Behavior Research Methods (2023) 55:364–416
https://doi.org/10.3758/s13428-021-01762-8

Eye tracking: empirical foundations for a minimal reporting guideline

Kenneth Holmqvist[1,2,3] · Saga Lee Örbom[3] · Ignace T. C. Hooge[4] · Diederick C. Niehorster[5] · Robert G. Alexander[6] · Richard Andersson[7] · Jeroen S. Benjamins[4,8] · Pieter Blignaut[2] · Anne-Marie Brouwer[9] · Lewis L. Chuang[10,11] · Kirsten A. Dalrymple[12] · Denis Drieghe[13] · Matt J. Dunn[14] · Ulrich Ettinger[15] · Susann Fiedler[16] · Tom Foulsham[17] · Jos N. van der Geest[18] · Dan Witzner Hansen[19] · Samuel B. Hutton[20] · Enkelejda Kasneci[21] · Alan Kingstone[22] · Paul C. Knox[23] · Ellen M. Kok[24,25] · Helena Lee[26] · Joy Yeonjoo Lee[27] · Jukka M. Leppänen[28] · Stephen Macknik[6] · Päivi Majaranta[29] · Susana Martinez-Conde[6] · Antje Nuthmann[30] · Marcus Nyström[31] · Jacob L. Orquin[32,33] · Jorge Otero-Millan[34] · Soon Young Park[35] · Stanislav Popelka[36] · Frank Proudlock[37] · Frank Renkewitz[38] · Austin Roorda[34] · Michael Schulte-Mecklenbeck[39,40] · Bonita Sharif[41] · Frederick Shic[42,43] · Mark Shovman[44,45] · Mervyn G. Thomas[37] · Ward Venrooij[46] · Raimondas Zemblys[47] · Roy S. Hessels[4]

Accepted: 29 November 2021 / Published online: 6 April 2022

Abstract

In this paper, we present a review of how the various aspects of any study using an eye tracker (such as the instrument, methodology, environment, participant, etc.) affect the quality of the recorded eye-tracking data and the obtained eye-movement and gaze measures. We take this review to represent the empirical foundation for reporting guidelines of any study involving an eye tracker. We compare this empirical foundation to five existing reporting guidelines and to a database of 207 published eye-tracking studies. We find that reporting guidelines vary substantially and do not match with actual reporting practices. We end by deriving a minimal, flexible reporting guideline based on empirical research (Section "An empirically based minimal reporting guideline").

Keywords Eye movements · Eye tracking · Data quality · Reporting guidelines · Reporting standards · Reporting practices · Replicability · Reproducibility

Introduction

Eye tracking is a method used to investigate eye movements, gaze behaviour, and pupil dilation in many different research fields (e.g. perception, attention, memory, reading, psychology, ophthalmology, neuroscience, human–computer interaction, animal research, human factors, consumer behaviour, optometry etc., see Duchowski, 2002; Kowler, 2011; Liversedge et al., 2011; Majaranta, 2011; Rayner, 1998, for overviews). In addition, there is a belief that eye tracking will soon become a ubiquitous technology in laptops and augmented reality headsets for consumers (e.g. Chuang et al., 2019; Clay et al., 2019). Eye tracking is widespread and likely to become more so in the future.

While eye tracking may be used in a variety of research fields to answer very different questions, many methodological aspects appear to be shared, such as the eye-tracker models used, or the algorithms for processing and analysing recorded eye-tracking data. One therefore might expect that the part of the method sections describing the eye-tracking setup, and the processing and analysis of the data it yields, are similar across very different research fields (such as human factors research and neuroscience).

However, recent research suggests that this is not necessarily the case. For example, in many studies using an eye tracker, reporting the quality of the eye-tracking data obtained is not common practice (see e.g. Hessels & Hooge, 2019; Holmqvist et al., 2012). Moreover, although

Authors Alexander through Zemblys are listed alphabetically.

✉ Kenneth Holmqvist
kenneth.holmqvist@ur.de

Extended author information available on the last page of the article.

there exist a number of reporting guidelines for research using eye trackers (e.g. Carter & Luke, Fiedler et al., 2019; McConkie, 1981; Oakes, 2010; Strohmaier et al., 2020), existing guidelines differ substantially from one another, are based on consensus decisions within a small group of authors or researchers, and to the best of our knowledge are not widely used.

The present paper was initiated after several large-scale meetings between eye-tracking researchers from many different disciplines. In these meetings, it was established that there is a need for guidance in what to report in a study using an eye tracker. However, the needs on what should be reported may differ substantially between research or applied fields. Therefore, the first step was to combine previous research into an empirical foundation for any future reporting standard.

Evidence-based reporting guidelines are essential for at least three reasons.

1. Being expected to report a specific set of features of the experiment may help researchers with *planning and designing* their studies, as they will be more aware of preparing and collecting information that needs to be reported at a later stage.
2. The adoption of reporting guidelines, leading to sufficient detail in the reported methods of a study, may allow reviewers and future readers to *assess the validity* of that study's claims.
3. Following reporting guidelines may assist authors in providing sufficient detail about a study to enable other researchers to *reproduce (and potentially replicate)* a study. A well-known study on replicability estimated that a mere one third of the findings in psychological science are replicable, qualifying this as a 'replication crisis' (Open Science Collaboration, 2015). In eye-tracking research specifically, replication may be particularly hampered by an over-reliance on the performance of the eye trackers and their default algorithms and settings.

Note that we distinguish between reporting guidelines, which offer researchers the possibility to make informed choices in what to report, and reporting standards which prescribe mandatory reporting items, approved by one or another authority. We here deliver the empirical foundation and derived from this what should minimally be reported according to empirical research. Our efforts may be followed up by e.g. consensus-based approaches to deliver formal reporting standards. We expect these to differ between for example fundamental research fields and clinical applications, due to different considerations with regard to e.g. safety, ethics, legal requirements, researcher background knowledge, and nature of the research field.

In what follows, we review the existing literature with regard to the following central question: How do the various aspects of a study using an eye tracker (such as the instrument, methodology, environment, participant, etc.) affect the quality of the eye-tracking data obtained, or the eye-movement and gaze measures? We contrast what has been shown to be relevant against what already existing reporting guidelines prescribe and against an existing database of 207 publications of what researchers have reported in eye-tracking research on judgement and decision-making (Fiedler et al., 2019). This review of empirical research forms the basis of our minimal reporting guideline.

As will become apparent, a large proportion of the studies that we discuss are conducted from the perspective of eye-tracking data quality. Why is that important? Better data quality may result in, for example, a lower attrition rate, fewer subjects, shorter experimental sessions, more statistical power, better diagnosis, etc. In other words, it means getting more out of each measurement, observation, or experimental session. The data quality approach entails scrutinising aspects of the procedure of an eye-tracking experiment and improving them such that the quality of the eye-tracking data may be increased. In studies on eye-tracking data quality, the eye trackers or aspects of the eye-tracking data analysis are the target of interest, analogous to the focus on specific traits of humans or animals in a psychological study. The goal can, for example, be to understand how the data from an eye tracker changes when the illuminance of the room, or the distance between a human's eye and the eye tracker, is varied. Likewise, researchers may be interested in the relationship between aspects of the eye-tracker signal and the age, eye colour, or eye physiology of the human or animal being tracked. Often, the effects of such environmental, setup-related or participant-related factors are quantified in terms of eye-tracking data quality (see e.g. Ehinger et al., 2019; Hessels et al., 2015; Holmqvist, 2015; Nyström et al., 2013).

Also, researchers may be interested in how the quality of eye-tracking data affects eye-movement measures when fed through a particular aspect of the eye-tracking data analysis pipeline (Fig. 1). For example, researchers may be interested in how a 'fixation duration' as reported by a fixation-detection algorithm is affected by the precision (Table 1) in the gaze-position signal, or how a measure derived from an area-of-interest (Table 1) analysis may be affected by the settings of a fixation-detection algorithm.

This paper may be useful for at least two types of readers: researchers interested in eye tracking *per se*, and researchers for whom eye tracking is not their core business but who use eye tracking as one of the tools in their research toolbox.

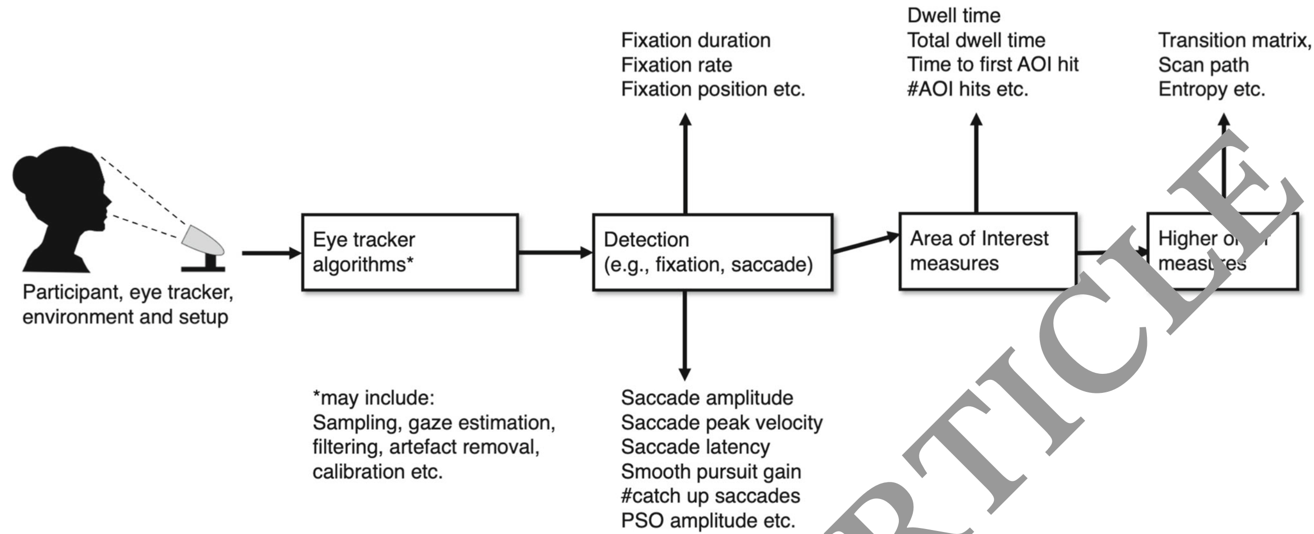

Fig. 1 From eye orientation to higher-order eye-tracking measures. This is a crude division of the process from eye orientation to higher-order eye-tracking measures. There may be cases where a more fine-grained division is applicable

Structure of this paper

For eye-tracking researchers at all levels of experience to follow along, it is vital that we clarify a number of important terms, among which are the characteristics of eye-tracking data quality, the various eye-tracking methods, and common terms in eye-tracking data processing and analysis. Table 1 lists some common terms and definitions. Figure 1 furthermore depicts a general flow from eye-tracking recording to eye-movement measure.

In Section "Measuring data quality of eye-tracker signals", we briefly explain how the fundamental data quality measures for eye-tracking data are operationalised and calculated. We will use the terms defined in Section "Measuring data quality of eye-tracker signals": accuracy, precision, data loss, latency etc., throughout the paper.

Section "A review of empirical eye-tracking studies as the basis for a reporting guideline", the first of the three subsequent content sections, consists of a scoping review of available research relevant to our question: How do the various aspects of a study using an eye tracker, such as the instrument, methodology, environment, and participant affect (or relate to) the quality of the eye-tracking data obtained, the properties of the eye-tracker signals, or the eye-movement and gaze measures? We furthermore review how the quality of the eye-tracking data and the data processing and analysis methods used may affect eye-movement and gaze measures.

In Section "Reporting practices and existing reporting guidelines", we compare the findings from our scoping review (Section "A review of empirical eye-tracking studies as the basis for a reporting guideline") against five existing reporting guidelines for research with an eye tracker, and against actual reporting practices. Conveniently for the latter, four of our co-authors have coded the frequencies of the actual reporting of 99 common aspects of eye-tracking experiments from 207 published studies using eye trackers in research on decision-making. See Fiedler et al. (2019) for an earlier presentation of the same data.

Finally, Section "An empirically based minimal reporting guideline" presents a summary of what is empirically relevant to report. This summary could serve as a flexible reporting guideline, offering researchers the ability to make informed choices about what to report for their particular study. This final section is written from the point of view that any aspect of a study that matters to the outcome of a study should be reported.

Measuring data quality of eye-tracker signals

Eye-tracking data quality is often characterised by three measures: accuracy, precision, and data loss (see Fig. 2). Accuracy refers to the difference between the true gaze position and the gaze position reported by the eye tracker. Precision refers to the reproducibility of a gaze position by the eye tracker when the true gaze position does not change. Finally, data loss refers to the amount of data lost in an eye-tracker signal. However, another data quality concept is sometimes reported: system latency, which refers to the time it takes to produce gaze coordinates from the sensor data (camera image, for instance). Below, we will give operationalisations for these data quality concepts.

Table 1 List of some common terms used in this paper

Term	Definition
Pupil-Corneal Reflection (P–CR)	A video-based eye-tracking method. Gaze direction is calculated from the corneal reflection (CR) coordinate and the pupil (P) coordinate in the eye camera image, originally by subtracting CR from P. The P–CR method currently dominates the eye-tracking market (Section "Eye-tracking methods: Similarities and differences").
Dual-Purkinje Imaging (DPI)	The DPI system is an analogue eye tracker that bases its estimation of gaze on the relative movement of an infrared reflection in the cornea (P1) vs. a reflection at the back of the crystalline lens (P4) (Section "Eye-tracking methods: Similarities and differences").
Electrooculography (EOG)	EOG is a method used to estimate eye orientation from the difference in electrical potential between the front and back of the human eye (Section "Eye-tracking methods: Similarities and differences").
Scleral search coils	Scleral search coil eye tracking involves attaching a silicon-enclosed copper wire onto the sclera of the eyeball. The participant is then placed in an oscillating magnetic field, and the amplitude of induction voltage can be taken to represent eye orientation with respect to the magnetic field (Section "Eye-tracking methods: Similarities and differences").
Limbus tracking	Limbus trackers use the border between the iris and the sclera to track gaze and eye movements (Section "Eye-tracking methods: Similarities and differences").
Retinal image-based tracking	Image processing of retinal features (such as blood vessels) are used to register very small eye movements (Section "Eye-tracking methods: Similarities and differences").
Calibration	The process of mapping eye-tracker measurements to physical gaze direction or gaze position. This can be done e.g. using fitting procedures or by using physically and biologically plausible models or a combination of the two (Section "Calibration and accuracy").
Gaze direction	The vectors from an eye, both eyes, or a cyclopean eye, which describe the line of sight.
Gaze position	The location of gaze in a measurement plane or space. Gaze is also termed point of regard.
Sampling frequency	For most eye trackers, the number of gaze direction or gaze position measurements made per second (Hz), determined by the eye camera or AD-convertor of the system. Sampling frequencies vary from around 10Hz in some web-camera eye trackers to 10000Hz in some recordings.
Accuracy	The difference between the true gaze position (assumed or instructed) and the gaze position reported by the eye tracker. Typically reported as error in degrees of visual angle where a larger error indicates poorer accuracy. Inaccuracy is also known as systematic error, while accuracy can be called trueness (Sections "Measuring data quality of eye-tracker signals" and "Calibration and accuracy").
Precision	The reproducibility of a gaze position from one sample to the next, assuming a stable gaze position. Typically reported as RMS or STD (or both). Greater error indicates poorer precision. Imprecision is also known as random or variable error (Sections "Measuring data quality of eye-tracker signals" and "Signal properties and processing").
Resolution	The just noticeable difference in a signal, which in eye tracking is measured as the smallest reliably detected eye movement (or rotation of an artificial eye) that an eye tracker can resolve (Holmqvist & Blignaut, 2020; Poletti & Rucci, 2016; Crane & Steele, 1985).
Data loss	The amount of eye-tracking data lost. It is the counterpart of what is commonly called tracking ratio or availability of eye-tracking data (Sections "Measuring data quality of eye-tracker signals" and "Signal properties and processing").
System latency	The duration from when an actual eye movement is made, until the corresponding gaze sample is output by the eye tracker, made accessible for instance to affect change on a monitor in a gaze-contingent study (Sections "Measuring data quality of eye-tracker signals" and "Signal properties and processing"). Latency is sometimes referred to as the *end-to-end delay* or *temporal accuracy* (Reingold, 2014, p. 641).

Table 1 (continued)

Term	Definition
Event detection, classification, and calculation	Usually refers to the segmentation of an eye-tracker signal into meaningful segments (or events) with a start, end, and duration, and calculation of event properties. Meaningful segments can be 'fixation', 'saccade', 'smooth pursuit', or other terms (Section "Fixation and saccade detection"). Terminology varies (Hessels et al., 2018).
Area of Interest (AOI)	A segment of a stimulus space (often defined by screen pixel boundaries in monitor-based eye tracking) that identifies a portion of the stimulus that is meaningful in the experimental design of a study (such as eyes and mouth areas in face perception, a pack shot in marketing research, or a target in a visual search experiment). In eye-tracking analysis AOIs allow for the calculation of commonly reported dependent measures; for instance, the number of times or amount of time gaze is within a specific AOI, or the number of transitions between AOIs (Section "Area-of-interest (AOI) measures").

Operationalizing accuracy requires that the participants look at a set of fixation targets on screen, often just after having completed the calibration. The accuracy measurement is commonly known as a validation procedure. Research on the positioning of validation points is lacking, but accuracy values may be underestimated (better) if the same points are used for validation as for calibration, or if only part of the stimulus is covered by validation points.

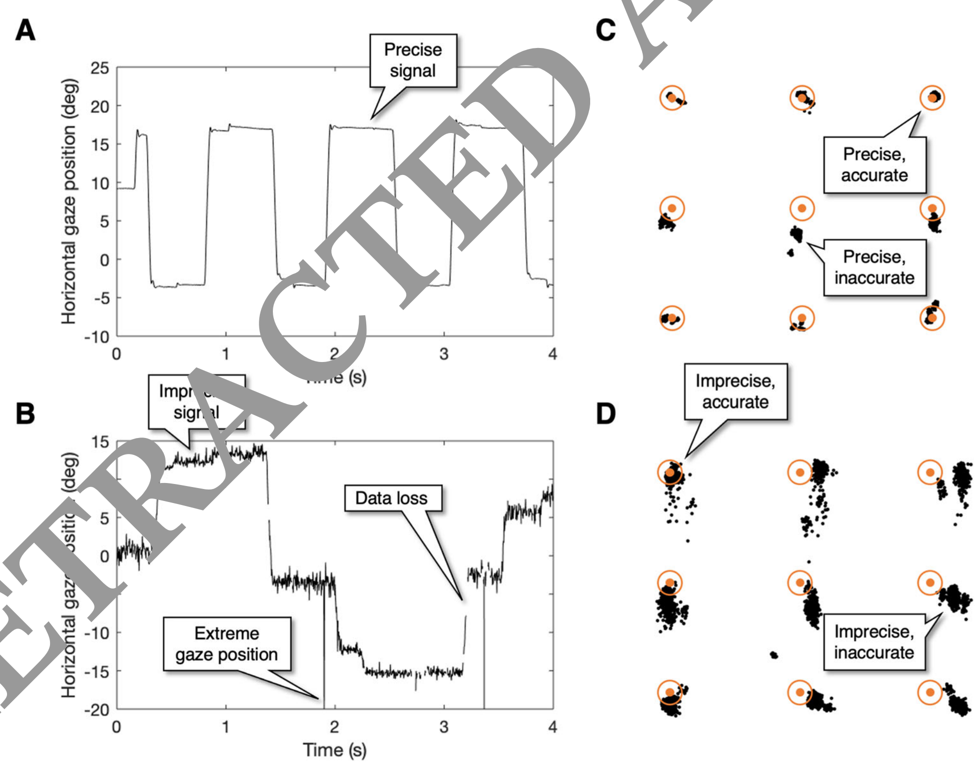

Fig. 2 Characteristics of eye-tracking data quality. **A** Horizontal gaze position (in Fick, 1854, coordinates, see Haslwanter (1995)) of the right eye as a function of time. The gaze position was recorded from an adult participant with an EyeLink 1000 by Hooge et al. (2015). Call-outs indicate the relatively precise gaze-position signal (compared with panel B). **B** Horizontal gaze position in Fick coordinates of the right eye as a function of time. The gaze position was recorded from an infant participant with the Tobii TX300 by Hessels et al. (2016). *Call-outs* indicate the relatively imprecise gaze-position signal (compared with panel A), short gaps in the gaze-position signal (data loss), and an extreme gaze position reported by the eye tracker. The extreme gaze position is interesting because it can be considered an aspect of eye-tracking data quality not captured in the measures accuracy, precision, or data loss. **C**, **D** Gaze position signals (*black dots*) in a 2D representation, i.e. as if on a screen. Gaze position signals were recorded from adult participants by Hooge et al. (2019). Gaze position samples with high velocity were removed such that saccades are not visible. Orange markers represent validation targets. They are positioned to illustrate good/poor accuracy and do not correspond to the location of the actual validation targets in the experiment by Hooge et al. (2019). Call-outs indicate validation targets with corresponding precise and accurate, precise and inaccurate, imprecise and accurate, and imprecise and inaccurate gaze position signals, respectively. Note that the qualifications 'precise', 'imprecise', 'accurate', and 'inaccurate' are relative here and are often quantified

Additionally, a second validation procedure and accuracy calculation at the end of the experiment might be beneficial to be able to detect changes in accuracy between experiment start and end.

Accuracy may be calculated as the mean difference between the reported gaze locations near a validation target and the actual position of that validation target. The achieved accuracy thus critically depends on participant gaze during calibration. Instructing the participant to confirm when s/he is looking at the target (Nyström et al., 2013) or letting the participant adjust the parameters of the calibration while getting feedback from online gaze data (Poletti & Rucci, 2016) may improve accuracy.

When participants produce a saccade to a validation target, they may under- or overshoot the target, make a small correction and only then fixate the target. A method is needed to find the period when the participant looks at the validation target. Manufacturers have built such selection methods into their software for calibration and validation, and some researchers have also investigated and used various sample selection principles (e.g. Hessels et al., 2015; Holmqvist, 2015; Niehorster et al., 2020c; Van der Stigchel et al., 2017). We refer to these studies for details.

Precision of the gaze position signal may be operationalised in different ways, such as the Root Mean Square sample-to-sample deviation (RMS-S2S) of a segment of gaze data collected when the participants' gaze is fixed on a validation target. Following Niehorster et al. (2020c), RMS-S2S is calculated as in Eq. 1:

$$\text{RMS-S2S} = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n-1}(x_i - x_{i+1})^2 + (y_i - y_{i+1})^2} \quad (1)$$

where (x_i, y_i) and (x_{i+1}, y_{i+1}) are successive gaze positions during a fixation. Another measure would be the standard deviation (STD) of that segment or the Bivariate Contour Ellipse Area (BCEA, Crossland and Rubin, 2002; Steinman, 1965). As detailed in Niehorster et al. (2020c), these calculations operationalise different aspects of the gaze signal. Given a stable sampling frequency, this makes the RMS-S2S value of the gaze signal an indicator of noise velocity, which can be compared to the velocity threshold in the event detectors (Section "Fixation and saccade detection"). In contrast, the STD calculation operationalises the dispersion of gaze samples in a segment of data. The dispersion measure STD is calculated as in Eq. 2, where $\overline{x}$ denote the mean of quantity x:

$$\text{STD} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2 + (y_i - \overline{y})^2} \quad (2)$$

The two calculations (1) and (2) can be applied not only to gaze data, but to any sequence of data from an eye tracker, such as pupil and CR position or pupil diameter data to investigate, for instance, the stability of a pupil dilation measurement.

