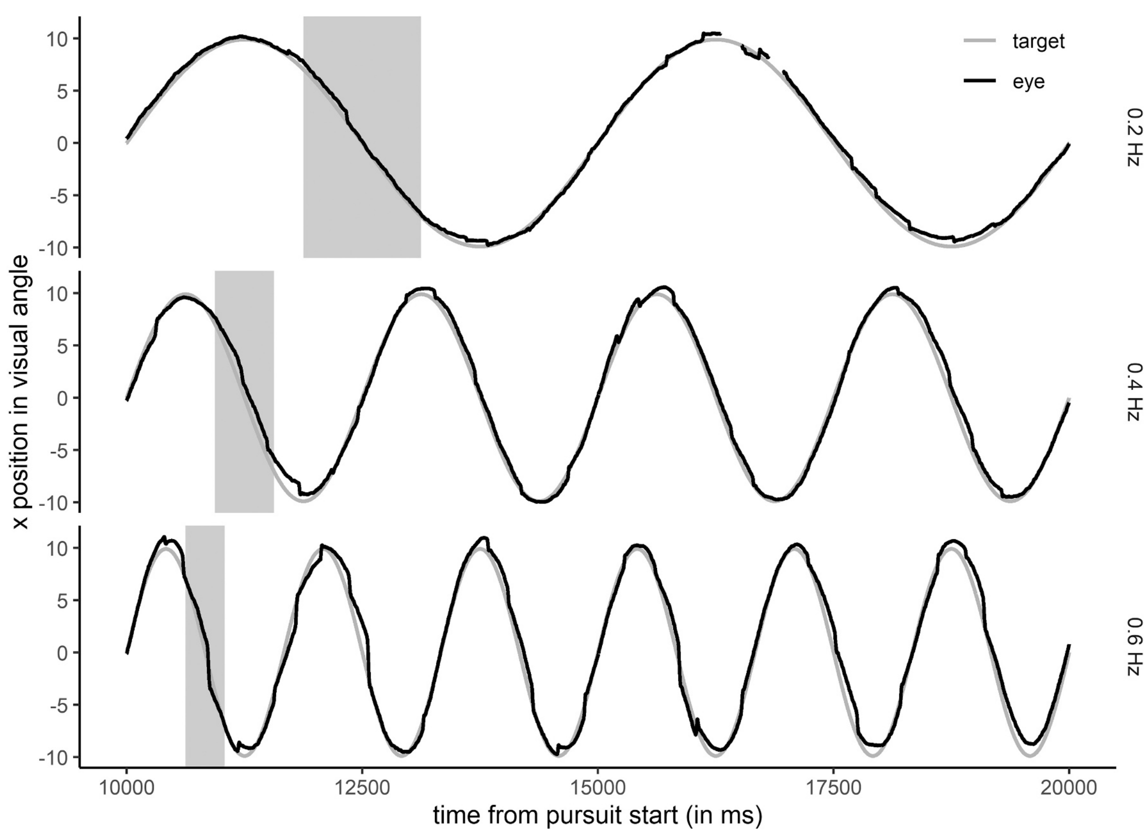

Fig. 1. Target and eye trajectories for the three different target velocities and window of gain analysis.
Legend: Trajectories of the target and exemplary eye data of one participant for the middle 10 s of each target velocity condition without background. Missing eye data indicate blinks. Grey shaded areas show the critical interval for gain analysis (only shown for the first half-ramp) for each of the target velocity conditions.

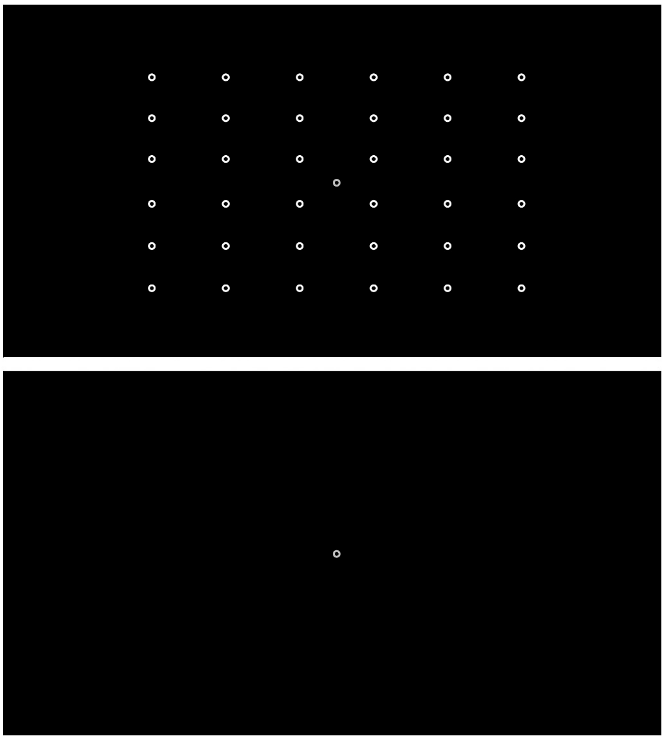

Fig. 2. Smooth pursuit target display.
Legend: The upper panel depicts the target display in the background condition. The lower panel depicts the target display in the no background condition. Both panels show the grey target at the center of the screen.

the beginning of each session, they performed a brief practice block in order to get familiarized with the task. The practise block consisted of four trials, each lasting 5 s in the following order: 0.4 Hz without structured background, 0.2 Hz with structured background, 0.4 Hz with structured background, 0.6 Hz without structured background.

2.5. Eye movement recording

To record eye movements, a desktop-mounted video-based combined pupil and corneal reflection eye-tracker (EyeLink 1000, SR Research, Ottawa, Ontario, Canada) was used. A centroid pupil-tracking algorithm was employed to detect pupil and corneal reflection of the right eye at a sampling rate of 1000 Hz. Prior to the task, a five-point horizontal-vertical calibration was performed. Distance from eye to monitor was approximately 70 cm.

2.6. Eye movement preprocessing

Eye movement data were preprocessed in Matlab R2019B. As we were not interested in the initiation phase of pursuit, the first excursion of the target from the centre to the peripheral turnaround point was excluded from the analyses.

Velocity gain was calculated as the average ratio of mean eye velocity and mean target velocity for the middle 50% of each half-cycle (the excursion of the target from one peripheral turnaround point to the other) for segments longer than 50 ms excluding blinks and saccades. The critical interval for gain analysis is depicted in Fig. 1. Saccades were identified using velocity ($\geq 22^\circ/s$) and acceleration ($\geq 3800^\circ/s^2$) criteria. Exclusion of blinks and saccades resulted in segments of different lengths that were time-weighted in the averaging procedure according to the duration of the segments.

To compare reliability between the short and long versions of the task, all dependent variables were also calculated for a joint version of blocks A and B of each session. To do so, data of relevant segments (i. e. after exclusion of blinks or saccades) of blocks A and B were again time-weighted according to the duration of the segment and then averaged.

Background effects were calculated separately for each target velocity condition by subtracting the gain values of the background condition from the no background condition. Target velocity effects were calculated separately for each background condition by subtracting the gain values of the 0.6 Hz condition from the 0.2 Hz condition.

The present paper focuses on pursuit gain as the primary outcome variable of pursuit. However, other outcomes are also important (Orban de Xivry & Lefèvre, 2007). Therefore, additional analyses of catch-up saccade rate and root mean square error (RMSE) are provided in the Supplementary Material.

2.7. Statistical analysis

Statistical analyses were carried out using R with the following packages: *ez* V.4.4.0 for analyses of variance (ANOVAs) (Lawrence, 2016), *rstatix* V.0.5.0 for pairwise *t*-tests (Kassambra, 2020), *irr* V.0.84.1 for ICCs (Gamer et al., 2019), *psych* V.1.9.12.31 for Pearson correlations (Revelle, 2019) and *tidyverse* V.1.3.0 for general data management (Wickham et al., 2019).

The significance threshold for all analyses was $\alpha = 0.05$. Outliers were identified separately for each condition. Values were defined as outliers if they exceeded the mean plus three times the interquartile range (IQR) criterion or fell below the mean minus three times the IQR criterion. If the IQR was 0, outliers were not defined. Participants were excluded from analyses of variance if their scores were defined as outliers in more than half of the conditions. Participants were excluded from reliability analyses of those condition pairs, where at least one of two scores were defined as an outlier.

