



Regulation of eye movements and pupil size in natural scenes[☆]

Alexander Hirsch^a, Aenne Briellmann^b, Niloufar Tabandeh^a, Manuel Spitschan^{a,c}^{*}

^a Max-Planck-Institute for Biological Cybernetics, Max-Planck-Ring 8, Tübingen, 72076, Germany

^b Liverpool Hope University, Department of Psychology, Hope Park, Taggart Avenue, Liverpool, L16 9JD, England, UK

^c Technical University of Munich, School of Medicine and Health, Chronobiology & Health, Am Olympiapark 11, München, 80809, Germany

ARTICLE INFO

Dataset link: <https://doi.org/10.6084/m9.figshare.30518291>, <https://doi.org/10.6084/m9.figshare.30520733>, <https://doi.org/10.6084/m9.figshare.30520892>, <https://doi.org/10.6084/m9.figshare.30520829>, <https://doi.org/10.6084/m9.figshare.30520925>, <https://doi.org/10.6084/m9.figshare.30520979>, https://github.com/tsnlab/HirschEtAl_VisRes_2026, <https://doi.org/10.5281/zenodo.19434785>

Keywords:

Eye tracking
Real-world
Scenes
Naturalistic task
Deception

ABSTRACT

The human visual environment is complex in its spatial, temporal, and spectral properties and, consequently, in the way it activates retinal photoreceptors. While laboratory studies have advanced our understanding of eye movements, how visual behavior unfolds in naturalistic settings remains less well characterized. In this study, we recorded eye movements and pupil size while participants viewed indoor and outdoor scenes under naturalistic conditions while remaining unaware of the eye-movement recordings. During the initial phase of viewing, saccade frequency, amplitude, and peak velocity were higher indoors and were accompanied by larger pupil size. Saccade metrics converged after this early exploration phase, whereas pupil size remained consistently larger indoors. Over the full viewing period, indoor and outdoor conditions exhibited distinct correlations between eye-movement metrics and subjective scene ratings. By minimizing task constraints and participant awareness of eye tracking while maintaining a controlled workflow, this study provides a dataset that reflects natural viewing behavior with reduced confounds. These findings suggest that eye-movement dynamics are particularly sensitive during initial scene exploration, whereas pupil size reflects persistent environmental differences. The present dataset offers a useful reference point for future research on eye movements in real-world contexts.

1. Introduction

Eye-movement control has been a topic of scientific discussion for decades. Classical eye-tracking studies demonstrated that visual exploration is strongly influenced by experimental context and task demands, as elegantly shown in pioneering work in the field (Buswell, 1935; Yarbus, 1967). This dependence of outcome measures on tasks and instructions is also apparent when more recent eye-movement datasets are compared (Henderson & Hollingworth, 1999; Zwierko et al., 2023). Since then, a large body of research has built on these foundational observations, mostly within highly controlled laboratory environments. Such studies have yielded valuable insights into gaze behavior, but their ecological validity remains limited. Laboratory-based eye tracking typically constrains participants with fixed head positions, artificial lighting, and predefined tasks. These restrictions may alter visual exploration patterns substantially, making it difficult to generalize findings to real-world behavior (Foulsham et al., 2011). Moreover, participants are usually aware of the monitoring process, which can further influence their gaze. More recent work has moved toward more naturalistic paradigms, for example by tracking eye movements during daily activities such as driving, sports, walking through a

supermarket, making a sandwich, navigating through a building, or interacting with complex environments (Kapitaniak et al., 2015; Hayhoe et al., 2012; Land, 2009; Gidlöf et al., 2017; Kredel et al., 2023; Drewes et al., 2021; Evans et al., 2012). While these approaches have greater ecological validity, they often remain task-driven and still suffer from methodological confounding factors that can shape gaze behavior in subtle ways.

In contrast, free visual exploration in real-world environments is shaped by complex interactions between the observer and the dynamic properties of the scene. One fundamental physiological mechanism influencing visual behavior is the pupillary light reflex (PLR): pupil size adjusts automatically to changes in luminance, constricting under bright light and dilating in dimmer environments, but is also affected by cognitive factors (DiCriscio et al., 2018). Additionally, while rapid pupil size adaptations are dependent on general photopic illumination, a delayed pupil response was found to be induced by intrinsic photosensitive retinal ganglion cells (ipRGC) (Lucas et al., 1999; Lucas et al., 2001), containing the photopigment melanopsin which is primarily activated at around 480 nm wavelength. The (PLR) helps regulate

[☆] This article is part of a Special issue entitled: 'Scene Processing' published in Vision Research.

^{*} Corresponding author at: Max-Planck-Institute for Biological Cybernetics, Max-Planck-Ring 8, Tübingen, 72076, Germany.

E-mail addresses: manuel.spitschan@tum.de, manuel.spitschan@tuebingen.mpg.de (M. Spitschan).

<https://doi.org/10.1016/j.visres.2026.108801>

Received 20 August 2024; Received in revised form 26 February 2026; Accepted 2 March 2026

Available online 6 April 2026

0042-6989/© 2026 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

retinal illumination and provides a sensitive index of environmental light exposure during natural viewing.

Beyond environmental factors, task structure remains a central consideration in naturalistic designs, since the direction of high-acuity foveal vision is motivated by behavioral goals (Tatler et al., 2011). Without such a goal or task, the participant is likely to generate their own intrinsic motivations and tasks, often subconscious in origin and impossible to predict for the investigator (Kay et al., 2023). An eye-tracking approach, with the ambition of naturalistic conditions, must therefore decide carefully on this matter. In recent studies that have not imposed specific tasks on their participants, the subjects were viewing specific scenes in a more artificial context (e.g., art works in a museum) (Fontoura & Menu, 2021; Stein et al., 2022). Other forms of naturalistic eye tracking involved looking at pictures of naturalistic scenes in the laboratory (Scott et al., 2020).

Another important consideration for naturalistic eye-tracking approaches is participant awareness of monitoring. Studies in this field often do not attempt to conceal the specific oculomotor measures being recorded (e.g., Drewes et al., 2021; Foulsham et al., 2011; blinding or deception are not mentioned in the eye tracking reporting guidelines by Holmqvist et al. (2022) & Dunn et al. (2023)). Such awareness may itself influence gaze behaviour, because participants can become conscious of their own gaze patterns (Risko & Kingstone, 2011; Nasiopoulos et al., 2015) and of their potential influence on the integrity of the results. For this reason, studies aiming to approximate natural viewing may benefit from minimizing participants' awareness of the specific eye-movement measures of interest, even though complete concealment is difficult when wearable eye-tracking equipment is used.

To the best of our knowledge, no systematic examination of how eye movement behavior unfolds in real-world settings without explicit task constraints has been performed yet, despite rapid advances in mobile eye tracking technology. In particular, little is known about how gaze dynamics differ between indoor and outdoor environments; likewise, evidence remains limited on how they behave during the initial phase of scene exploration, when the visual system rapidly extracts global information. Most existing datasets either focus on a small subset of metrics, are collected under task conditions, or lack sufficient control to minimize awareness of tracking (Henderson & Hollingworth, 1999; Zwierko et al., 2023).

Here, we address these gaps by performing eye tracking in naturalistic indoor and outdoor real-world scenes while minimizing experimental constraints and awareness of the specific eye-movement measures of interest. We recorded saccade and fixation metrics, pupil size, photopic and melanopic irradiance during measurements, as well as subjective scene ratings. By contrasting the initial exploration phase with the full viewing period, we aim to identify how environmental context shapes early and sustained visual behavior. Our approach is designed to provide a reference dataset for controlled yet naturalistic eye movement research, offering insights into how luminance, environmental structure, and visual exploration interact outside the laboratory.

2. Methods

In preparing the methodology section, we drew on the minimal reporting guidelines for eye-tracking research by Holmqvist et al. (2022) and Dunn et al. (2023).

2.1. Participants

2.1.1. Study sample

The study sample consisted of 18 participants (13 female; mean age 24.2 years, range 20–29 years) who gave full consent for data use after debriefing. The mean refractive error across participants was -1.1 dpt (range -3.75 to 0 dpt). The larger number of female participants reflected the larger number of applications from women. Although we aimed for a numerically balanced gender distribution, we do not

consider data quality to have been compromised as a result. Coors et al. (2022) reported gender differences in certain eye-movement characteristics (blink rate, smooth-pursuit gain), but these differences were small in magnitude.

2.1.2. Recruitment

The first instance of participant recruitment was self-selection through advertisements, which included the attachment of flyers in Tübingen and a circular mail through the newsletter of the University of Tübingen. Individuals interested in participating in the study were asked to fill out an online questionnaire on REDCap (Harris et al., 2019, 2009), an online data-collection tool, to assess inclusion and exclusion criteria. Participants were compensated at the end of the study according to their compliance with the experimental procedure: for every hour that the volunteers participated in the study procedure (measurement session + preceding screening), they received €10. All participants agreed to make their eye movement data accessible.

2.1.3. Screening procedure

Eligible participants were selected according to inclusion and exclusion criteria based on mental and physical health information. The age limit of the adult participants was 40 years due to the increase in optical density of the lens with age (Pokorny et al., 1987), which causes different spectral sensitivities of retinal photopigment melanopsin (Spitschan, 2019). Participants were required to have normal binocular vision and color vision, as well as normal or corrected to normal visual acuity between $+5$ and -5 diopters. People with impairments, diseases or previous surgeries of their ocular apparatus were excluded from participation in the study. Moreover, individuals suffering from diagnosed sleep or psychiatric disorders were excluded from participation. Additionally, sleep quality was assessed via self-report on the (PSQI) (Pittsburgh Sleep Quality Index) questionnaire. Participants who exceeded a total PSQI score of 5 were excluded from study participation.

The intake of any drugs and/or medications known to influence photosensitivity, pupil responses or the ability to concentrate was considered a criterion for exclusion.

2.2. General measurement procedure

Participants were scheduled individually or in pairs, based on the availability of the participants and investigators. For an overview, a workflow chart of the study and individual appointments visualizes this procedure (Fig. 1). At each scene, participants first completed 5 min of on-site light adaptation while wearing a custom diffusing face shield (Luminit Light Shaping Diffuser, polycarbonate $0.010''$, 40° , 85%–95% transmission), which preserved retinal illumination while blurring scene details. This was done to account for the sluggish influence of ipRGC photoreceptors on pupil size (McDougal & Gamlin, 2010) and the saccade behavior (Lee & Yeh, 2021). Despite the shield obscuring fine details, participants were likely still picking up the low-frequency gist – for example, overall green hues suggesting vegetation – providing some limited prior knowledge. The shield was then removed, and the eye tracker was calibrated immediately before each recording (~ 25 – 40 s). The participants viewed each scene freely for 4 min with the simple instruction to keep their attention on the environment in front of them; spectral irradiance/illuminance were measured with the spectroradiometer 10 s after recording onset, and high-definition video was recorded continuously. Participant and instrument positions were standardized with floor markings to maintain geometry across sessions (see Fig. 3); during outdoor recordings, a brimmed hat minimized direct sunlight on the eyes and tracker. After each measurement, the setup was disassembled, the team moved to the next location, and participants completed a post-measurement questionnaire. The eye tracker and face shield were kept on between measurements to maintain habituation to the devices.