Data loss may be operationalised as the percentage (or proportion) of samples which lack coordinates for the gaze signal. An example of the latter would be an eye tracker that has an advertised sampling frequency of 250Hz but reports only 2000 gaze coordinates during 10 s; this would represent a data loss of 20%. However, there are other operationalisations of data loss that may be useful in some situations: for instance, in some cases, the researcher might wish to count gaze or pupil samples that are missing due to blinks as data loss. Blinks may account for about 2% loss of the total data set (Holmqvist & Andersson, 2017, p. 167). In some cases, gaze shifts outside the tracking range of the eye tracker may count as data loss. In developmental research, where young children are prone to look away from a monitor when they are no longer interested, researchers might wish to exclude periods of looking away from the calculation of data loss (see e.g. Hessels et al., 2015; Wass et al., 2014, for operationalisations of data loss in developmental research).

System latency (also known as temporal accuracy and end-to-end delay, e.g. Reingold, 2014, p. 641) may be operationalised as the average duration from the time of an actual movement of the tracked eye until the recording computer signals that the eye movement has taken place. In a video-based P–CR tracker, the optimal latency is the time from image acquisition to calculated gaze, which takes 1–3 samples (1–3ms in a 1000Hz recording, see Holmqvist & Andersson, 2017, p. 85). Any timing issues in the processes run by the computers involved in the data recording may add latencies. A large variability in the latency may be characterised as poor temporal precision.

Long and variable latencies are problematic for the interpretation of measurements that are assumed to be synchronised: eye tracker and EEG, for instance, or eye tracker and stimulus monitor. The latter is very important in gaze-contingent research, where latencies are reported to be 10–60ms, including the delay to the next retrace of the monitor (Section "Signal properties and processing").

Latencies can be measured in at least the following five ways, some of which require specific equipment and/or software. The first method measures the time until there is an update in the gaze signal. Methods three to five measure latency until a display change has been completed. The second method can be used for either of these two types of measurements.

1. Compare the file of the raw data stream against a video output of the participant's eye (Leppänen et al., 2015) or gaze scanpath (Morgante et al., 2012).
2. Equip an artificial eye with two diodes that act as artificial corneal reflections per IR illuminator, and turn one off while the other diode is turned on, so that the eye appears to move, and then measure the time until a movement is seen in the gaze signal, or until the display changes (Bernard et al., 2007; Holmqvist et al., 2012; Reingold, 2014).
3. Shukla et al. (2011) used a mirror positioned next to the participant's face and a 300Hz high-speed camera, which captured the participant's eye and, through the mirror, the monitor where the stimuli appeared and disappeared.
4. Saunders and Woods (2014) tested gaze-contingent monitors with the EyeLink 1000, by blinding the eye tracker with an infrared pulse and measuring the time until the gaze-contingent monitor changed by recording both the infrared pulse and the monitor with a 1000Hz camera.
5. Hohenstein and Kliegl (2014) measured the latency between saccades and display changes in a gaze contingent study, with a light sensor attached on to the monitor.

As is evident from the operationalisations above, lower values for accuracy, precision, data loss and system latency are better: The ideal value is 0 for each data quality measure. Worse data quality manifests as higher values.

Examples of procedures, formulas, (pseudo)code or links to software for estimating some measures of data quality and effects thereof may be found in e.g. Crossland and Rubin (2002), Blignaut and Beelders (2012), Akkil et al. (2014), Dalrymple et al. (2018), Hessels et al. (2017), Orquin and Holmqvist (2018), Kangas et al. (2020), Niehorster et al., (2020a, b).

A review of empirical eye-tracking studies as the basis for a reporting guideline

We will present our review ordered by the categories Eye-tracking methods, Environment, Setup and geometry, Participant, Calibration, Features of the experiment, Signal processing, Event detection, Area-of-Interest measures, and Higher-order measures. The minimal reporting guideline itself can be found in Section "An empirically based minimal reporting guideline".

Eye-tracking methods: Similarities and differences

Over the past 130 years (e.g. Delabarre, 1898; Lamare, 1892), many methods for eye movement registration have been developed. A recent comprehensive overview is provided by Holmqvist and Andersson (2017, Ch. 4). For other overviews of eye trackers and methods for measuring eye movements, see Hansen and Ji (2010), Duchowski, (2007, pp. 51–59), Ciuffreda & Tannen (1995, pp. 184–205), Young and Sheena (1975), and Ditchburn, (1973, pp. 36–77).

In this section, we describe how characteristics of the eye-tracker signal differ between the measurement techniques and between various eye-tracker models. From the perspective of a researcher embarking on a new project, with a limited budget, each measurement technique is likely to have some advantages and some disadvantages. Within each technique, differences between manufacturer models in data quality and other properties may be found to be large enough to determine the success or failure of the upcoming study.

Table 2 summarises 42 existing cross-comparative benchmarking studies of eye trackers, which we refer the reader to for specific details. In short, these 42 studies inform their readers that data quality often differs very considerably, in very many ways, between eye trackers, while other eye trackers record data with similar quality. The studies in Table 2 may assist in assessing whether an eye tracker can actually produce data of the desired quality, either in preparation for acquiring a system, or when preparing a replication where the eye tracker in the intended replication study differs from the eye tracker in the original publication.

Summarising studies on accuracy and precision, particularly, Holmqvist and Andersson (2017) point out that the difference in distribution of RMS-S2S *precision* values between eye trackers may be up to two orders of magnitude, while in comparison between-subjects differences in precision within each eye tracker tend to be relatively small. In contrast, the distributions of *accuracy* values for each eye tracker overlap considerably between eye trackers (i.e. they have similar accuracy), but exhibit a very wide range within each eye tracker which represents data from people with different eye physiologies, spectacles, and data obtained during fixations in the corner vs central positions of monitors. This suggests that for precision, the eye tracker matters more, while for accuracy: the participant, the calibration and the geometrical setup matter more. This was found for adult human participants in the lab and may differ for infants, animals and difficult recording environments.

As we outline below, irrespective of measurement method: *anything that interferes with obtaining or process-*

Table 2 Comparative benchmarking studies

Eye trackers	Examined	Publication
EyeLink 1000+, EyeLink II, Tobii Pro Spectrum, SMI HiSpeed240, SMI RED250mobile	Small head movements, chinrest, precision, filter on/off, oculomotor drift	Holmqvist et al. (2021)
Fove-0, Varjo VR-1, HTC Vive Pro Eye, EyeLink 1000, EOG	Latency	Stein et al. (2021)
EyeLink 1000+, SMI RED250, SMI REDm, Tobii TX300, Tobii X2-60	Power spectral density, colour of noise, filter on/off, oculomotor drift	Niehorster et al. (2021)
EyeLink1000+, Tobii Pro Spectrum, FLEX	Pupil-size artefact	Hooge et al. (2021)
SMI ETG2, SMI HTC Vive	Accuracy, precision	Pastel et al. (2021)
Tobii Pro Spectrum, EyeLink 1000+	Microsaccades, precision	Nyström et al. (2021)
DPI Gen5.5, SMI HiSpeed 240, SMI HiSpeed 1250, SMI RED250mobile, SMI ETG 2, EyeLink II, EyeLink 1000+, Tobii X2-60, Tobii T120, Tobii TX300, Tobii Pro Spectrum	Accuracy of movement amplitudes, measurement resolution, precision	Holmqvist and Blignaut (2020)
DPI Gen5.5, Tobii Pro Spectrum, EWET1	Accuracy, precision, pupil-size artefact, microsaccade detection, resolution	Holmqvist et al. (2020)
EyeLink 1000+, SMI RED250, SMI REDm, Tobii TX300, Tobii X2-60	Five different precision measures, filters on/off	Niehorster et al. (2020c)
SMI ETG2, Tobii Pro Glasses 2, PupilLabs	Accuracy, precision, data loss	Niehorster et al. (2020b)
EyeLink 1000, Pupil Labs	Accuracy, precision, drift, temporal precision, number of blinks, blink duration	Ehinger et al. (2019)
EyeLink 1000, SMI HiSpeed 1250	Vergence	Hooge et al. (2019)
EyeLink 1000, Stereotracker	Accuracy, precision, main sequence	Barsingerhorn et al. (2018)
EyeLink 1000, EyeTribe	Main sequence modelling, number of saccades	Raynowska et al. (2018)
SMI RED250, Tobii TX300	Fixation durations, number of fixations	van Renswoude et al. (2018)
EyeLink1000+, EyeTribe, SMI REDm, Tobii T60XL, Tobii TX300	Data loss and recovery during head movements	Niehorster et al. (2018)
DPI Gen 5.5, EyeLink1000, SMI HiSpeed 240, HiSpeed 1250, RED250, RED500 and REDm, Tobii TX300, T60 XL, X2-60, LC EyeFollower, EyeTribe	Precision, power spectral density, colour of noise	Wang et al. (2017)
Tobii T60, Tobii T120, Tobii TX300	Saccadic reaction time	Kenward et al. (2017)
DPI Gen 6, Scleral Search Coil	Precision, resolution, drift, noise colour	Ko et al. (2016)
EyeLink 1000, SMI ETG 2	Main sequence, saccade amplitudes	Engbert et al. (2016)
EyeTribe, Tobii EyeX, SeeingMachines faceLAB, SmartEye Pro, SmartEye Aurora	Accuracy, precision, data loss	Funke et al. (2016)
EyeTribe, GazePoint GP3	Pupil diameter	Coyne and Sibley (2016)
EyeTribe, SMI RED250	Accuracy, data loss, fixation count	Popelka et al. (2016)
SMI RED250, EyeTribe	Accuracy, temporal precision	Ooms et al. (2015)

Table 2 (continued)

Eye trackers	Examined	Publication
DPI Gen 5.5, EyeLink1000, SMI HiSpeed 240, HiSpeed 1250, RED250, RED500 and REDm, Tobii TX300, T60 XL, X2-60, LC EyeFollower, EyeTribe	Accuracy, precision, data loss	Holmqvist (2015)
EyeLink 1000, Scleral Search Coil	Precision, microsaccades, oculomotor drift	McCamy et al. (2015)
LC Technologies EyeFollower, SMI RED250, SMI REDm, Tobii T120, TX300 and X2-60	Data loss and recovery during head movements	Hessels et al. (2015)
SMI RED250, SMI RED500, SMI HiSpeed1250, Tobii TX300	Accuracy, calibration	Blignaut et al. (2014)
EyeLink remote, SMI RED60	Number of fixations	Wang et al. (2014)
EyeLink 1000, EyeTribe	Precision, drift	Dalmaijer (2014)
Tobii X120, T120, EyeLink 1000	Pupil foreshortening	Brisson et al. (2013)
EyeLink II and a piezoelectric sensor	Microsaccade amplitudes	McCamy et al. (2013)
SMI RED250, Tobii TX300	Three precision measures, seating distance	Blignaut and Beelders (2012)
EyeLink 1000, Scleral Search Coil	Accuracy, saccade dynamics, microsaccades, pupil-size artefact	Kimmel et al. (2012)
EyeLink 1000, Scleral Search Coil	Pupil-size artefact	Drewes et al. (2012)
EyeLink II, Scleral Search Coil	Saccade dynamics	Lappe-Osthege et al. (2010)
SMI RED50, SMI HED50	Accuracy, drift	Komínková et al. (2008)
Search Coil, Chronos Vision	Accuracy, main sequence, torsion	Houben et al. (2006)
Tobii 1750, ASL, 501, ASL 504	Accuracy, data loss	Nevalainen and Sajaniemi (2004)
Tobii ET-17, LC EyeGaze	Accuracy, data loss, drift	Cheng and Vertegaal (2004)
EyeLink I, Scleral Search Coil	Saccade dynamics	Frens and van der Geest (2002)
DPI Gen 5, Scleral Search Coil	Saccade dynamics	Deubel and Bridgeman (1995)

ing of a feature used in estimating gaze direction (P, CR, P1, P4, limbus, magnetic induction or retinal features) *will affect the data quality of the signal* in the data reported by the eye tracker.

P–CR eye tracking

Video-based P–CR eye tracking was introduced by Merchant (1967). In 2021, camera-based P–CR eye trackers dominate the market almost completely. The P of P–CR eye trackers refers to the pupil centre in the camera image, and the CR to one or more reflection centre(s) in the cornea from infrared illuminators in the eye tracker. P–CR eye trackers estimate gaze direction as a function of the relative positions of P and CR coordinates in the pixel coordinate system of the video image, for instance by subtracting the CR coordinate from the P coordinate. Note that more advanced models have been developed (Hansen & Ji, 2010).

More types and models of P–CR eye trackers are available than for any other measurement technique, and prices vary over a wide range. There exists plenty of software for stimulus presentation, data processing and analysis, and the learning threshold for beginning researchers is lower than for other eye-tracking methods.

Many studies have examined aspects of P–CR eye trackers (Table 2). A host of issues with the feature detection of both pupil and corneal reflection may impair quality of gaze and pupil-size data. As we point out elsewhere, P–CR trackers suffer from the pupil-size artefact (Section "Environment") and the pupil foreshortening artefact (Section "Setup and geometry"). Refraction in the cornea alters the pupil size in the camera image and its position with respect to the limbus (Villanueva & Cabeza, 2008). Pupil occlusion and mascara can interfere with pupil detection. Blue irises tend to result in poorer precision (in dark-pupil eye trackers), which is due to poor contrast between (a dark) pupil and iris in the infra-red light of video-based eye trackers (Section "Participants", and Figure 4.13 in Holmqvist & Andersson, 2017). Combining the pupil with the CR signal to form the P–CR gaze signal may amplify post-saccadic oscillations and overestimate peak saccadic velocity (Hooge et al., 2016).

P–CR eye trackers exhibit clear post-saccadic oscillations (PSOs) (Hooge et al., 2015; Nyström et al., 2013), which make it difficult to draw a clear border between saccade and subsequent fixation, and which has led to the development of event detection algorithms that include PSO detection (Larsson et al., 2013; Nyström & Holmqvist, 2010; Zemblys et al., 2019).

Discussing which technologies could be used for future studies of saccade dynamics, Hooge et al. (2016) reason that variants of CR-tracking without the involvement of the pupil feature could be the preferred future method. However, Holmqvist and Blignaut (2020) reported incorrectly measured amplitudes of small eye movements (below 2°) in all 11 P–CR eye trackers they tested, and suggest that it is due to erroneous calculations of the CR centre by the image processing algorithms in the eye trackers, interacting with the resolution of the eye camera sensor. Other artefacts in the CR signal arise from changes in head position (relative to the eye tracker), which may alter the size and the shape of corneal reflections (Guestrin & Eizenman, 2006). Patterns in the iris may interact with the CR image and change the calculated CR center ([illegible] & Kaufman, 2003). Illumination levels, sampling frequency and the optic lenses in the camera may all affect the CR. Droege and Paulus (2009) point out that the use of low-quality eye cameras may further degrade precision in the gaze signal, due to the slower pixel updating, which makes pixels retain some of the brightness of the passing corneal reflection, leaving a bright trace behind the real reflection, making centre calculation of the CR image more perilous.

DPI eye tracking

The Dual-Purkinje Imaging (DPI) system is an analogue eye tracker that bases its estimation of gaze on the relative movement of the infrared reflection off the cornea (P1) versus the reflection at the back of the crystalline lens (P4), and reports P1, gaze and head translation as voltages (Crane & Steele, 1985). At present, there are around 60 DPI trackers left in the world (Personal communication; Warren Ward). As the DPI produces a continuous signal, it can be digitised to the desired sampling frequency in an AD-converter. Internal bandwidth restrictions limit the maximum sampling frequency to 39.06kHz (Personal communication; Warren Ward).

The DPI used to be the main workhorse of many psychology laboratories and features in many influential publications such as Frazier and Rayner (1982) and Deubel and Schneider (1996). The learning threshold is clearly higher than for P–CR trackers, but the major drawback of the DPI is that it is a bulky and sensitive machine built using optoelectronics from the 1970s that are serviced commercially by only one person. However, the camera-based DPI built by Rucci et al. (2020) has a data quality comparable to the original analogue system and is built with modern electronics, which may revive the DPI measurement technique.

The P1 in DPI eye tracking is the same reflection as the CR of P–CR trackers, with the important distinction that P–CR eye trackers estimate the center of the CR from a small portion of a pixelated camera image, while the DPI finds the centre of an analogue light beam. This has been proposed to be the reason that the DPI does not mismeasure the amplitudes of small eye movements (Holmqvist & Blignaut, 2020).

The DPI records gaze signals with a quality sufficient to detect tremor, oculomotor drift, microsaccades, and smooth pursuit with good reliability (see Holmqvist & Blignaut, 2020; Ko et al., 2016; Poletti & Rucci, 2016, for details). Holmqvist (2015) report a median precision of 0.008° and an accuracy of 0.4° across 192 participants, both better than any video-based P–CR system. The quality of DPI data is generally lower when recording participants with small pupils that cover the P4 reflection, which causes inaccuracies and data loss (Crane & Steele, 1985; Holmqvist et al., 2020). A DPI is best recorded with participants who have large pupils, either in dark rooms or with artificially dilated pupils. The reliance on the P4 reflection furthermore results in the largest measured amplitudes of post-saccadic oscillations in any eye tracker (Deubel & Bridgeman, 1995).

Scleral search coils

Scleral search coils were introduced by Robinson (1963) and adapted for use with human participants by Collewijn et al. (1975). The scleral search coil method involves placing a copper wire coil, embedded in an annulus or contact lens, onto the sclera. The participant is placed in oscillating magnetic fields and the induced voltage in the eye coil is taken to represent the orientation of the eye with respect to the magnetic fields. This technique was dubbed the gold standard of eye tracking by Collewijn (1998). Reulen and Bakker (1982) presented the double magnetic induction principle, improved by Bour et al. (1984). Like the DPI, scleral search coil systems are analogue trackers, and data can be digitised at very high sampling frequencies. Coils can even record combined eye and head rotation for the same participant (Collewijn et al., 1985).

Houben et al. (2006) compared a coil system with a torsion-capable video eye tracker, finding that the gaze signal from the coil system was ten times more precise, and Ko et al. (2016) compared a coil system to a DPI, finding that although data from a coil system are somewhat more precise, both systems provide a data resolution sufficient for reliable detection of intersaccadic (fixational) eye movements. Collewijn (2001) sampled data at 10000Hz, and additionally reported a tracking range of 20° in all directions with a resolution of 1', while Malpeli (1998) reports a precision of 1' (0.017°) and Collewijn et al. (1988) recorded saccades with amplitudes of up to 80°.

All studies in Table 2 that have compared EyeLink systems with scleral search coils reported substantial agreement in precision and detection of microsaccades and oculomotor drift in both systems (McCamy et al., 2015, for a review). Note however that coils have been suspected to slow down the saccades of participants who wear them (Frens & van der Geest, 2002; Träisk et al., 2005). However, coils probably estimate the velocity more accurately than P–CR eye trackers, which overestimate saccadic velocity (Hooge et al., 2016).

The scleral coil tracking method is distinctly invasive, and evidence exists that older coils systems in combination with the anaesthetics that were applied caused temporary reductions in visual acuity (Irving et al., 2003, but see Murphy et al. 2001), deformation of the visual field (Duwaer et al., 1982), and blurred vision (Arend & Skavenski, 1979). Contemporary search coils are embedded on flexible contact lenses and used for research and clinical diagnostic purposes in neuro-ophthalmology and neurology, due to their high precision, and the fact that patients often suffer from uncontrolled head and body movements.

EOG

Schott (1922) and Meyers (1929) could produce recordings of the horizontal component of gaze, based on the corneo-retinal potential principle discovered in 1849 by Du Bois-Raymond. An EOG system records eye movements using electrodes on the side of the eyes that pick up an electromagnetic field produced by this corneo-retinal electrical potential of 10–30mV (Brown et al., 2006). The signal is then taken through an isolated instrumentation amplifier connected to a chart recorder or a computer. EOG is an analogue method. EOG systems are often part of other recording devices. For instance, electroencephalogram (EEG) systems often have extra electrodes for the eyes that can be used for EOG recordings.

Brown et al. (2006) proposed a standardized measurement procedure for clinical EOG measurements, aiming at acquiring high-quality EOG data. Their procedure includes dilating the pupil, preparing the skin of the participant, and then applying two electrodes on the sides of each eye and a reference electrode to the forehead. The corneo-retinal potential is mainly derived from the retinal pigment epithelium, and it changes in response to retinal illumination. Hence, in a totally dark environment, the participant spends 15 minutes looking at dim fixation targets, followed by a light phase of similar duration. This darkness-light sequence maximizes the corneo-retinal potential. The actual data recording then commences.

EOGs can be a useful variety of eye tracking when studying larger movements of the eye. Small movements will drown in the noise of EOG data (compare Fig. 2). One specific advantage of EOGs is that they can be used when the eyes are closed, for instance to study REM sleep (Aserinsky & Kleitman, 1953). However, EOG eye tracking comes with a poor accuracy, compared with most other eye trackers: Young and Sheena (1975) report a 1.5–2° inaccuracy on average.

Limbus tracking

The first published implementation of a (photo-electric) limbus tracker was by Török et al. (1951). Limbus trackers estimate the limbus border between the iris and sclera, either from video or photosensors. Limbus eye trackers based on photodiodes were sold for research up until the year 2000 by the Skalar company, but are now only known for controlling the laser during refractive surgery of the eye (Arba-Mosquera & Aslanides, 2012). The Ober Saccadometer is not a limbus tracker, but a corneal bulge tracker (Holmqvist & Andersson, 2017, p. 73), although like the Skalar limbus tracker, the Saccadometer uses photosensors to track the corneal bulge.

Video-based limbus trackers use the fact that the limbus border (between iris and sclera) has a contrast comparable to the pupil-iris border. However, limbus trackers do not suffer from pupil-based artefacts, which affect both DPI and P–CR systems. Refraction in the cornea is also not a problem. Eye trackers with low-resolution cameras may benefit from using the limbus method. The drawback is that a large portion of the limbus may be covered by the eyelid, which puts challenges on image processing.

Piezoelectric eye tracking

The piezoelectric transduction method, first introduced by Bengi and Thomas (1968), involves bringing a silicone-tipped piezoelectric bimorph into contact with the sclera, typically in the interpalpebral region near the temporal limbus. It outputs voltage signals, in which horizontal microsaccades and oculomotor tremor can be detected. This analogue eye tracker has not been used for purposes other than measuring intrafixational eye movements. There is a suspicion that the introduced pressure on the sclera affects the microsaccade behaviour (see McCamy et al., 2013, for a discussion).

Retinal image-based eye tracking

Computational tracking of retinal features involves finding the optic disk, blood vessels and smaller features, and was first done by Cornsweet (1958). A computer vision algorithm provides an analysis of the movement of features in the camera view, and infers eye movements.

Retinal image-based eye trackers are the most accurate and precise of all existing eye trackers. An early system by Cornsweet (1958), albeit limited in that it only tracked features along one axis, could detect eye movements (microsaccades) down to amplitudes of 10 seconds of arc (0.0028°). Putnam et al. (2005) presented very impressive numbers on gaze position accuracy (5" which is 0.0014°) based on snapshots taken with an adaptive optics retinal camera.

The retinal-based eye trackers with the highest speed and best accuracy are preferably built from scanning imagery, specifically from scanning laser ophthalmoscopes (SLO). These rely on the so-called 'rolling shutter' principle to recover eye motion (Mulligan, 1997), and are especially effective in SLOs that use adaptive optics that offer high resolution, high magnification and densely sampled retinal video (Stevenson & Roorda, 2005). Stevenson et al. (2016) introduced the first binocular system, which optically divided a single SLO image field between two eyes.

Retinal imaging systems also generally occlude forward viewing, impeding stimulus presentation. This may however change: Bartuzel et al. (2020) describe a MEMS-based retinal imaging system that allows for presentation of stimuli while recording with a high sampling frequency (1240Hz). Even then, the measurement range (also "trackable range") tends to be smaller than with other eye trackers: Bartuzel et al. (2020) report an 16° range (8° left, 8° right), which we can compare to 20–40° for the DPI and many video-based P–CR trackers, and 90° or more for scleral coils.