2.7.1. Test-retest reliability

To assess test-retest reliability, both Pearson and intraclass correlations (ICC) between T1 and T2 were calculated separately for the short (only block A) and long (blocks A and B combined) versions of the task. Additionally, they were also calculated for the background and target velocity effects. For ICCs, we used the two-way mixed-effect model for single measurements as a measure of absolute agreement (McGraw & Wong, 1996). ICCs have been widely used to assess test-retest reliabilities for a variety of outcomes (Bargary et al., 2017; Ettinger et al., 2003; Hedge et al., 2018).

To facilitate comparison, 95% confidence intervals were calculated. Reliability coefficients are interpreted according to the guidelines proposed by Cicchetti (1994), indicating that values less than 0.40 are poor, values between 0.40 and 0.59 are fair, values between 0.60 and 0.74 are good and values between 0.75 and 1.00 are excellent.

2.7.2. Split-half reliability

To assess split-half reliability, we calculated Pearson correlations and ICCs between block A and block B separately for T1 and T2. Additionally, they were also calculated for the background and target velocity effects.

2.7.3. Analyses of variance

To assess the effects of background, target velocity, block and time on velocity gain, we carried out a four-way repeated-measures ANOVA with the within-subjects factors background (present, absent), target velocity (0.2 Hz, 0.4 Hz, 0.6 Hz), block (A, B) and time (T1, T2) for pursuit gain as the dependent variable. Effect sizes were calculated as partial eta squared. If the sphericity assumption was violated, Greenhouse-Geisser correction was applied. Uncorrected degrees of freedom and Greenhouse-Geisser ε were calculated. Bonferroni-corrected *t*-tests were calculated as post hoc tests with d_{av} (Lakens, 2013) as effect size. Uncorrected *p*-values were obtained but significance was inferred from corrected alpha-thresholds.

Study data and the analysis code are available at https://osf.io/qbtcf.

3. Results

3.1. Participants

The final sample consisted of $N = 45$ participants (17 males, 28 females), aged $M = 23.00$ ($SD = 3.02$) years. Six additional participants completed T1 but did not return for T2. Those participants are not included in the analyses. Additionally, one participant (male) was excluded from the ANOVA because his scores were outliers in more than half of the conditions.

The mean absolute difference between T1 and T2 was 7.07 days ($SD = 0.33$ days, minimum = 7 days, maximum = 9 days). The mean absolute difference between the starting times of the two sessions was 7.78 min ($SD = 18.43$ min, minimum = 0 min, maximum = 110 min).

3.2. Descriptive results

Descriptive statistics of all dependent variables are shown in Table 1. As can be seen descriptively, both the presence of a stationary background and an increase in target velocity substantially affected performance, indicative of the expected background and target velocity effects (see Table 2; for detailed statistical analyses of these effects, see 3.4).

3.3. Reliability analyses

Results of the reliability analyses are shown in Supplementary Table 1 and Fig. 3 (ICCs only, Pearson correlations can be found in Supplementary Fig. 1). The number of outliers in each condition that were removed from the analyses can be found in Supplementary Table 4.

For the direct performance measures, Pearson correlations ranged from 0.51 to 0.94. ICCs were very similar to Pearson correlations and ranged from 0.50 to 0.92. Reliabilities were fair to good for the 0.2 Hz no background condition and good to excellent for all other conditions. Descriptively, in almost all conditions, reliability was higher in the presence of a stationary background and at higher target velocities.

Reliability indicators of the background and target velocity effects of velocity gain can be found in Supplementary Tables 2 and 3. For the background effect, reliability was good to excellent, ranging from 0.60 to 0.86 (Pearson correlations) and 0.59 and 0.85 (ICCs). For the target velocity effect, reliability was fair to excellent, ranging from 0.52 to 0.85 (Pearson correlations) and 0.51 to 0.86 (ICCs). Fig. 4 shows the correlations between T1 and T2 of the background effects of pursuit gain for the long version. Fig. 5 shows the correlations between T1 and T2 of the target velocity effects of pursuit gain for the long version.

For the direct performance measures, the longer task version always achieved higher test-retest Pearson correlations and ICCs than the shorter version except for the 0.4 Hz no background condition, where the opposite pattern was found. On average, the longer version was more reliable than the shorter version by 0.037 (Pearson correlation) and 0.042 (ICCs). The same pattern of results was found for the difference scores (background effect, target velocity effect). For the background effect, the longer version outperformed the shorter version on average by 0.08 (Pearson correlations) or 0.07 (ICCs). For the target velocity effect, the average difference between the short and long version was 0.07 (both for Pearson correlations and ICCs). For all conditions, confidence intervals of the short and long version of the task overlapped.

Split-half reliabilities reached results similar to test-retest reliabilities. In all conditions, split-half reliability was higher at T2 than at T1. The average difference between split-half reliabilities at T1 and T2 was 0.130 (Pearson correlations) and 0.135 (ICCs) for the direct performance measures. However, confidence intervals overlapped for all conditions except for the 0.2 Hz background condition. This pattern of results was similar for the difference scores. For the background effect, T2 split-half reliability outperformed T1 split-half reliability on average by 0.113 (Pearson) or 0.127 (ICCs). For the target velocity effect, T2 split-half reliability outperformed T1 split-half reliability on average by 0.195 (Pearson) or 0.210 (ICCs). All confidence intervals of T1 and T2 split-half reliability overlapped for the difference scores.

Table 1
Descriptive statistics of velocity gain.

Target velocity	Time	Block	Background		No background	
			M	*SD*	*M*	*SD*
0.2 Hz	T1	A	0.87	0.11	0.96	0.04
		B	0.84	0.14	0.96	0.04
		A + B	0.86	0.11	0.96	0.04
	T2	A	0.86	0.14	0.95	0.05
		B	0.87	0.13	0.96	0.05
		A + B	0.86	0.13	0.96	0.05
0.4 Hz	T1	A	0.75	0.17	0.90	0.08
		B	0.76	0.18	0.90	0.08
		A + B	0.75	0.17	0.90	0.07
	T2	A	0.77	0.18	0.90	0.07
		B	0.78	0.18	0.91	0.07
		A + B	0.77	0.17	0.90	0.07
0.6 Hz	T1	A	0.60	0.21	0.79	0.15
		B	0.61	0.23	0.81	0.13
		A + B	0.60	0.21	0.80	0.13
	T2	A	0.66	0.21	0.82	0.13
		B	0.68	0.21	0.83	0.12
		A + B	0.67	0.20	0.83	0.12

Descriptive statistics (*M* mean and *SD* standard deviation) of velocity gain of the background and no background conditions at three different target velocities of two sessions (T1, T2) one week apart, separately for blocks A and B and a joint version of the blocks (A + B) in a sample of $N = 44$ participants.

3.4. Task and time effects

Analyses of velocity gain revealed main effects of background ($F_{(1, 43)} = 69.87, p < .001, \eta_p^2 = 0.619$), target velocity ($F_{(2, 86)} = 114.69, p < .001, \eta_p^2 = 0.727, \epsilon = 0.62$) and time ($F_{(1, 43)} = 7.16, p = .011, \eta_p^2 = 0.143$). Gain was higher without a stationary background, at lower target velocities and at T2.

In addition, we found a significant two-way interaction between background and target velocity ($F_{(2, 86)} = 20.77, p < .001, \eta_p^2 = 0.326$). Bonferroni-corrected *t*-tests revealed significant differences between the background conditions at all target velocities (0.2 Hz: background vs. no background $t_{(43)} = -6.47, p < .001, d_{av} = -1.11$; 0.4 Hz: background vs. no background $t_{(43)} = -7.16, p < .001, d_{av} = -1.10$; 0.6 Hz: background vs. no background $t_{(43)} = -9.43, p < .001, d_{av} = -1.01$). Qualitatively, the interaction suggests that the effect of target velocity was stronger in the presence of a structured background.

We also found a significant two-way interaction between background and time ($F_{(1, 43)} = 8.44, p = .006, \eta_p^2 = 0.164$). Bonferroni-corrected *t*-tests showed that the improvement in gain from T1 to T2 was significant only in the background condition but not in the absence of a background (background: T1 vs. T2 $t_{(43)} = -3.49, p < .001, d_{av} = -0.15$; no background: T1 vs. T2 $t_{(43)} = -1.09, p = .28$, n.s. at the Bonferroni-corrected alpha-level, $d_{av} = -0.08$). The background effect was smaller at T2 than at T1, but achieved significance in both sessions (T1: background vs. no background $t_{(43)} = -9.09, p < .001, d_{av} = -0.92$; T2: background vs. no background $t_{(43)} = -7.25, p < .001, d_{av} = -0.86$).