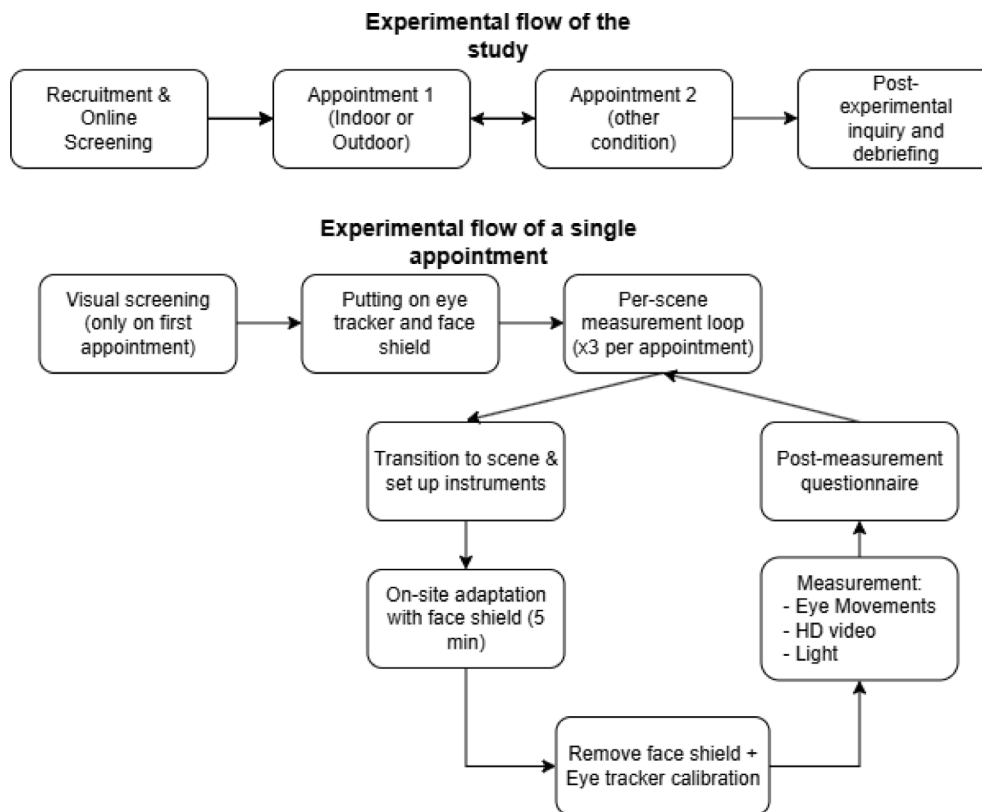


Fig. 1. Schematic overview of the study workflow (top) and of individual appointments (bottom).

2.3. Deception approach

Before and during data collection, certain information about the scope of the study was withheld from participants. Participants were informed that their pupil size would be monitored, but not that their eye movements would also be tracked. The purpose of this procedure was to reduce the potential influence of monitoring awareness on gaze behavior. Accordingly, all advertisement materials, information sheets used during screening, and instructions given during the measurement procedure referred only to pupil-size monitoring while participants viewed different indoor and outdoor scenes.

2.4. Outdoor and indoor environments

Participants were exposed to six real-world measurement environments, comprising three indoor environments (IME) and three outdoor environments (OME) (Fig. 2). All scenes were selected to reflect everyday settings commonly encountered by individuals in and around Tübingen. The indoor scenes were located within the Cyberneum building of the Max-Planck Institute for Biological Cybernetics and represented typical work-related environments (basement, hallway, office). The outdoor scenes similarly reflected everyday surroundings, including views toward institutional buildings and open urban-natural landscapes. During indoor recordings, all available ceiling lights were turned on to ensure stable artificial illumination. IME 1 (basement) contained no windows, IME 2 (hallway) included a single window, and IME 3 (office) featured multiple windows directly within the participants' field of view. When present, windows were double-glazed, closed, and uncovered.

To enable objective comparison across scenes, we characterized viewing geometry, light properties, composition, and dynamics (see Table A.6). All indoor scenes were located at ground level and did not include a visible horizon from the participant's point of view, whereas outdoor scenes varied in elevation and horizon presence.

Scene depth was defined as the ratio between the furthest and closest salient surfaces within the scene (excluding floors), with outdoor scenes generally exhibiting much greater depth. Photopic illuminance and melanopic irradiance were substantially higher outdoors, while the photopic-to-melanopic ratio was comparatively higher indoors. In terms of composition, indoor scenes were largely dominated by artificial structures, with window presence documented only for indoor environments. Outdoor scenes displayed more heterogeneous composition, including vegetation, sky, and built elements. Finally, outdoor environments were considered more dynamic overall, given the possibility of pedestrians, vehicles, or animals entering the scene, whereas indoor scenes were comparatively static.

2.5. Measurement instruments

The spectroradiometer and camera were mounted on a fixed holder that was positioned at the scenes with a tripod. This was done in order to guarantee the same geometry between instruments for every measurement (see Fig. 3).

2.5.1. Eye tracker

Eye movements and pupil size were recorded with Tobii Pro Glasses 3, a pupil-centred corneal-reflection (P-CR) mobile eye tracker with eight infrared illuminators (~850 nm) and two cameras in each lens sampling alternately at an effective 100 Hz. The device records both eyes simultaneously, provides a field of view of 95° (horizontal) × 63° (vertical), and spatial accuracy of ~0.6°. Participants with refractive error used device-specific corrective lenses (+5 to -5 dpt). The manufacturer's one-point calibration procedure was used, involving fixation on a standardized calibration card at 50–100 cm eye level; nose pads were used to stabilize the device, minimize slippage and position the eyes centrally in the lenses. Slippage was further minimized by a thin and adjustable headband.



Fig. 2. GoPro Hero 10 images (lens mode: wide) of all measurement environments. The images show only the part of each scene that was directly in front of the participants. Because head movements were unrestricted, participants could also view parts of the scene that are not depicted here.

2.5.2. Spectroradiometer

For the purpose of this study, the spectroradiometer MSC15 from Giga-hertz-Optik was used to capture photopic illuminance (lux) and melanopic irradiance. During data acquisition, the calibration certificate of the device was active, meaning the device was producing accurate results with high confidence.

2.6. Questionnaires

2.6.1. Post-measurement questionnaire

After each eye-tracking measurement, participants completed a questionnaire (see Appendix). The questionnaire assessed subjective characteristics of the scene on rating scales from 1 to 10, including beauty, brightness, visual comfort, sleepiness, and complexity. In addition, the Karolinska Sleepiness Scale (KSS) was included to quantify participants' sleepiness during the measurements.

2.6.2. Post-experimental questionnaire

The effectiveness of the deception was assessed by a post-experimental inquiry (see Appendix) immediately after the last measurement. The questionnaire was completed on a tablet rather than in a face-to-face verbal interview because this format is likely to increase the honesty of participants' responses in such inquiries (Blackhart et al., 2012). Among other things, participants were asked whether they had engaged in cross-talk, that is, whether they had divulged important experimental details to other individuals who might participate in the study. Cross-talk could have been detrimental to the study procedure if one participant had already been debriefed while another had not yet completed all measurements. Before participants answered the post-experimental inquiry, the experimenter emphasized the scientific

rationale and integrity of the study, because such a statement has also been suggested to increase the truthfulness of responses (Blackhart et al., 2012).

To eliminate the risk of cross-talk entirely, participants would have had to be debriefed later by phone call or email after the data-acquisition phase. However, this would have increased the interval between participation and explanation of the study procedure, especially because some participants would have been debriefed more than a week after their participation had ended. We therefore chose immediate debriefing as a pragmatic compromise between limiting cross-talk and maintaining transparency with participants.

2.7. Data processing and analysis

Eye-tracking data were exported from Tobii Pro Lab (v1.217) using the I-VT (Attention) filter (velocity threshold = 100 °/s; fixations < 60 ms removed; effective sampling 100 Hz), specialized to analyze dynamic eye movement data during head movements. For each recording we computed: average saccade frequency (s^{-1}), amplitude (°), peak velocity (°/s); average fixation frequency (s^{-1}), duration (ms); and average pupil size (mm). Saccade amplitudes were computed as head-centered angular displacements between successive gaze positions, rather than world- or scene-referenced coordinates. Moreover, under the operational definition of fixations in Tobii's I-VT gaze filter, fixations correspond to sequences of gaze samples with estimated velocities below the threshold (100°/s) and may therefore include periods of smooth pursuit or tracking movements, particularly in more dynamic environments. Velocity-based classification of fixations and saccades is a widely established approach in eye-movement research, relying on the distinct velocity profiles of these events (Duchowski,



Fig. 3. Example of the setup and a participant in IME 2 (A & B) and OME 1 (C & D). Panels A and C show the participant during the light-adaptation phase, during which the face shield preserved retinal adaptation while diffusing scene details. Panels B and D show the measurement itself. In outdoor measurement environments, participants wore a hat to shield themselves and the eye tracker from direct sunlight. The blue arrows and red circles on the ground indicate standardized distances and positions.

2011; Holmqvist et al., 2011; Stuart, Parrington, et al., 2019). Saccades typically exhibit peak velocities well above $100^\circ/\text{s}$ (e.g., a 5° saccade often exceeds $240^\circ/\text{s}$), whereas fixations are characterized by substantially lower velocities. The chosen threshold therefore lies far below typical saccadic peak velocities and allows robust discrimination between saccade and fixation events in dynamic recordings (Stuart, Parrington, et al., 2019; Stuart, Hickey, et al., 2019), without artificially reducing the frequency of saccades. Light data (spectral irradiance/illuminance) were processed to obtain photopic illuminance and melanopic measures. Quality control followed an ETSSR $\geq 70\%$ rule; recordings below this threshold were excluded. Due to spectroradiometer data loss, the original 18 datasets were reduced to 16 OMEs and 17 IMEs, yielding 15 complete datasets (eye + light) for combined analysis. After ETSSR filtering, usable OME datasets were 14 (OME1), 13 (OME2), 12 (OME3); no IME recordings were discarded for ETSSR. Primary comparisons tested indoor vs outdoor over the entire 4-min viewing period (independent-samples t-tests; Cohen's d reported) and across the six scenes (one-way ANOVA per metric; Bonferroni-adjusted $\alpha = 0.01$ for the five gaze metrics). To probe early exploration, the first 15 s were analyzed in three 5-second bins with independent-samples t-tests for each metric.

Pairwise Pearson's correlations (ρ) among eye-movement metrics, pupil size, light measures, and questionnaire ratings were computed separately for indoor, outdoor, and combined conditions, as well as for individual scenes. Multiple testing was controlled with Bonferroni (α set to 0.0045 based on the number of pairs).

Questionnaire ratings (beauty, brightness, visual comfort, complexity, sleepiness) were evaluated with Tukey's HSD for post-hoc contrasts to investigate differences between individual scenes ($\alpha = 0.01$); indoor vs outdoor differences in ratings used Welch's t-tests ($\alpha = 0.05$). The post-experimental inquiry included all 18 participants to assess deception effectiveness and subjective experience.

Because sample size was constrained by the logistics of the multi-location field protocol, we conducted a sensitivity analysis to estimate the range of effect sizes that could be detected with the final analyzable sample. The primary inferential contrast of interest was the indoor-outdoor condition difference at the participant level, while accounting for repeated measurements across scenes in the statistical models. To provide an interpretable approximation of inferential sensitivity, we therefore estimated power under a paired-samples framework with $n = 15$ participants and a two-sided $\alpha = 0.05$. Under this approximation, standardized mean differences of approximately Cohen's $d \approx 0.70$ – 0.80 correspond to $\sim 80\%$ power, whereas large effects ($d \geq 1.1$) correspond to $\leq 99\%$ power. Accordingly, moderate effects ($d \approx 0.55$ – 0.60), such as those observed for early exploration saccade measures, fall within a region of moderate sensitivity, whereas the observed pupil-size effects ($d \approx 1.1$ – 1.4) lie in a range of high statistical sensitivity.