Retinal image-based eye-tracking systems typically rely on a reference frame which, in a scanning system, is a single retinal image upon which to register strips of all movie frames to compute the eye motion. This process generally yields two outputs; a stabilised movie and an eye motion trace. If the reference frame is perfect and every strip from each scanned frame is perfectly registered to it, then it follows that the eye motion trace will also be perfect. However, distortions in the reference remain a challenge to overcome and these distortions yield artefacts in the eye motion trace. Recent efforts have been made to correct for these (Azimipour et al., 2018; Bedggood & Metha, 2017) but, if uncorrected, these artefacts are evident as peaks in the power spectrum of eye motion (Bowers et al., 2019).

To date, however, retinal-image-based eye trackers have had a limited scope of application. The intrinsic trade-off between accuracy and range has rendered them most useful to study eye movements during steady fixation (Bowers et al., 2019). Retinal eye trackers have predominately been used in ophthalmology applications, often relating to disease in the retina and how that expresses itself in vision and miniature eye movements (Godara et al., 2010).

Binocular vs monocular eye tracking

The different technologies above can be constructed or set up to record either monocularly or binocularly. A common use of binocular eye tracking, particularly in remote eye trackers, is to combine the left and right signal by averaging synchronous data samples from the two eyes in the recording software, sometimes referred to as "cyclopean gaze". Cui and Hondzinski (2006) report that averaging left and right signals improves accuracy, but Hooge et al. (2019) found that averaging the gaze positions from the two eyes improved accuracy only for some of the participants.

Furthermore, head-mounted eye trackers may suffer from parallax errors, which happens because the vantage point of the eye and the scene camera do not coincide, typically when the measurement is not confined to a single plane. Binocular averaging is regularly done in glasses-based eye trackers (SMI ETG, Tobii Glasses, for instance), and in the Ober Saccadometer, which helps to alleviate the parallax issue. A thorough investigation of the geometry of the parallax error is provided by Mardanbegi and Hansen (2012), Narcizo et al. (2017), and Narcizo and Hansen (2015), and Tatler et al. (2019).

Alternatively, the two signals from the two eyes can be used to measure vergence (e.g. Liversedge et al., 2006). Jaschinski et al. (2010) showed that the EyeLink II, assuming no environmental and participant artefacts, can resolve vergence eye movements of just below 40mm in depth at a 60cm viewing distance. However, vergence measurements with P–CR eye trackers are sensitive to artefacts that affect accuracy: Hooge et al. (2019) and

Jaschinski (2016) both report effects of the pupil-size artefact on vergence. Calibration for binocular recordings introduces the choice whether to calibrate both eyes at once, or separately (Kirkby et al., 2013; Nuthmann & Kliegl, 2009; Švede et al., 2015). Additionally, Wang et al. (2019) found that the calculation of the vergence point (intersection between the gaze direction vectors of left and right eye) may show a large deviation to the fixated point, with a wide distribution in depth and a misestimation of the vergence mean point towards the participant.

Environment

Eye tracking may take place in various environments–such as an MRI scanner, cars, fighter jets, behind a desk, in VR, and during sports. These environments may differ in light conditions, vibrations and sound, temperature and the presence of other people.

Light conditions

Direct sunlight has a critical impact on data quality in video-based P–CR and DPI eye trackers. Hansen and Pece (2005) and Holmqvist & Andersson, 2017, p. 138–139) show several examples of how infrared radiation from sunlight and hot light bulbs undermine tracking in video-based P–CR trackers. The importance of a controlled light environment is exemplified by Wang et al. (2010), who excluded 32% of participants, recorded while driving a real car from one of their analyses due to poor data quality, but only had to remove 17% of participants recorded in a car simulator. The authors attributed the difference in data quality to the variable lighting conditions encountered during real driving. In a study of six pupil centre calculation algorithms for video-based outdoor eye tracking, Fuhl et al. (2016) note that pupil algorithms have good average performance, but there are still problems in obtaining robust pupil centres in the case of poor illumination conditions. Rapid changes in illumination, common in car driving and flight deck research, can be detrimental to data quality and lead to a time-consuming investment in manual post-processing (Kasneci et al., 2014). Non-commercial algorithms to improve tracking in sunlight have been developed by Santini et al. (2018) and Hansen and Pece (2005).

Even moderate changes in light levels can indirectly affect data quality. Multiple studies have established the existence of the *pupil-size artefact*, in which changes in pupil size affects gaze position accuracy in both video-based P–CR systems (Choe et al., 2016; Drewes et al., 2012, 2014, 2011; Hooge et al., 2021, Hooge et al., 2019; Jaschinski, 2016; Wildenmann & Schaeffel, 2013; Wyatt, 2010) and for the DPI (Holmqvist et al., 2020; Holmqvist, 2015). Manipulating light levels to affect pupil size typically results in increased gaze inaccuracy of 1 to 5°. The reason that changes in pupil-size affect reported gaze direction is that the pupil constricts and dilates asymmetrically, altering the pupil shape, and hence the calculated centre of the pupil image shifts position. In any video-based P–CR eye tracker, this implies a shift in gaze, even though the eyeball has not rotated with respect to the head. In a DPI, a small pupil may result in the P4 reflection at the back of the crystalline lens to be obstructed. The geometry of the setup, gaze direction and distance to the eye camera have also been found to influence the magnitude of pupil-based errors (Ahmed et al., 2016; Hooge et al., 2021; Wilson et al., 1992; Wyatt, 2010, 1995). In addition, it has been reported that pupil size in P–CR eye trackers is also related to some eye-movement measures, such as the saccadic peak velocity (Nyström et al., 2016).

Accuracy in video-based P–CR trackers is generally better for participants who have smaller baseline pupils (before calibration), measured under controlled illumination, as reported by Ahmed et al. (2016) and Holmqvist (2015). For the DPI eye tracker, the opposite is true: a large baseline pupil size results in better accuracy (Holmqvist, 2015). The signals of EOG systems and scleral coils are likely independent of pupil size, while data from retinal trackers benefit from a large pupil.

The pupil-size artefact may affect other measures. For instance, Hooge et al. (2019) found that light levels affect vergence estimations, with an error of 0.36–0.75°/mm change in pupil size (and similar findings were reported by Jaschinski, 2016). We can expect that gaze position errors induced by the pupil-size artefact will inevitably propagate to many AOI- and other higher-order measures.

Environmental vibrations and ambient noise

Sources of vibration in the recording environment contribute to increased variation in the gaze signal, as exemplified by Figure 6.24 in Holmqvist and Andersson (2017), showing how transients in the signal appear when a person walks in a room where an artificial eye is being measured with a tower eye tracker. Vibrations could be expected to matter particularly on flight decks, in cars, and during sports. For instance, De Reus et al. (2012) report that alignment shifts of the eye tracker inside the flight helmet due to external motion frequently caused inaccuracies of gaze (see also Niehorster et al., 2020b). For lab studies, a nearby elevator shaft, a powerful air conditioning unit, or vibrations caused by someone walking nearby on hard floors may add measurable noise to a sensitive eye-tracking recording. Sound in the recording situation is another form of oscillation that could make the eye tracker vibrate and affect the quality of recorded data. However, Hooge et al. (2019) recorded Tobii TX300 data at an indoor science festival with moderately loud music and found accuracy values close to manufacturer specifications. Controlled studies of the effect of vibrations on eye-tracking data quality appear to be lacking.

Presence of others

The presence of other people during the recordings may affect measures of eye movements and gaze behaviour in ways that are little understood. *Social appropriateness* may matter: The very presence of an eye tracker can impact head and eye movements, with people looking only at what they feel is socially appropriate when they believe that an eye tracker is recording (Risko & Kingstone, 2011; Nasiopoulos et al., 2015). *Distraction* is another possible factor: For instance, infants are easily distracted, looking at nearby people rather than at the monitor (Tomalski & Malinowska-Korczak, 2020). *Accidental mismeasurements* may happen when the infant is seated in the lap of a parent, and the eye tracker finds and records the parent's eyes. Additionally, Oliva et al. (2017) found longer latencies in the antisaccade task when adult participants were recorded in proximity to one another, for reasons that are not well understood.

Special recording environments

The MRI scanner environment consists of a dark and noisy tunnel, with powerful magnetic fields, in which participants must lie down. The duration of experiments and pacing of stimuli often differs from outside the MRI. Importantly, data quality from video-based P–CR tracking in MRI (SR Research, SMI, Arrington, Gaze Intelligence) generally appears to be lower than outside the MRI: poorer precision and accuracy, and more frequent data loss (Dar et al., 2021). For infrared limbus trackers (MR-Eyetracker, Cambridge Research Systems) attached to the headcoil, even small movements of the head may over time result in data loss. MRI trackers also exist that use a multicore fiber to transmit light back to outside the MRI machine where they process the reflections of the corneal bulge. The Ober MRI-tracker exhibits crosstalk (i.e. correlation) between horizontal and vertical signals, which makes the gaze signal useful only for horizontal tracking.

A curious observation is that saccadic latencies are longer when obtained in an MRI scanner than outside the MRI scanner, which could reflect the long fixation periods between saccades required in scanners, or other differences, such as participants laying down and potentially feeling drowsy (e.g. Talanow et al., 2020, their Table 1). Furthermore, the magnetic field of 7T MRIs has been reported to induce nystagmus in some participants (Roberts et al., 2011).

Head-mounted virtual-reality sets allow exclusive control over the visual stimulation provided to a subject, while shutting out any visual references provided by the outside world. Little is known of the data quality of eye trackers integrated into VR goggles, but Pastel et al. (2021) found that precision is significantly poorer in the SMI Vive VR goggles compared to the SMI glasses. Accuracy however differs only in some conditions, mostly when the distance to the fixation point changes. Stein et al. (2021) found that the end-to-end latency of common VR headsets ranged from 45ms to 81ms (compare Section "Signal properties and processing").

Setup and geometry

When preparing a manuscript about an experiment involving an eye tracker, it is important to realise that an eye-tracking setup is more than just the eye tracker itself. Hessels and Hooge (2019) point out that a screen-based eye-tracking setup may consist of at least an eye tracker, computer screen, a seat for the participant, and a table or mounting device for positioning the eye tracker. For wearable eye trackers, the setup includes the participant, eye tracker, and whatever frame, headbands, helmets or straps are used to position the eye tracker relative to the participant's eyes. With geometry, we mean the "absolute position and orientations of the eye, the eye-tracker camera, and the IR illuminator" (Hooge et al., 2021), and in the case of screen-based eye tracking, the screen. The geometry can thus (partially) be described by the distances between eye tracker (camera and/or IR illuminator), participant, and screen, and their relative orientations. A picture or schematic can be useful in providing this information, as done in Choe et al. (2016, Figure 1), Hessels & Hooge (2019, Figure 2), Valtakari et al. (2021, Figure 1), and our Fig. 3.

Gaze direction, measurement space and monitor size

Relevant properties of the setup may include the distance and relative orientation between participant and eye tracker, participant and computer screen, and the size and resolution of the computer screen. Most video eye trackers report gaze position in pixels on a screen. For some research this is sufficient (e.g. area-of-interest research in marketing). For other studies, one may wish to report the orientation and rotation of the eye in angular measurements (e.g. Haslwanter, 1995). In order to convert a gaze position on a screen in pixels to an angular measurement, it is necessary to know the distance and relative orientation between participant and eye tracker, participant and computer screen, and the size and resolution of the computer screen. If the

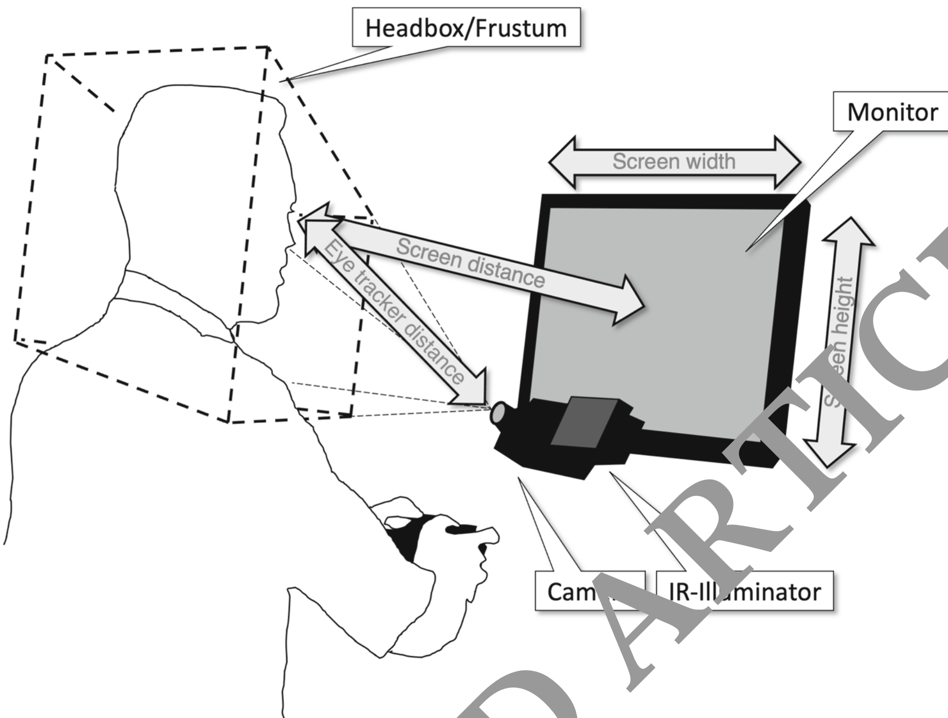

Fig. 3 Example of a head-boxed eye-tracking setup. The setup consists of a participant, eye tracker (camera and IR illuminator) and a computer screen. The geometry of this setup can be described by the relative orientations and distances of the monitor, camera and IR illuminator, and participant. Some eye trackers have a fixed relation with the computer screen (e.g. Tobii Pro Spectrum), while others do not and allow for more adjustments (e.g. SR Research EyeLink 1000). Note that the eye-tracker distance and screen distance are not identical. Screen height and width refer to both the physical and the pixel measures

width and height of the screen are smaller than 20° (10° to the left and 10° to the right), the small angle approximation may be applied. For example, this allows one to transform gaze positions in centimetres or pixels on screen to angles with a simple multiplication factor. For a general and more accurate method for this transformation, see Holmqvist & Andersson (2017, p. 21).

When the monitor is larger than the measurement range of the eye tracker (Section "Eye-tracking methods: Similarities and differences"), data quality will be poorer in the outer parts. Niehorster et al. (2020b), Schlegelmilch and Wertz (2019), Popelka et al. (2016), Holmqvist (2015), and Guestrin and Eizenman (2006) all found that data recorded in the corners of the monitor (or measurement plane) are of poorer quality than those recorded at the monitor's centre. Generally, recordings made while looking at corner positions exhibit a precision that might be worsened by a factor of 3, and accuracy by an average 1–10°, depending on the system. Such findings led Majaranta et al. (2009) to suggest putting important information in gaze-controlled systems in the centre of the screen, to give the user a better perceived accuracy.

As most P–CR eye trackers do not report physical pupil size, but pupil size in the eye image, the pupil-size signal is susceptible to viewing direction and distance. Therefore, in experimental designs in which the participant is required to look around the screen, researchers should also be aware of the *pupil foreshortening artefact* (Brisson et al., 2013; Mathur et al., 2013; Young & Sheena, 1975). As the gaze direction deviates from the eye-tracker camera axis, the image of the pupil in the eye-camera sensor deforms, making the pupil shape appear more oval and the pupil diameter – a common basis for pupil-size measurements – artificially shorter, and pupil area measurements artificially smaller. This is of particular importance for experiments using the pupil size as a measurement for estimates of the participant's psychological state (e.g. cognitive load or arousal) during free-viewing.

Various compensation algorithms have been developed to decrease the pupil foreshorting artefact, for instance relying on a geometrical model (Gagl et al., 2011), or using data from an artificial eye rotating horizontally in front of the screen (Hayes & Petrov, 2016).

Distance between participant and eye tracker

The distance between participant and eye tracker needs to be given attention, for all eye trackers, remote as well as head mounted systems. Chatelain et al. (2020) report that when participants are allowed to choose for themselves where to sit in front of a remote eye tracker, the distance to the eye tracker ranges from 40–120cm. This self-preferred range of seating distances is larger than what eye trackers can handle. Most manufacturers of remote eye trackers recommend having the distance between the participant and the eye tracker to be within a narrow range, defined by the optics of the system, with its centre at around 60–70cm (the LC EyeFollower being an exception with a specified range of 46–97cm). When a participant moves outside of the tracking range, the inaccuracies and noise levels in data can quickly triple and data loss also increases (Blignaut & Beelders, 2012; Blignaut & Wium, 2014; Kolakowski & Pelz, 2006; Schlegelmilch & Wertz, 2019).

Restrained vs. free head movements

The history of eye-movement research includes numerous examples of attempts to minimize the participants' head movements. Often, the use of head restriction is based on assumptions that the recorded data will be of better quality with a restricted head (e.g. van der Laan et al., 2017). Although overall there is a lack of studies on the effect of using chinrests, there are a few indications that they may be useful: For instance, Hermens (2015) concluded that in some cases, the EyeLink II may produce artificial microsaccades due to small head movements, and Cerrolaza et al. (2012) showed that inaccuracies may originate from small stabilizing head movements that participants make. Additionally, Holmqvist et al. (2021) found that recording participants in a chinrest increased the level of noise in some eye trackers.

Head restriction methods can be roughly divided into chinrest, forehead rest, and bite bar/board, the three of which can be combined to prevent both rotation and translation of the head. For some animal participants that take part in concurrent eye-movement and neurophysiological measurements, such as the rhesus macaque, the desire for head-movement restriction from both measurement methods has led to head restraints being surgically attached to the animal's skull for data collection with video-based eye trackers (McFarland et al., 2013) or they may have scleral coils implanted in their eyes for use with magnetic coil trackers (Kimmel et al., 2012).

The P–CR technique found in the vast majority of eye trackers today, originally came about to allow some head movement by the participant (Merchant, 1967). While the original P–CR method may handle small movements of the head, at the size of a few millimetres up to a centimetre, recent remote video-based eye trackers are designed to allow for free head movements in a much larger space (the headbox, see Fig. 3), tens of centimetres or more across.

One way to accomplish room for larger head movements is to use a wide-angled eye camera that covers a large space around the participant, and use a trade-off: The sampling frequency of the eye camera can be increased by reducing the size of the recording window on the camera sensor so it just samples the eye region. When the participant moves, this recording window on the camera sensor must be moved in real-time (or physically, using a pan-tilt camera as in the LC EyeFollower). Although moving the recording window allows for larger head-movements, this window motion introduces sample dropping (data loss) in some eye trackers (Holmqvist & Andersson, 2017, p. 1[illegible]8). Studying the effect on accuracy, precision, latency and loss of data, Blignaut (2018) found that one or two headbox adjustments per second would have no effect on accuracy, but it did on spatial and temporal precision (in the author's custom-built eye tracker). However, some eye trackers change sampling frequency altogether when the eye is lost in the recording window of the camera sensor and the eye tracker goes into full-sensor search mode (Hessels et al., 2015, Figure 3).

When participant eyes are at the center of the headbox eye-tracking data quality is best. When located away from the headbox center, data quality is negatively affected, as experienced by many infancy researchers and investigated experimentally by Hessels et al. (2015) and Niehorster et al. (2018), who found a strong effect of rotating the head on the quality of eye-tracking data on a number of eye trackers. In fact, any relative movement between eye and the eye camera of the eye tracker can reduce data quality, also in eye-tracking glasses (Niehorster et al., 2020b).

During gaze interaction, the human–computer interaction technique of controlling a computer with gaze, the participant/user has immediate cursor feedback of where the eye tracker thinks that gaze is located. Gaze inaccuracy originating from the users' movements undermines effective usage. Chinrests are not a solution here, because many users have involuntary head movements or seating positions that make a simple head restriction impossible, requiring a different user interface design (Donegan, 2012). Some users (try to) actively use head movements to adjust gaze pointing inaccuracies (Špakov et al., 2014). The authors speculate that this can be common among people with disabilities who actually use gaze control in their everyday life.

For infants, adults with certain disabilities, and animals, head restriction methods are not always practically usable, and alternative methods for head movement reduction are often used. Hessels et al. (2015) compared the eye-tracking data quality of infants recorded in a reclining car seat versus that of infants sitting on the parent's lap or in a highchair. Accuracy was worse (higher) for infants seated on the parent's lap or in the highchair than for infants in the car seat. Yet, a participant's positioning puts additional constraints on the placement of the eye tracker. Hessels and Hooge (2019) found that placing infants in a car seat required the eye tracker to be tilted forward substantially, which that might not be possible for some eye trackers without extensive modifications and additional equipment. Similarly, for patients confined to the bed, mounting the eye tracker on an adjustable arm allowed for effective gaze interaction for disabled users lying on their back (Blignaut, 2017; Hansen et al., 2011).

Participants

In this section, we review how certain characteristics of participants are related to the quality of recorded eye-tracking data, to eye-movement measures and high-order measures of gaze behaviour. The characteristics we discuss include gender, age, visual acuity, visual aids, physiology of the eye region, mental state (e.g. sleep deprivation, mental fatigue, cognitive workload), expertise, and psychopathology. A complete review of all these characteristics – particularly expertise and psychopathology – is beyond the scope of the present paper. However, our goal here is to show that these characteristics may be relevant, which researchers may use when defining their participant group and exclusion criteria. Whenever possible, we direct readers to more in-depth reviews on the specific topics.

Attrition rate

Attrition rate is operationalised as the proportion (or percentage) of participants who were not included in the analysis. Attrition rate exhibits a large variation between studies. For instance, Dalveren and Cagiltay (2019) report an attrition rate of 17.9% for the EyeTribe, while Holmqvist (2015) report 1.0% for the same eye tracker. The reported attrition rates appear to be lower in studies with adult participants in light-controlled labs, for instance 0–8.2% in Holmqvist (2015), compared to recordings made in sun-lit environments, for instance Wang et al. (2010), who report 32% attrition rate during outdoor driving. Attrition rates may be high for infant studies, for instance: 59–64% in Burmester and Mast (2010), and for children in the autism spectrum (100% in Birmingham et al., 2017).

Older remote video-based eye trackers have been reported to have higher attrition values also for lab studies with adults. For instance, Sibert and Jacob (2000) reported 38% attrition rate for ASL Model 3250R, while Schnipke and Todd (2000) reported 62.5% for the ASL 504.

52.2% of the publications in the reporting database (see Section "Reporting practices and existing reporting guidelines" for details) report the number of participants excluded from analysis. Their main reasons for excluding participants were "data quality" (44.1% of the publications), "impossible to calibrate" (19.8%), "the participant" (12.6%), "other" (7.2%), "error in the experimental procedure" (5.4%), and "failed to follow the instructions" (0.9%). This suggests that poor data quality is the major reason for excluding participants from analysis.

Alternatively, attrition rate can refer to the number or proportion of trials or events per participant that were excluded, for those participants included in the analysis. In the reporting database, 30.9% of the studies reported excluding trials or fixations. Each study reported a slightly different reason for exclusion, many of which relate to data quality, outliers, technical failures or behavioural mishaps.

Gender

There are some reports of differences between genders in gaze behaviour towards other people (Coutrot et al., 2016; Gluckman & Johnson, 2013; Rupp & Wallen, 2007), and in pupil reactions to pain (Ellermeier & Westphal, 1995). Coors et al. (2021) found that although gender-related differences in eye-movement measures (blink rate, smooth pursuit gain) do exist, most are negligible in magnitude.