Moreover, there was a two-way interaction between target velocity and time ($F_{(2, 86)} = 25.29, p < .001, \eta_p^2 = 0.370, \epsilon = 0.82$). Post hoc *t*-tests results showed that gain scores significantly increased from T1 to T2 only in the 0.6 Hz condition, but not in the lower target velocity conditions (0.2 Hz: T1 vs. T2 $t_{(43)} = 0.27, p = .79$, n.s. at the Bonferroni-corrected alpha-level, $d_{av} = 0.02$; 0.4 Hz: T1 vs. T2 $t_{(43)} = -1.37, p = .18$, n.s. at the Bonferroni-corrected alpha-level, $d_{av} = -0.08$; 0.6 Hz: T1 vs. T2 $t_{(43)} = -5.21, p < .001, d_{av} = -0.24$). Significant main effects of target velocity were found in both sessions, when analyzing them

Table 2
Descriptive statistics of the background and target velocity effects.

Time	Block	Background effect						Target velocity effect			
		0.2 Hz		0.4 Hz		0.6 Hz		Background		No background	
		M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
T1	A	0.09	0.10	0.15	0.15	0.18	0.13	0.27	0.15	0.18	0.13
T1	B	0.12	0.12	0.14	0.14	0.20	0.16	0.24	0.17	0.16	0.12
T1	A + B	0.11	0.10	0.14	0.14	0.19	0.13	0.26	0.15	0.17	0.12
T2	A	0.09	0.12	0.13	0.14	0.16	0.14	0.20	0.15	0.13	0.11
T2	B	0.09	0.11	0.13	0.15	0.14	0.14	0.19	0.14	0.13	0.10
T2	A + B	0.09	0.11	0.13	0.13	0.15	0.13	0.19	0.13	0.13	0.10

Descriptive statistics (*M* mean and *SD* standard deviation) of velocity gain background effect (no background condition minus background condition) at three different target velocities and velocity gain target velocity effect (0.2 Hz condition minus 0.6 Hz condition) of the two background conditions of two sessions (T1, T2) one week apart, separately for blocks A and B and a joint version of the blocks (A + B) in a sample of $N = 44$ participants.

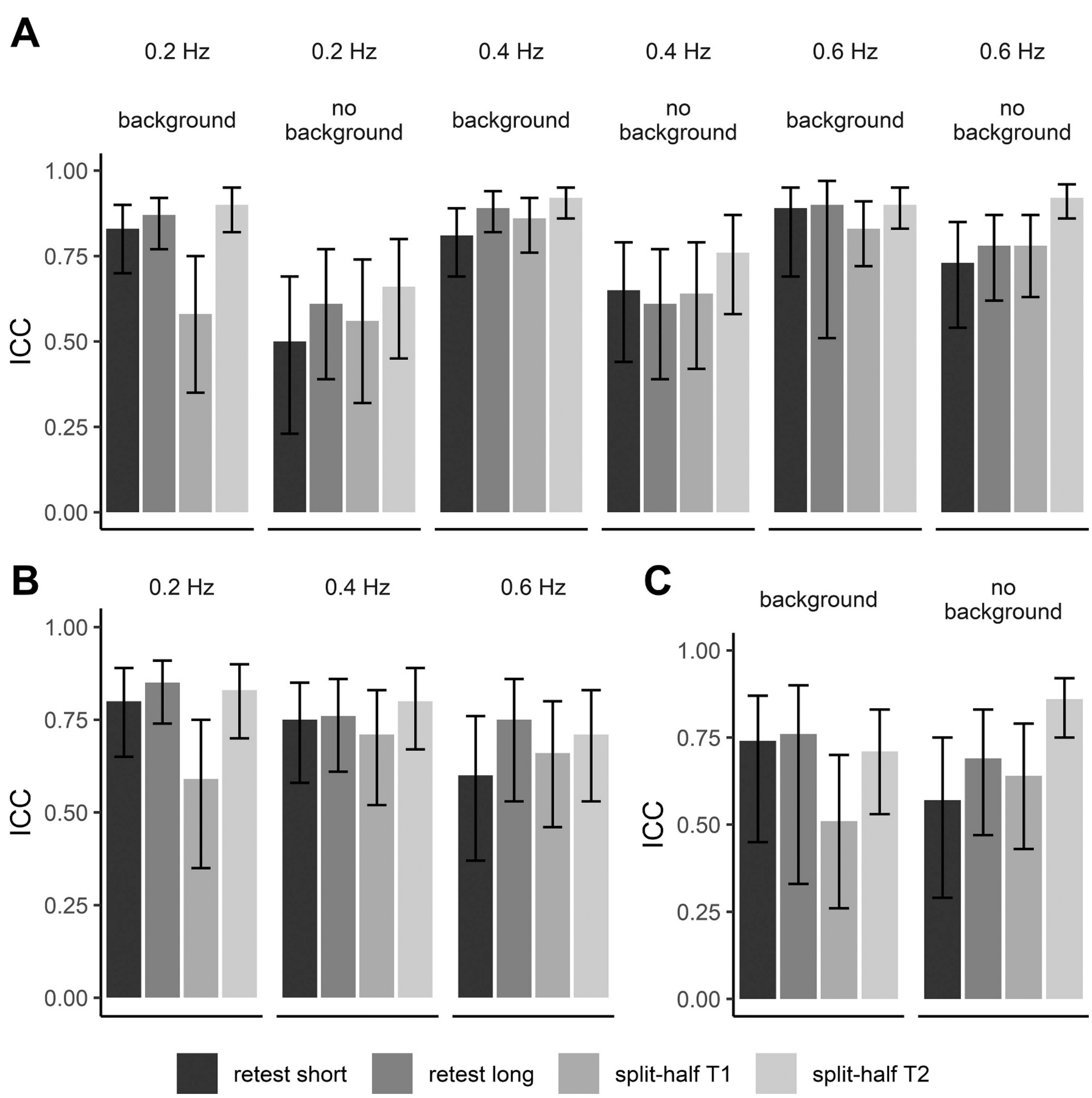

Fig. 3. Results of the reliability analyses (intraclass correlations).
Legend: Reliabilities for pursuit gain (Panel A) for the six background and target velocity conditions, pursuit gain background effect (Panel B; no background condition minus background condition) and pursuit gain target velocity effect (Panel C; 0.2 Hz condition minus 0.6 Hz condition). ICC = intraclass correlation. Error bars represent the upper and lower limit of the 95% confidence interval.

separately (T1: $F_{(2, 86)} = 124.75$, $p < .001$, $\eta_p^2 = 0.744$, $\epsilon = 0.62$; T2: $F_{(2, 86)} = 87.03$, $p < .001$, $\eta_p^2 = 0.669$, $\epsilon = 0.67$). However, the target velocity effect was smaller at T2.

There was no main effect of block and there were no further interactions (all $p > .05$). Supplementary Fig. 2 gives an overview of the task and time effects. Fig. 6 depicts the significant interactions.

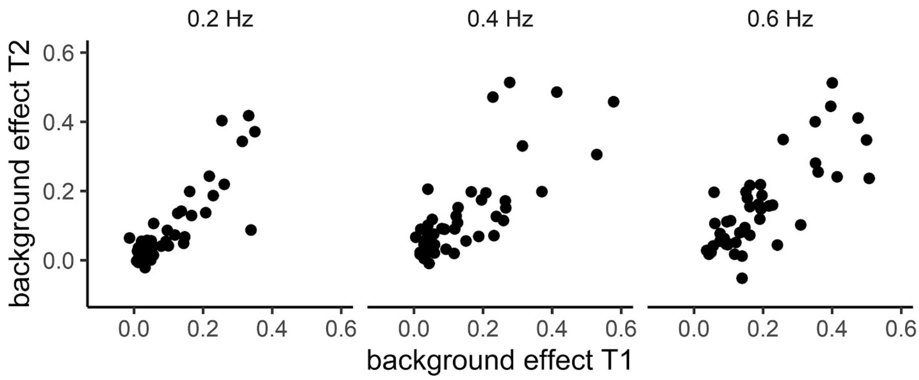

Fig. 4. Scatter plot of the background effect of velocity gain (long version) at T1 and T2 for the three target velocity conditions.