3. Results

3.1. Comparison of eye tracking metrics between conditions

Statistical analysis of the initial 5 s of eye tracking revealed higher saccade metrics and larger pupil size in the “Indoor” conditions (IME 1–3) than in the “Outdoor” conditions (OME 1–3) (Table 1 & Fig. 4B). None of the fixation metrics displayed corresponding differences between indoor and outdoor conditions. While pupil size remained significantly larger indoors, the indoor-outdoor differences in saccade metrics were no longer significant in the subsequent bins (Fig. 4B).

Pairwise comparisons of the time bins did not yield significant changes in any metric over time ($0.17 \leq p\text{-value} \leq 0.63$). Across the entire 4-minute measurement duration, no significant differences in saccade or fixation metrics between the indoor and outdoor conditions were detected (Table 2 & Fig. 4A). The only exception was

Table 1

Summarized data of initial 5 s of eye tracking. CI: Confidence interval, AFS: Average frequency of saccades, AAS: Average amplitude of saccades, APVS: Average peak velocity of saccades, ADF: Average duration of fixations, AFF: Average frequency of fixations.

Metric (Unit)	Mean “Indoor”	CI “Indoor”	Mean “Outdoor”	CI “Outdoor”	p-value	Cohen’s d
AFS (s^{-1})	1.63	[1.45, 1.81]	1.27	[1.04, 1.50]	0.04	0.55
AAS ($^{\circ}$)	8.23	[7.40, 9.06]	6.75	[5.99, 7.52]	< 0.01	0.56
APVS ($^{\circ}/s$)	251.38	[234.47, 268.28]	217.85	[200.54, 235.16]	< 0.01	0.60
ADF (ms)	400.56	[342.84, 458.28]	418.76	[313.89, 523.63]	0.75	0.07
AFF (s^{-1})	2.43	[2.21, 2.65]	2.28	[1.99, 2.57]	0.40	0.19
Pupil size (mm)	2.91	[2.64, 3.18]	2.01	[1.98, 2.20]	< 0.01	1.11

Table 2

Summarized eye tracking data of whole 4-minute measurement period. CI: Confidence interval, AFS: Average frequency of saccades, AAS: Average amplitude of saccades, APVS: Average peak velocity of saccades, ADF: Average duration of fixations, AFF: Average frequency of fixations.

Metric (Unit)	Mean “Indoor”	CI “Indoor”	Mean “Outdoor”	CI “Outdoor”	p-value	Cohen’s d
AFS (s^{-1})	1.07	[0.97, 1.17]	1.03	[0.93, 1.13]	0.56	0.13
AAS ($^{\circ}$)	7.35	[7.02, 7.68]	6.94	[6.49, 7.39]	0.14	0.33
APVS ($^{\circ}/s$)	220.41	[214.8, 226.01]	217.80	[210.94, 224.66]	0.55	0.13
ADF (ms)	653.73	[535.29, 772.17]	524.01	[454.48, 593.52]	0.08	0.37
AFF (s^{-1})	1.59	[1.44, 1.74]	1.73	[1.58, 1.91]	0.15	0.31
Pupil size (mm)	2.88	[2.68, 3.09]	2.07	[1.96, 2.17]	< 0.01	1.41

“Average Pupil Size”, which remained larger indoors throughout the measurement period (Fig. 4A).

Across the six scenes, no statistically significant difference was detected for any eye-movement metric over the full 4-minute measurement duration ($0.08 \leq p\text{-value} \leq 0.58$; $0.76 \leq F\text{-value} \leq 2.03$), with “Average Pupil Size” again being the only exception ($F = 63.47$; $p < 0.01$) (one-way ANOVA; Fig. 4A). In addition, Fig. 5 shows that pupil size decreased with increasing luminance across environments, consistent with the well-established inverse relationship between luminance and pupil size.

3.2. Pairwise correlations of metrics

The pairwise correlation analysis of 11 metrics amounted to 66 correlation pairings (Fig. 6). For each pair of metrics, a separate correlation analysis of indoor, outdoor and combined conditions was performed. The number of significant pairwise correlations amounted to 16 for the indoor condition, 12 for the outdoor condition and 13 for the combination of indoor & outdoor conditions.

Several of the observed correlations were unsurprising because the corresponding metrics are reciprocal (e.g. duration of fixations and frequency of fixations (Indoor: $\rho = -0.75$; Outdoor: $\rho = -0.78$)). Other correlations differed between indoor and outdoor conditions. For instance, average saccade peak velocity and average saccade amplitude were positively correlated in the outdoor condition ($\rho = 0.65$) but not in the indoor condition ($\rho = 0.09$) (Fig. A.11). Because our dataset generated only mean values of saccade peak velocities and amplitudes, this pattern should be interpreted cautiously and not as a direct test of the classical main sequence at the level of individual saccades. Similarly, a strong positive correlation between peak velocity of saccades and saccade frequency was observed only in the outdoor setting ($\rho = 0.56$). Additionally, we found a significant correlation of pupil size and light metrics in the indoor condition ($\rho = -0.80$ & -0.81) but not for outdoor environments ($\rho = -0.01$ & -0.03), although the measured light intensity was much higher outdoors. Interestingly, the subjective perception of scene beauty showed a strong positive correlation with perceived brightness, a finding that was consistent in both indoor ($\rho = 0.51$) and outdoor conditions ($\rho = 0.60$). Corresponding correlations between perceived beauty and measured light metrics were found only for the indoor environment ($\rho = 0.41$ & 0.42). Correlation matrices

of the individual scenes can be found in the supplementary material (Fig. A.10). Additionally, principal component analysis (PCA) suggested separation between indoor and outdoor conditions (Fig. 7). The first two components explained 44.5% of the total variance and were mainly driven by oculomotor (PC1) and luminance-related (PC2) metrics, indicating that both eye movement control and visual scene characteristics contributed to the observed clustering.

3.3. Post-measurement questionnaire

The post-measurement questionnaire was completed by participants immediately after each recording and included several questions about their subjective experience of the measurement environment. The responses are visualized in Fig. 8. Results for perceived visual comfort are not shown because several participants later reported that they had misunderstood the rating scale for this item.

Perceived beauty differed significantly between indoor and outdoor scenes (Welch’s t-test: $p < 0.05$; $d = 0.93$). Tukey’s HSD revealed that the highest beauty rating was attributed to OME 2 ($p\text{-adj} < 0.01$); however, the difference between OME 2 and OME 3 is not statistically significant (Tukey’s HSD: $p\text{-adj} = 0.13$). Moreover, IME 1 was significantly less beautiful than any other scenes (Tukey’s: $p\text{-adj} < 0.01$).

Although the photopic illuminance and melanopic irradiance measured during the experiments varied substantially between environments, the subjective differences in perceived brightness were minor. The only significant difference in brightness ratings occurred between IME 1 and OME 2 (Tukey’s: $p\text{-adj} < 0.01$). In summary, the perceived brightness of IMEs and OMEs was not significantly different (Welch’s t-test; $p = 0.11$).

The perceived complexity of indoor and outdoor environments was not significantly different (Welch’s t-test; $p = 0.07$). The only significant difference in perceived complexity between individual scenes was observed for the comparison between IME 2 and OME 2 (Tukey’s: $p\text{-adj} < 0.01$).

The comparison of sleepiness ratings between indoor and outdoor conditions did not reveal a statistically significant difference (Welch’s t-test; $p = 0.06$). Scene-specific comparisons of sleepiness (Tukey’s HSD: $p\text{-adj} = 0.17$) likewise did not reveal significant differences.

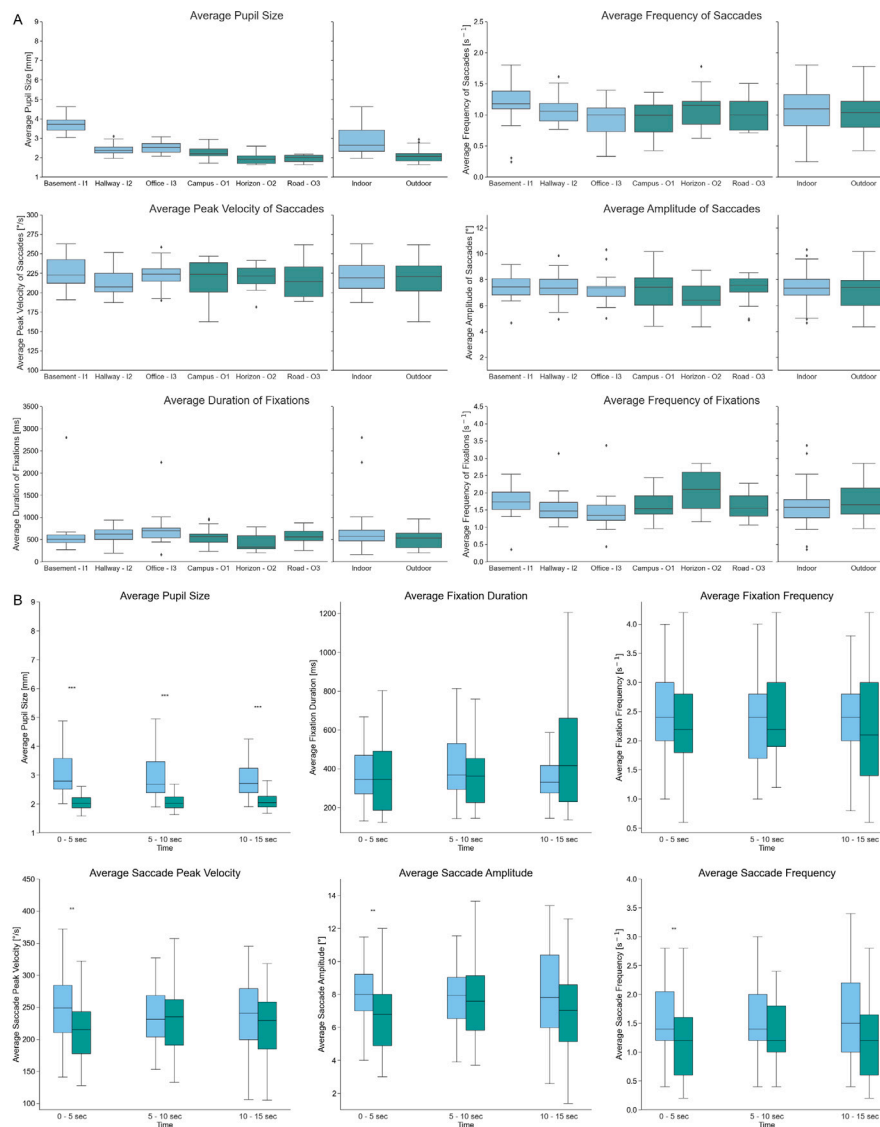


Fig. 4. Visualization of all analysed gaze metrics (average frequency of saccades, average pupil size, average peak velocity of saccades, average amplitude of saccades, average duration of fixations, and average frequency of fixations). Boxplots show the median, 25th and 75th percentiles (box edges), and $1.5\times$ the interquartile range (IQR; whiskers). Data points beyond $1.5\times$ IQR are outliers. A: Each subplot depicts one metric for all six measurement environments and for the pooled indoor and outdoor conditions. B: Each subplot depicts the first 15 s of one metric for the pooled indoor and outdoor conditions, shown in three 5-s bins.