Ethnicity

Blignaut and Wium (2014) report that, statistically, Asian participants are more difficult to track, and the resulting data are on average of worse quality than for participants of European or African ethnicity (see also Holmqvist, 2015). These findings reflect the generally narrower palpebral aperture in the east Asian population. Amatya et al. (2011) found a larger proportion of express saccade makers in the Asian participant group, indicative of faster saccadic reaction times.

Age

Data quality as well as many eye movement measures covary with the age of the participant. Firstly, infant researchers have consistently shown that eye-tracking data quality tends to be worse for younger children than for adults. For example, accuracy and precision are generally worse, and data loss is generally poorer, for infants and toddlers than for school-aged children and adults (Dalrymple et al., 2018; Hessels et al., 2016, 2019). Interestingly, worse precision in infant eye-tracking data is not due to fixation instability (Seemiller et al., 2018). Moreover, higher amounts of data loss with infant participants are not only due to infants looking away more from the screen, as it is often characterised by short periods of data loss (less than 100ms: Hessels et al., 2015; Wass et al., 2014). Neither is this due to blinking, as young children blink significantly less than adults (Stern et al., 1994). In addition, it seems that individual differences in data quality are larger for the younger participants (5–10 months) than for the older participants (3–9 years, Hessels & Hooge, 2019). The latter is particularly problematic when analysis methods are used that are susceptible to differences in data quality.

The oculomotor system develops into adulthood and old age. The resting pupil diameter has been found to be larger for young adults (around 20 years) than for older (around 70 years), independent of luminance level (Bitsios et al., 1996). Saccadic amplitudes have been found to be shorter both for children (below 10 years) and older adults (above 60), compared to young adults (30–40 years, Helo et al., 2014; Açik et al., 2009; Mackworth & Bruner, 1970; Açık et al., 2010). The latencies of said saccades follow the same pattern, decreasing from childhood into adulthood (Luna & Velanova, 2011; Salman et al., 2006), and then increasing again as participants grow older (Moschner & Baloh, 1994). Smooth pursuit parameters such as latency (time until the movement is initiated) and gain (how closely gaze follows the target velocity) also have been found to be related to age. While latency is longer for older than for younger adults (Sharpe & Sylvester, 1978), gain is closer to the ideal value in young adults compared to children (Luna & Velanova, 2011; Salman et al., 2006).

Binocular coordination during reading is also poorer in children than in adults (Blythe et al., 2006). In a review of the eye movements of the aging reader, Paterson et al. (2020) point out changes both on lexical (e.g. the word frequency effect), and orthographic levels (e.g. sensitivity to removal of inter-word spacing). Age variation in fixations and blinks has not been systematically explored outside reading research (Marandi & Gazerani, 2019).

Also, with older age, it is more likely that the participant will wear spectacles or lenses, have droopy eyelids, have cataracts, or an artificial lens from cataract surgery, macular degeneration and peripheral scotomas, as well as several neurodegenerative ailments, which tend to make either data quality worse or alter eye movements, or both.

Visual acuity and visual impairment

For readers with low acuity, the fixation durations are longer, saccades shorter, and consequently text reading takes much longer (Legge et al., 1997). Furthermore, blurred vision caused by, for instance, myopic refractive error results in an increase of the amplitude of microsaccades (Ghasia & Shaikh, 2015). Eye movements are dramatically different for participants with low vision, i.e. a loss of vision that cannot be corrected by medical or surgical treatments or conventional eyeglasses, such as macular degeneration, scotomas, cataracts, or nystagmus (Leigh & Zee, 2006).

Spectacles, lenses and makeup

Nyström et al. (2013) investigated the effect of eye-region physiology, spectacles and other factors on accuracy, precision and data loss in the SMI HiSpeed1250, finding poorer precision when participants wear spectacles, and poorer accuracy, precision and data loss when contact lenses are worn. In a large follow-up using 12 eye trackers, Holmqvist (2015) reports up to 10° worse accuracy and up to three times (300%) poorer precision for recordings where the participants wore spectacles that were scratched or dirty or that had an anti-reflective coating, compared to recordings where no visual aids were used. Data recorded from participants wearing soft contact lenses exhibited 0.5–3° poorer accuracy and on average 20–40% poorer precision, compared to when participants wore no visual aid. Asking a participant to remove the spectacles to record data of better quality might result in poorer acuity that may alter the eye movements (see above).

Makeup (eyeliner, eye shadow and mascara) result in a poorer accuracy by 0.2–3°, and up to three times poorer precision (Holmqvist, 2015). For participants with forward- and downward-pointing eyelashes, makeup results in poor data quality (see also Nyström et al., 2013). Mascara is black in both infrared and visible light, and Holmqvist & Andersson (2017, Figure 5.5) show eye images from actual recordings that depict how the dark mascara may interact with the pupil center calculation.

Physical properties of the eye region

Differences in eye physiology refers to eye colour, lash direction, ocular dominance, baseline pupil size and more. Holmqvist (2015), Hessels et al. (2015), and Nyström et al. (2013) investigated the relation of data quality to physical properties of eyes, from large groups ranging between 75 and 194 participants, in up to 12 eye trackers, and reported compatible findings. In this subsection, we report effect sizes from these three studies, as ranges from the many eye trackers.

Holmqvist (2015) found that darker pigmentation in hair, eyes and skin correlate positively with better (lower) accuracy on most video-based eye trackers (0.5–1°), and also better precision (20–80% lower RMS-S2S). The advantage of dark iris pigmentation over blue eyes has been hypothesised to result from poor contrast between pupil and iris when the eye image is recorded in infrared light: A blue iris is dark, while a brown iris is bright (Holmqvist and Andersson, 2017, Figure 4.13), providing a clearer contrast between iris and the dark pupil, which the image processing algorithms can make better use of.

Clinical participant groups may have features in their irises that may make tracking more difficult for some eye trackers. For instance, participants who lack an iris, known as aniridia (Beby et al., 2011), are likely difficult to record with P–CR trackers. Participants with William's Syndrome have a stellate pattern in the iris (Tran & Kaufman, 2003) that could interfere with the CR image of P–CR trackers. These iris features are often associated with specific eye-movements. For instance, participants with albinism may have transillumination effects in their irises, and their lack of pigmentation in skin and in the retina is associated with congenital nystagmus (Collewijn et al., 1985).

A smaller baseline pupil results in better accuracy (up to 2°) and up to three times poorer precision (Holmqvist, 2015). Interocular distance is defined as the distance between pupil centres when looking straight ahead. Holmqvist (2015) found poorer accuracy (0.5–1.0°) for small interocular distances, but only in remote eye trackers.

A larger eye opening (also 'palpebral fissure' or 'eye cleft') correlates with better accuracy: up to 1° better in fully open compared to eyes with the smallest palpebral fissure. Forward or upward-pointing lashes show the best accuracy, while downward-pointing eye lashes, which Holmqvist (2015) found in about 10% of their 194 participants, exhibit a poorer accuracy (up to 4°) and precision, although some eye trackers are more affected than others. A more closed eye is more likely to block the eye tracker's view of pupil and CR features, but this depends on the geometry of the setup, both in remote and head-mounted systems.

Arousal, mental fatigue and cognitive workload

Ayres et al. (2021) present a meta-study of 33 experiments and conclude that eye-movement measures of cognitive load are more sensitive than heart, skin, and brain measures. Mental workload and arousal are positively associated with pupil dilation as shown in a large number of controlled studies and life-like human factors studies, measured using high- or low-end eye trackers (Einhäuser, 2017). Examples include performing a memory task (Kahneman & Beatty, 1966), arithmetic tasks (Ahern & Beatty, 1979; Hess & Polt, 1964), Air Traffic Control (Ahlstrom & Friedman-Berg, 2006), (simulated) driving (Čegovnik et al., 2018), tasting a disgusting drink (Kaneko et al., 2019) and social stress caused by having to sing a song (Toet et al., 2017). Other parameters of eye movement behaviour can be affected as well, but this seems to be context or task dependent. For instance, for blinking rate, Recarte et al. (2008) and Čegovnik et al. (2018) found an increase with increasing workload, whereas Brouwer et al. (2014) found no effect; and Bauer et al. (1987) and Fogarty and Stern (1989) found a decrease in blinking rate with increasing workload. This variation in results may be caused by the differences in the workload-inducing task across these studies.

Workload has also been reported to decrease microsaccade rates but increase their amplitudes (Siegenthaler et al., 2014), increase fixation duration (Rayner & Pollatsek, 1989) and decrease horizontal scanning during driving (Recarte & Nunes, 2003). Mental fatigue and workload have been found to affect saccade and microsaccade dynamics during visual search (Di Stasi et al., 2013), surgery (Di Stasi et al., 2014) and for pilots suffering from low levels of oxygen (Di Stasi et al., 2014). When researchers investigate workload, these eye-movement measures are often combined. For instance, Van Orden et al. (2000) developed a model using regression analysis from eye movement data on a surveillance tracking task, showing that fixation duration, blink duration and mean pupil dilation combined to a robust and reliable predictor of the performance of surveillance tracking.

Sleep deprivation

Many studies have reported effects of partial and total sleep deprivation on eye movements. Sleep deprivation is known to result in increased saccadic latency and reduced saccadic peak velocity and smooth pursuit velocity, as well as more antisaccade errors (Ahlstrom et al., 2013; Fransson et al., 2008; Meyhöfer et al., 2017). Furthermore, Schalén et al. (1983) present data showing that saccadic and smooth pursuit peak velocity may vary with the circadian rhythm.

Moreover, sleep deprivation has been shown to cause mental fatigue and affect a myriad of cognitive domains such as memory (Van Der Werf et al., 2009), cognitive speed (Van Dongen & Dinges, 2005) and arousal (Gunzelmann et al., 2007), which in turn may affect eye movements.

Expertise

Many eye-tracking studies of expertise have been made. Good overall reviews are provided by Reingold and Sheridan (2011) and Gegenfurtner et al. (2011). For instance, expert chess players tend to have fewer, longer fixations in the middle, while novices scan more (Charness et al., 2001). Expert radiologists tend to fixate abnormalities earlier than novices (Nodine et al., 2002; Alexander et al., 2020). Even the ability to keep one's eye still is affected by training and experience (Cherici et al., 2012; Di Russo et al., 2003). In medical expertise research, a lack of experience or familiarity in the task has been correlated with blink rate and duration, fixation duration, transition rate, and pupil dilation (Lee et al., 2019, 2020). Machine learning approaches have been used to differentiate between levels of language proficiency (Karolus et al., 2017). Findings in expertise studies do not easily transfer to other domains of expertise. The one and same participant can be an expert in one task while having no expertise in a very related task (Kevic et al., 2015). In fact, it is important to understand that the participant's field of expertise, the task, and the stimulus are crucial determinants of what effect can be expected in terms of eye movements.

Pathology and personality

Several different psychiatric disorders have independently been found to coincide with oculomotor impairments with medium-to-large effect sizes, although these depend on diagnosis and experimental task (Alexander et al., 2018; Smyrnis et al., 2019). For instance, patients with schizophrenia reliably show reduced smooth pursuit accuracy (reduced gain, increased root-mean-square error of the signal, increased frequency of saccades during pursuit). In a meta-study on the eye movements of patients with schizophrenia, O'Driscoll and Callahan (2008) stated that "Average effect sizes and confidence limits for global measures of pursuit and for maintenance of gain place these measures alongside the very strongest neurocognitive measures in the literature." (p. 359). Patients with schizophrenia also reliably show increased rates of direction errors on the antisaccade task. Similar impairments, albeit with smaller effect size, are observed in patients with bipolar disorder or major depressive disorder (Katsanis et al., 1997).

Differences in gaze behaviour between individuals with and without a diagnosis of autism spectrum disorder (ASD) have also been substantially investigated (see e.g. Bast et al., 2021; Guillon et al., 2014; Sasson et al., 2011). One often-reported finding is differences in gaze behaviour to the eyes of a face between individuals with and without an ASD diagnosis (e.g. Dalton et al., 2005; Jones et al., 2008, 2013; Klin et al., 2002; Rice et al., 2012). However, these findings are not unequivocal (see e.g. Dapretto et al., 2006; McPartland et al., 2011; van der Geest et al., 2002). Several potential explanations have been posited for the inconsistent findings, including the presence of alexithymia (Bird et al., 2011) and the cognitive demand required in the experimental setting (Senju & Johnson, 2009). A meta-analysis of 122 studies on gaze differences to social and non-social information between people with and without autism is given by Frazier et al. (2017). Other reported differences include eye movements during visual search (e.g. Keehn and Joseph, 2016; Kemner et al., 2008) and attentional disengagement (e.g. Keehn et al., 2013).

Furthermore, Alzheimer's (Kapoula et al., 2014), Parkinson's (Otero-Millan et al., 2018) and Huntington's are known to affect several characteristics of eye movements (Leigh & Zee, 2006).

Variation in human personality has been associated with eye movements (Bargary et al., 2017) and with gaze patterns to social stimuli (Wu et al., 2014).

Medication and drugs

For studies that investigate differences in eye-movement measures between clinical and control groups, recording patients who may be under medication, the question may arise whether it is the psychopathological state or the medication that drives the difference. For example, benzodiazepine drugs cause reduced saccade peak velocity (De Visser et al., 2003) as well as increased saccade latency and reduced spatial accuracy of saccades (Ettinger et al., 2018). Measures of intra-individual variability of saccades are also increased. Benzodiazepines also reliably reduce smooth pursuit velocity (Karpouzian et al., 2019).

Even in non-clinical trials, drug use may be a consideration. Acute consumption of nicotine may improve smooth pursuit accuracy, reduce catch-up saccades (Meyhöfer et al., 2019; Avila et al., 2003), and may reduce antisaccade latencies as well as the rates of direction errors in the antisaccade task (Ettinger & Kumari, 2019). Cannabis has the opposite effects to nicotine: latencies and errors in the antisaccade and memory-guided saccade tasks are increased, and saccade peak velocity is lower (Huestegge et al., 2009). Pupil size is affected by some drugs (Newmeyer et al., 2017). Increased blood alcohol levels impair the quality of smooth pursuit (Flom et al., 1976; Wilkinson et al., 1974), decrease saccade velocity (Lehtinen et al., 1979) and increase fixation durations (Moser et al., 1998). Alcohol also has effects on gaze behaviour. For instance, Buikhuisen and Jongman (1972) presented a traffic film containing 86 important events to participants, while tracking their eye movements. Those who were alcohol-intoxicated fixated on fewer events, especially when located away from the centre of the display, than non-intoxicated participants.

Calibration and accuracy

Calibrating the eye tracker for the specific participant is a prerequisite for recording gaze in some eye trackers and for optimal accuracy on all eye trackers. In this section, we first describe the procedure and principles of calibration generally, how to assess calibration, and correct for poor accuracy, and then we describe methods for calibrating challenging participants, such as infants, dogs, and people with nystagmus. These methods all aim to ensure the best possible accuracy.

How is calibration done?

Just before or at the beginning of a recording session, participants typically need to perform a small initial task of looking at a set of pre-defined targets that either appear on, or smoothly move across the stimulus monitor, or are otherwise presented in front of the participant. If the recording is made within the software of a video-based P–CR tracker, when the participant fixates the point, the eye tracker registers the relative positions of features (such as P and CR) for each calibration point. Quite often, the researcher may choose how many targets (often points) will be shown during this initial phase, and in some cases, where targets appear, and what the target will look like. For most other technologies (DPI, coils, EOG, etc.), calibration needs to be done with custom software and will likely also involve looking at or following fixation targets.

Fixation targets

The choice of calibration target may have an effect on the data quality in the subsequent recording. Thaler et al. (2013) examined which fixation target results in the least dispersion during fixation for adult participants, while Schlegelmilch and Wertz (2019) investigated the effects of calibration targets on the dispersion of the gaze position signal of the EyeLink 1000 Plus, for infant research. Whether showing a calibration target that minimises dispersion will result in better accuracy is unknown.

Colour and luminance of the background

Previously referenced studies on the pupil-size artefact (Section "Environment") tell us that changes in pupil size will affect the accuracy of the gaze position signal. Thus, calibrating at a different luminance from the luminances displayed during data collection is likely to affect the accuracy of the measurement. If stimuli vary in luminance, it may be useful to calibrate for a range of pupil sizes (Drewes et al., 2012).

Which data segment to use for the calibration?

The eye-tracking software, manufacturer-based or custom tailored, selects a segment of data for when it estimates that the participant is looking at the calibration target. The exact decision which segment of data is used for calibration is mostly made by the software itself (Hansen & Ji, 2010). Nyström et al. (2013), however, showed accuracy is higher when the participant indicates s/he is looking at the fixation target, than leaving this decision up to the system. This finding also relates to the idea behind the participant-controlled post-calibration by Ko et al. (2016). However, participant-controlled calibration does not appear to be the standard in most eye-tracking software today.

Number of targets and the mathematics of calibration

Akkil et al. (2014) reported for the Tobii T60 that calibrating with 9 points result in a better accuracy compared to using 5 or 2 points, with a difference of about 0.2° between the 9-point and the 2-point calibrations.

In a number of video-based eye trackers (most SMIs, all EyeLinks, and many Tobiis, for instance the T60), the calibration involves finding a best fit between the sensor values (P and CR positions in the eye camera, for instance) and the spatial positions of calibration points. The exact polynomials used in these equations vary by the manufacturers, but also by the number of calibration points. Thus, it is important to realise that the choice of a specific number of calibration points in the eye-tracker manufacturer software is also a choice of a specific set of equations used for the calibration procedure. Each set of polynomial equations may result in different accuracy values for the same eye movement data (Blignaut & Wium, 2013; Blignaut, 2016; Cerrolaza et al., 2012).

Modelling the 3D shape of the eyeball is possible when multiple cameras and/or multiple corneal reflections are employed. Theoretically, the minimum number of calibration points is one, and this point is needed to measure the difference between optical and visual axes (Guestrin & Eizenman, 2006; Hansen & Ji, 2010). Recently, some manufacturers have developed calibration methods that model the eyeball more extensively. In particular, the curvature of the cornea is an important part in these calibration models, which have been used in eye trackers such as the SMI glasses, many Tobii eye trackers (US Patent US7,572,008), and in the open-source eye tracker by Barsingerhorn et al. (2018).

Calibration software is not supplied with every eye tracker. For instance, the DPI eye trackers require the researcher to employ custom-built calibration algorithms to establish the mapping between sensor values and points on the monitor. Holmqvist (2015) used a RANSAC fit (Fischler & Bolles, 1981) followed by a linear shift to calibrate the DPI.

Using the calibration of another participant

There are also examples of researchers calibrating their eye tracker on a person other than their actual participant, when the actual participant is difficult to calibrate. For example, Kulke (2015) calibrated on adults, and then recorded infants by reusing that adult calibration, arguing that this procedure improved data quality compared to calibrating for infants. Indeed, Harrar et al. (2018) present data showing that this practice does not introduce non-linearities (variations in accuracy over space), and also find that calibrating on one person and recording on another led to a poorer accuracy by 2–4°. Similarly, researchers recording with artificial eyes also calibrate on themselves before recording with the artificial eye. Holmqvist and Blignaut (2020) show that no noticeable non-linearities appear in the data when using the human calibration for a subsequent recording with artificial eyes, but also note that accuracy is likely to be poorer.

Validation of the calibration

Present eye-tracker vendor software almost always reports accuracy after each calibration, recorded on validation points immediately after the calibration sequence. If the accuracy is not sufficient after the first calibration, commercial recording software may allow the operator to recalibrate several times, and select the calibration with the best accuracy in the validation test.

Post-calibration correction

Although it is rarely done, a poor accuracy after calibration can also be improved using a post-calibration correction. This procedure involves a second round of looking at points. For instance, Blignaut et al. (2014) used a regression model to improve accuracy by 0.3–0.6°. Correction can also be made by letting the participant manually guide an online, calibrated, gaze-contingent visualisation of raw gaze samples to fall exactly in line of his/her gaze (Poletti & Rucci, 2016), i.e. until these samples are projected onto the centre of the fovea, and then push a recalibration button, which in their study improved the already very accurate DPI by a factor of 2.

Drift, and methods for drift correction

Accuracy that worsens over time is often called drift (not to be confused with oculomotor drift), irrespective of its source: small body adjustments, head-mount slippage, changes in pupil size, or some change in the hardware or software setup. Head-mount slippage could be the reason that the SMI EyeLink I and the SR Research EyeLink II were known to be so drift-prone that most researchers used to adjust their calibration, with a one-point drift correction, once before each trial (e.g. Greene & Rayner, 2001). Although drift refers to accuracy, other measures may also be affected by long recordings. For instance, Hessels et al. (2015) and Wass (2014) report a decline in precision from an early trial to a later one.

It is not known how much drift there is in current eye trackers, which are often sold as "drift free" (S. R. Research, 2017, p. 24), but a certain drift still exists in some instruments. Nyström et al. (2013) report a 0.2° drift during a 15-min reading task with the SMI HiSpeed 1250, and Choe et al. (2016, Figure 2) show drift due to the pupil-size artefact. Ko et al. (2016) found that the DPI and coils recording artificial eyes drift by around 0.03' per minute. Drift happens not only in long recordings, but also in cases where the recording does not immediately follow calibration: Chatelain et al. (2020) found that when recording participants on the Tobii 4C in sessions over one month with no recalibrations, accuracy degraded by 0.30° + 0.13°/month, i.e. the initial drop in accuracy is the largest.

Drift correction procedures involve re-calibrating with a single point, shifting all subsequent data by the measured offset. Later EyeLink models offer drift checks in which the offset between gaze cursor and target is assessed, and the experimenter can optionally make a linear shift of estimated gaze. In infant research, Constantino et al. (2017) implemented *automatic* drift correction on the fly, using an appearing fixation target and a criterion on accuracy. Jones et al. (2014) instead used a happy face and a probability calculation that decided whether the infant had looked on the face, even if the eye tracker records the contrary, in which case an automatic drift correction was made. The threshold for when to perform drift correction may impose a maximum allowed accuracy. However, this is not the same as the empirically determined accuracy, and there is no guarantee that a central drift correction will improve accuracy in more peripheral points. When the user has a visible gaze cursor, as with users of gaze-controlled computers, Graupner and Pannasch (2014) show that they can learn to take advantage of the visible cursor as a cue to understand variations in accuracy over space, and choose to recalibrate when it is needed for the functionality they want.

If accuracy is found to be poor after the recordings are completed, while inspecting the data as scanpath plots, the EyeLink Data Viewer by SR Research allows the possibility of 'performing drift correction on fixations' by simply grabbing any fixation or group of fixations and pulling it to a new position. A simple test reveals that saccade amplitudes and velocities also change during these data editing operations, not only the fixation positions themselves (Data Viewer 3.1.97). The Data Viewer manual states that when batch-moving fixations like this, a movement of more than 30 pixel is not acceptable; however, for those users who want to move fixations more than this, the 30 pixel setting can easily be changed. Later, SMI also started offering this feature in the BeGaze software, and it is also possible in OGAMA (ogama.net). Note that the researcher has to be very careful not to move fixations in favour of a hypothesis to avoid subsequently arriving at faulty conclusions.

This practice is mostly relevant for text reading, in particular when participants read more than one line of text. Cohen (2013, p. 677) comment on practices in reading research that "Fixations are typically corrected manually, sometimes within a program such as EyeDoctor" (https://blogs.umass.edu/eyelab/software/, accessed 10-03-2021). Alternative software solutions for re-aligning inaccurate gaze data to lines of text are offered by Cohen (2013), Hyrskykari (2006) and Špakov et al. (2019).

Dragging fixations in place has also been applied in infant research (Frank et al., 2012; Kooiker et al., 2016). Manual post hoc calibration was commonplace in nystagmus research in the past, and tended to be based on finding the fixation periods of the nystagmus waveform and using those gaze locations for the re-alignment (Dell'Osso, 2005).