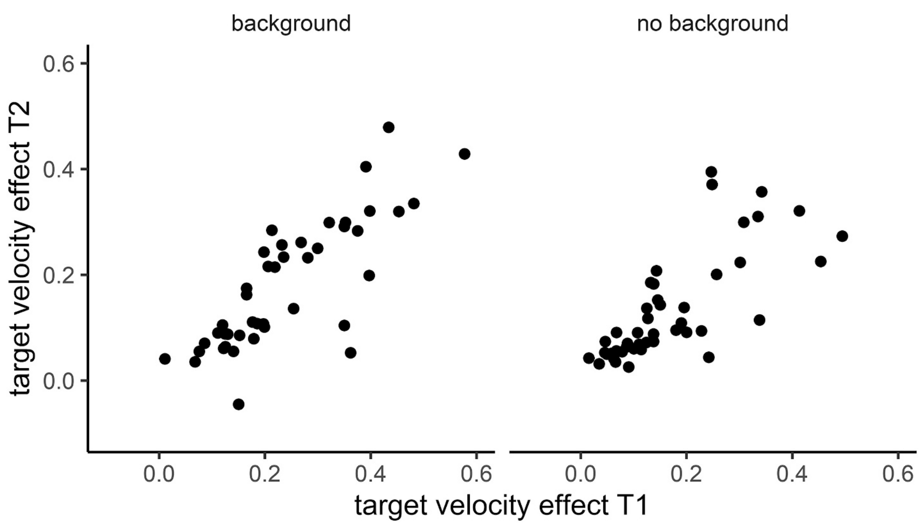

Fig. 5. Scatter plot of the target velocity effect of velocity gain (long version) at T1 and T2 for the two background conditions.

4. Discussion

The present study aimed to contribute to the controversial matter of the reliability of task effects in experimental psychology, termed the "reliability paradox" by Hedge et al. (2018). Specifically, we investigated the replicability and reliability of two well-established experimental effects in the domain of oculomotor control, viz. the detrimental effects of a stationary structured background and target velocity on smooth pursuit eye movement performance. To do so, healthy participants performed smooth pursuit tasks in a repeated-measures design in two sessions one week apart. The main results are as follows.

The presence of background distractors and higher target velocity impaired velocity gain, thereby replicating the background and target velocity effects, respectively.

Analyses of reliability revealed heterogeneous findings, with good to excellent reliabilities in the more challenging task conditions (higher target velocity and/or stationary background) and lower reliabilities in the absence of a stationary background or with lower target velocity. Importantly, the background effect revealed good to excellent reliability scores at all target velocities. Similarly, the target velocity effect reached predominantly good and excellent reliability scores.

Descriptively, the longer task version reached higher reliability than the shorter version, and split-half reliability was higher at T2 than at T1 for most task conditions and variables. However, differences were not significant and reliability scores fell in the same categories of interpretation (Cicchetti, 1994). Generally, Pearson correlations and ICCs reached converging results.

Repeated task exposure led to increased velocity gain at T2 compared to T1. However, this effect was driven by an increase in performance only in the presence of a structured background and at higher target velocities.

4.1. Reliability

The reliability indices obtained in this study were predominantly good or even excellent (Cicchetti, 1994).

For velocity gain, the primary measure of smooth pursuit performance, the study revealed excellent reliability scores in the presence of a stationary background and at higher target velocity. For slower targets and in the absence of a stationary background, however, reliability scores were lower, replicating earlier findings (Ettinger et al., 2003). The poorest reliability outcomes occurred in the putatively easiest version of the task, implying a correlation between task difficulty and reliability. As performance in those easier conditions was excellent and standard deviations were low, these results can be interpreted in terms of the reliability paradox (Hedge et al., 2018) since the sample was very homogenous in their responses in these conditions.

An interesting observation was that the decrease in performance (and increase in variance; Table 1) from the easier to the more difficult task conditions was comparably large for the target velocity and background manipulations. Strikingly, however, the improvement in reliability from the no background condition to the background condition was larger than from lowest to highest target velocity. This suggests that the higher reliability in the background condition was not merely due to

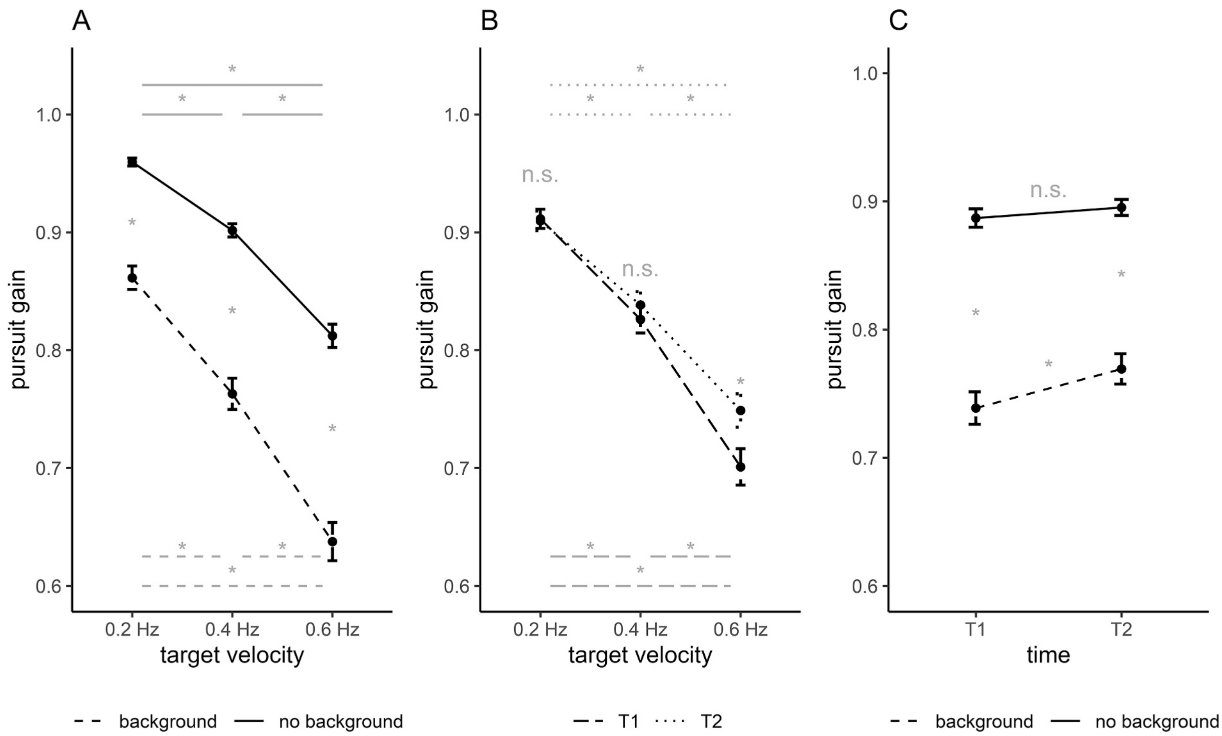

Fig. 6. Interactive effects of target velocity, background and time on pursuit velocity gain.
Legend: Two-way interaction of target velocity and background (Panel A), two-way interaction of target velocity and time (Panel B), two-way interaction of background and time (Panel C) on pursuit velocity gain. All interactions are significant at $\alpha = 0.05$. Significant *t*-tests after Bonferroni-correction are marked with an asterisk. Data are presented as mean ± standard errors.

greater variance in that condition compared to the no background condition. Instead, pursuing a target over a structured background appears to bring out inter-individual differences, possibly linked to enhanced recruitment of inhibitory processes (Lindner et al., 2001), that are highly reliable, and more so than in a high target velocity condition that yields comparable overall performance levels and variance.

For both the background and the target velocity effects, excellent reliability scores were obtained for the long version of the task for all levels of the other factor. These results, combined with the large magnitudes of the background and target velocity effect sizes at group level, are at odds with major claims of the reliability paradox. Specifically, in our study, between-subject variance was in fact so high (see Figs. 4 and 5) that excellent reliability scores could be ensured despite clear experimental effects at group level.

Earlier studies have revealed heterogeneous results for the reliability of difference measures depending on the specific task used (Hedge et al., 2018; Paap & Sawi, 2016; Soveri et al., 2018). Our results match the assumptions by Zimmerman and Williams (1998) concerning the circumstances under which difference scores can achieve good reliability. Specifically, variance and reliability were larger in the background condition than in the no background condition and at higher compared to lower target velocities (see Figs. 4 and 5, Table 1 and Supplementary Tables 1 to 3).

A potentially major conclusion from our findings is that oculomotor data might be better suited for individual difference research than outcomes from standard cognitive tasks focusing on manual-motor reaction times or error rates (Hedge et al., 2018). For instance, pursuit gain is less susceptible to common problems with reaction times or accuracy metrics such as speed-accuracy trade-offs or impurity of the measures (Draheim et al., 2019; Miller & Ulrich, 2013). Pursuit gain might therefore reflect the main function of smooth pursuit, the matching of eye velocity to target velocity, more accurately than reaction time measures reflect processes of inhibitory control, suggesting lower task impurity and higher ecological validity (Burgess et al., 2006; Miyake et al., 2000). In addition, smooth pursuit over a structured background is a behaviour shown regularly when exploring the environment, e.g., when looking at naturalistic scenes (Agtzidis et al., 2020; Startsev et al., 2019), indicative of a better match between the behaviour in the laboratory and in free viewing circumstances, compared to other inhibitory or visual tasks (Kristjánsson & Draschkow, 2021). Indeed, laboratory tasks of inhibitory control such as the Stroop task have been criticized for their lack of ecological validity (Burgess et al., 2006).