3.4. Light metrics

Light intensity varied substantially across scenes, particularly between indoor and outdoor conditions (Fig. A.9). Outdoor measurement environments were generally more brightly illuminated, exhibiting higher photopic illuminance (mean = 8883 lux; range 1395–56290 lux) and higher melanopic irradiance (mean = 8812 W/m²; range 1297–49360 W/m²) compared to indoor environments (mean photopic illuminance = 520 lux; range 120–1149 lux; mean melanopic irradiance = 388 W/m²; range 65–894 W/m²). Although absolute photopic and melanopic values were higher outdoors, the photopic-to-melanopic ratio was comparatively higher in indoor environments than in outdoor environments (Table A.6).

3.5. Eye tracking quality in different conditions

Here, we investigated the relationship between ETSSR and the present levels of photopic illuminance during the measurements (Fig. A.9). For this analysis, only OME were considered since illuminance

and ETSSR varied much more in the outdoor condition (in the indoor condition, every eye tracking measurement amounted to $> 85\%$ ETSSR). Outdoor measurements conducted at photopic illuminances below 10000 lux (ϕ ETSSR = 90.54%) were significantly more successful than measurements during which the equipment was subjected to photopic illuminance above 10000 lux (ϕ ETSSR = 77.86%) (two independent samples t-test with $\alpha = 0.05$; $p = 0.02$; $d = 0.68$).

3.6. Post-experimental inquiry

A total of five participants reported that they had suspected eye-movement recordings during the experiment. In addition, five participants reported physical discomfort and one participant reported psychological discomfort during the measurements. The physical discomfort was always attributed to pain on the nose and ears caused by the pressure of the glasses, while the psychological discomfort was based on people being in the same room as the participant during the measurement. Participants who stated that the task felt unnatural ($n = 2$) specified that either the four-minute measurement procedure of

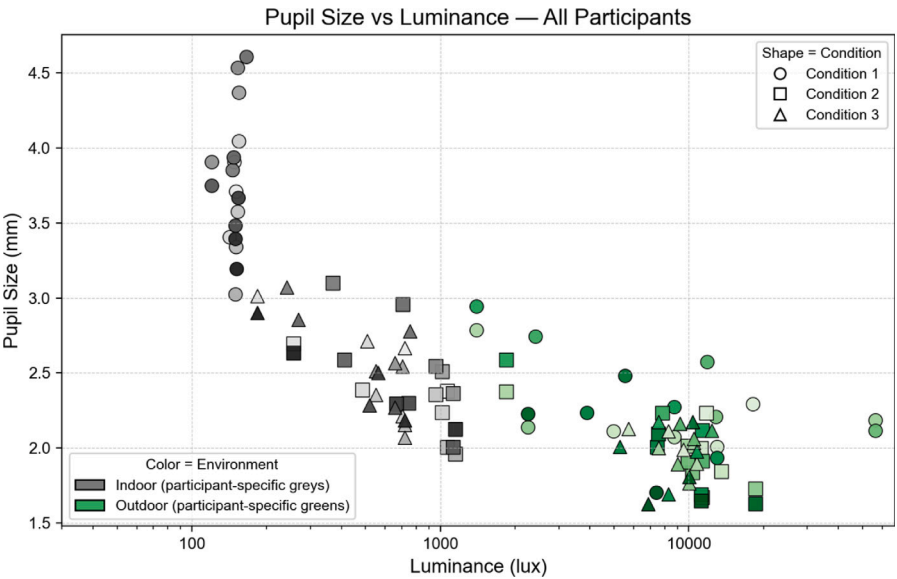


Fig. 5. Relationship between pupil size and luminance across all participants. Each point represents a single recording, colour-coded by environment and shaded by scene. The inverse trend shows that pupil size decreases systematically with increasing luminance, indicating that pupil diameter was closely related to luminance across environments.

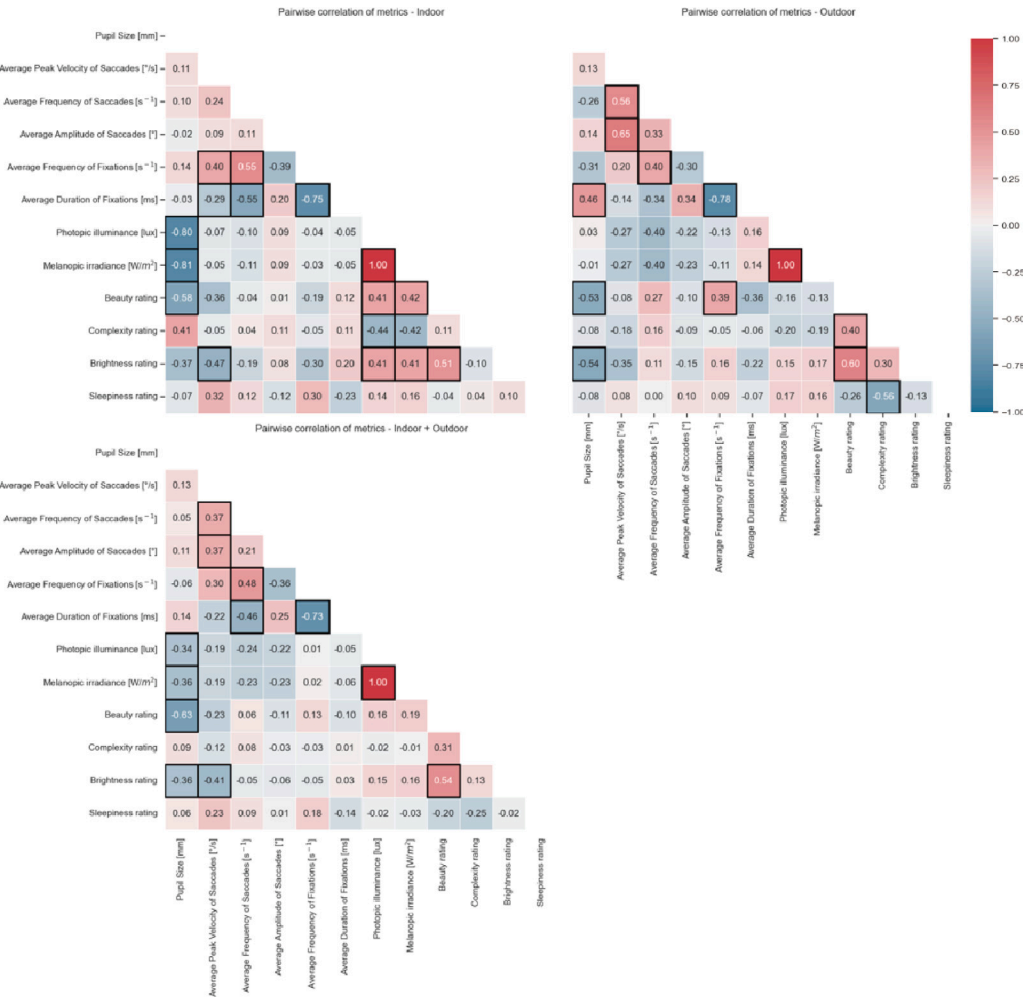


Fig. 6. Heatmaps that visualize pairwise correlations between all analyzed metrics in three different conditions. Correlations of statistical significance are signified with a black frame around the square.

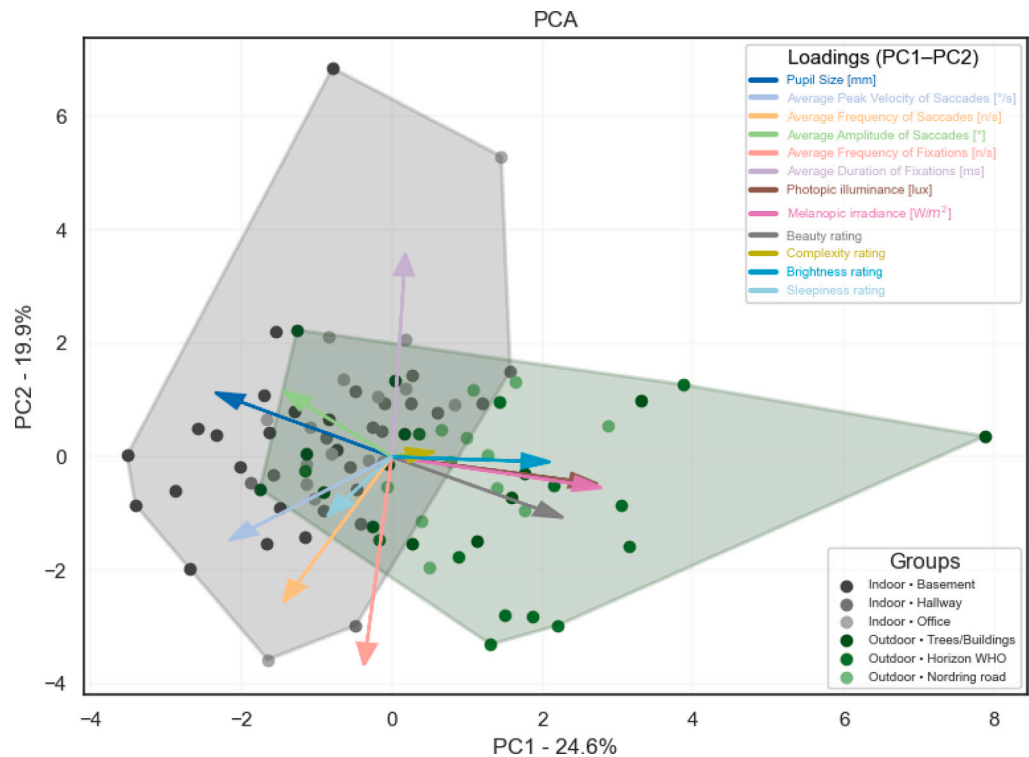


Fig. 7. Principal component analysis (PCA) of eye-tracking metrics across indoor and outdoor viewing conditions. Each data point corresponds to one measurement. The first two principal components (PC1 and PC2) capture 24.6% and 19.9% of the total variance, respectively. Partial separation is visible for indoor (gray shades) and outdoor (green shades) conditions, with each shade representing a different scene type. Filled convex hulls outline the distribution of measurements within each subgroup, illustrating the clustering of eye movement behavior by environmental context.

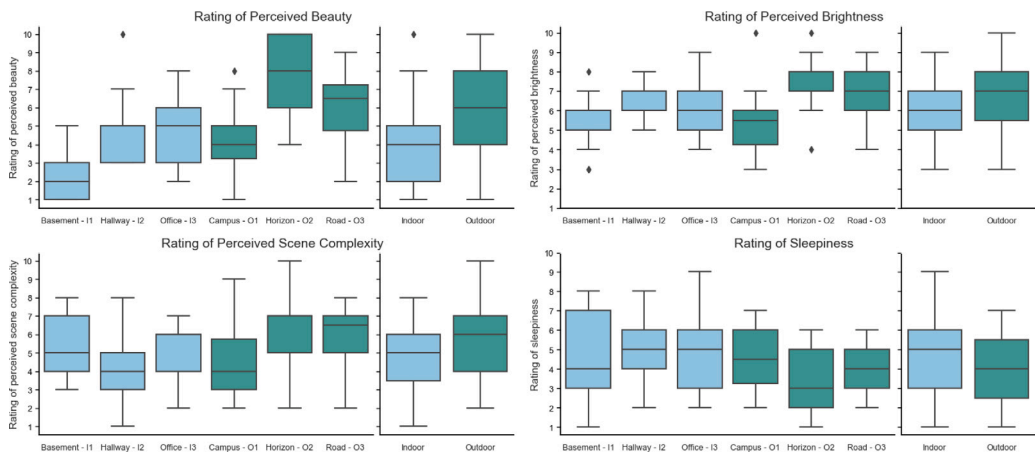


Fig. 8. Ratings from the post-measurement questionnaire for perceived beauty, perceived brightness, perceived complexity, and sleepiness across all measurement environments. Boxplots show the median, 25th and 75th percentiles (box edges), and 1.5× the interquartile range (IQR; whiskers). Data points beyond 1.5× IQR are outliers. Each subplot depicts one rating for all six measurement environments and for the pooled indoor and outdoor conditions.