Binocular calibration

Recording from the participant's dominant eye results on average in 0.2° better accuracy and also better precision (Holmqvist, 2015; Nyström et al., 2013), as compared to recordings from the non-dominant eye. This difference in data quality between the dominant and non-dominant eye leads to one consideration when calibrating for binocular recordings: whether to calibrate both eyes simultaneously or to instead calibrate the two eyes separately, patching one while calibrating the other. Calibrating both eyes at once, binocularly, may give an erroneous (absolute) disparity value because the calibration procedure assumes that both eyes are directed towards the calibration point, when in fact one eye may be slightly off. Nuthmann and Kliegl (2009) nevertheless calibrate for both eyes simultaneously, arguing that they can still correctly measure *relative changes* in disparity. Švede et al. (2015) and Liversedge et al. (2006) recommended a separate monocular calibration for each eye when using binocular recordings, for investigating the *absolute* disparity between the two lines of gaze. This should be done by covering one eye, calibrating the other, and then switching.

Calibration of special populations

Researchers working with participant populations other than young adults, such as infants or animals, will likely be faced with additional challenges during calibration. This may be due, for example, to these participants not being able to respond to verbal instructions. While some animals can to a degree be trained to remain still and to look at the desired calibration target (Park et al., 2020), infants and some monkeys can be nudged to look at the desired point by using contracting and dilating images, or by using transient appearances of calibration targets on screen (e.g. Hessels et al., 2017; Jones et al., 2014).

Patients with age-related macular degeneration have difficulty foveating calibration targets (because they have no or reduced foveal vision). Harrar et al. (2018, p. 9) suggest using the calibration of another person and found that accuracy degrades by 4–8° with this method, but that it does not introduce non-linearities.

Calibrating an eye tracker for participants with an unstable gaze, such as nystagmus or continuous square wave jerks, presents the problem that as they look at a calibration point their eyes will not be still. For these participant groups, researchers have developed dedicated calibration routines specific to the particular oculomotor condition (Dunn et al., 2019; Rosengren et al., 2020). Note that not all eye trackers allow for these calibration routines, e.g. when a standard calibration procedure has to be performed before a recording can commence. Eye trackers that can record without explicit calibration include the DPI and scleral coils (Holmqvist & Andersson, 2017, pp. 214–217) and some P–CR eye trackers.

Features of the experiment

Here, we address only those aspects of experimental design that may be specifically relevant or problematic in the context of eye-tracking research, such as the operator skill level, eye-movement measures used as dependent variables, the number of trials and experiment duration.

Operator skill level

By operator we mean the person (researcher or research assistant) who records data from the participant. Nyström et al. (2013) report an advantage of 0.2° in the accuracy recorded by experienced operators, compared to inexperienced, whereas Hessels and Hooge (2019) report experienced operators tend to succeed calibrating difficult participants where inexperienced operators give up, and point out that training of operators could have a beneficial effect on data quality.

The instruction to participants

Task instructions have a strong influence on eye movement behaviour, as elegantly shown by Buswell (1935, p. 136) and Yarbus (1967, p. 174). The instruction to the participants is part of the experimental design, and can be used actively to drive participant behaviour. However, the small differences in wording may have unexpected effects, and the exact instruction may need to be verified during piloting. For instance, asking participants to "fixate" rather than "hold the eyes still" reduces the rate of microsaccades (Poletti & Rucci, 2016), and Enright and Hendriks (1994) found that "staring" differs from "scrutinizing", in that the latter involves a larger net muscular force exerted on the eye from the opposing rectus muscles, pulling the eyeball backward in its socket.

Trial durations and trial-by-trial effects

Besides the fact that data quality seems to be worse after longer periods of time (Section "Calibration and accuracy"), the duration of trials and experiments is relevant also for other reasons. For instance, during scene viewing, fixations tend to be shorter and saccade amplitudes longer during the first second or two of a trial. This can be interpreted as an initial overview/ambient scan followed by detailed/focal inspection, shown by Tatler and Vincent (2008), Unema et al. (2005) and Buswell (1935) for free-viewing, by Scinto et al. (1986) for visual search and by Over et al. (2007) for visual search and free viewing. This would imply that when trials vary in duration, mean fixation duration for long-lasting trials may be longer than mean fixation duration for short trials, irrespective of other factors. Also, when trials are short, comparing mean fixation durations for short sequences of saccades, one should consider not including initial fixation durations because initial fixation durations are longer than subsequent fixation durations (Hooge & Erkelens, 1996; Zingale & Kowler, 1987). This also holds for infant participants (Hessels et al., 2016).

A technical trial-by-trial effect is that the duration of the initial fixation of a sequence of fixations may not reflect the whole duration of that initial fixation, because it started before the trial started, and was cut in two by the change of trial. In the visual-cognition literature, when analysing fixation durations, the first and last fixations are typically discarded (e.g. Nuthmann, 2013).

Tatler and Hutton (2007) found trial-by-trial effects in the antisaccade task: Both the error rates and latencies increased on trials following a trial with an erroneous anti-saccade. Switching from making an antisaccade in one trial to making a prosaccade in the next trial involves a cost in increased saccade latency of the prosaccade (Tari et al., 2019). Similarly, a saccade to a location that was fixated at the end of the previous trial may be preceded by a prolonged fixation (Carpenter, 2001), and may affect latencies and fixation durations in the current trial.

Eye-movement measures as dependent variables

In some research fields the choice of the appropriate eye-movement measure and the range of task parameters, for the study at hand is either straightforward or very well established. This is for instance the case in reading research (Clifton et al., 2007), and for studies employing the anti-saccade paradigm (Antoniades et al., 2013).

In some applied research fields, measure selection is all but obvious and terminology of measures confusing (e.g. Sharafi et al., 2015). A line of publications may get accustomed to a choice of measures that later turns out to be unfortunate. See for instance Šmideková et al. (2020) for a discussion of the selection of measures for research in classroom management.

Naming of events is also variable. What some know as saccade latency (Holmqvist and Andersson, 2017, p. 580) is sometimes termed saccade reaction time or calculated as time to first fixation (Tatham et al., 2020). Fixation duration is sometimes called 'fixation time', but also 'dwell time', or 'dwell time of the fixation'. Oster and Stern (1980) used the terms saccadic reaction time and intersaccadic interval for fixation duration. The original term was 'pause time' (Erdmann & Dodge, 1898), and the term 'pause duration' was used long into the 1940s.

Terminology for the dwell time measure also varies. In some parts of human factors research, the dwell time measure is called 'glance duration' (Horrey & Wickens, 2007), while Loftus and Mackworth (1978) used the term 'duration of the first fixation' for the first dwell time in an AOI. Terms like 'observation' and 'visit' can also be found. In reading and some parts of scene perception research, dwell time is often called 'gaze duration' or 'regional gaze duration', and 'first-pass fixation time' when the AOI consists of two words (Clifton et al., 2007).

Signal properties and processing

In this section, we discuss the properties and processing of the stream of data from the eye tracker, such as gaze position signals, time stamps, pupil-size signal, and more.

Sampling frequency

Sampling frequency (also temporal resolution) is the number of measurements per second. The sampling frequency of modern video-based eye trackers ranges from 30 to over 2000Hz. Some eye trackers, like the DPI, scleral search coils and some other analogue systems have no sampling frequency. Instead, their analogue signals may be digitized to any desirable frequency up to at least 10000Hz (Collewijn, 2001), who remarked that "The choice of 10000Hz followed from the general rule that the (temporal) resolution of a measurement should preferably be an order of magnitude better than the expected effect." (p. 3417). For video-based eye trackers, the video camera and its settings determine the sampling frequency.

Sampling frequency is one of the most highlighted properties of modern eye trackers, often being either a part of, or mentioned directly in connection to the model name. The competition for higher sampling frequencies has made some manufacturers of video-based eye-tracking systems with multiple cameras interleave image acquisition to achieve higher effective sampling rates. For instance, the Tobii Glasses 2 have two cameras per eye, each sampling the eye at 50Hz. This system is made into a 100Hz eye tracker by alternately sampling each camera. However, the alternating samples are offset in the resulting data, yielding a zigzag pattern that is very common in 100Hz data from Tobii Glasses but does not happen in 50Hz data (see Figure 11 in Niehorster et al., 2020b). The EyeFollower from LC Technologies uses two 60Hz cameras, one per eye, to achieve a net gaze sampling rate of 120Hz by alternatingly sampling the right and left eyes.

In theory, high sampling rates *when combined with low velocity noise* would allow for very precise determination of velocity and acceleration, and therefore facilitate more precise determination of on- and offset of fixation, saccades and other events. This would obviate the need for filtering and for averaging metrics such as saccade latency / fixation duration over large numbers of trials, which are difficult to record with patients and other groups that only provide small samples.

In practice, however, the many different eye trackers exhibit a large variation of both sampling frequencies and precision levels. Research on the relation between eye-tracking measures and sampling frequency shows that some outcome measures (e.g. fixation durations) are less sensitive to sampling frequency, whereas others (saccadic peak velocity) are more so.

For instance, Andersson et al. (2010) quantified the effect of sampling frequency on event durations, such as fixation durations, in a series of simulations and tests on human eye-movement data. They also provided estimates of the number of measurements that are required to average out the mis-estimations of the on- and offset of fixations due to a low sampling frequency.

Saccadic peak velocity measures are more dependent on sampling frequency, but exactly how much more is a matter of debate. Wierts et al. (2008) showed that although a 50Hz eye tracker cannot provide accurate saccadic peak acceleration/deceleration values, it can be used to accurately measure peak velocities without aliasing if saccades are at least 5°. Inchingolo and Spanio (1985) used a 200Hz EOG system and found that saccade duration and velocity values in that data were comparable to those obtained in data of a 1000Hz system, as long as the saccades were larger than 5° in size. However, using EOG- and photoelectric eye-tracking systems to study 20 saccades, Juhola et al. (1985) provided evidence that sampling frequency should preferably be higher than 300Hz in order to reliably calculate the peak saccade velocity. Mack et al. (2017) replicate the finding that the peak saccade velocity estimation is more inaccurate for lower sampling frequencies. Unfortunately, these somewhat contradictory results are made more difficult to interpret because of differences in the precision of the eye trackers, how velocity is calculated, and whether filters were involved in the velocity calculation. The observations that both DPI and P–CR technologies misestimate saccade velocity (e.g. Hooge et al., 2016) add complication to the interpretation of these studies.

Temporal precision

Temporal precision is the variation in the inter-sample durations. A perfect temporal precision means that samples always arrive after exactly the same time interval. However, when temporal precision is poor, there could sometimes be, for instance, 33ms between samples, and other times 43ms (actual intervals found in data from an EyeTribe, Holmqvist and Andersson, 2017, p. 193). This is indicative of an unstable sampling frequency, the explanation for which could be in small head movements, the camera type and transfer protocols as well as image processing. Examples of eye trackers with unstable sampling frequencies include the EyeTribe (Ooms et al., 2015), the Pupil Labs 240Hz (Ehinger et al., 2019), the Tobii 1750 (Shukla et al., 2011), and the SMI REDm 60/120 and the SMI RED 250 (Hessels et al., 2015). Some implementations of algorithms for filtering, velocity and acceleration calculation, as well as event detectors, may assume a stable sampling frequency, and may thus not be suitable for data with unstable sampling frequencies.

Spatial precision

Precision ranges reported in the publications of Table 2 vary between eye trackers with a factor of 100 or more (median RMS-S2S deviation 0.001–0.75°). Precision ranges vary little with calibration, and can be calculated from participants (and artificial eyes) without their cooperation. Precision calculations can be made in many different ways (Niehorster et al., 2020c). The resulting precision values change when filtering the gaze signal with the built-in manufacturer filters (Niehorster et al., 2021).

Precision recorded with human eyes is often worse (e.g. higher RMS-S2S deviation) than precision recorded with artificial eyes (Holmqvist et al., 2021; Niehorster et al., 2020c), but different artificial eyes may also result in different precision levels.

Niehorster et al. (2020c) investigated how four different precision measures correlate, depend on sampling frequency and express different properties of the signal. In particular, RMS-S2S deviation reflects the noise velocity in the signal, while STD (standard deviation) and BCEA of the gaze signal (bivariate contour ellipse area, Steinman, 1965; Crossland and Rubin, 2002) are measures of the dispersion of gaze samples. The slope α of the power spectrum density instead measures the colour of the noise, as does RMS-S2S divided by STD (for the same gaze data).

Together, these four measures allow for a more complete characterization of the precision in gaze data from an eye tracker. Niehorster et al. (2020c) provide code to generate noise based on this characterization. Adding synthetic noise to data is a method to test event detectors, and can also be used to provide identification privacy in future consumer products with inbuilt eye-tracking systems (Liu et al., 2019).

Filters

The most common way to reduce (improve) precision values is to employ a filter. McConkie (1981) proposes that all filters should be reported. Filtering of the resulting data stream compensates for noise generated earlier at the level of sensors, light, fans and more. However, filtering affects various characteristics of the signal differently, and using the four different measures above allows researchers to investigate whether filters are present (Niehorster et al., 2021).

Ko et al. (2016) remarked that an optimal filter should be based on (a) a characterization of the noise level and (b) the component of eye movements one is interested in examining. Most other design criteria of filters seem to be guided by heuristics, or 'rules of thumb', motivated by visual inspection of the data (e.g. Stampe, 1993). Notice that pattern matching filters, such as those described by Stampe (1993, p. 138, known as the heuristic filter in EyeLink and SMI trackers) and Duchowski (2007) amplify parts of the gaze signal with a similar appearance to the filter pattern, while attenuating other portions. Špakov (2012) compared several noise filters, and revealed that finite-impulse response filters with triangular or Gaussian kernel (weighting) functions, and parameters dependent on signal state, show the best performance, as judged by a comparison to idealised saccade models using multiple criteria.

Derivatives of the gaze position signals are used by both researchers and event detection algorithms. Numerical differentiation of a signal however amplifies high frequency content (which is usually noise) in the signal. Specific filters are therefore often used to counteract the increased high frequency noise resulting from differentiation. The most detailed investigations of these filters were conducted by Inchingolo and Spanio (1985) and Larsson (2010), who showed how saccade parameters (e.g. duration and peak velocity) were affected by the type of differentiation filter and peak velocity threshold in the event detector. Larsson (2010) concluded that the Savitzky–Golay filter used by Nyström and Holmqvist (2010) and the differential filter used by Engbert and Kliegl (2003) produced eye movement velocity and acceleration most like those found in literature. Unlike the pattern-matching filters, these two filters make no strong assumptions on the overall shape of the velocity curve.

Data loss and interpolation

Several studies have shown that average data loss differs between eye trackers. Holmqvist (2015) report that the video-based eye trackers SMI HiSpeed 1250 and the EyeLink 1000 had the lowest data loss with around 3% of the raw data samples lost on average, while the Tobii T60 XL and the TX300 lost 15% or more. Nevalainen and Sajaniemi (2004) report 3.0–8.7% data loss for the Tobii 1750 and two ASL trackers, while Funke et al. (2016) found 22% in EyeTribe and 24% data loss in Tobii EyeX. For reference, around 2% of the data are lost due to blinks (Holmqvist and Andersson, 2017, p. 167). In contrast to the values reported for the TX300 by Holmqvist & Andersson (2017, p. 167), Hessels et al. (2015, Figure 6) reported less than 3% data loss for the TX300 for upright head orientations, and Hessels & Hooge (2019, Figure 9) reported less than 10% data loss in 10-year old children measured with the TX300. There is thus a large range in the reported data loss values for each eye-tracker model. This suggests that not only the eye-tracker hardware itself plays a role, but also operator experience, participant groups, lighting conditions, stimuli and experimental procedures, and laboratory protocols. This should be taken into account when interpreting data loss values reported in the literature.

Furthermore, Castner et al. (2020) reported that data loss values produced by manufacturer software are not always reliable. They found that for a participant with a reported tracking ratio of 98% (a data loss of 2%), an additional large gap in the left eye gaze signal–approximately 3.5s out of a 90s recording–appeared as data loss, but was labelled as a blink.

Fixation points positioned in the corner of the monitor, as well as recording participants with downward-pointing eye lashes and large head movements tend to result in higher data loss (Hessels et al., 2015; Holmqvist et al., 2011; Niehorster et al., 2018), though the operator might have a significant influence as well (Hessels & Hooge, 2019).

Data loss may affect the output of event detection, if the event detector terminates fixations and other events whenever a period of data loss is encountered. Holmqvist et al. (2012) added increasing amounts of data loss (as short segments) into data with no data loss, and found that 18% data loss reduces the number of fixations by about one quarter, and increases their average duration by around 50ms, when using the Nyström and Holmqvist (2010) algorithm. Hessels et al. (2017) found that adding periods of data loss to eye-tracking data affected the number of fixations and corresponding fixation durations for different event detection algorithms strongly and idiosyncratically.

Some algorithms merge fixations close in time and space where there are small bursts of data loss (Komogortsev et al., 2010; Wass et al., 2013; Zemblys et al., 2018), reducing some of the effect of periods of data loss. The solution to gaps in data in the Tobii Pro Lab software is to allow users to fill the gaps of data loss using a linear interpolation with synthetic data. This interpolation is selected in the event detection dialog menu in the Tobii software. The I2MC algorithm (Hessels et al., 2017) also employs interpolation of gaps up to a certain duration, but instead uses a non-linear Steffen interpolation (Steffen, 1990).

Latency, gaze contingency

Latency (also known as temporal accuracy and end-to-end delay, e.g. Reingold, 2014) is often defined as the average end-to-end delay from the time of an actual movement of the tracked eye until the recording computer signals the eye movement. Theoretically, there is always a latency of a few milliseconds, and in the optimal case, it is constant. Any processes run by the computers involved in the data recording may add to this basic latency.

A known constant latency is uncritical for most research (except closed-loop, gaze-contingent experiments). A variable latency, which translates to high temporal imprecision, is much more critical, as it cannot be easily compensated for, particularly if the eye tracker does not provide reliable timestamps.

A large and variable latency is somewhat tricky to detect, measure, and prevent, and may come as an unpleasant surprise long after data were recorded. McConkie (1997) looked back at the foundational work on reading using gaze-contingency (McConkie & Rayner, 1975), and remarked that they were unaware of a filter in the eye-tracker circuitry that increased the latency by 25ms between the eye movement and the registered signal, potentially undermining their conclusions.

Table 3 lists existing measurements of eye-tracker latencies. Measurement type 1 concerns the time from when an eye movement is made until the output gaze coordinates change, while measurement types 2–5 include the time needed to update the monitor.

Gaze-contingent paradigms and latencies Whether a gaze-contingent paradigm – for instance, boundary and moving window paradigms (Hohenstein & Kliegl, 2014; McConkie & Rayner, 1975; Nuthmann, 2013) or saccadic adaptation paradigms (McLaughlin, 1967; Pélisson et al., 2010) – can be run without exceeding the maximum allowed latency depends on how quickly a gaze coordinate can be fed back to the stimulus program so that the stimulus monitor can be changed without the participant realizing (facilitated by saccadic suppression, Campbell & Wurtz, 1978; Holt, 1903). Loschky and Wolverton (2007) reported that it is enough to update the stimulus image within 60ms after the onset of the eye movement. However, Slattery et al. (2011) point out that the position of gaze during the display change has an effect on fixation durations (for the next word after the boundary) that can be seen already at 15–25ms delay of the signal. This behavioural change indicates detection of the manipulation, and the delay can be compared to the measured latencies in Table 3. Note that a single detection may be enough to affect behaviour, which means that maximum latency, rather than the mean, would be the most relevant comparison.

Saccade latency measurements versus system latencies In other cases, researchers are concerned whether their eye-movement recording was properly synchronized to stimulus onsets on their displays. Improper synchronization would for instance affect eye latency measures, such as saccadic latencies. One method to check this has been to compare the eye video to the file of the raw data stream or gaze scanpath (Morgante et al., 2012). This however has the drawback that both data streams are generated by the same software, and could be affected by the same latencies. Also, the video is

Table 3 Studies of eye-tracker mean latencies. While measurement type 1 compares the duration from an eye movement starts until a change in gaze coordinate, measurements 2–5 include the time needed to update the monitor in a gaze-contingent setup. Numbers in brackets denote standard deviations

Type of latency measurement	Eye tracker	Mean latency	Reference
1: Compare raw data file against video of participant eye	TX300	20.3–24.1ms	Leppänen et al. (2015)
	TX60XL	44.5ms (7.3)	Morgante et al. (2012)
Timing of gaze data vs network hub time	T120	33ms (8.9)	Creel (2012)
Comparisons of VOG against EOG-system baseline	Five VR-eye trackers	45–81ms	Stein et al. (2021)
2: Artificial eye with diodes until display change	EL1000, screen 160 Hz	4.82ms (1.86)	Reingold (2014)
	EL1000, screen 60 Hz	9.69ms (4.79)	Reingold (2014)
Artificial eye with diodes / constituent latencies	ELII	10.5ms (0.7)	Bernard et al. (2007)
3: High-speed camera films eye and monitor through mirror	T1750	27ms	Shukla et al. (2011)
4: Blinding the eye tracker + high-speed camera	EL1000	12–40ms	Saunders and Woods (2014)
5: Measure display changes against saccade onset	EL1000	10ms	Hohenstein and Kliegl (2014)

usually of a low temporal resolution in comparison to the eye-tracking data, which limits detection of synchronization issues to the temporal resolution of the video recording. As an alternative method of measuring synchronisation, Shukla et al. (2011) used a mirror positioned next to the participant and a 300Hz high-speed camera, which made a recording of the participant's eye and, through the mirror, the monitor where the stimuli appeared and disappeared. Results revealed a variable latency with a mean of 27ms on their Tobii 1750, similar to the latencies reported by Leppänen et al. (2015) in a study using the same approach with a low temporal resolution camera and a Tobii TX300, while Morgante et al. (2012) reported latencies of up to 54ms for the Tobii TX60XL.

Fixation and saccade detection

Historically, fixation and saccade detection were conducted manually and was very time-consuming. For instance, Hartridge and Thomson (1948) presented a novel method to process eye movements at a rate of approximately 10000s (almost three hours) of manual work for 1s of recorded data. Decades later, Monty (1975) remarked: "It is not uncommon to spend days processing data that took only minutes to collect" (p. 331–332). Today, software can run a similar analysis in a matter of minutes, even for several hours of recorded data. Potential reasons for still doing manual analysis include that it allows for better general monitoring of data quality as well as participant performance and engagement.

Event detection algorithms (or event classification, see Hessels et al., 2018) are used to process a time series signal (gaze position, pupil size, etc.) into labelled, meaningful units, such as fixations, saccades, blinks, etc. What happens inside the event detection algorithms was considered important enough by McConkie (1981) that he recommended that details about these algorithms should be published in the paper presenting the processed events.

Note that operationalisations for fixations may depend on the frame of reference (i.e. whether the eye tracker is fixed to the world or to the head). A moving observer that fixates a static object in the world, produces a gaze point in the world that is stationary with respect to the object, but slowly moving with respect to the head. This point is extensively discussed in Lappi (2016), Holmqvist and Andersson (2017, Chapter 7) and Hessels et al. (2018).

There are many different event detection algorithms available. Here, we describe a select number of them to give an idea of the breadth and scope. The I-DT finds fixations using a spatial threshold on maximum gaze dispersion (typically 0.75–1.5°) and a temporal threshold on minimum fixation duration (typically 50–150ms). What remains are assumed to be saccades. The I-VT instead finds saccades using a minimum peak velocity criterion (such as 20–100°/s), and assumes that everything in between saccades are fixations. The I-DT and I-VT were described by Salvucci and Goldberg (2000), and later appeared in software from manufacturers. For instance, BeGaze by SMI offers both the I-VT and the I-DT algorithms, whereas Tobii Pro Lab provides a version of the I-VT, and the Data Viewer by SR Research has an I-VT-related saccade detector with both velocity and acceleration thresholds.