Castelhano and Henderson (2008) demonstrated that saccadic and fixational behaviour when viewing natural scenes is highly stable within participants across different stimuli. This might suggest that trial-to-trial variability in pursuit gain is smaller compared to classic inhibitory tasks, which in turn would facilitate higher reliability values as high trial variance has been identified as a major driver of low reliability in inhibition tasks (Rouder et al., 2019). Recently, de Haas and colleagues provided evidence not only for substantial interindividual differences in fixation pattern when viewing natural scenes but also for excellent test-retest and split-half reliabilities of these patterns in two independent samples (De Haas et al., 2019; Linka & de Haas, 2020), in line with the findings of the current study. However, the nature of smooth pursuit variability and its relation to reliability has to be explored in much more detail in future investigations. Still, our results suggest that robust experimental effects can indeed translate into reliable individual difference measures.

Interestingly, the reliability of the background effect declined at higher target velocities (especially when looking at the difference between the lowest and the intermediate and the lowest and highest target velocity, respectively). Conversely, reliability increased with higher target velocities for the direct performance measures. Possibly, reliability was poor for the lowest target velocity in the absence of a stationary background because of a ceiling effect of performance. At both sessions, participants were so accurate at pursuing the slow target that fluctuations between the sessions were random, resulting in low reliabilities. Higher target velocities decreased performance and simultaneously increased between-subject variance, thereby facilitating higher reliabilities. Similarly, the presentation of a stationary background led to more between-subjects variance, enabling higher reliability values. The decrease in reliability of the background effect at higher target velocities might be due to the fact that the reliabilities mainly depend on variance in the background condition. As performance was quite homogeneous at 0.2 Hz in the absence of a stationary background and consequently the baseline was very similar between participants, the difference measures mainly reflect the background condition. A similar rationale has been argued by Hedge et al. (2018) for the explanation of the different reliabilities for reaction time and accuracy data.

The good reliabilities under the more difficult task conditions are promising, considering that in everyday life smooth pursuit occurs to more complex stimuli than in the present laboratory study. Future studies could therefore turn to investigating reliability of natural viewing behaviour (Agtzidis et al., 2020).

Descriptively, the longer version of the task reached higher reliability scores than the shorter version for almost all conditions, in accordance with previous investigations (Hedge et al., 2018; Wöstmann et al., 2013). However, for most variables, confidence intervals overlapped between the two versions, which consequently does not permit the conclusion of significant superiority of the longer task. In the future, therefore, the task length should be selected according to the specific study objectives, with the longer version recommended when individual differences are the main focus.

For most variables, split-half reliability scores for T2 were descriptively higher than for T1. The observed pattern of results can be interpreted as evidence of a reduction in error variance at T2, possibly due to better familiarity with the task. However, again, in most cases the confidence intervals overlapped, allowing for no firm conclusions to be drawn.

Pearson correlations and ICCs generally led to consistent or only slightly diverging results. Future investigations might therefore choose to report only one measure, preferably ICCs, as they not only rely on relative consistency but also contain additional information on absolute agreement (McGraw & Wong, 1996).

Our results extend previous studies on smooth pursuit reliability due to differences in the methodological approach. Firstly, in contrast to Bargary et al. (2017) and Ettinger et al. (2003), we did not use a triangular, constant-velocity target but a target following a sinusoidal velocity pattern, as is common in clinical smooth pursuit research (Barnes, 2008; Levy et al., 2010; Meyhöfer et al., 2015; Nkam et al., 2010; Ohlendorf et al., 2007). Secondly, we reported reliability scores separately for each of the target velocity and background conditions. This approach proved to be very valuable as our results suggest that reliability scores widely differed between different task configurations. Lastly, the present study was the first to directly assess the reliability of the background and target velocity effects, which had not yet been investigated despite these effects being well replicated experimental group-level findings in the pursuit literature.

The observed pattern of results speaks for the general stability and likely trait nature of eye movement performance, which has previously been shown for saccades using latent-state-trait modelling (Meyhöfer et al., 2016). Similar modelling approaches should also be adopted in the future to delineate the contribution of stable person effects and situational influences on smooth pursuit performance.

4.2. Task and time effects

4.2.1. Background

For all levels of target velocity, we found strong background effects, indicative of worse performance in the presence of background stimuli, in line with earlier findings (Barnes & Crombie, 1985; Collewijn & Tamminga, 1984; Hutton et al., 2000; Meyhöfer et al., 2019). This pattern of results can be explained by background induced optokinetic drive that slowed eye movements (Barnes, 2008; Lawden et al., 1995). The highly structured background stimuli provided a strong signal to the optokinetic reflex, interfering with the smooth pursuit tracking of the target. As a consequence, participants had to actively engage cognitive resources to selectively enhance processing of the target stimulus (Barnes & Crombie, 1985) or inhibit background processing in order to perform stable smooth pursuit. The latter might be achieved by attenuating sensitivity to global motion signals in the opposite direction of the pursuit target, which also corresponds to the direction of self-induced retinal motion (Lindner et al., 2001).

Lower pursuit performance in the presence of a structured background might depend on gradual attention shifts between target and background stimuli (Kerzel et al., 2008). These shifts may be modulated by feature similarity between target and background distractors (Störmer et al., 2011). For example, it has been shown that increased attention to the target due to a target-related secondary task may even improve smooth pursuit performance despite a general increase in demands due to the secondary task (Stubbs et al., 2018). Future studies should further explore the role of attention on target and distractor or background stimuli. In this context, partial processing of the background may even be beneficial in some cases, as it can also provide valuable information about future events such as target trajectory (Eggert et al., 2009; Ladda et al., 2007).

Results from a lesion study suggest that inhibition of irrelevant background distractors might depend on parietal cortex and frontoparietal white-matter connections (Lawden et al., 1995). Interestingly, the posterior parietal cortex has also been associated with divided attention between a SPEM target stimulus and an additional auditory target (Baumann & Greenlee, 2009) as well as the dissociation between the focus of attention and gaze (Ohlendorf et al., 2007). The specific role of this area in the context of smooth pursuit against a structured background might concern motion processing relative to a frame of reference (Ohlendorf et al., 2010) or attentional control (Baumann & Greenlee, 2009). However, the studies discussed above reported different peak locations in the parietal cortex, so further research is needed concerning the exact functions of the parietal cortex and its subregions during this complex sensorimotor process. In addition, it has been suggested that intact pathways of the basal ganglia are necessary for successful suppression of irrelevant stimuli during smooth pursuit (Henderson et al., 2011).

We also detected significant background × target velocity interactions, indicating that the adverse effects of a stationary background were particularly pronounced at higher target velocities, in line with previous findings (Howard & Marton, 1992; Hutton et al., 2000; Meyhöfer et al., 2019). Thus, if the pursuit system is challenged – by fast targets, structured backgrounds or the combination of both – performance deteriorates substantially. These effects might indicate that both factors rely on the same underlying process, for example spatial attention, which is important for both the matching of eye to target velocity and the inhibition of background processing. Interestingly, it has been shown that the focus of attention is modulated by target velocity (Van Donkelaar & Drew, 2002). The faster the target moves, the further ahead of the target attention shifts. If the target is no longer at the center of attention at high velocities, pursuit gain might be reduced as the matching of eye to target velocity is impeded. However, the assumed asymmetrical distribution of attention biased ahead of the target is not supported by all lines of evidence (Souto & Kerzel, 2021). The processing benefit at locations ahead of the target has been found in reaction time paradigms and with frequency-tagged steady-state visual evoked potentials (Chen et al., 2017; Van Donkelaar, 1999; Van Donkelaar & Drew, 2002) but not in perceptual discriminations tasks (Lovejoy et al., 2009; Watamaniuk & Heinen, 2015). It is debated whether these effects in fact represent spatial attention or rather visual processes such as the suppression of visual signals opposite of target direction (Souto & Kerzel, 2021). Therefore, more research is needed to investigate the relationship between smooth pursuit and attention and how they relate to target velocity and the background the target is presented on.

Another important aspect to consider in explaining the background × target velocity interaction is velocity perception during pursuit, which has been shown to depend on the background the target is presented on (Raymond et al., 1984). However, this has been probed relative to a reference stimulus presented before the target and should be investigated in more detail.