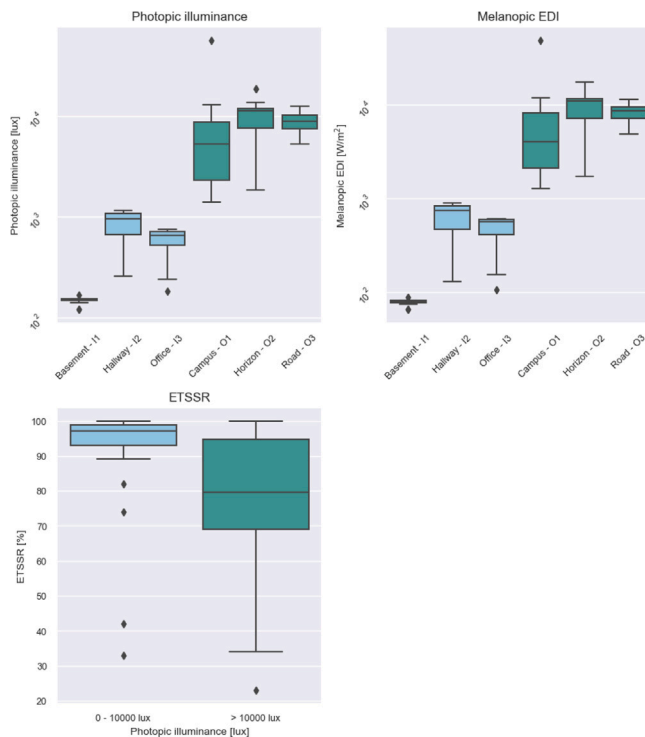


Fig. A.9. A: Photopic illuminance and melanopic irradiance in the six measurement environments. B: Percentages of eye tracking sampling success rate (ETSSR) in relation to outdoor illuminance levels above and below/equal 10000 lux.

sitting and viewing felt unusual or wearing glasses was foreign to them. Moreover, participants were asked whether the glasses affected their viewing experience. If they did not answer “Not at all” ($n = 12$), they were asked to describe the effect. Most commonly, the given reason for an altered viewing experience was the weight and physical feeling of the glasses on the nose and ears ($n = 3$), as well as the fear of impairing the success of the measurement by larger head movements ($n = 2$). The post-experimental inquiry responses of all 18 participants are presented in Table A.3.

4. Discussion

In this study, we performed eye tracking under naturalistic conditions to investigate potential differences between indoor and outdoor environments. Overall, descriptive analyses of the recorded eye-movement metrics did not reveal substantial differences between conditions. The only notable descriptive divergence was an elevation of saccade-related metrics in indoor environments during the first 5 s of measurement, suggesting differences in the initial exploration phase. Beyond this early effect, metrics were largely comparable across conditions. An additional exploratory analysis of pairwise relationships between metrics, however, revealed condition-specific differences in correlation structure between indoor and outdoor environments.

4.1. Comparison of eye tracking metrics between conditions

We analyzed the first 15 s of our measurements to compare exploratory eye movement patterns between indoor and outdoor environments. Similar strategies for binning the initial eye movements to assess information processing during scene perception had been done before in less naturalistic and task-dependent approaches (Unema et al., 2005; Pannasch et al., 2008; Guo et al., 2022). These approaches investigated exploratory eye movements with varying lengths (3–20 s)

and bin sizes (200–1000 ms). Interestingly, our measurements did not reveal changes for any eye movement characteristic over time. This comes as a surprise, since a time-dependent increase in fixation duration and decrease in saccade amplitudes have been reported previously (Unema et al., 2005; Pannasch et al., 2008). The lack of this characteristic exploratory eye movement pattern in our dataset, could be explained by the eye tracker calibration, during which the participant’s gaze could briefly explore the scene without being measured. Future studies should aim to reduce or abolish the time required for eye tracker calibration to minimize this confounding factor. In addition, we observed a comparatively low saccade frequency both during the initial viewing phase and across the entire 4-minute recording period. In laboratory-based free-viewing paradigms, typical saccade rates range from approximately $2\text{--}3\text{ s}^{-1}$ (Yarbus, 1967; Gray et al., 2014), making the present values ($\sim 1\text{ s}^{-1}$) notably lower. However, the applied velocity threshold of $100^\circ/\text{s}$ lies well below typical saccadic peak velocities (Stuart, Parrington, et al., 2019), and is therefore unlikely to miss voluntary saccades. Beyond methodological factors, the reduced saccade rate may reflect behavioral adaptation: following an initial exploratory phase (including brief pre-calibration exposure), gaze behavior may stabilize once the global structure of the scene has been extracted, resulting in fewer exploratory saccades. This is indicated by a slightly elevated saccade frequency within the initial 5 s of measurement. Additionally, within this first time bin of recording, indoor environments showed significantly elevated saccade metrics. This pattern suggests that the earliest phase of scene exploration differs systematically between indoor and outdoor environments. Although not reflected in the subjective complexity ratings, indoor scenes may have contained more salient targets. As early saccades are known to be influenced by image salience (Anderson et al., 2015), this pattern may reflect that mechanism. Furthermore, over the course of the entire measurement duration, we did not find significant differences between the indoor and outdoor conditions, as well as between the individual scenes. This suggests that, despite large differences in scene content, eye movement patterns are largely similar over an elongated exploration period.

Previous research has demonstrated that tasks exert a significant influence on eye movement (Kay et al., 2023). The aim of the study, creating conditions for unbiased eye movement behavior, could be achieved solely with a very naturalistic task and careful instructions. As expected, the metrics that we have chosen to analyze in this study deviated from the results of different studies, since every study imposes different instructions and tasks on its participants. The utilization of a naturalistic task (“View the scene that is in front of you and feel free to look wherever you want”) gives this dataset the advantage of reduced susceptibility to confounding by specific task instructions. While data collection with loosely defined task instructions may present additional challenges (e.g., drift of attention; emergence of intrinsic tasks or motivations), we assert that this dataset holds substantial value for the domain of eye movement research.

4.2. Pairwise correlations of metrics

This exploratory correlation analysis highlights several pairwise correlations that were either exclusive to indoor or outdoor condition, or universally present in both conditions. Correlations between reciprocal metrics will not be discussed in further detail, as their origins are evident.

Several strikingly different gaze-metric correlations between indoor and outdoor conditions were revealed by this analysis. For instance, despite the robustness of the saccadic main sequence in laboratory paradigms (Bahill et al., 1975; Gibaldi & Sabatini, 2021), we did not observe a clear amplitude–peak velocity relationship in the indoor condition. To further examine this unexpected finding, we plotted “Average Saccade Peak Velocity” against “Average Saccade Amplitude” separately for indoor and outdoor environments. A linear relationship consistent with the classical main sequence (Bahill et al., 1975) was

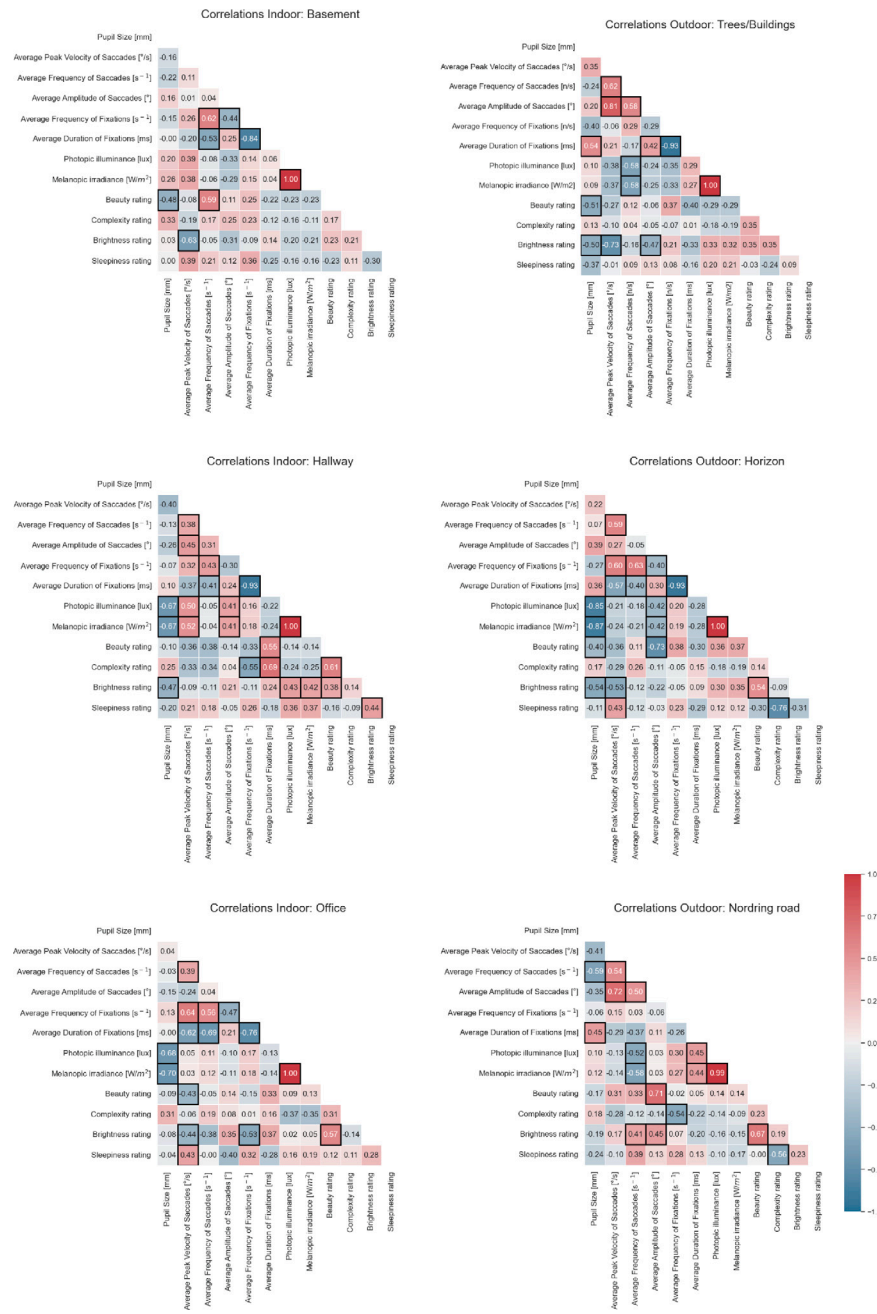


Fig. A.10. Heatmaps that visualize pairwise correlations between all analyzed metrics in all scenes individually. Correlations of statistical significance are signified with a black frame around the square.

apparent in the outdoor condition but largely absent indoors. However, this result should be interpreted with caution. Our analysis relied on averaged saccade metrics across the entire measurement period rather than on individual saccade events. The main sequence relationship is known to be most robust within a limited amplitude range, particularly below approximately 10° (Bahill et al., 1975). It is therefore conceivable that, in the indoor condition, especially very large exploratory saccades may have obscured the expected relationship when only mean values were considered. Averaging across heterogeneous saccade distributions could attenuate the underlying amplitude–velocity coupling. Consequently, confirmation of this finding will require analyses at the

level of individual saccades using an eye-tracking system capable of reliably capturing event-wise amplitude and velocity under comparable naturalistic conditions.