The NH2010 algorithm by Nyström and Holmqvist (2010) is an improvement of the I-VT algorithm which adapts the peak velocity threshold to the level of noise in the data, and additionally outputs detected post-saccadic oscillations. The I2MC by Hessels et al. (2017) is an algorithm designed to be robust against increasing levels of noise and data loss common in infant research.

GazeNet by Zemblys et al. (2019) is a fully end-to-end machine learning-based event detector that learns from examples, and detects fixations, saccades, and post-saccadic oscillations with very high resemblance to human expert coders. The Deep eye movement classifier by (Startsev et al., 2019) is another recent machine-learning algorithm that also detects periods of smooth pursuit in data.

There also exist dedicated event detection algorithms for data from head-mounted eye trackers, used to describe gaze behaviour during e.g. navigation in real environments (Hessels et al., 2020; Niehorster et al., 2020a). For researchers interested in labelling eye-tracking data from head-mounted eye trackers into smooth pursuit, fixations during head movements, OKN, vergence etc, no automated techniques exist at the moment. However, this is a quickly evolving field, in which relevant work is done on some of the problems it involves (Kothari et al., 2020; Larsson et al., 2014).

Furthermore, there are many other special-purpose event detectors (for instance, blink detectors, microsaccade detectors, algorithms for desaccading smooth pursuit or nystagmus data, and smooth pursuit detectors), summarised by Holmqvist & Andersson (2017, Section 7.4).

Most event detection algorithms are offline, operating on already recorded data. However, for gaze-contingent research, event detection algorithms have to be fast and online, operating in real-time when saccades happen (Holmqvist & Andersson, 2017, p. 234–235). This online algorithm is necessary in the Fixation-Contingent Scene Quality Paradigm (Henderson et al., 2013; Walshe & Nuthmann, 2014). In the boundary paradigm, however, there is just a simple check whether raw data (typically one eye only, see discussion in Nuthmann & Kliegl, 2009, p. 23) have crossed the boundary, assuming such a crossing to mean that a saccade is in progress (see also Slattery et al., 2011).

The risk that poor precision poses for the detection of small eye movements

Small eye movements may be hidden in the noisy, imprecise parts of data. For instance, Fig. 2A shows how the large saccades are often followed by small saccades which are clearly seen and reasonably easy to detect by algorithms. In Fig. 2B, the big saccades are visible, but the small saccades, if they were made during the recording, have left a trace that is harder to distinguish from noise, for human data inspectors and algorithms alike.

The degree to which outcome measures of event-detection algorithms are sensitive to the noise level has been systematically investigated by Hessels et al. (2017), Holmqvist (2016), and Holmqvist et al. (2012), who all investigated the effect of artificially increasing noise levels (degrading precision) on the outcome of event detectors, and by van Renswoude et al. (2018), who investigated correlations between precision and outcome measures. Effect sizes are large; for instance, using the algorithm by Nyström and Holmqvist (2010), Holmqvist et al. (2012) compared the precision levels 0.03–0.37° and found an increase of average fixation durations from 430ms to 630ms and a reduction of the number of fixations by about one third, for the same eye-movement data. Hessels et al. (2017) and Holmqvist (2016) report (and illustrate in figures) how for some algorithms, no fixations whatsoever are found when imprecision increases beyond a certain level.

Algorithm settings

Event detection algorithms have a variety of settings, some examples of which are the minimal peak velocity threshold for saccade detection (I-VT, EyeLink), the minimal fixation duration and the maximum gaze dispersion for fixations (I-DT). Changing the settings of these algorithms can have large effects on measures such as number and duration of fixations and saccades (Blignaut, 2009; Holmqvist, 2016; Manor & Gordon, 2003; Shic et al., 2008). For some experimental designs, in particular between-subjects comparisons and when comparing between studies, or when conducting replication studies, a change of algorithm settings may have an impact on the rejection of a hypothesis (see for instance, Shic et al., 2008, for a within-subjects design with comparison between different stimulus types).

Settings can be manually adapted based on for instance the precision of the data. Holmqvist (2016), and (Holmqvist and Andersson, 2017, Ch 7) provide practical advice on the relationship between precision and settings and the outcome measures, for two commonly used algorithms: I-DT and I-VT. The larger the saccades are in the task, the higher the thresholds can be. Studies with a focus on small saccades need good precision and low thresholds.

There are also adaptive algorithms that change the thresholds based on the precision in the data (e.g. Braunagel et al., 2016; Engbert & Kliegl, 2003; Hooge & Camps, 2013; Mould et al., 2012; Nyström & Holmqvist, 2010). However, an adaptive algorithm does not solve the problem of variable precision, as it may adapt the parameters to the level of noise, but changed parameters have consequences in the fixation and saccade output by the algorithm. Hessels et al. (2017) developed an algorithm which had the explicit goal to be robust to differences in data quality and enable comparisons across conditions when there are differences in data quality. Note, however, that although noise-resilient algorithms may produce fixations that result in the same average fixation duration from data of varying precision, further investigations are needed to assess the extent to which the individual events (their on- and offsets) change as precision varies.

Algorithm comparisons

Not everyone is free to choose which event detection algorithm to use, but for those who are and want an algorithm adapted to their wishes, there are many algorithms to choose from. The many existing event-detection algorithms do not necessarily produce the same output measures when given the same eye-tracking data. In fact, several algorithm comparisons have reported large differences in fixation and saccade measures between algorithms (Andersson et al., 2017; Benjamins et al., 2018; Dalveren & Cagiltay, 2019; Komogortsev et al., 2010; Salvucci & Goldberg, 2000; Stuart et al., 2019). This research suggests that differences in, for instance, average fixation durations between studies that use different algorithms may in part stem for differences between the algorithms.

It has become common that developers of algorithms benchmark their novel algorithm against previous ones (e.g. Hessels et al., 2017; Otero-Millan et al., 2014; Zemblys et al., 2018, 2019). Event detectors based on machine learning have started to appear, whose behaviour cannot be fully described in terms of rules that relate to concepts humans have about the eye-movement signal. Consequently, trust in the algorithm derives from benchmarking against human coders or existing algorithms (Zemblys et al., 2019).

There is an ongoing discussion around the methods in building and evaluating event detectors, in particular how to calculate inter-rater reliability, used to compare algorithms against algorithms or against human coders (e.g. Friedman, 2020; Startsev et al., 2019; Zemblys et al., 2019, 2021). Other current topics concern whether human coding of events is a good benchmark to test the algorithms against (Hooge et al., 2018), or build algorithms from (Zemblys et al., 2019), and what kind of noise to add to the data when testing the noise-robustness of an event detector (Niehorster et al., 2020c).

Event operationalisation

Fixations, saccade latencies, amplitudes, and curvature have been operationalised in more than one way. For instance, a common way to calculate saccade amplitudes is to calculate the Euclidian distance between start and end of a saccade (e.g. van der Geest et al., 2002). Alternatively, the amplitude can be measured as the distance along the saccade path (calculated, for instance, as duration multiplied by average velocity). These two amplitude calculations will differ for curved saccades (Holmqvist & Andersson, 2017, p. 613).

Different algorithms calculate fixation durations and other measures in different ways (Andersson et al., 2017). In particular, some algorithms exclude the post-saccadic oscillation (PSO) from both the saccade and the following fixation event (e.g. Nyström & Holmqvist, 2010; Zemblys et al., 2019), while the I-VT algorithm and the EyeLink algorithm have no separate detection of PSOs and assign parts of the PSO either to the saccade or the fixation, largely depending on the amplitude of the PSO.

Area-of-interest (AOI) measures

Areas of Interest (AOIs, also known as Regions of Interest, ROI, and Interest Areas, IA) are employed when the researcher's interest is in the relation between gaze behaviour and the visual world (e.g. Buswell, 1935; Viviani, 1990). Researchers may be interested in what parts of a webpage attract gaze most effectively, and in what order (Goldberg et al., 2002), or interested in gaze behaviour while listening to ambiguous sentences about a scene (Allopenna et al., 1998). AOI-measures such as absolute or relative time spent in AOIs or the number of transitions between various AOIs may be used for this.

Areas of Interest provide fundamental processing tools for the analysis of eye-tracking data, and are used in many branches of cognitive psychology, architecture, marketing, clinical research, neuroscience, educational science and many other fields. Multiple methods exist to relate the AOIs to the stimulus, presented by Holmqvist & Andersson (2017, Ch 8), Hessels et al. (2016), and Orquin et al. (2016).

There are methods that assist with the same function that AOIs are used for, but that are not referred to as AOIs: Reading researchers use non-proportional fonts and oftentimes study single sentences only. This way, fixation-to-word and/or fixation-to-letter assignment is easily done post-hoc; all they need to know is the horizontal offset of the sentence and the PPC value (pixel per character), along with the actual sentence. This also makes gaze-contingent reading research (moving window and boundary paradigms) technically easier to implement. For reading researchers who prefer to use AOIs, both BeGaze from SMI and the SR Research stimulus presentation software automatically segments text into AOIs at the word, sentence, and character level.

When the stimulus consists of animated material or videos, a static segmentation of space into AOIs may not suffice. Dynamic interest areas can be made to move in synch with the underlying object, but may require AOI measures to be calculated based on raw data samples rather than using fixations (e.g. because event detectors often are not reliable when smooth pursuit is present).

AOI size

The size of the AOI is of great importance. If the accuracy of the gaze data is poor, the eye tracker might report a gaze position that is outside the AOI, even though the participant was looking in that area, and vice versa (Holmqvist et al., 2012).

Hessels et al. (2016) report the effects of altering the size of AOIs (face stimuli) on important AOI measures (dwell time, total dwell time, time to first AOI hit), pointing out that effect sizes are large and the relationship is *non-linear*. Below a certain AOI size, the total dwell times are no longer significantly different between the two AOIs (eyes vs mouth) used in their study. Orquin et al. (2016) reanalysed four experiments using different AOI sizes, and found only some effects of varying AOI size on the outcome of the statistical analysis. Orquin et al. (2016) also note that one third of the researchers in their survey reported conducting analyses with multiple AOI sizes, which may help confirming that the result is robust over all AOI sizes.

Orquin and Holmqvist (2018) present simulations where they vary AOI size, the shape and position of the AOIs, and accuracy and precision, and investigate the effect on the AOI measure hit rate. They report complex, non-linear interactions between data quality measures and AOI properties.

Not only the inaccuracy of the eye tracker matters when calculating AOI measures from AOIs of different sizes. The minimum size of an AOI that encircles a target stimulus is also limited by the inaccuracy of the visuo-oculomotor system when targeting small objects, which can be larger for some participant groups (Clayden et al., 2020; Pajak & Nuthmann, 2013).

It has been suggested that margins should be added around AOIs to compensate for inaccuracy (Holmqvist &

Andersson, 2017; Orquin et al., 2016), which may or may not be possible depending on how densely populated the stimulus is. Hooge and Camps (2013) point out that if the visual stimulus is sparse, AOIs could be made as large as possible, sharing the remaining empty space between nearby AOIs. Their argument is that in sparse stimuli, there is not much crowding, and the functional visual field is large (Engel, 1971; Toet & Levi, 1992). A large functional visual field implies that objects are visible at larger eccentricities (or larger distance from the gaze point), allowing observers to overview larger areas around the gaze point.

Higher-order measures

Outcome measures that build upon or are derived from AOI or fixation and saccade measures could be referred to as higher-order measures. As a rule of thumb, the higher-order measures have a large number of settings that can be varied, whether in

- *scan path analysis* (Anderson et al., 2015; Cristino et al., 2010; Dewhurst et al., 2012; Duchowski et al., 2010; Jarodzka et al., 2010; Kübler et al., 2014)
- *(hidden) Markov models* (Chuk et al., 2014; Coutrot et al., 2018; Ellis & Stark, 1986)
- *recurrence quantification analysis* (Anderson et al., 2013; Pérez et al., 2018)
- *entropy analyses* (Allsop & Gray, 2014, 2017; Hessels et al., 2019; Hooge & Camps, 2013; Krejtz et al., 2014; Niehorster et al., 2019)
- *heatmap-based analysis* (Caldara & Miellet, 2011)

It is reasonable to expect that data loss, as well as poor precision and accuracy, will be carried through event detection and AOI procedures, and propagate into these higher-order measures. Similarly, settings in the event detector and choices of AOI sizes may also have strong effects on the higher-order measures.

To date, very few studies have been made of the effect on higher-order analyses of changing settings and varying data quality. One example is Krejtz et al. (2015), who show that the size of gridded AOIs affect gaze transition entropy results, with non-linear relationships and large effect sizes in outcome entropy.

Summary

We have reviewed research on how the eye tracker, methodology, environment, participant, settings of event detectors and AOI tools, etc., affect (or relate to) the quality of the eye-tracking data obtained, the properties of the eye-tracker signals, and the eye-movement and gaze measures. Our review has shown that there exists a significant body of research that has investigated the quality of data from eye trackers and what this quality relates to.

These studies have reported that sunlight and luminance (*environment*) have large effects on gaze, that the accuracy, precision and data loss often vary significantly between different *eye trackers*, and that the *setup and geometry* of the recording situation is of great importance to the quality of the data.

These studies have also shown, for instance, that accuracy, precision and data loss vary between *participants*, depending on age, eye-region physiology and many other factors. We have seen that *calibration* matters for accuracy, and that operator skill and trial structures may influence outcome measures. We have learnt that some researchers use filters to counter poor precision, interpolation across gaps of data loss, and manual methods for re-aligning inaccurate gaze data.

The reviewed literature suggests that algorithms for *event detection* vary dramatically between studies and most algorithms are highly influenced by both precision and settings. Other research has quantified the large non-linear effect of data quality on *area-of-interest* and *higher-order* measures.

In the next section, we will examine how the various factors reviewed above are reflected in current reporting practices and guidelines.

Reporting practices and existing reporting guidelines

The many studies reviewed in the previous section show that the knowledge exists to help make good choices when conducting a study with an eye tracker. Is this knowledge readily applied by researchers using eye trackers? How does our literature review (Section "A review of empirical eye-tracking studies as the basis for a reporting guideline") of important aspects of an eye-tracking study compare to the reporting practices of researchers using eye trackers? In the current section, we first summarise reporting practices from a database of 207 eye-tracking studies of judgement and decision-making (see Fiedler et al., 2019, for details) and discuss this in the light of our literature review. We then discuss reporting practices in light of five existing reporting guidelines, which attempt to make explicit what researchers are expected to report.

Reporting practices

The reporting database used here was first made public on https://decisionlab.shinyapps.io/iGuidelines/ on June 13, 2018, and later on https://osf.io/ysvzk/?view_ only=1be57d949dff43e99189ec6ad13f8a23 as supplementary material to the present paper. Table 4 present a comprehensive synopsis of this section.

Environment Only 12.5% of the 207 publications in the database report the location and setting where data were recorded.

Eye tracker Table 4 presents data showing, for instance, that the eye-tracker model (90.8% of studies) and eye-tracker sampling frequency (75.8%) are often reported. While ranges of data quality values differ radically between eye trackers, sampling frequency is of importance only in some cases (Section "Signal properties and processing"). In contrast, the fundamental data quality measures–accuracy, precision, data loss and latency–are virtually never reported in the 207 publications of the database. Only 4.3% of the studies reported precision, and only 3.8% reported data loss. Only 0.5% of the studies were found to have reported a (measured or reiterated) latency value. Studies report the manufacturer's specified accuracy (29.3%) almost ten times more often than self-measured accuracy (3.5%).

Table 4 Synopsis of reporting frequencies of different aspects of studies derived from the reporting database

Aspect	%
Environment	12.5
Eye-tracker model	90.8
Sampling frequency	75.8
Number of calibration targets	41.1
Accuracy, vendor specification	29.3
Accuracy, self-measured	3.4
Precision	4.4
Data loss	3.8
Latency	0.5
Monitor resolution	56.6
Monitor size	29.6
Participant to eye-tracker distance	56.5
Use of chinrest	27.5
Method of stimulus presentation	17.9
Analysis software	44.9
Gender	77.8
Age	67.0
Visual aids	40.6
Number of participants	99.0
Number of trials	94.2
Duration of recording	31.0
Attrition rate	51.2
Dependent variables	100.0
Operator skill	0.5
Recalibration	16.4
Event detector	27.0
Provide example stimulus figure (with AOIs)	76.3
AOI size	24.0

Geometry and setup 56.5% of studies reported monitor resolution, while only 29.6% reported its physical size. Furthermore, 56.5% of studies reported the distance between participant and eye tracker (range 18–280cm, with 60 and 70cm being most common). To make full use of one of these measures, the other two are usually also required. Reporting all three measures is done in 20.3% of the studies. In comparison, 27.7% of the studies report that the authors applied a chinrest during recordings, their reasons are not revealed by the database.

The software used for stimulus presentation was reported by 17.9% of the studies. 44.9% of the studies reported which software was used for data processing and analysis. The most commonly reported processing tools were SMI BeGaze, Tobii, and SR Research Data Viewer, while the most common statistical tools were SPSS, R and Matlab. Papers that investigate the relationship between software tools and data quality are to the best of our knowledge currently lacking.

Participants The gender distribution is reported in 77.8% of the publications in the reporting database. Although gender is potentially relevant to certain aspects of some studies, there is no clear evidence that it is related to eye-tracking data quality and only to a small extent to aspects of eye movement behaviour. Age is reported by 67% of the studies in the reporting database, and in contrast to gender, age was found to relate to smaller pupil, more frequent use of spectacles, droopier eye lid and other issues that affect data quality as well as changes in the eye movements themselves (Section "Participants"). Of those studies that report age, the average age is below 25 years in 67.4% of the publications, and between 26 and 46 years in the remaining 32.6%. Use of spectacles or lenses for correction for poor visual acuity of participants is reported by 40.6% of authors. Reports of having excluded recorded participants from further analysis were found in 51.2% of the publications, in which case exclusion criteria were always given.

Calibration 59.4% of the studies report having calibrated only at the beginning, versus 16.4% who reported having recalibrated at some point during the study. 41.1% reported the number of calibration targets, with 9 points being most common (67% of those studies that report number of targets), and 5 (17.6%), 13 (5.8%), and 3 (3.5%)

occasionally used. Only 2.4% of all studies reported the background colour of the screen during calibration.

Features of the experiment 99% of the studies in the reporting database report the number of participants (on average just above 40). As an example of the range, Noton and Stark (1971) used data of 2 participants in the first, and 4 in the second experiment, whereas Coors et al. (2021) compared eye-movement data of almost 4000 people to draw their conclusions. 94.2% report the number of trials (on average just below 60). 31% of the authors report the duration of the total recording. Of those who report this duration, 31% report durations of 16–30 minutes, 28% 31–45 minutes and 20% 46–60 minutes. Only one study (0.5%) reported who recorded their data.

Unsurprisingly, 100% of the authors reported which dependent variables were used. This number does not necessarily mean that reporting dependent variables is straightforward. Naming of dependent variable is often unclear. For instance, dwell time is also called gaze duration and glance duration, depending on the research field. Sometimes terminology is confused, as when fixation duration is called dwell time, or when time to first fixation is named saccade latency or saccadic reaction time.

Exclusion criteria Exclusion criteria for trials and events were reported by 30.9% of the authors, while 53.[illegible]% report having used exclusion criteria for participants. Exclusion criteria are composed from conditions for data quality and event values, personal characteristics, behavioural mishaps by the participants or operators, technical issues, and more.

Event detector Overall, 27.[illegible]% reported the event detector that was used. Among those authors who used fixation-based or saccade-based measures as their dependent variables, 37.0% reported their event detector. However, only 2.1% of those authors who used event detectors in their analysis reported precision, compared to 4.3% of the authors overall.

Areas of interest 76.3% of the authors in the reporting database included a figure with a stimulus image in their publication (which may have included an AOI drawn onto it). Of those who use AOI analyses, 28.7% report accuracy, although these authors always reiterated values from manufacturer specifications and never measured accuracy in their own data. 24% reported the size of their AOIs. 33% of the authors in the reporting database stated that the AOI was larger than the stimulus object (margin included), 27.5% that the AOI and the stimulus object were the same size (no margin), and 5% that the AOI was smaller than the stimulus object (negative margin). 1% of the authors used overlapping AOIs, 68% made clear that their AOIs do not overlap, while 31% failed to mention either. Only 8% mentioned the distance between AOIs, whereof 3% stated a zero distance between AOIs, and the rest reported distances between 5 and 241cm.

Summary Many authors in the database report dependent variables, number of participants, eye-tracker sampling frequency, and eye-tracker model, which are readily available in most studies, but often fail to report measures and settings that we have found to be relevant from a data quality perspective. We can only speculate as to why this is: Lack of knowledge of what is relevant to report may play a role. Some researchers may find it unclear how to measure and calculate accuracy and precision. An over-reliance on the eye tracker and its software may add to that, as evidenced by the large proportion of authors reporting manufacturer-specified accuracy (29.3%) rather than measured accuracy (3.5%). In sum, we conclude that there is a discrepancy between reporting practices (the current section) and what is relevant to report for a study using an eye tracker (Section 4).

Existing reporting guidelines

The discrepancy between what is relevant from a data quality perspective, and the actual reporting practices, raises the question whether it is difficult for the users of eye trackers to find out what they need to report.

There are at least five existing reporting guidelines (Carter & Luke, 2020; Fiedler et al., 2019; McConkie, 1981; Oakes, 2010; Strohmaier et al., 2020). McConkie (1981) provides an early but still remarkably relevant example of general publishing guidance for eye-movement research, from an era when researchers often built their own eye trackers, and there were only a few manufacturers who sold them. In 2010, the journal Infancy adopted a policy for what to report in eye-tracking studies (Oakes, 2010). In the field of eye tracking in decision-making studies, Fiedler et al. (2019) proposed a reporting standard aimed to support replicability, based on suggestions from a panel of researchers. Carter and Luke (2020) provide a standard for reporting for eye-tracking studies, as part of a broader goal to describe best practices in a variety of disciplines around psychophysiology. In a review of eye-tracking research on mathematics education, in preparation of their guideline, Strohmaier et al. (2020) reported that "Although studies necessarily vary in the specific eye-tracking method they use, we found large inconsistencies in the reporting of these methods" (p. 165).

In Table 5, we summarise all recommendations that are common to at least two of the existing guidelines. Table 5 shows that there are also inconsistencies between the existing reporting guidelines. Although all five guidelines

Table 5 Features of eye-tracking experiments that were common to at least two existing reporting guidelines. Terminology used in this table is by necessity reduced, but as closely as possible quoted from each guideline publication. See original publications for details

	McConkie (1981)	Oakes (2010)	Fiedler et al. (2019)	Strohmeier et al. (2020)	Carter & Luke (2020)
Eye tracker		Details about the eye-tracking system	Model / brand	A precise description	Model and make
Sampling rate	Sampling rate	Sampling rate		Sampling rate	Sampling rate
Seating	Viewing distance	Camera-Participant distance		Stimulus-Participant distance	Set-up, viewing distance
Head stabilization		How head movements are dealt with		Movement restrictions	Chin-/Headrest used?
Calibration	Calibration task & table	Number & location of calibration points		Calibration procedure	Calibration, recalibrations
Accuracy	Accuracy (multiple tests)	Available information about accuracy		Average accuracy	Accuracy
Precision	Noise characteristics				Precision
Data loss			Percentage	Amount & reason	Amount & reason
Exclusion		Criteria & rationale	Number & reason		Criteria
Monitor	Visual angle of display	Visual angle (width/height)	Size & resolution	Framerate	Make, model, size & resolution
Blink algorithm	Blinks, squints & irregularities	Blinks, head movements etc.			Report & justify data cleaning
Event detection	How saccades are detected	Procedures & parameters	Aggregations & transformations	Algorithm & thresholds	Software used for event detection
AOIs		How defined, in pixels? Same over trials?	Absolute & relative size	Position & size	Size (pixels & °), matching between conditions

recognise the importance of reporting monitor properties and procedures for event detection, they diverge in everything else. Even when several existing guidelines recommend the same feature to be reported, they differ in the details such as which operationalisations and terminology they use.