4.2.2. Target velocity

Additionally, we aimed to replicate the effects of target velocity on pursuit performance. In line with previous research and our hypotheses, performance was affected by target velocity (Collewijn & Tamminga, 1984; Lisberger et al., 1981; Meyhöfer et al., 2019), with pursuit gain being lower at higher target velocity. Effect size measures indicated that target velocity accounted for a substantial part of pursuit gain variance. The manipulation of target velocity can therefore be regarded as a valuable tool to control overall task demands. Our results add to the literature by highlighting the importance of using more than a single level of target velocity since some task manipulations only become evident at a certain level of target velocity (e.g. background × target velocity interactions). As a consequence, studies that employ only a single level of target velocity might be limited in their explanatory power.

Although the effects of manipulating target velocity have been studied extensively, to our knowledge, target velocity effects as the difference in performance between two levels of target velocity are not typically reported. Therefore, here, we demonstrate the feasibility of this approach. Individual target velocity effects may prove to be particularly useful for later use in individual differences research to better understand between-subjects variance.

4.2.3. Block

In order to examine the course of performance over time, data were analyzed for within-session performance changes from the first to the second half of the experiment. However, no differences between the two blocks could be observed. This absence of effects suggests that performance is relatively stable over a duration of several minutes. This could be due to the fact that at the beginning of each session, participants had the opportunity to familiarize themselves with the task through a practice session. This way, understanding of instructions and familiarity with the study setup was ensured and did not have to be acquired during task performance, which may have led to differences between blocks.

4.2.4. Time

We found a significant main effect of time for pursuit gain, indicative of improvement of performance from the first (T1) to the second (T2) session day over a period of one week.

Interactions between time and background conditions as well as time and target velocity revealed that these effects only occurred in the more challenging conditions, namely in the presence of a stationary background and at higher target velocities. This pattern indicates that repetition or practice effects might depend on baseline performance, which is related to task difficulty. Performance in the easier task conditions may already be near ceiling at T1, with gain scores close to 1. Thus, the performance level could not improve any further. Performance improvements in the more difficult task conditions might be due to learning processes, which is consistent with previous results from healthy and neurological samples (Eibenberger et al., 2012; Kerkhoff et al., 2013). If the task is to be used in study designs with multiple assessments, the issue of performance changes due to multiple exposure to the task needs to be considered. Importantly, however, the background effect proved to be highly robust and could be observed in both sessions with large effect sizes. The same pattern of results was observed for the target velocity effect. Thus, time only influenced the size of the background or target velocity effect but not the effect as such.

Future investigations should further explore whether or at what point performance stabilizes over time by employing multiple assessments over several days.

We aimed to control the influence of previous exposure to the experimental stimulus sets by introducing a brief practice block at the beginning of each session. However, familiarity with the assessment setting might still influence performance, which is problematic especially in repeated-measures designs, where participants undergo experimental manipulations at different time points. In order to further reduce this influence in the future, the practice block could be presented for a longer time or participants could get familiarized with the task during an initial baseline measurement. The latter may also help to examine possible baseline effects on learning as stabilization might occur earlier or later dependent on initial performance levels.

4.3. Limitations

The study was limited in the number of sessions. In order to explore the effects of repeated exposure to the same task in more detail, more measurement sessions would have been informative. Also, latent state trait modelling (Geiser et al., 2015) of smooth pursuit data is called for to formally estimate trait and state components of variance as well as measurement error.

4.4. Conclusion

In this study, we have presented reliability scores and task effects for a smooth pursuit paradigm employing three target velocities and a background vs. no background condition in two sessions one week apart in a sample of 45 participants. Presenting background distractors and increasing target velocity impaired smooth pursuit performance consistent with previous investigations.

Background as well as target velocity effects were robustly replicated in both sessions. However, practice effects between the two sessions were observed in the more difficult task conditions and should be considered in future investigations using longitudinal designs.

Reliability results were good, especially at higher target velocities and/or in the presence of a stationary background. Reliability scores were largely in line with previous literature but also revealed interesting novel insights, especially concerning the role of task difficulty. Background as well as target velocity effects proved to be highly reliable. These findings demonstrate, in contrast to some other cognitive tasks (Hedge et al., 2018), that a task can produce both robust experimental effects and reliable individual difference outcomes.

Declaration of competing interest

None.

Acknowledgements

The PhD project of Rebekka Schröder is funded by the German Academic Scholarship Foundation.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.actpsy.2021.103364.

References

Agtzidis, I., Meyhöfer, I., Dorr, M., & Lencer, R. (2020). Following Forrest Gump: Smooth pursuit related brain activation during free movie viewing. *NeuroImage, 216*, Article 116491. https://doi.org/10.1016/j.neuroimage.2019.116491

Asendorpf, J. B., Conner, M., Fruyt, F.d., Houwer, J.d., Denissen, J. J. A., Fiedler, K., … Wicherts, J. M. (2013). Recommendations for increasing replicability in psychology. *European Journal of Personality, 27*(2), 108–119. https://doi.org/10.1002/per.1919

Bargary, G., Bosten, J. M., Goodbourn, P. T., Lawrance-Owen, A. J., Hogg, R. E., & Mollon, J. D. (2017). Individual differences in human eye movements: An oculomotor signature? *Vision Research, 141*, 157–169. https://doi.org/10.1016/j.visres.2017.03.001

Barnes, G. R. (2008). Cognitive processes involved in smooth pursuit eye movements. *Brain and Cognition, 68*(3), 309–326. https://doi.org/10.1016/j.bandc.2008.08.020

Barnes, G. R., & Asselman, P. T. (1991). The mechanism of prediction in human smooth pursuit eye movements. *The Journal of Physiology, 439*, 439–461. https://doi.org/10.1113/jphysiol.1991.sp018675

Barnes, G. R., & Crombie, J. W. (1985). The interaction of conflicting retinal motion stimuli in oculomotor control. *Experimental Brain Research, 59*(3), 548–558. https://doi.org/10.1007/BF00261346

Baumann, O., & Greenlee, M. W. (2009). Effects of attention to auditory motion on cortical activations during smooth pursuit eye tracking. *PLoS One, 4*(9), Article e7110. https://doi.org/10.1371/journal.pone.0007110

Buizza, A., & Schmid, R. (1986). Velocity characteristics of smooth pursuit eye movements to different patterns of target motion. *Experimental Brain Research, 63*(2), 395–401. https://doi.org/10.1007/BF00236858

Burgess, P. W., Alderman, N., Forbes, C., Costello, A., Coates, L. M.-A., Dawson, D. R., … Channon, S. (2006). The case for the development and use of "ecologically valid" measures of executive function in experimental and clinical neuropsychology. *Journal of the International Neuropsychological Society : JINS, 12*(2), 194–209. https://doi.org/10.1017/S1355617706060310

Calkins, M. E., Iacono, W. G., & Curtis, C. E. (2003). Smooth pursuit and antisaccade performance evidence trait stability in schizophrenia patients and their relatives. *International Journal of Psychophysiology : Official Journal of the International Organization of Psychophysiology, 49*(2), 139–146. https://doi.org/10.1016/S0167-8760(03)00101-6

Castelhano, M. S., & Henderson, J. M. (2008). Stable individual differences across images in human saccadic eye movements. *Canadian Journal of Experimental Psychology = Revue Canadienne De Psychologie Experimentale, 62*(1), 1–14. https://doi.org/10.1037/1196-1961.62.1.1

Censor, N., Sagi, D., & Cohen, L. G. (2012). Common mechanisms of human perceptual and motor learning. *Nature Reviews Neuroscience, 13*(9), 658–664. https://doi.org/10.1038/nrn3315

Chen, J., Valsecchi, M., & Gegenfurtner, K. R. (2017). Attention is allocated closely ahead of the target during smooth pursuit eye movements: Evidence from EEG frequency tagging. *Neuropsychologia, 102*, 206–216. https://doi.org/10.1016/j.neuropsychologia.2017.06.024

Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment, 6*(4), 284–290. https://doi.org/10.1037/1040-3590.6.4.284

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*, 217–250. https://doi.org/10.1113/jphysiol.1984.sp015242

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 of the United States of America, 116*(24), 11687–11692. https://doi.org/10.1073/pnas.1820553116

Draheim, C., Mashburn, C. A., Martin, J. D., & Engle, R. W. (2019). Reaction time in differential and developmental research: A review and commentary on the problems and alternatives. *Psychological Bulletin, 145*(5), 508–535. https://doi.org/10.1037/bul0000192