Furthermore, despite substantially higher light intensities outdoors, we did not observe a strong correlation between pupil size and light metrics in outdoor environments. A plausible explanation is a ceiling effect: once luminance exceeds a certain level, further increases produce little additional constriction, attenuating the correlation at high intensities. In addition, squinting and the generally lower ETSSR outdoors may have introduced measurement artifacts that contributed to this counterintuitive result. Another interesting finding is the rather strong

Table A.3

Answers in the post-experimental inquiry used to assess the experiences of the participants during the experiment. Q1: *Did you experience physical and/or psychological discomfort during either the first or second appointment?*; Q2: *To what extent do you believe that wearing the glasses modified how you looked at the indoor scenes?*; Q3: *How likely would you be exposed to indoor scenes similar to the ones you were exposed to in your daily life?*; Q4: *How likely would you be exposed to outdoor scenes similar to the ones you were exposed to in your daily life?*; Q5: *How natural was the task to you?*; Q6: *Were the instructions by the study personnel clear?*; Q7: *Do you know what types of data were collected throughout the study?*; Q8: *Are you aware of anyone you know, who participated in the study (besides your team partner)?*; Q9: *Did you and the other person share information about the purpose of the study?*.

ID	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9
1	No	Not at all	Likely	Neutral	Unnatural	Yes	Yes	No	No
2	Yes	Little	Neutral	Likely	Natural	Yes	Partly yes	No	No
3	No	Little	Neutral	Likely	Natural	Yes	Yes	No	No
4	No	Little	Likely	Very likely	Neutral	Yes	Partly yes	No	No
5	Yes	Little	Neutral	Very likely	Natural	Yes	Partly yes	Yes	No
6	No	Not at all	Neutral	Very likely	Natural	Yes	Partly yes	No	No
7	Yes	Not at all	Likely	Likely	Very natural	Yes	Partly yes	No	No
8	No	Not at all	Neutral	Very likely	Unnatural	Yes	Partly yes	No	No
9	No	Not at all	Very likely	Very likely	Natural	Yes	Yes	No	No
10	Yes	Not at all	Likely	Likely	Natural	Yes	Yes	No	No
11	Yes	A bit	Likely	Very likely	Natural	Yes	Yes	Yes	No
12	No	Not at all	Very likely	Likely	Natural	Yes	Partly yes	No	No
13	No	A bit	Likely	Likely	Natural	Yes	Partly yes	No	No
14	No	Not at all	Likely	Very likely	Natural	Yes	Yes	No	No
15	No	Not at all	Likely	Neutral	Natural	Yes	No	No	No
16	No	Little	Likely	Very likely	Neutral	Yes	Partly yes	No	No
17	Yes	A bit	Unlikely	Neutral	Neutral	Yes	Partly yes	No	No
18	No	Little	Unlikely	Likely	Natural	Yes	Partly yes	No	No

Table A.4

Metadata for all appointments conducted indoors. IME3 p.p. stands for *people present* at the scene.

Appointment	Number of participants	Date	Time	Temperature [° C]	Notes
1	One	14.07.23	08:30	n.a.	IME 3 03:45: Person walked into the field of view; IME3 p.p.
2	Two	15.07.23	09:30	n.a.	IME 2: Measurement had to be stopped and repeated after 45 s due to loss of connection
3	Two	17.07.23	09:00	n.a.	IME 3 people present (p.p.)
4	One	18.07.23	09:30	n.a.	IME 3 p.p.
5	Two	19.07.23	10:30	n.a.	OME 2&3: Calibration issues with both participants; IME 3 p.p.
6	One	20.07.23	09:30	n.a.	IME 3 p.p.
7	Two	21.07.23	09:30	n.a.	IME 3 p.p.
8	Two	22.07.23	09:00	n.a.	
9	One	24.07.23	09:30	n.a.	
10	One	25.07.23	09:30	n.a.	
11	One	26.07.23	09:30	n.a.	
12	One	28.07.23	09:00	n.a.	
13	One	29.07.23	09:30	n.a.	

positive correlation between perceived scene beauty and perceived scene brightness in both conditions, implying that the brightness of a scene is a major driver of perceived beauty. Several previous studies have shown that brighter appearance of colors (lightness) positively correlates with perceived beauty/pleasantness of colors (Guilford & Smith, 1959; McManus et al., 1981), but to the best of our knowledge, the present study is the first to suggest this in real-world environments. In addition, the strong negative correlation between pupil size and beauty is likely a by-product of the positive correlation between

beauty and brightness. Beyond individual metric differences, the observed divergence in correlation structure between indoor and outdoor conditions suggests environment-dependent coupling between oculomotor control, luminance regulation, and subjective scene appraisal. Rather than reflecting isolated metric changes, these patterns indicate that visual environments may shape how eye-movement dynamics interact with perceptual and evaluative processes. Importantly, such coupling structures may be less apparent in tightly controlled laboratory paradigms, where task demands constrain gaze allocation and

Table A.5

Metadata for all appointments conducted outdoors. Temperature data for Tübingen were added retrospectively to the dataset using the website ([Meteostat](https://meteostat.de/), 2024).

Appointment	Number of participants	Date	Time	Temperature [° C]	Notes
1	One	14.07.23	15:00	min:13.9 max:28.0 avg:21.2	
2	Two	15.07.23	15:00	min:18.0 max:33.3 avg:25.5	OME 2 02:10: Protective hat blown off by wind
3	One	17.07.23	15:30	min:15.3 max:25.2 avg:20.4	
4	One	18.07.23	15:00	min:13.3 max:25.3 avg:20:0	
5	Two	19.07.23	15:00	min:17.9 max:26.5 avg:21.9	OME 2&3: Calibration issues with both participants
6	Two	20.07.23	15:00	min:15.9 max:24.1 avg:19.9	OME 2: Gathering of 5-7 people ca. 10 m behind us
7	One	21.07.23	14:00	min: 13.8 max:23.9 avg: 18.8	
8	Two	22.07.23	15:00	min:11.5 max:23.6 avg:18.3	
9	Two	23.07.23	15:00	min:15.6 max:26.5 avg:21.0	OME 2&1: Calibration issues; 01:55: participant was bothered by flying insect
10	One	27.07.23	15:00	min:11.4 max:21.4 avg:16.4	OME 1 loud noise from construction site
11	One	28.07.23	16:00	min:14.8 max:23.0 avg:17.9	
12	One	29.07.23	15:00	min:15.1 max:24.6 avg:19.2	
13	One	30.07.23	15:00	min:15.7 max:21.6 avg:18.3	

reduce variability in scene-driven behavior. Naturalistic approaches therefore provide an opportunity to examine how oculomotor control, physiological regulation, and subjective valuation co-vary under conditions that more closely resemble everyday visual experience.

4.3. Eye tracking quality in different conditions

Generally, in an indoor environment, eye tracking was more successful than outdoors. The reasons underlying this were lower intensities of light and more stable lighting circumstances during the measurements. In the outdoor environment, eye tracking was significantly more successful at lower light levels (<10000 lux). Although the spectroradiometer did not measure infrared light, we assume much higher irradiances of infrared light in outdoor conditions due to the composition of sunlight. This was likely an important influence on the tracking ability of our video-based eye tracker, which illuminates the pupil with infrared light at approximately 850 nm wavelength. Furthermore, some participants reported that they squinted their eyes when subjected to high illuminance in the outdoor condition. Other potential issues with mobile eye tracking data like head-set slippage (Niehorster et al. 2020) or variable luminance (Evans et al. 2012) could be largely reduced in this study. Here, the participants did not walk during measurements and individually fitted nose pads and head-bands

minimized slippage of the eye tracker. Other methodological adaptations, like the protective hat and stationary measurement positions, reduced the effects of variable light conditions.

4.4. Post-measurement questionnaire

In the present study, outdoor environments were perceived as significantly more beautiful than indoor environments. This is consistent with previous studies showing that people generally believe beauty to lie more in nature than in artificial objects (Brielmann et al., 2021) and that ratings of beauty/attractiveness are increased when spatial structures of images resemble structures found in natural scenes (Palmer et al., 2012). Especially OME 2, which offered a wide-ranging view with heterogeneous content, was perceived as attractive. Contrary to that, the windowless room in IME 1 was considered to be significantly more unattractive. Interestingly, the rating of brightness did not significantly differ between indoor- and outdoor conditions, even though the measured light metrics varied drastically. The likely reason for that is the high surface reflectance of the predominantly white indoor surfaces. The visual system factors out the physical properties of the light (illumination), and the awareness shifts toward the nature of the observed objects, rather than its physical properties (proximal mode of perception).

Table A.6

Scene feature table allowing for objective scene comparison. All indoor scenes display no elevation from the ground floor as seen from the point of view, while the outdoor scenes show variability in this parameter. Similarly, indoor scenes did not include a horizon. Scene depth was assessed by the ratio between the furthest salient surface and the closest salient surface within the scene (except floors). Outdoor scenes generally showed higher scene depth than indoor scenes. Photopic illuminance and melanopic irradiance were much higher in the outdoor scenes, but the ratio between the two was higher in indoor scenes. Composition of outdoor scenes was heterogeneous, while indoor scenes were largely dominated by artificial objects (presence of windows was only accounted for in indoor scenes). In terms of scene dynamics, outdoor scenes were considered to be generally more dynamic, with the possibility of cars, humans, or animals entering the scenes)

Scene parameters		Indoor 1	Indoor 2	Indoor 3	Outdoor 1	Outdoor 2	Outdoor 3
Semantic scene description		Basement storage room	Office building hallway	Office space	Urban vegetation	Panoramic view	Cross road with vegetation
Viewing Geometry	Elevation	0 m	0 m	0 m	0 m	15 m	7 m
	Horizon visible (yes/no)	no	no	no	no	yes	yes
	Scene depth	7/3= 2.33	8/2= 4	7/2= 3.5	60/3= 20	900/20= 45	250/10= 25
Light	Mean photopic illuminance (lux)	147.6	840.6	572.2	9378.4	10219.9	8659
	Mean melanopic irradiance (W/m2)	77.4	628.3	458.1	8270.4	9906.6	8206.3
	Photopic/Melanopic ratio	1.91	1.34	1.25	1.13	1.03	1.05
Scene composition (%)	Vegetation/Plants	0	3	1	79	14	46
	Sky	0	0	0	3	46	27
	Artificial features	100	95	95	18	40	27
	Windows (Indoor scenes)	0	2	4	-	-	-
Dynamics	Moving entities (yes/no)	no	no	no	yes	yes	yes
	Human presence possible (yes/no)	no	no	yes	yes	yes	yes
	Noise (yes/no)	no	no	no	yes	no	yes

Interestingly, participants did not report a significant difference in sleepiness between indoor and outdoor conditions. Given the cognitive-enhancing effects of bright light (Vandewalle et al., 2006; Shishegar & Boubekri, 2016) and the reported increase in alertness under blue-enriched light (Chellappa et al., 2011), we had expected lower sleepiness ratings in the outdoor condition.

We excluded the results from question three (*Perceived Visual Comfort*) because three participants reported later in the experiment that they had misunderstood the rating scale for this question. We cannot confidently estimate how many other participants may have had the same issue without mentioning it to the investigators.

4.5. Post-experimental inquiry

The responses to this questionnaire help estimate the effectiveness of the deception approach and the extent to which participants may have been distracted during the measurements. Five of the eighteen participants suspected that eye tracking was being performed, even though it had not been mentioned to them explicitly. This may have been due to previous experience with eye-tracking studies, familiarity with the Tobii device, general suspicion about the acquired datasets, or a combination of these factors.