For instance, McConkie (1981) presents three separate tests of accuracy that each researcher should conduct and report, while Oakes (2010) only requires available information about accuracy to be reported, which may suggest the accuracy specification by the manufacturer, and Strohmaier et al. (2020) ask for average accuracy, i.e. accuracy measured by the researchers in their own experiment. While Strohmaier et al. (2020) specifically ask for event detection algorithms and thresholds, Oakes (2010) asks that future papers provide specifics concerning the definitions of saccades and fixations.

Furthermore, each guideline appears to have its own specific focus, which may reflect the field it originated from. For instance, the guideline by Oakes (2010) requires that filtering and interpolation algorithms for post-processing of eye-tracking data be reported. This presumably reflects the fact that eye-tracking data in infant research tends to have poor precision and frequent periods of data loss, that may need interpolation and filtering (Section "Signal properties and processing"). Also, Oakes (2010) is the only guideline to ask for recovery time data to be reported: the time it takes to resume tracking when the eyes reappear in the eye-tracker camera view after a period of track loss.

The guideline by Strohmaier et al. (2020) asks for "correlation between all used measures", which is presumably intended to detect cases where multiple eye-movement measures are used as separate, independent corroborations of a single hypothesis, for instance number of fixations in an AOI, and dwell time in the AOI. A similar argument is made by Orquin and Holmqvist (2018).

The guideline by McConkie (1981) is the only one to emphasize measurement and reporting of linearity, system latency, drift, and multiple tests of accuracy.

Table 6 Detailed descriptions of each aspect are found in Section "A review of empirical eye-tracking studies as the basis for a reporting guideline'. Please note that these suggestions comprise neither a mandatory, nor an exhaustive list; common sense is highly recommended

Section	Aspect	Examples and units	Example reporting
Environment	Illuminance	lx	Ashby et al. (2009, p. 5348)
	Vibrations	Hz, amplitude	
	Infrared light sources	Sun light, hot light bulbs	
	Ambient noise	dB, distance from sound source, test tone	Webb et al. (2020)
	Presence of others	Adoption of standard protocol for others' positions, and documentation of protocol deviations	Webb et al. (2020, p. 6)
Eye tracker	Manufacturer		Holleman et al. (2020, p. 6)
	Model name		Nuthmann (2013, p. 809)
	Technique	P–CR, coil, EOG, etc.	Hooge and Van den Berg (2000, p. 2759)
	Filters (built-in)	Name of filter, on or off	Niehorster et al. (2021, p. 6)
	Soft- and firmware versions of any software used	Recording, stimulus presentation, SDK, processing and analysis	Niehorster et al. (2020b, p. 4), Nyström et al. (2021, p. 3)
Monitor	Dimensions	Physical size (cm), retinal size (°)	Erens et al. (1993, p. 146), Kok et al. (2017)
	Luminance	cd/m^2	Holmqvist et al. (2020, p. 4)
	Contrast	Michelson contrast, Weber contrast, or RMS contrast.	Paffen et al. (2005, p. 573)
	Other properties	Brand, type, resolution (in px)	Nuthmann and Kliegl (2009, p. 5)
	Participant to eye-tracker distance	Distance (cm)	Kok et al. (2017) and Nuthmann and Kliegl (2009)
	Head movement restrictions	Chin- and forehead rest, bite bar	Erens et al. (1993, p. 146), Pajak & Nuthmann (2013, p. 4)
Participant characteristics	Age		
	Visual acuity	LogMAR acuity or Snellen acuity	
	Visual aids / corrections	Spectacles, soft or hard lenses	Holmqvist (2015, p. 7)
	Makeup	Mascara, eye liner etc.	Holmqvist (2015, p. 7)
	Eye colour and specific iris features		
	Baseline pupil size	Physical size (mm)	Holmqvist (2015, p. 7)
	Interocular distance	Physical size (mm)	Holmqvist (2015, p. 7)
	Palpebral fissure (eye cleft)	Physical size (mm)	Holmqvist (2015, p. 7)
	Eye-lash direction	Up, down, forward	Holmqvist (2015, p. 7)
	Eye dominance	Miles' test	
	Eye recorded	Left, right, dominant, both	Kok et al. (2017) and Nuthmann and Kliegl (2009)
	Facial movements; due to speech etc.		Niehorster et al. (2020b)
	Sleep deprivation	Duration (e.g. h)	Ahlstrom et al. (2013, p. 22)
	Expertise		Kok et al. (2016) and Emhardt et al. (2020)
	Oculomotor atypicalities	(e.g. amblyopia, strabismus, nystagmus)	Thomas et al. (2011)
	Clinical characteristics	IQ, psychiatric conditions DSM or ICD; describe	Kapoula et al. (2014)
	Substance use (state effects)	Medicine, alcohol, nicotine etc.	Ettinger and Kumari (2019)

Table 6 (continued)

Section	Aspect	Examples and units	Example reporting
Calibration	Target		Holmqvist et al. (2020, p. 4)
	Number/position of calibration points		Nuthmann and Kliegl (2009, p. 6)
	Number/position of validation points	Same as calibrations point or not?	
	Binocular or monocular		Nuthmann & Kliegl (2009, p. 6)
	Control	Participant, operator, software	
	Operator expertise	Duration (yr) or # of participants	Hessels & Hooge (2019, p. 4)
	Drift compensation and recalibrations	Frequency, onset conditions	Greene & Rayner (2001), Vlaskamp et al. (2005, p. 248)
Experiment	Trials	Definition, number and durations	
	Instruction		
	Recording	Duration	
	Breaks	Plan for breaks and rest periods	
	Exclusion criteria	Trials or participants, due to characteristics, performance or data quality	Walcher et al. (2017) and Chisari et al. (2020)
Signal	Sampling frequency	Hz	
	Latency	ms	Hohenstein and Kliegl (2014), Saunders and Woods (2014), and Bernard et al. (2007)
	Data loss	%	Holleman et al. (2020)
	Attrition rate	Number, percentage or proportion	Burmester and Mast (2010)
	Spatial accuracy		Räihä (2015)
	Spatial precision	RMS-S2S (°), STD (°), or BCEA (°^{2})	Niehorster et al. (2020c, Table 1)
Events and AOIs	Name of event detection algorithm	Citation of publication	Nuthmann and Kliegl (2009, p. 24), Hessels et al. (2018)
	Algorithm thresholds	Duration (s), dispersion (°), or velocity (°/s)	Wenzlaff et al. (2018, p. 5)
	AOI size	Physical size (cm), retinal size (°), size on screen or in scene camera image (px)	Hessels et al. (2020) and Emhardt et al. (2020)
	AOI margins	–”–	Orquin et al. (2016), Clayden et al. (2020, p. 47)
	Distances between AOIs	–”–	Sun et al. (2016)
	AOI overlap	No, yes? How and why?	
	Sample or event-based AOI metrics		

The guideline by Fiedler et al. (2019) is the only one to ask authors to report on many experimental design parameters, such as inter-stimulus interval, counterbalancing of the position of AOIs, number of trials and the location where the data were collected. Furthermore, Fiedler et al. (2019) provide the most specific recommendations on reporting AOI details, which makes it surprising that Fiedler et al. (2019) do not recommend that accuracy be reported.

Carter and Luke (2020) is the only existing guideline to ask for basic demographic information, and the only one to also ask for "A list of the dependent variables selected for analysis, and a justification for that selection".

Existing guidelines may not be the obvious choice for all researchers. For instance, Uesbeck et al. (2020) did not make use of any of the guidelines above, but opted to report according to the CONSORT reporting guideline, which is used in the medical sciences (http://www.consort-statement.org).

Our summary suggests that previous reporting guidelines are incomplete and inconsistent, and often biased towards specific research fields. Therefore, in the next section, we will design a minimal reporting guideline based on empirical research, which may be used as is or form the basis for developing mandatory reporting standards.

An empirically based minimal reporting guideline

We have presented research showing how various aspects of a study with an eye tracker, such as the instrument, methodology, environment, participant, etc., affect the quality of the eye-tracking data obtained, the properties of the eye-tracker signals, and the eye-movement and gaze measures. We have summarised these aspects in Table 6. We have then shown that this body of research has not made any major imprint on current reporting practices. We have also shown that existing reporting guidelines for research using an eye tracker leave much to be desired.

Considerations on reporting guidelines for eye-tracking research

Ideally, all details necessary to replicate a study, or assess the validity of a study's claims, should be reported. The review above forms the empirical foundation for what may need to be reported in studies using an eye tracker.

Guidelines may also include requirements on, for instance, data formats and data sharing principles that make collaboration more convenient. Similarly, researchers publishing in an APA journal that requires gender and socioeconomic status of participants to be reported should report those. Such items are not specific to eye tracking, and are therefore not considered in this paper. For specific research fields, there may exist additional considerations for conducting eye-tracking studies (see e.g. Sharafi et al., 2020, for software engineering). Our minimal guideline can be appended with such items for use in specific contexts.

Furthermore, previous guidelines have presented a single list with everything each author must report. We believe that this is counterproductive. Eye trackers are used in many different research fields, and not all of the many aspects in Table 6 are relevant for each and every study. For example, reporting monitor properties is nonsensical for studies that do not use screens, such as a wearable eye-tracking study during locomotion. Reporting the interocular distance does not make sense for monocular eye-tracking studies, nor do firmware versions apply to analogue eye trackers. The exact reporting in each study needs to take the study's particularities into account. The purpose of a reporting guideline should be to provide authors with the information that allows them to make an *informed* selection of which specific aspects to report, how to measure them, and how to describe them.

Table 7 Reporting aspects common to all studies. We consider this a strongly recommended list of aspects to report, albeit not exhaustive

1	Details about the eye tracker (manufacturer, type, technique [video-based, EOG, dark pupil/bright pupil, etc.]). For wearable eye trackers, provide details about the scene camera (e.g. sampling frequency, resolution).
2	Sampling frequency, either of eye tracker itself, or for analogue systems the sampling frequency of any AD conversion (or some such). We recommend determining sampling frequency empirically from the recorded signals.
3	A description of the setup and geometry, including details of head stabilization if used.
4	A description of the recording environment.
5	The instructions given to participants.
6	Empirically determined data quality for the analysed eye-tracker signals. For gaze-position signals, this includes at least accuracy, precision, and data loss. For pupil-size signals, this includes at least data loss, and we recommend a measure of precision. Also report the software used to calculate these values.
7	An adequate description of the data processing and analysis steps with all relevant parameters. For example, the pupil-size estimation algorithm, velocity determination, event-detection algorithms, AOI analysis.
8	Some of the processing and analysis steps may be hidden in the software or firmware provided by the eye-tracker manufacturer. Thus, it is important to also report firmware and software versions whenever applicable.
9	Exclusion criteria, pre- and post-recording.

Recommendations for making informed choices

Based on these considerations, we have arrived at a flexible reporting structure with three parts. Firstly, Table 6 provides a list of reporting items that may be useful to report, depending on the specifics of each particular study. Secondly, we deem certain central aspects found in our review to be essential to report in any study. These are found in Table 7 and could form some kind of minimal core of future reporting guidelines. The third part of our recommendations for reporting guidelines comprises a list of prototypical situations and contexts (Tables 8, 9, 10, 11, 12, 13, 14, 15, 16, 17 and 18), that may assist readers to select the reporting items from Tables 6 for studies in specific research areas.

Table 8 Research comparing specific groups of participants

1	Can differences in e.g. gaze behaviour or pupil size be attributed to differences in eye-tracking data quality between groups? Determine and report data quality for the different groups and compare them, including data and trial loss for each group reported separately.
2	Determine whether there is a relation between data quality and the outcome measure of interest (within and between groups).
3	For pupil-size studies: consider that the baseline pupil size may differ between older and younger participants.

Table 9 Clinical studies or case studies in neuropsychology, psychiatry, rehabilitation or ophthalmology

1	For case studies, give an extensive description of participant physiology, such as palpebral fissure, eye colour, and direction of eye lashes.
2	Give a detailed description of the underlying pathology and its potential for affecting eye movement recordings. For instance, the presence of nystagmus (or fixational instability), iris transillumination defects (for example in albinism) and other iris defects such as aniridia (because of trauma, or congenital).
3	Describe the level of cooperation of patients/children with the procedure.
4	Provide any details regarding group characteristics that may impact data quality, including attentional and behavioural (compliance) issues, or that may impact interpretation (e.g. cognitive ability differences between groups).

Table 10 Eye tracking for **fixation control**, i.e. do participants look where they are instructed to look?

1	For studies using screens, report monitor properties and measurement area.
2	Report details of participant instructions.

Table 11 Pupil-size estimation

1	Illuminance.
2	Stimulus luminance and contrast.
3	Give a description of the calibration display and content (incl. luminance).
4	The baseline pupil size.
5	The angle of permitted gaze during pupil measurement, and if applicable the method to compensate for the pupil foreshortening artefact.

Table 12 AOI research on a single screen

1	An elaborate description of AOIs (size, margin, distances).
2	How are AOI measures computed from the gaze-position signal (sample-based or fixation-based)?

Table 13 Research with **more than one screen** (for instance vehicle simulators)

1	Dimensions of display(s) and their function (e.g. external front-facing world, mini-displays to simulate mirrors) as well as respective viewing distances.
2	If multiple displays are used, explain how eye-tracking data recorded across multiple displays are fused.
3	Fluctuating illuminance is important given that vehicle simulators are typically dimly lit with dynamic scenes.
4	AOI description, are they based on display segments or gaze intersections with physically separate displays.

Table 14 Wearable eye-tracking studies in unconstrained situations, e.g. in supermarkets, cars, and flight decks, or during locomotion and sports

1	The illuminance may be important. Particularly changes in illuminance throughout a measurement.
2	Are there potential vibrations that may affect data quality?
3	Automatic analysis methods for transforming gaze location from a head-centered to a world-centered description are often problematic. If used, describe how these methods were validated by the authors for their use case.
4	If manual analysis is done, provide information on e.g. the annotation protocol, instructions to the coders, and report inter-rater reliability.

Table 15 Development, evaluation or validation of **eye-tracker methodology**

1	Give an elaborate description of any relevant hardware, software, the settings, and the recording environment.
2	Give a complete description of the characteristics of participants, when used as independent variables, or the specifics of artificial eyes being used.

Table 16 **Gaze interaction** in applications and experimental studies

1	Rules governing gaze-contingent stimulus exposure e.g. "remove stimulus when gaze exceeds 250ms" or "commence trial when gaze is detected"
2	Rules for gaze button activation in gaze interaction applications.
3	Rules governing the detection of gaze within button area (AOIs) e.g. calculation of gaze position, blink removal/interpolation, and button size/margin.

Table 17 Gaze-contingent research

1	Quantify lag between eye-tracking data acquisition and display of the contingent data. Specify this with associated variability; some systems might enforce a fixed non-variable transport lag at the cost of higher lags. This can be done by estimating the delay added by each component of the gaze-contingent system, or by measuring the end-to-end system latency (Table 3).
2	Describe computing pipeline of the gaze-contingency system. Is this only graphical rendering or does it involve actuations (e.g. gaze-controlled prosthetics).
3	Provide information on the stimulus display (typically a monitor), monitor refresh rate, pixel update latency (pixel response time), monitor model, and input delay (if some input devices like mouse is used to trigger something).
4	Rules governing the detection of gaze within AOIs e.g. calculation of gaze position, blink removal/interpolation, and AOI size/margin.

Table 18 Saccade reaction time studies

1	Quantify lag between eye-tracking data acquisition and display of the stimulus. Specify this with associated variability; some systems might enforce a fixed non-variable transport lag at the cost of higher lags. This can be done by estimating the delay added by each component of the eye tracker and monitor system (Table 3).

Conclusions

What is reported in eye-tracking publications is decided on a case-by-case negotiation between authors, reviewers, and action editors of the journal/venue in question, which appears to lead to a large variation in reporting practices (Table 4).

Our review of the existing literature showed that many factors in the environment, setup, participant, eye tracker, experimental design, event detectors, and area of interest settings may impact the conclusions of any eye-tracking study. We examined a separate database on what is reported in published research on decision-making using eye trackers, which suggests that actual reporting is variable and may be in need of guidance. We examine five existing reporting guidelines for eye-tracking research and concluded that they are inconsistent, incomplete, and little used.

We have proposed a flexible, minimal reporting guideline with a core set of aspects which everyone should aim to report (Table 7), a large list of suggestions that may apply to many or some studies (Table 6), and several scenarios for specific uses of eye trackers (Tables 8-18). This information may help in making informed decisions what to report.

The reporting items that we have listed may also be used as a checklist by researchers when designing and conducting their eye-tracking experiments, and when analysing their eye-tracking data. Moreover, reviewers and journal editors may use Table 6 when assessing research during peer-review to ensure that sufficient detail is provided for replication.

Our proposal of reporting aspects may also be taken as the empirical component for a future process to develop a formalised and mandatory reporting standard (using the EQUATOR approach[1] or similar). It is possible that potential future mandatory standards would differ between clinical practice and research, or between research fields. However, we urge all such future endeavours to consider including the suggestions for reporting that we present in our empirical approach.

[1] https://www.equator-network.org/

Open Practices Statement

The reporting database has been made available at https://osf.io/ysvzk/?view_only=1be57d949dff43e99189ec6ad13f8a23.

References

Açik, A., Onat, S., Schumann, F., Einhäuser, W., & König, P. (2009). Effects of luminance contrast and its modifications on fixation behavior during free viewing of images from different categories. *Vision Research*, *49*(12), 1541–1553. https://doi.org/10.1016/j.visres.2009.03.011

Açık, A., Sarwary, A., Schultze-Kraft, R., Onat, S., & König, P. (2010). Developmental changes in natural viewing behavior: bottom-up and top-down differences between children, young adults and older adults. *Frontiers in Psychology*, *1*, 207. https://doi.org/10.3389/fpsyg.2010.00207.

Ahern, S., & Beatty, J. (1979). Pupillary responses during information processing vary with scholastic aptitude test scores. *Science*, *205*(4412), 1289–1292. https://doi.org/10.1126/science.472746

Ahlstrom, C., Nyström, M., Holmqvist, K., Fors, C., Sandberg, D., Anund, A., ..., Åkerstedt, T. (2013). Fit-for-duty test for estimation of drivers' sleepiness level: Eye movements improve the sleep/wake predictor. *Transportation Research Part C: Emerging Technologies*, *26*, 20–32. https://doi.org/10.1016/j.trc.2012.07.008

Ahlstrom, U., & Friedman-Berg, F. J. (2006). Using eye movement activity as a correlate of cognitive workload. *International Journal of Industrial Ergonomics*, *36*(7), 623–636. https://doi.org/10.1016/j.ergon.2006.04.002.

Ahmed, Z., Mardanbegi, D., & Hansen, D. W. (2016). Pupil center as a function of pupil diameter. In *Proceedings of the Symposium on Eye Tracking Research and Applications* (pp. 283–286).

Akkil, D., Isokoski, P., Kangas, J., Rantala, J., & Raisamo, R. (2014). TraQuMe: a tool for measuring the gaze tracking quality. In *Proceedings of the Symposium on Eye Tracking Research and Applications* (pp. 327–330).

Alexander, R. G., Macknik, S. L., & Martinez-Conde, S. (2018). Microsaccade characteristics in neurological and ophthalmic disease. *Frontiers in Neurology*, *9*, 144. https://doi.org/10.3389/fneur.2018.00144.

Alexander, R. G., Waite, S., Macknik, S. L., & Martinez-Conde, S. (2020). What do radiologists look for? advances and limitations of perceptual learning in radiologic search. *Journal of Vision*, *20*(10), 1–13. https://doi.org/10.1167/jov.20.10.17

Allopenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory and Language*, *38*(4), 419–439. https://doi.org/10.1006/1997.2558.

Allsop, J., & Gray, R. (2014). Flying under pressure: Effects of anxiety on attention and gaze behavior in aviation. *Journal of Applied Research in Memory and Cognition*, *3*(2), 63–71. https://doi.org/10.1016/j.jarmac.2014.04.010

Allsop, J., Gray, R., Bülthoff, H. H., & Chuang, L. (2017). Eye movement planning on single-sensor-single-indicator displays is vulnerable to user anxiety and cognitive load. *Journal of Eye Movement Research*, *10*(5), 1–15. https://doi.org/10.16910/jemr.10.5.8

Amatya, N., Gong, Q., & Knox, P. C. (2011). Differing proportions of express saccade makers in different human populations. *Experimental Brain Research*, *210*(1), 117–129. https://doi.org/10.1007/s00221-011-2609-z.

Anderson, N. C., Anderson, F., Kingstone, A., & Bischof, W. F. (2015). A comparison of scanpath comparison methods. *Behavior Research Methods*, *47*(4), 1377–1392. https://doi.org/10.3758/s13428-014-0550-3.

Anderson, N. C., Bischof, W. F., Laidlaw, K. E. W., Risko, E. F., & Kingstone, A. (2013). Recurrence quantification analysis of eye movements. *Behavior Research Methods*, *45*(3), 842–856. https://doi.org/10.3758/s13428-012-0299-5

Andersson, R., Nyström, M., & Holmqvist, K. (2010). Sampling frequency and eye-tracking measures: How speed affects durations, latencies, and more. *Journal of Eye Movement Research*, *3*(3), 1–12. https://doi.org/10.16910/jemr.3.3.6

Andersson, R., Larsson, L., Holmqvist, K., Stridh, M., & Nyström, M. (2017). One algorithm to rule them all? An evaluation and discussion of ten eye movement event-detection algorithms. *Behavior Research Methods*, *49*(2), 616–637. https://doi.org/10.3758/s13428-016-0738-9.

Antoniades, C., Ettinger, U., Gaymard, B., Gilchrist, I., Kristjánsson, A., Kennard, C., . . . , Carpenter, R. H. S. (2013). An internationally standardised antisaccade protocol. *Vision Research*, *84*, 1–5. https://doi.org/10.1016/j.visres.2013.02.007

Arba-Mosquera, S., & Aslanides, I. M. (2012). Analysis of the effects of Eye-Tracker performance on the pulse positioning errors during refractive surgery. *Journal of Optometry*, *5*(1), 31–37. https://doi.org/10.1016/j.optom.2011.11.002

Arend, L. E., & Skavenski, A. A. (1979). Free scanning of gratings produces patterned retinal exposure. *Vision Research*, *19*(12), 1413–1419. https://doi.org/10.1016/0042-6989(79)90216-5

Aserinsky, E., & Kleitman, N. (1953). Regularly occurring periods of eye motility, and concomitant phenomena, during sleep. *Science*, *118*(3062), 273–274. https://doi.org/10.1126/science.118.3062.273.

Ashby, R., Ohlendorf, A., & Schaeffel, F. (2009). The effect of ambient illuminance on the development of deprivation myopia in chicks. *Investigative Ophthalmology & Visual Science*, *50*(11), 5348–5354.

Avila, M. T., Sherr, J. D., Hong, E., Myers, C. S., & Thaker, G. K. (2003). Effects of nicotine on leading saccades during smooth pursuit eye movements in smokers and nonsmokers with schizophrenia. *Neuropsychopharmacology*, *28*(12), 2184–2191. https://doi.org/10.1038/sj.npp.1300265

Ayres, P., Lee, J. Y., Paas, F., & van Merriënboer, J. (2021). The validity of physiological measures to identify differences in intrinsic cognitive load. *Frontiers in Psychology, 12*, 702538. https://doi.org/10.3389/fpsyg.2021.702538.

Azimipour, M., Zawadzki, R. J., Gorczynska, I., Migacz, J., Werner, J. S., & Jonnal, R. S. (2018). Intraframe motion correction for raster-scanned adaptive optics images using strip-based cross-correlation lag biases. *PLOS ONE*, *13*(10), e0206052. https://doi.org/10.1371/journal.pone.0206052.

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.