Eggert, T., Ladda, J., & Straube, A. (2009). Inferring the future target trajectory from visual context: Is visual background structure used for anticipatory smooth pursuit? *Experimental Brain Research, 196*(2), 205–215. https://doi.org/10.1007/s00221-009-1840-3

Eibenberger, K., Ring, M., & Haslwanter, T. (2012). Sustained effects for training of smooth pursuit plasticity. *Experimental Brain Research, 218*(1), 81–89. https://doi.org/10.1007/s00221-012-3009-8

Ettinger, U., Kumari, V., Crawford, T. J., Davis, R. E., Sharma, T., & Corr, P. J. (2003). Reliability of smooth pursuit, fixation, and saccadic eye movements. *Psychophysiology, 40*(4), 620–628. https://doi.org/10.1111/1469-8986.00063

Ettinger, U., Meyhöfer, I., Steffens, M., Wagner, M., & Koutsouleris, N. (2014). Genetics, cognition, and neurobiology of schizotypal personality: A review of the overlap with schizophrenia. *Frontiers in Psychiatry, 5*, 18. https://doi.org/10.3389/fpsyt.2014.00018

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*(2), 175–191. https://doi.org/10.3758/bf03193146

Friedman, N. P., & Miyake, A. (2004). The relations among inhibition and interference control functions: A latent-variable analysis. *Journal of Experimental Psychology. General, 133*(1), 101–135. https://doi.org/10.1037/0096-3445.133.1.101

Gamer, M., Lemon, J., & Puspendra Singh, I. F. (2019). irr: Various coefficients of interrater reliability and agreement. Retrieved from https://CRAN.R-project.org/package=irr.

Geiser, C., Litson, K., Bishop, J., Keller, B. T., Burns, G. L., Servera, M., & Shiffman, S. (2015). Analyzing person, situation and person × situation interaction effects: Latent state-trait models for the combination of random and fixed situations. *Psychological Methods, 20*(2), 165–192. https://doi.org/10.1037/met0000026

Haraldsson, H. M., Ettinger, U., Magnusdottir, B. B., Sigmundsson, T., Sigurdsson, E., & Petursson, H. (2008). Eye movement deficits in schizophrenia: Investigation of a genetically homogenous Icelandic sample. *European Archives of Psychiatry and Clinical Neuroscience, 258*(6), 373–383. https://doi.org/10.1007/s00406-008-0806-y

Haslam, N., McGrath, M. J., Viechtbauer, W., & Kuppens, P. (2020). Dimensions over categories: A meta-analysis of taxometric research. *Psychological Medicine*, 1–15. https://doi.org/10.1017/S003329172000183X

Hedge, C., Powell, G., & Sumner, P. (2018). The reliability paradox: Why robust cognitive tasks do not produce reliable individual differences. *Behavior Research Methods, 50*(3), 1166–1186. https://doi.org/10.3758/s13428-017-0935-1

Henderson, T., Georgiou-Karistianis, N., White, O., Millist, L., Williams, D. R., Churchyard, A., & Fielding, J. (2011). Inhibitory control during smooth pursuit in Parkinson's disease and Huntington's disease. *Movement Disorders : Official Journal of the Movement Disorder Society, 26*(10), 1893–1899. https://doi.org/10.1002/mds.23757

Holzman, P. S., Proctor, L. R., & Hughes, D. W. (1973). Eye-tracking patterns in schizophrenia. *Science (New York, N.Y.), 181*(4095), 179–181. https://doi.org/10.1126/science.181.4095.179

Howard, I. P., & Marton, C. (1992). Visual pursuit over textured backgrounds in different depth planes. *Experimental Brain Research, 90*(3), 625–629. https://doi.org/10.1007/BF00230947

Hutton, S. B., Crawford, T. J., Kennard, C., Barnes, T. R., & Joyce, E. M. (2000). Smooth pursuit eye tracking over a structured background in first-episode schizophrenic patients. *European Archives of Psychiatry and Clinical Neuroscience, 250*(5), 221–225. https://doi.org/10.1007/s004060070011

Kassambra, A. (2020). Rstatix: Pipe-friendly framework for basic statistical tests. Retrieved from https://CRAN.R-project.org/package=rstatix.

Kaufman, S. R., & Abel, L. A. (1986). The effects of distraction on smooth pursuit in normal subjects. *Acta Oto-Laryngologica, 102*(1–2), 57–64. https://doi.org/10.3109/00016488609108647

Kerkhoff, G., Reinhart, S., Ziegler, W., Artinger, F., Marquardt, C., & Keller, I. (2013). Smooth pursuit eye movement training promotes recovery from auditory and visual neglect: A randomized controlled study. *Neurorehabilitation and Neural Repair, 27*(9), 789–798. https://doi.org/10.1177/1545968313491012

Kerzel, D., Souto, D., & Ziegler, N. E. (2008). Effects of attention shifts to stationary objects during steady-state smooth pursuit eye movements. *Vision Research, 48*(7), 958–969. https://doi.org/10.1016/j.visres.2008.01.015

Kristjánsson, Á., & Draschkow, D. (2021). Keeping it real: Looking beyond capacity limits in visual cognition. *Attention, Perception, & Psychophysics, 83*, 1375–1390. https://doi.org/10.3758/s13414-021-02256-7

Ladda, J., Eggert, T., Glasauer, S., & Straube, A. (2007). Velocity scaling of cue-induced smooth pursuit acceleration obeys constraints of natural motion. *Experimental Brain Research, 182*(3), 343–356. https://doi.org/10.1007/s00221-007-0988-y

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for t-tests and ANOVAs. *Frontiers in Psychology, 4*, 863. https://doi.org/10.3389/fpsyg.2013.00863

Lawden, M. C., Bagelmann, H., Crawford, T. J., Matthews, T. D., & Kennard, C. (1995). An effect of structured backgrounds on smooth pursuit eye movements in patients with cerebral lesions. *Brain: A Journal of Neurology, 118*(Pt 1), 37–48. https://doi.org/10.1093/brain/118.1.37

Lawrence, M. A. (2016). ez: Easy analysis and visualization of factorial experiments. Retrieved from https://CRAN.R-project.org/package=ez.

Leigh, R. J., & Zee, D. S. (2015). *The neurology of eye movements* (5 ed.). New York: Oxford University Press. https://doi.org/10.1093/med/9780199969289.001.0001

Lencer, R., & Trillenberg, P. (2008). Neurophysiology and neuroanatomy of smooth pursuit in humans. *Brain and Cognition, 68*(3), 219–228. https://doi.org/10.1016/j.bandc.2008.08.013

Lenzenweger, M. F., & O'Driscoll, G. A. (2006). Smooth pursuit eye movement and schizotypy in the community. *Journal of Abnormal Psychology, 115*(4), 779–786. https://doi.org/10.1037/0021-843X.115.4.779

Levy, D. L., Sereno, A. B., Gooding, D. C., & O'Driscoll, G. A. (2010). Eye tracking dysfunction in schizophrenia: Characterization and pathophysiology. *Current Topics in Behavioral Neurosciences, 4*, 311–347. https://doi.org/10.1007/7854_2010_60

Lindner, A., Schwarz, U., & Ilg, U. J. (2001). Cancellation of self-induced retinal image motion during smooth pursuit eye movements. *Vision Research, 41*(13), 1685–1694. https://doi.org/10.1016/S0042-6989(01)00050-5

Linka, M., & de Haas, B. (2020). Osieshort: A small stimulus set can reliably estimate individual differences in semantic salience. *Journal of Vision, 20*(9), 13. https://doi.org/10.1167/jov.20.9.13

Lisberger, S. G. (2015). Visual guidance of smooth pursuit eye movements. *Annual Review of Vision Science, 1*, 447–468. https://doi.org/10.1146/annurev-vision-082114-035349

Lisberger, S. G., Evinger, C., Johanson, G. W., & Fuchs, A. F. (1981). Relationship between eye acceleration and retinal image velocity during foveal smooth pursuit in man and monkey. *Journal of Neurophysiology, 46*(2), 229–249. https://doi.org/10.1152/jn.1981.46.2.229

Lisberger, S. G., Morris, E. J., & Tychsen, L. (1987). Visual motion processing and sensory-motor integration for smooth pursuit eye movements. *Annual Review of Neuroscience, 10*, 97–129. https://doi.org/10.1146/annurev.ne.10.030187.000525

Lovejoy, L. P., Fowler, G. A., & Krauzlis, R. J. (2009). Spatial allocation of attention during smooth pursuit eye movements. *Vision Research, 49*(10), 1275–1285. https://doi.org/10.1016/j.visres.2009.01.011