Out of 18 participants, one responded that they felt distracted by other people being in the same room as them during the measurements. This situation occurred regularly in IME 2, especially during the weekdays, suggesting that more people could have felt distracted, although they did not point it out in the inquiry. This circumstance might have influenced the eye movement pattern of some participants in this particular scene (Risko & Kingstone, 2011). A related limitation concerns the fluctuating number of participants present during measurement sessions. Depending on availability, participants completed the recordings either individually or in pairs. This variation may have influenced viewing behavior, an effect that remains poorly understood (Holmqvist et al., 2022). Future studies on similar topics should select measurement environments without other people present and only schedule participants individually to minimize potential social influences on gaze behavior.

4.6. Limitations

While the study was designed to minimize major confounding factors associated with naturalistic eye tracking, several limitations should be considered. First, although analyses focused on the initial 15 s of viewing, participants had already been exposed to the scenes during the preceding calibration period (ca. 25–40 s), meaning this window does not represent fully naïve exploration. Moreover, the absence of an explicit task reduced task-driven bias but increased behavioral variability and limits causal attribution of effects to specific scene or physiological mechanisms. Second, scene sampling was limited in number and heterogeneity. Only six environments were assessed, and variability existed within conditions—particularly among indoor scenes differing in window presence and view, which affected luminance and structural composition. Indoor and outdoor scenes also differed in expansiveness, with outdoor views generally having broader perspectives than eye-level indoor views. Such structural differences, as well as potential variations in visual salience, may have influenced gaze patterns independently of the indoor–outdoor categorization. Systematic manipulation of scene parameters would be necessary to isolate the contribution of specific environmental features. Third, constraints inherent to mobile eye tracking must be acknowledged. Although participants were seated and movement was minimized, saccade amplitudes were computed in a head-centered reference frame, and residual head movements cannot be fully excluded as a source of variability. High outdoor luminance increased data loss and noise, leading to exclusion of recordings that did not meet the predefined ETSSR threshold of 70%, and resulting in unequal sample sizes across conditions. In addition, fixation detection relied on a velocity-based algorithm, meaning that fixation periods may have included smooth pursuit or other slow tracking components, particularly in dynamic outdoor settings. Fourth, analytical limitations apply. Because only averaged saccade metrics were available, event-level amplitude–velocity coupling (main sequence) could not be directly assessed. Although sensitivity analysis indicated adequate power for large effects, moderate effects may be less reliably detectable given the sample size. Finally, social and contextual factors may have influenced gaze behavior. Despite the use of deception, some participants suspected eye-movement recording, and variation in social context (e.g., presence of others) may have contributed to behavioral variability.

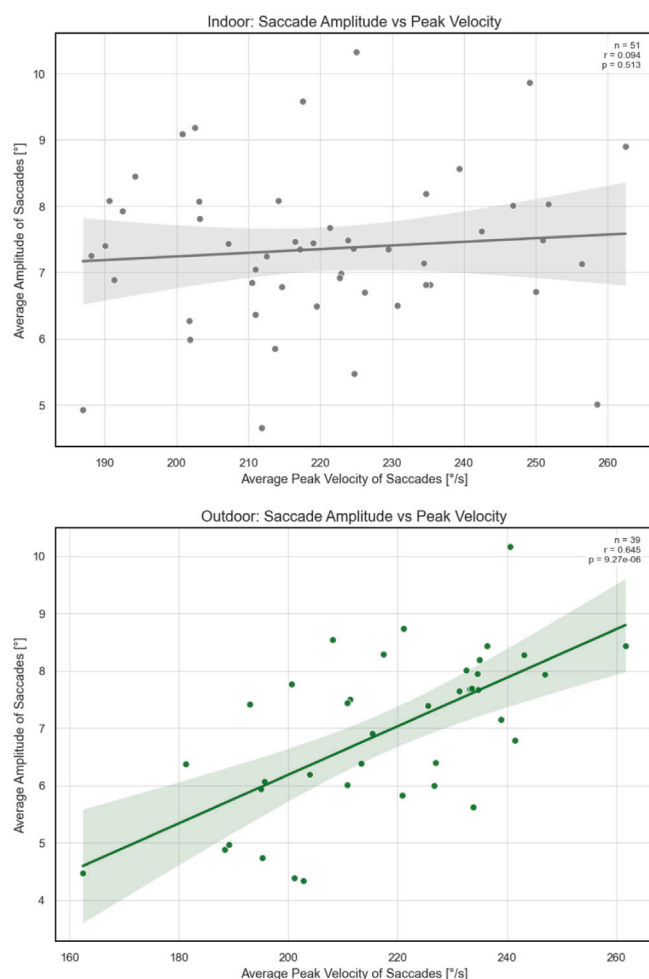


Fig. A.11. Relationship between average saccade peak velocity and average saccade amplitude in indoor and outdoor environments. Each data point represents one measurement, and solid lines indicate linear regression fits. The upper panel shows combined indoor scenes, where no clear linear relationship between average peak velocity and amplitude is observable, despite the classical main-sequence prediction of a positive association between these metrics. In contrast, the lower panel depicts outdoor environments, where a positive linear relationship consistent with the saccadic main sequence is evident. These findings suggest that, when averaged metrics are considered, an amplitude–velocity association is evident in outdoor conditions but not in indoor recordings.

5. Conclusion and outlook

With this study, we set out to examine whether indoor and outdoor environments influence eye-movement behavior under naturalistic conditions. In summary, our findings indicate differences in saccade behavior during the initial scene-exploration phase, which were not evident later in the viewing period, while fixation metrics showed no comparable pattern. Considering the whole measurement period, the two conditions differed primarily in terms of metric correlations and pupil size. The larger pupil size observed indoors is consistent with the substantial conditional differences in measured light characteristics.

Beyond its specific findings, the present study contributes methodologically to naturalistic eye-tracking research. It shows that oculomotor metrics can be measured using wearable eye-tracking equipment while participants experience real-world indoor and outdoor stimuli, enabling physiological and behavioral investigations under controlled

field conditions. Using a reproducible workflow with reduced confounding variables under natural viewing conditions, this study illustrates an approach that balances ecological validity with experimental control, thereby helping to bridge laboratory and field research.

CRedit authorship contribution statement

Alexander Hirsch: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Data curation, Conceptualization. **Aenne Briemann:** Writing – review & editing, Methodology, Investigation, Conceptualization. **Niloufar Tabandeh:** Writing – review & editing, Resources, Conceptualization. **Manuel Spitschan:** Writing – review & editing, Supervision, Resources, Project administration, Methodology, Funding acquisition, Conceptualization.

Ethical approval

This study was reviewed and approved by the TUM Ethics Committee (approval no. 2023-231-1-S-KH).

Acknowledgments

This work was supported by a Max Planck Research Group awarded to M.S.

The authors thank the “Translational Sensory and Circadian Neuroscience” research group for valuable advice and helpful discussions that helped shape this work.

Appendix. Additional plots, tables, questionnaires

See the supplementary figures and tables above.

List of abbreviations

- AAS** Average amplitude of saccades.
- ADF** Average duration of fixations.
- AFF** Average frequency of fixations.
- AFS** Average frequency of saccades.
- APVS** Average peak velocity of saccades.
- CI** Confidence interval.
- dpt** Diopters.
- ETSSR** Eye tracking sample success rate.
- IME** Indoor measurement environment.
- ipRGC** intrinsically photosensitive retinal ganglion cells.
- KSS** Karolinska Sleepiness Scale.
- OME** Outdoor measurement environment.
- P-CR** pupil-centred corneal reflection.
- PLR** Pupillary light reflex.
- PSQI** Pittsburgh Sleep Quality Index.

Post-measurement questionnaire

1. How beautiful was the scene to you?

1 = not beautiful at all

10 = extremely beautiful

1	2	3	4	5	6	7	8	9	10
---	---	---	---	---	---	---	---	---	----

2. How would you rate the brightness of the scene?

1 = dimly lit room

10 = bright sunny day

1	2	3	4	5	6	7	8	9	10
---	---	---	---	---	---	---	---	---	----

3. How visually comfortable did you find the scene?

1 = No discomfort

3 = Some discomfort

5 = Moderate discomfort

7 = Uncomfortable

10 = Very uncomfortable

1	2	3	4	5	6	7	8	9	10
---	---	---	---	---	---	---	---	---	----

4. How would you rate the complexity of the scene (e.g. colour-contrasts, different object-identities, moving objects)?

1 = Very simple (e.g., homogeneous white wall)

10 = Very complex

1	2	3	4	5	6	7	8	9	10
---	---	---	---	---	---	---	---	---	----

5. How sleepy do you currently feel?

1 = Extremely alert

2 = Very alert

3 = Alert

4 = Rather alert

5 = Neither

6 = Some signs of sleepiness

7 = Sleepy, but no effort to keep alert

8 = Sleepy, some effort to keep alert

9 = Very sleepy, great effort to keep alert, fighting sleep

10 = Extremely sleep, can't keep awake

1	2	3	4	5	6	7	8	9	10
---	---	---	---	---	---	---	---	---	----

Data availability

Data are available openly. Videos, images and metadata of each measurement are available on FigShare (Indoor scene 1: <https://doi.org/10.6084/m9.figshare.30518291>; Indoor scene 2: <https://doi.org/10.6084/m9.figshare.30520733>; Indoor scene 3: <https://doi.org/10.6084/m9.figshare.30520892>; Outdoor scene 1: <https://doi.org/10.6084/m9.figshare.30520829>; Outdoor scene 2: <https://doi.org/10.6084/m9.figshare.30520925>; Outdoor scene 3: <https://doi.org/10.6084/m9.figshare.30520979>). Code is available under the GPL-3.0 License on GitHub (https://github.com/tscnlab/HirschEtAl_VisRes_2026) and archived on Zenodo (<https://doi.org/10.5281/zenodo.19434785>).

References

- Anderson, N., Ort, E., Wouter, K., Meeter, M., & Donk, M. (2015). It depends on when you look at it: salience influences eye movements in natural scene viewing and search early in time. *Journal of Vision*, 15(5), <http://dx.doi.org/10.1167/15.5.9>.
- Bahill, A., Clark, M., & Stark, L. (1975). The main sequence, a tool for studying human eye movements. *Mathematical Bioscience*, 24(3-4), 191–204.
- Blackhart, G., Brown, K., Clark, T., Pierce, D., & Kelsys, S. (2012). Assessing the adequacy of postexperimental inquiries in deception research and the factors that promote participant honesty. *Behavioural Research*, 44, 24–40. <http://dx.doi.org/10.3758/s13428-011-0132-6>.
- Brielmann, A., Nuzzo, A., & Pelli, D. (2021). Beauty, the feeling. *Acta Psychologica*, 219, <http://dx.doi.org/10.1016/j.actpsy.2021.103365>.
- Buswell, G. (1935). *How people look at pictures: A Study of the psychology and perception in art* (1st ed.). Chicago: University of Chicago Press.
- Chellappa, S., Steiner, R., Blattner, P., Oelhafen, P., Götz, T., & Cajochen, C. (2011). Non-visual effects of light on melatonin, alertness and cognitive performance: Can blue-enriched light keep us alert. *PLoS One*, 6(1), <http://dx.doi.org/10.1371/journal.pone.0016429>.
- Coors, A., Merten, N., D., Ward, Schmidt, M., & Bretler, U. (2022). Strong age but weak sex effects in eye movement performance in the general adult population: Evidence from the Rhineland study. *Vision Research*, 178, 124–133. <http://dx.doi.org/10.1016/j.visres.2020.10.004>.
- DiCriscio, A., Hu, Y., & Troiani, V. (2018). Task-induced pupil response and visual perception in adults. *PLoS ONE*, 13(12), <http://dx.doi.org/10.1371/journal.pone.0209556>.
- Drewes, J., Feder, S., & Einhaeuser, W. (2021). Gaze during locomotion in virtual reality and the real world. *Frontiers in Neuroscience*, 15, <http://dx.doi.org/10.3389/fnins.2021.656913>.
- Duchowski, A. (2011). *Eye tracking methodology: theory and practice* (3rd ed.). Springer.
- Dunn, M., Alexander, R., Amiebenomo, O., Arblaster, G., Atan, D., Erichsen, J., Ettinger, U., Giardini, M., Gilchrist, I., Hamilton, R., Hessels, R., Hodgins, S., Hooge, I., Jackson, B., Lee, H., Macknik, S., Martinez-Conde, S., McIlreavy, L., Muratori, L., ... Sprenger, A. (2023). Eye tracking: empirical foundations for a minimal reporting guideline (2023 edition). *Behavior Research Methods*, <http://dx.doi.org/10.3758/s13428-023-02187-1>.
- Evans, K., Jacobs, R., Tarduno, J., & Pelz, J. (2012). Collecting and analyzing eye-tracking data in outdoor environments. *Journal of Eye Movement Research*, 5(2):6, 1–19. <http://dx.doi.org/10.16910/jemr.5.2.6>.
- Fontoura, P., & Menu, M. (2021). Visual perception of natural colours in paintings: An eye-tracking study of grünewald's resurrection. *Color Research and Application*, 46, 582–594. <http://dx.doi.org/10.1002/col.22641>.
- Foulsham, T., Walker, E., & Kinstone, A. (2011). The where, what and when of gaze allocation in the lab and the natural environments. *Vision Research*, 51, 1920–1931. <http://dx.doi.org/10.1016/j.visres.2011.07.002>.

- Gibaldi, A., & Sabatini, S. (2021). The saccade main sequence revisited: A fast and repeatable tool for oculomotor analysis. *Behavior Research Methods*, 53(1), 167–187. <http://dx.doi.org/10.3758/s13428-020-01388-2>.
- Gidlöf, K., Anikin, A., Lingonblad, M., & Wallin, A. (2017). Looking is buying. How visual attention and choice are affected by consumer preferences and properties of the supermarket shelf. *Appetite*, 116, 29–38. <http://dx.doi.org/10.1016/j.appet.2017.04.020>.
- Gray, M., Blangero, A., Herman, J., Wallman, J., & Harwood, M. (2014). Adaptation of naturally paced saccades. *Journal of Neuroscience*, 111(11), 2343–2354. <http://dx.doi.org/10.1152/jn.00905.2013>.
- Guilford, J., & Smith, P. (1959). A system of color-preferences. *American Journal of Psychology*, 72(4), 487–502. <http://dx.doi.org/10.2307/1419491>.
- Guo, Y., Helmer, J., Graupner, S., & Pannasch, S. (2022). Eye movement patterns in complex tasks: Characteristics of ambient and focal processing. *PLoS ONE*, 17(11), <http://dx.doi.org/10.1371/journal.pone.0277099>.
- Harris, P., Taylor, R., Minor, B., Elliot, V., & Kirby, J. (2019). The REDCap consortium: Building an international community of software platform partners. *Journal of Biomedical Informatics*, 95, <http://dx.doi.org/10.1016/j.jbi.2019.103208>.
- Harris, P., Taylor, R., Thielke, R., Fernandez, M., & Conde, J. (2009). Research electronic data capture (REDCap) – a metadata-driven methodology and workflow process for providing translational research informatics support. *Journal of Biomedical Informatics*, 42(2), 377–381. <http://dx.doi.org/10.1016/j.jbi.2008.08.010>.
- Hayhoe, M., McKinney, T., Chajka, K., & Pelz, J. (2012). Predictive eye movements in natural vision. *Experimental Brain Research*, 217, 125–136. <http://dx.doi.org/10.1007/s00221-011-2979-2>.
- Henderson, J., & Hollingworth, A. (1999). The role of fixation position in detecting scene changes across saccades. *Psychological Science*, 10(5), 438–443. <http://dx.doi.org/10.1111/1467-9280.00183>.
- Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., & Van de Weijer, J. (2011). *Eye tracking: A comprehensive guide to methods and measures* (1st ed.). Oxford University Press.
- Holmqvist, K., Örbom, S., Hooge, I., Niehorster, D., Alexander, R., Andersson, R., Benjamins, J., Blignaut, P., Brouwer, A., Chuang, L., Dalrymple, K., Drieghe, D., Dunn, M., Ettinger, S., Fiedler, S., Foulsham, T., van der Geest, J., Witzner Hansen, D., Hutton, S., ... Hessels, R. (2022). Eye tracking: empirical foundations for a minimal reporting guideline. *Behavior Research Methods*, 55, 364–416. <http://dx.doi.org/10.3758/s13428-021-01762-8>.
- Kapitaniak, B., Walczak, M., Kosobudzki, M., Józwiak, Z., & Bortkiewicz, P. (2015). Application of eye-tracking in the testing of drivers: A review of research. *International Journal of Occupational Medicine and Environmental Health*, 28(6), 941–954. <http://dx.doi.org/10.13075/ijomh.1896.00317>.
- Kay, K., Bonnen, K., Denison, R., Arcaro, M., & Barack, D. (2023). Tasks and their role in visual neuroscience. *Neuron*, 111(11), 1697–1713. <http://dx.doi.org/10.1016/j.neuron.2023.03.022>.
- Kredel, R., Hernandez, J. E., Hossner, & Zahno, S. (2023). Eye-tracking technology and the dynamics of natural gaze behavior in sports: an update 2016–2022. *Frontiers in Psychology*, 14, <http://dx.doi.org/10.3389/fpsyg.2023.1130051>.
- Land, M. (2009). Vision, eye movements, and natural behavior. *Visual Neuroscience*, 26(1), 51–62. <http://dx.doi.org/10.1017/S0952523808080899>.
- Lee, H., & Yeh, S. (2021). Blue-light effects on saccadic eye movements and attentional disengagement. *Attention, Perception, & Psychophysics*, 83, 1713–1728. <http://dx.doi.org/10.3758/s13414-021-02250-z>.
- Lucas, R. J., Douglas, R., & Foster, R. (2001). Characterization of an ocular photopigment capable of driving pupillary constriction in mice. *Nature Neuroscience*, 4, 621–626. <http://dx.doi.org/10.1038/88443>.
- Lucas, R. J., Freedman, M., Munoz, M., & Foster, R. (1999). Regulation of the mammalian pineal by non-rod, non-cone, ocular photoreceptors. *Science*, 284(5413), 505–507. <http://dx.doi.org/10.1126/science.284.5413.505>.
- McDougal, D., & Gamlin, P. (2010). The influence of intrinsically photosensitive retinal ganglion cells on the spectral sensitivity and response dynamics of the human pupillary light reflex. *Vision Research*, 50(1), 72–87. <http://dx.doi.org/10.1016/j.visres.2009.10.012>.
- McManus, I., Jones, A., & Cottrell, J. (1981). The aesthetics of colour. *Perception*, 10(6), 651–666. <http://dx.doi.org/10.1068/p100651>.
- Meteostat 2024. <https://meteostat.net/de/place/de/tubingen?s=D4294&t=2024-07-14/2024-07-30>. Accessed: 2024-08-15.
- Nasiopoulos, E., Risko, E., Foulsham, T., & Kingstone, A. (2015). Wearable computing: Will it make people prosocial? *British Journal of Psychology*, 106(2), 209–216. <http://dx.doi.org/10.1111/bjop.12080>.
- Palmer, S., Schloss, K., & Sammartino, J. (2012). Visual aesthetics and human preference. *Annual Review of Psychology*, 64, 77–107. <http://dx.doi.org/10.1146/annurev-psych-120710-100504>.
- Pannasch, S., Helmer, J., Roth, K., Herbold, A., & Walter, H. (2008). Visual fixation durations and saccade amplitudes: Shifting relationship in a variety of conditions. *Journal of Eye Movement Research*, 2(2), 1–19. <http://dx.doi.org/10.16910/jemr.2.2.4>.
- Pokorny, J., Smith, V., & Lutze, M. (1987). Aging of the human lens. *Applied Optics*, 26(8), 1437–1440. <http://dx.doi.org/10.1364/AO.26.001437>.
- Risko, E., & Kingstone, A. (2011). Eyes wide shut: implied social presence, eye tracking and attention. *Attention, Perception, & Psychophysics*, 73(2), 291–296. <http://dx.doi.org/10.3758/s13414-010-0042-1>.
- Scott, N., Le, D., Becken, S., & Connolly, R. (2020). Measuring perceived beauty of the great barrier reef using eye-tracking technology. *Current Issues in Tourism*, 23(20), 2492–2502. <http://dx.doi.org/10.1080/13683500.2019.1626812>.
- Shishegar, N., & Boubekri, M. (2016). Natural light and productivity: Analyzing the impacts of daylighting on students' health and alertness. *International Journal of Advances in Chemical Engineering and Biological Sciences*, 3(1), 72–77.
- Spitschan, M. (2019). Melanopsin contributions to non-visual and visual function. *Current Opinion in Behavioral Sciences*, 30, 67–72. <http://dx.doi.org/10.1016/j.cobeha.2019.06.004>.
- Stein, I., Jossberg, H., & Gruber, H. (2022). Investigating visual expertise in sculpture: A methodological approach using eye tracking. *Journal of Eye Movement Research*, 15(2):5, <http://dx.doi.org/10.16910/jemr.15.2.5>.
- Stuart, S., Hickey, A., Vitorio, R., Welman, K., Foo, S., Keen, D., & A., Godfrey (2019). Eye-tracker algorithms to detect saccades during static and dynamic tasks: a structured review. *Physiological Measurement*, 40(2), <http://dx.doi.org/10.1088/1361-6579/ab02ab>.
- Stuart, S., Parrington, L., Martini, D., Popa, B., Fino, P., & King, L. (2019). Validation of a velocity-based algorithm to quantify saccades during walking and turning in mild traumatic brain injury and healthy controls. *Physiological Measurement*, 40(4), <http://dx.doi.org/10.1088/1361-6579/ab159d>.
- Tatler, B., Hayhoe, M., Land, M., & Ballard, D. (2011). Eye guidance in natural vision: Reinterpreting salience. *Journal of Vision*, 11(5):5, 1–23. <http://dx.doi.org/10.1167/11.5.5>.
- Unema, P., Pannasch, S., Joos, M., & Velichkovsky, B. (2005). Time course of information processing during scene perception: The relationship between saccade amplitude and fixation duration. *Visual Cognition*, 12(3), 473–494. <http://dx.doi.org/10.1080/1350628044000409>.
- Vandewalle, G., Baeteau, E., Phillips, C., Degueldre, C., Moreau, V., Sterpenich, V., Albouy, G., Darsaud, A., Desseilles, M., Dang-Vu, T., Peigneux, P., Luxen, A., Dijk, D., & Maquet, P. (2006). Daytime light exposure dynamically enhances brain responses. *Current Biology*, 16, 1616–1621. <http://dx.doi.org/10.1016/j.cub.2006.06.031>.
- Yarbus, A. (1967). *Eye movements and vision* (1st ed.). NY: Plenum Press.
- Zwierko, T., Redondo, B., Jedziniak, W., Molina, R., Jiménez, R., & Vera, J. (2023). Gaze behaviour during multiple object tracking is dependent on binocular vision integrity. *Ophthalmic and Physiological Optics*, 44, 1–9. <http://dx.doi.org/10.1111/opo.13225>.