McGraw, K. O., & Wong, S. P. (1996). Forming inferences about some intraclass correlation coefficients. *Psychological Methods, 1*(1), 30–46. https://doi.org/10.1037/1082-989X.1.1.30

Meyhöfer, I., Bertsch, K., Esser, M., & Ettinger, U. (2016). Variance in saccadic eye movements reflects stable traits. *Psychophysiology, 53*(4), 566–578. https://doi.org/10.1111/psyp.12592

Meyhöfer, I., Kasparbauer, A.-M., Steffens, M., & Ettinger, U. (2019). Effects of nicotine on smooth pursuit eye movements in healthy non-smokers. *Psychopharmacology, 236*, 2259–2271. https://doi.org/10.1007/s00213-019-05223-1
Meyhöfer, I., Steffens, M., Kasparbauer, A.-M., Grant, P., Weber, B., & Ettinger, U. (2015). Neural mechanisms of smooth pursuit eye movements in schizotypy. *Human Brain Mapping, 36*(1), 340–353. https://doi.org/10.1002/hbm.22632
Miller, J., & Ulrich, R. (2013). Mental chronometry and individual differences: Modeling reliabilities and correlations of reaction time means and effect sizes. *Psychonomic Bulletin & Review, 20*(5), 819–858. https://doi.org/10.3758/s13423-013-0404-5
Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: A latent variable analysis. *Cognitive Psychology, 41*(1), 49–100. https://doi.org/10.1006/cogp.1999.0734
Mohrmann, H., & Thier, P. (1995). The influence of structured visual backgrounds on smooth-pursuit initiation, steady-state pursuit and smooth-pursuit termination. *Biological Cybernetics, 73*(1), 83–93. https://doi.org/10.1007/BF00199058
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. https://doi.org/10.1007/PL00005723
Nkam, I., Bocca, M.-L., Denise, P., Paoletti, X., Dollfus, S., Levillain, D., & Thibaut, F. (2010). Impaired smooth pursuit in schizophrenia results from prediction impairment only. *Biological Psychiatry, 67*(10), 992–997. https://doi.org/10.1016/j.biopsych.2009.11.029
Ohlendorf, S., Kimmig, H., Glauche, V., & Haller, S. (2007). Gaze pursuit, 'attention pursuit' and their effects on cortical activations. *The European Journal of Neuroscience, 26*(7), 2096–2108. https://doi.org/10.1111/j.1460-9568.2007.05824.x
Ohlendorf, S., Sprenger, A., Speck, O., Glauche, V., Haller, S., & Kimmig, H. (2010). Visual motion, eye motion, and relative motion: A parametric fMRI study of functional specializations of smooth pursuit eye movement network areas. *Journal of Vision, 10*(14), 21. https://doi.org/10.1167/10.14.21
Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia, 9*(1), 97–113. https://doi.org/10.1016/0028-3932(71)90067-4
Orban de Xivry, J.-J., & Lefèvre, P. (2007). Saccades and pursuit: Two outcomes of a single sensorimotor process. The Journal of Physiology, 584(Pt 1), 11–23. doi:https://doi.org/10.1113/jphysiol.2007.139881.
Ostry, D. J., & Gribble, P. L. (2016). Sensory plasticity in human motor learning. *Trends in Neurosciences, 39*(2), 114–123. https://doi.org/10.1016/j.tins.2015.12.006
Paap, K. R., & Sawi, O. (2016). The role of test-retest reliability in measuring individual and group differences in executive functioning. *Journal of Neuroscience Methods, 274*, 81–93. https://doi.org/10.1016/j.jneumeth.2016.10.002
Raymond, J. E., Shapiro, K. L., & Rose, D. J. (1984). Optokinetic backgrounds affect perceived velocity during ocular tracking. *Perception & Psychophysics, 36*(3), 221–224. https://doi.org/10.3758/BF03206362
Revelle, W. (2019). *psych: Procedures for Psychological, Psychometric, and Personality Research*. Evanston, Illinois. Retrieved from https://CRAN.R-project.org/package=psych.
Rouder, J., Kumar, A., & Haaf, J. M. (2019). *Why most studies of individual differences with inhibition tasks are bound to fail*. PsyArXiv. Advance online publication. https://doi.org/10.31234/osf.io/3cjr5.
Roy-Byrne, P., Radant, A., Wingerson, D., & Cowley, D. S. (1995). Human oculomotor function: Reliability and diurnal variation. *Biological Psychiatry, 38*(2), 92–97. https://doi.org/10.1016/0006-3223(94)00225-R
Smyrnis, N. (2008). Metric issues in the study of eye movements in psychiatry. *Brain and Cognition, 68*(3), 341–358. https://doi.org/10.1016/j.bandc.2008.08.022
Smyrnis, N., Evdokimidis, I., Mantas, A., Kattoulas, E., Stefanis, N. C., Constantinidis, T. S., … Stefanis, C. N. (2007). Smooth pursuit eye movements in 1,087 men: Effects of schizotypy, anxiety, and depression. *Experimental Brain Research, 179*(3), 397–408. https://doi.org/10.1007/s00221-006-0797-8
Souto, D., & Kerzel, D. (2021). Visual selective attention and the control of tracking eye movements: A critical review. *Journal of Neurophysiology*. https://doi.org/10.1152/jn.00145.2019. Advance online publication.
Soveri, A., Lehtonen, M., Karlsson, L. C., Lukasik, K., Antfolk, J., & Laine, M. (2018). Test-retest reliability of five frequently used executive tasks in healthy adults. *Applied Neuropsychology. Adult, 25*(2), 155–165. https://doi.org/10.1080/23279095.2016.1263795
Spering, M., Gegenfurtner, K. R., & Kerzel, D. (2006). Distractor interference during smooth pursuit eye movements. *Journal of Experimental Psychology. Human Perception and Performance, 32*(5), 1136–1154. https://doi.org/10.1037/0096-1523.32.5.1136
Spering, M., & Montagnini, A. (2011). Do we track what we see? Common versus independent processing for motion perception and smooth pursuit eye movements: A review. *Vision Research, 51*(8), 836–852. https://doi.org/10.1016/j.visres.2010.10.017
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. https://doi.org/10.1167/19.14.10
Störmer, V. S., Li, S.-C., Heekeren, H. R., & Lindenberger, U. (2011). Feature-based interference from unattended visual field during attentional tracking in younger and older adults. *Journal of Vision, 11*(2). https://doi.org/10.1167/11.2.1
Stubbs, J. L., Corrow, S. L., Kiang, B., Panenka, W. J., & Barton, J. J. S. (2018). The effects of enhanced attention and working memory on smooth pursuit eye movement. *Experimental Brain Research, 236*(2), 485–495. https://doi.org/10.1007/s00221-017-5146-6
Van Donkelaar, P. (1999). Spatiotemporal modulation of attention during smooth pursuit eye movements. *Neuroreport, 10*(12), 2523–2526. https://doi.org/10.1097/00001756-199908200-00016
Van Donkelaar, P., & Drew, A. S. (2002). The allocation of attention during smooth pursuit eye movements. In *, Vol. 140. Progress in brain research. The brain's eye: Neurobiological and clinical aspects of oculomotor research* (pp. 267–277). Elsevier. https://doi.org/10.1016/S0079-6123(02)40056-8.
Versino, M., Castelnovo, G., Bergamaschi, R., Romani, A., Beltrami, G., Zambarbieri, D., & Cosi, V. (1993). Quantitative evaluation of saccadic and smooth pursuit eye movements. Is it reliable? *Investigative Ophthalmology & Visual Science, 34*(5), 1702–1709.
Walter, S. D., Eliasziw, M., & Donner, A. (1998). Sample size and optimal designs for reliability studies. *Statistics in Medicine, 17*(1), 101–110. https://doi.org/10.1002/(SICI)1097-0258(19980115)17:1<101::AID-SIM727>3.0.CO;2-E
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. https://doi.org/10.1167/15.9.9
Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., … Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software, 4*(43), 1686. https://doi.org/10.21105/joss.01686
Wöstmann, N. M., Aichert, D. S., Costa, A., Rubia, K., Möller, H.-J., & Ettinger, U. (2013). Reliability and plasticity of response inhibition and interference control. *Brain and Cognition, 81*(1), 82–94. https://doi.org/10.1016/j.bandc.2012.09.010
Zimmerman, D. W., & Williams, R. H. (1998). Reliability of gain scores under realistic assumptions about properties of pre-test and post-test scores. *British Journal of Mathematical and Statistical Psychology, 51*(2), 343–351. https://doi.org/10.1111/j.2044-8317.1998.tb00685.x