Barsingerhorn, A. D., Boonstra, F. N., & Goossens, J. (2018). Development and validation of a high-speed stereoscopic eyetracker. *Behavior Research Methods*, *50*(6), 2480–2497. https://doi.org/10.3758/s13428-018-1026-7.

Bartuzel, M. M., Wróbel, K., Tamborski, S., Meina, M., Nowakowski, M., Dalasiński, K., . . . , Szkulmowski, M. (2020). High-resolution, ultrafast, wide-field retinal eye-tracking for enhanced quantification of fixational and saccadic motion. *Biomedical Optics Express*, *11*(6), 3164–3180. https://doi.org/10.1364/BOE.392849

Bast, N., Mason, L., Freitag, C. M., Smith, T., Portugal, A. M., Poustka, L., . . . , EU-AIMS LEAP Group (2021). Saccade dysmetria indicates attenuated visual exploration in autism spectrum disorder. *Journal of Child Psychology and Psychiatry*, *62*(2), 149–159. https://doi.org/10.1111/jcpp.13267

Narcizo, F. B., & Hansen, D. W. (2015). Depth compensation model for gaze estimation in sport analysis. In *Proceedings of the IEEE International Conference on Computer Vision Workshops* (pp. 71–78).

Bauer, L. O., Goldstein, R., & Stern, J. A. (1987). Effects of information-processing demands on physiological response patterns. *Human Factors*, *29*(2), 213–234. https://doi.org/10.1177/001872088702900208.

Beby, F., Dieterich, K., & Calvas, P. (2011). A [c.566-2A>G] heterozygous mutation in the *PAX6* gene causes aniridia with mild visual impairment. *Eye*, *25*(5), 657–658. https://doi.org/10.1038/eye.2010.234.

Bedggood, P., & Metha, A. (2017). De-warping of images and improved eye tracking for the scanning laser ophthalmoscope. *PLOS ONE*, *12*(4), e0174617. https://doi.org/10.1371/journal.pone.0174617.

Bengi, H., & Thomas, J. G. (1968). Three electronic methods for recording ocular tremor. *Medical and Biological Engineering*, *6*(2), 171–179. https://doi.org/10.1007/bf02474271

Benjamins, J. S., Hessels, R. S., & Hooge, I. T. C. (2018). GazeCode: open-source software for manual mapping of mobile eye-tracking data. In *Proceedings of the Symposium on Eye Tracking Research and Applications* (pp. 1–4).

Bernard, J.-B., Scherlen, A.-C., & Castet, E. (2007). Page mode reading with simulated scotomas: A modest effect of interline spacing on reading speed. *Vision Research*, *47*(28), 3447–3459. https://doi.org/10.1016/j.visres.2007.10.005

Bird, G., Press, C., & Richardson, D. C. (2011). The role of alexithymia in reduced eye-fixation in autism spectrum conditions. *Journal of Autism and Developmental Disorders*, *41*(11), 1556–1564. https://doi.org/10.1007/s10803-011-1183-3

Birmingham, E., Johnston, K. H. S., & Iarocci, G. (2017). Spontaneous gaze selection and following during naturalistic social interactions in school-aged children and adolescents with autism spectrum disorder. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, *71*(3), 243–257. https://doi.org/10.1037/cep0000131.

Bitsios, P., Prettyman, R., & Szabadi, E. (1996). Changes in autonomic function with age: A study of pupillary kinetics in healthy young and old people. *Age and Ageing*, *25*(6), 432–438. https://doi.org/10.1093/ageing/25.6.432.

Blignaut, P., & Beelders, T. (2012). The precision of eye-trackers: a case for a new measure. In Spencer, S. N. (Ed.) *Proceedings of the Symposium on Eye Tracking Research and Applications* (pp. 289–292). New York: ACM.

Blignaut, P. (2009). Fixation identification: The optimum threshold for a dispersion algorithm. *Attention, Perception, & Psychophysics*, *71*(4), 881–895. https://doi.org/10.3758/app.71.4.881

Blignaut, P. (2014). Mapping the pupil-glint vector to gaze coordinates in a simple video-based eye tracker. *Journal of Eye Movement Research*, *7*(1), 1–11. https://doi.org/10.16910/jemr.7.1.4

Blignaut, P. (2017). Development of a gaze-controlled support system for a person in an advanced stage of multiple sclerosis: a case study. *Universal Access in the Information Society*, *16*(4), 1003–1016. https://doi.org/10.1007/s10209-016-0493-9

Blignaut, P. (2018). The effect of real-time headbox adjustments on data quality. *Journal of Eye Movement Research*, *11*(1), 1–18. https://doi.org/10.16910/jemr.11.1.4

Blignaut, P., & Beelders, T. (2012). TrackStick: a data quality measuring tool for Tobii eye trackers. In *Proceedings of the Symposium on Eye Tracking Research and Applications* (pp. 293–296).

Blignaut, P., Holmqvist, K., Nyström, M., & Dewhurst, R. (2014). Improving the accuracy of video-based eye tracking in real time through post-calibration regression. In Horsley, M., Toon, N., Knight, B. A., & Reilly, R. (Eds.) *Current Trends in Eye Tracking Research* (pp. 77–100). Switzerland: Springer.

Blignaut, P., & Wium, D. (2013). The effect of mapping function on the accuracy of a video-based eye tracker. In *Proceedings of the 2013 conference on eye tracking South Africa* (pp. 39–46).

Blignaut, P., & Wium, D. (2014). Eye-tracking data quality as affected by ethnicity and experimental design. *Behavior Research Methods*, *46*(1), 67–80. https://doi.org/10.3758/s13428-013-0343-0

Blythe, H. I., Liversedge, S. P., Joseph, H. S. S. L., White, S. J., Findlay, J. M., & Rayner, K. (2006). The binocular coordination of eye movements during reading in children and adults. *Vision Research*, *46*(22), 3898–3908. https://doi.org/10.1016/j.visres.2006.06.011

Bour, L. J., Van Gisbergen, J. A. M., Bruijns, J., & Ottes, F. P. (1984). The double magnetic induction method for measuring eye movement-results in monkey and man. *IEEE Transactions on Biomedical Engineering*, *31*(5), 419–427. https://doi.org/10.1109/TBME.1984.325281.

Bowers, N. R., Boehm, A. E., & Roorda, A. (2019). The effects of fixational tremor on the retinal image. *Journal of Vision*, *19*(11), 1–16. https://doi.org/10.1167/19.11.3

Braunagel, C., Geisler, D., Stolzmann, W., Rosenstiel, W., & Kasneci, E. (2016). On the necessity of adaptive eye movement classification in conditionally automated driving scenarios. In *Proceedings of the Symposium on Eye Tracking Research and Applications* (pp. 19–26).

Brisson, J., Mainville, M., Mailloux, D., Beaulieu, C., Serres, J., & Sirois, S. (2013). Pupil diameter measurement errors as a function of gaze direction in corneal reflection eyetrackers. *Behavior Research Methods*, *45*(4), 1322–1331. https://doi.org/10.3758/s13428-013-0327-0.

Brouwer, A.-M., Hogervorst, M. A., Holewijn, M., & van Erp, J. B. F. (2014). Evidence for effects of task difficulty but not learning on neurophysiological variables associated with effort. *International Journal of Psychophysiology*, *93*(2), 242–252. https://doi.org/10.1016/j.ijpsycho.2014.05.004

Brown, M., Marmor, M., Vaegan, Zrenner, E., Brigell, M., & Bach, M. (2006). ISCEV standard for clinical electro-oculography (EOG) 2006. *Documenta Ophthalmologica*, *113*(3), 205–212. https://doi.org/10.1007/s10633-006-9030-0

Buikhuisen, W., & Jongman, R. W. (1972). Traffic perception under the influence of alcohol. *Quarterly Journal of Studies on Alcohol*, *33*(3), 800–806. https://doi.org/10.15288/qjsa.1972.33.800

Burmester, M., & Mast, M. (2010). Repeated web page visits and the scanpath theory: A recurrent pattern detection approach. *Journal of Eye Movement Research*, *3*(4), 1–20. https://doi.org/10.16910/jemr.3.4.5.

Buswell, G. T. (1935). *How people look at pictures*. University of Chicago Press.

Caldara, R., & Miellet, S. (2011). iMap: A novel method for statistical fixation mapping of eye movement data. *Behavior Research Methods*, *43*(3), 864–878. https://doi.org/10.3758/s13428-011-0092-x

Campbell, F. W., & Wurtz, R. H. (1978). Saccadic omission: Why we do not see a grey-out during a saccadic eye movement. *Vision Research*, *18*(10), 1297–1303. https://doi.org/10.1016/0042-6989(78)90219-5.

Carpenter, R. H. S. (2001). Express saccades: is bimodality a result of the order of stimulus presentation? *Vision Research*, *41*(9), 1145–1151. https://doi.org/10.1016/s0042-6989(01)00047-4

Carter, B. T., & Luke, S. G. (2020). Best practices in eye tracking research. *International Journal of Psychophysiology*, *155*, 49–62. https://doi.org/10.1016/j.ijpsycho.2020.05.010

Castner, N., Appel, T., Eder, T., Richter, J., Scheiter, K., Keutel, C., …, Kasneci, E. (2020). Pupil diameter differentiates expertise in dental radiography visual search. *PLOS ONE*, *15*(5), e0223941. https://doi.org/10.1371/journal.pone.0223941

Čegovnik, T., Stojmenova, K., Jakus, G., & Sodnik, J. (2018). An analysis of the suitability of a low-cost eye tracker for assessing the cognitive load of drivers. *Applied Ergonomics*, *68*, 1–11. https://doi.org/10.1016/j.apergo.2017.10.011

Cerrolaza, J. J., Villanueva, A., & Cabeza, R. (2012). Study of polynomial mapping functions in video-oculography eye trackers. *ACM Transactions on Computer-Human Interaction*, *19*(2), 1–25. https://doi.org/10.1145/2240156.2240158

Charness, N., Reingold, E. M., Pomplun, M., & Stampe, D. M. (2001). The perceptual aspect of skilled performance in chess: Evidence from eye movements. *Memory & Cognition*, *29*(8), 1146–1152. https://doi.org/10.3758/bf03206384

Chatelain, P., Sharma, H., Drukker, L., Papageorghiou, A. T., & Noble, J. A. (2020). Evaluation of gaze tracking calibration for longitudinal biomedical imaging studies. *IEEE Transactions on Cybernetics*, *50*(1), 153–163. https://doi.org/10.1109/tcyb.2018.2866274

Cheng, D., & Vertegaal, R. (2004). An eye for an eye: a performance evaluation comparison of the LC Technologies and Tobii eye trackers. In *Proceedings of the Symposium on Eye Tracking Research and Applications* (pp. 61–61).

Cherici, C., Kuang, X., Poletti, M., & Rucci, M. (2012). Precision of sustained fixation in trained and untrained observers. *Journal of Vision*, *12*(6), 1–16. https://doi.org/10.1167/12.6.31

Chisari, L. B., Mockevičiūtė, A., Ruitenburg, S. K., van Vemde, L., Kok, E. M., & van Gog, T. (2020). Effects of prior knowledge and joint attention on learning from eye movement modelling examples. *Journal of Computer Assisted Learning*, *36*(4), 569–579. https://doi.org/10.1111/jcal.12428

Choe, K. W., Blake, R., & Lee, S.-H. (2016). Pupil size dynamics during fixation impact the accuracy and precision of video-based gaze estimation. *Vision Research*, *118*, 48–59. https://doi.org/10.1016/j.visres.2014.12.018.

Chuang, L., Duchowski, A., Qvarfordt, P., & Weiskopf, D. (2019). Ubiquitous gaze sensing and interaction (dagstuhl seminar 18252). In *Dagstuhl Reports*, Vol. 8: Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik.

Chuk, T., Chan, A. B., & Hsiao, J. H. (2014). Understanding eye movements in face recognition using hidden Markov models. *Journal of Vision*, *14*(11), 1–14. https://doi.org/10.1167/14.11.8

Ciuffreda, K. J., & Tannen, B. (1995). Eye movement basics for the clinician. Mosby.

Clay, V., König, P., & König, S. U. (2019). Eye tracking in virtual reality. *Journal of Eye Movement Research*, *12*(1), 1–18. https://doi.org/10.16910/jemr.12.1.3

Clayden, A. C., Fisher, R. B., & Nuthmann, A. (2020). On the relative (un)importance of foveal vision during letter search in naturalistic scenes. *Vision Research*, *177*, 41–55. https://doi.org/10.1016/j.visres.2020.07.005.

Clifton, C., Staub, A., & Rayner, K. (2007). Eye movements in reading words and sentences. In Van Gompel, R. P. G., Fischer, M. H., Murray, W. S., & Hill, R. L. (Eds.) *Eye Movements: A window on mind and brain* (pp. 341–371). Oxford: Elsevier.

Cohen, A. L. (2013). Software for the automatic correction of recorded eye fixation locations in reading experiments. *Behavior Research Methods*, *45*(3), 679–683. https://doi.org/10.3758/s13428-012-0280-3.

Collewijn, H. (1998). Eye movement recording. In Carpenter, R. H. S., & Robson, J. G. (Eds.) *Vision Research: A Practical Guide to Laboratory Methods* (pp. 245–285): Oxford: Oxford University Press.

Collewijn, H., Apkarian, P., & Spekreijse, H. (1985). The oculomotor behaviour of human albinos. *Brain*, *108*(1), 1–28. https://doi.org/10.1093/brain/108.1.1

Collewijn, H., van der Mark, F., & Jansen, T. C. (1975). Precise recording of human eye movements. *Vision Research*, *15*(3), 447–450. https://doi.org/10.1016/0042-6989(75)90098-X

Collewijn, H., Van der Steen, J., Ferman, L., & Jansen, T. C. (1985). Human ocular counterroll: assessment of static and dynamic properties from electromagnetic scleral coil recordings. *Experimental Brain Research*, *59*(1), 185–196. https://doi.org/10.1007/BF00237678.

Collewijn, H. (2001). Interocular timing differences in the horizontal components of human saccades. *Vision Research*, *41*(25–26), 3413–3423. https://doi.org/10.1016/s0042-6989(01)00047-5

Collewijn, H., Erkelens, C. J., & Steinman, R. M. (1988). Binocular co-ordination of human horizontal saccadic eye movements. *Journal of Physiology*, *404*(1), 157–182.

Constantino, J. N., Kennon-McGill, S., Weichselbaum, C., Marrus, N., Haider, A., Glowinski, A. L., . . . , Jones, W. (2017). Infant viewing of social scenes is under genetic control and is atypical in autism. *Nature*, *547*(7663), 340–344. https://doi.org/10.1038/nature22999

Coors, A., Merten, N., Ward, D. D., Schmid, M., Breteler, M. M. B., & Ettinger, U. (2021). 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. https://doi.org/10.1016/j.visres.2020.10.004

Cornsweet, T. N. (1958). New technique for the measurement of small eye movements. *Journal of the Optical Society of America*, *48*, 808–811. https://doi.org/10.1364/josa.48.000808

Coutrot, A., Binetti, N., Harrison, C., Mareschal, I., & Johnston, A. (2016). Face exploration dynamics differentiate men. *Journal of Vision*, *16*(14), 1–19. https://doi.org/10.1167/16.14.16

Coutrot, A., Hsiao, J. H., & Chan, A. B. (2018). Scanpath modeling and classification with hidden Markov models. *Behavior Research Methods*, *50*(1), 362–379. https://doi.org/10.3758/s13428-017-0876-8.

Coyne, J., & Sibley, C. (2016). Investigating the use of two low cost eye tracking systems for detecting pupillary response to changes in mental workload. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, (Vol. 60, pp. 37–41). Los Angeles: SAGE Publications Sage CA.

Crane, H. D., & Steele, C. M. (1985). Generation-V dual-Purkinje-image eyetracker. *Applied Optics*, *24*(4), 527–537. https://doi.org/10.1364/AO.24.000527.

Creel, S. C. (2012). Looking forward: Comment on Morgante, Zolfaghari, and Johnson. *Infancy*, *17*(2), 141–158. https://doi.org/10.1111/j.1532-7078.2011.00106.x.

Cristino, F., Mathôt, S., Theeuwes, J., & Gilchrist, I. D. (2010). Scanmatch: A novel method for comparing fixation sequences. *Behavior Research Methods*, *42*(3), 692–700. https://doi.org/10.3758/BRM.42.3.692.

Crossland, M. D., & Rubin, G. S. (2002). The use of an infrared eyetracker to measure fixation stability. *Optometry and Vision Science*, *79*(11), 735–739. https://doi.org/10.1097/00006324-200211000-00011.

Cui, Y., & Hondzinski, J. M. (2006). Gaze tracking accuracy in humans: Two eyes are better than one. *Neuroscience Letters*, *396*(3), 257–262. https://doi.org/10.1016/j.neulet.2005.11.071

Dalmaijer, E. (2014). *Is the low-cost eyetribe eye tracker any good for research?* (Technical Report). PeerJ PrePrints.

Dalrymple, K. A., Manner, M. D., Harmelink, K. A., Teska, E. P., & Elison, J. T. (2018). An examination of recording accuracy and precision from eye tracking data from toddlerhood to adulthood. *Frontiers in Psychology*, *9*, 803. https://doi.org/10.3389/fpsyg.2018.00803.

Dalton, K. M., Nacewicz, B. M., Johnstone, T., Schaefer, H. S., Gernsbacher, M. A., Goldsmith, H. H., . . . , Davidson, R. J. (2005). Gaze fixation and the neural circuitry of face processing in autism. *Nature Neuroscience*, *8*(4), 519–526. https://doi.org/10.1038/nn1421.

Dalveren, G. G. M., & Cagiltay, N. E. (2019). Evaluation of ten open-source eye-movement classification algorithms in simulated surgical scenarios. *IEEE Access*, *7*, 161794–161804. https://doi.org/10.1109/access.2019.2951506

Dapretto, M., Davies, M. S., Pfeifer, J. H., Scott, A. A., Sigman, M., Bookheimer, S. Y., & Iacoboni, M. (2006). Understanding emotions in others: mirror neuron dysfunction in children with autism spectrum disorders. *Nature Neuroscience*, *9*(1), 28–30. https://doi.org/10.1038/nn1611

Dar, A. H., Wagner, A. S., & Hanke, M. (2021). REMoDNaV: robust eye-movement classification for dynamic stimulation. *Behavior Research Methods*, *53*(1), 399–414. https://doi.org/10.3758/s13428-020-01428-x.

De Reus, A. J. C., Zon, R., & Ouwerkerk, R. (2012). *Exploring the use of an eye tracker in a helmet mounted display* (Technical Report No. NLR-TP-2012-001).

De Visser, S. J., Van Der Post, J. P., De Waal, P. P., Cornet, F., Cohen, A. F., & Van Gerven, J. M. A. (2003). Biomarkers for the effects of benzodiazepines in healthy volunteers. *British Journal of Clinical Pharmacology*, *55*(1), 39–50. https://doi.org/10.1046/j.1365-2125.2002.t01-10-01714.x.

Delabarre, E. B. (1898). A method of recording eye movements. *American Journal of Psychology*, *9*(4), 572–574. https://doi.org/10.2307/1412191.

Dell'Osso, L. F. (2005). *Recording and calibrating the eye movements of nystagmus subjects* (Tech. Rep. No. OMLAB Report #011105). Retrieved from http://www.omlab.org/Teaching/teaching.html.

Deubel, H., & Bridgeman, B. (1995). Fourth purkinje image signals reveal eye-lens deviations and retinal image distortions during saccades. *Vision Research*, *35*(4), 529–538. https://doi.org/10.1016/0042-6989(94)00146-D.

Deubel, H., & Schneider, W. X. (1996). Saccade target selection and object recognition: Evidence for a common attentional mechanism. *Vision Research*, *36*(12), 1827–1837. https://doi.org/10.1016/0042-6989(95)00294-4.

Dewhurst, R., Nyström, M., Jarodzka, H., Foulsham, T., Johansson, R., & Holmqvist, K. (2012). It depends on how you look at it: Scanpath comparison in multiple dimensions with multimatch, a vector-based approach. *Behavior Research Methods*, *44*(4), 1079–1100. https://doi.org/10.3758/s13428-012-0212-2

Di Russo, F., Pitzalis, S., & Spinelli, D. (2003). Fixation stability and saccadic latency in élite shooters. *Vision Research*, *43*(17), 1837–1845. https://doi.org/10.1016/s0042-6989(03)00299-2

Di Stasi, L. L., Cabestrero, R., McCamy, M. B., Ríos, F., Catena, A., Quirós, P., . . . , Martinez-Conde, S. (2014). Intersaccadic drift velocity is sensitive to short-term hypobaric hypoxia. *European Journal of Neuroscience*, *39*(8), 1384–1390. https://doi.org/10.1111/ejn.12482.

Di Stasi, L. L., McCamy, M. B., Catena, A., Macknik, S. L., Cañas, J. J., & Martinez-Conde, S. (2013). Microsaccade and drift dynamics reflect mental fatigue. *European Journal of Neuroscience*, *38*(3), 2389–2398. https://doi.org/10.1111/ejn.12248

Di Stasi, L. L., McCamy, M. B., Macknik, S. L., Mankin, J. A., Hooft, N., Catena, A., & Martinez-Conde, S. (2014). Saccadic eye movement metrics reflect surgical residents' fatigue. *Annals of Surgery*, *259*(4), 824–829. https://doi.org/10.1097/sla.0000000000000260

Ditchburn, R. W. (1973). *Eye movements and visual perception*. Oxford University Press.

Donegan, M. (2012). Features of gaze control systems. In *Gaze interaction and applications of eye tracking: Advances in assistive technologies* (pp. 28–35): IGI Global.

Drewes, J., Zhu, W., Hu, Y., & Hu, X. (2014). Smaller is better: Drift in gaze measurements due to pupil dynamics. *PLOS ONE*, *9*(10), e111197. https://doi.org/10.1371/journal.pone.0111197

Drewes, J., Masson, G. S., & Montagnini, A. (2012). Shifts in reported gaze position due to changes in pupil size: Ground truth and compensation. In *Proceedings of the Symposium on Eye Tracking Research and Applications* (pp. 209–212): ACM.

Drewes, J., Montagnini, A., & Masson, G. S. (2011). Effects of pupil size on recorded gaze position: a live comparison of two eyetracking systems. *Journal of Vision*, *11*(11), 494–494.

Droege, D., & Paulus, D. (2009). Improved pupil center detection in low resolution images. In Liversedge, S. (Ed.) *Proceedings of the 15th European Conference on Eye Movements*. Southampton.

Duchowski, A. T. (2002). A breadth-first survey of eye-tracking applications. *Behavior Research Methods, Instruments, & Computers*, *34*(4), 455–470. https://doi.org/10.3758/BF03195475

Duchowski, A. T. (2007). *Eye tracking methodology: Theory and practice*. New York: Springer-Verlag.

Duchowski, A. T., Driver, J., Jolaoso, S., Tan, W., Ramey, B. N., & Robbins, A. (2010). Scanpath comparison revisited. In *Proceedings of the Symposium on Eye Tracking Research and Applications* (pp. 219–226).

Dunn, M. J., Harris, C. M., Ennis, F. A., Margrain, T. H., Woodhouse, J. M., McIlreavy, L., & Erichsen, J. T. (2019). An automated segmentation approach to calibrating infantile nystagmus waveforms. *Behavior Research Methods*, *51*(5), 2074–2084. https://doi.org/10.3758/s13428-018-1178-5

Duwaer, A. L., Van Den Brink, G., Van Antwerpen, G., & Keemink, C. J. (1982). Comparison of subjective and objective measurements of ocular alignment in the vertical direction. *Vision Research*, *22*(8), 983–989. https://doi.org/10.1016/0042-6989(82)90034-7.

Ehinger, B. V., Groß, K., Ibs, I., & König, P. (2019). A new comprehensive eye-tracking test battery concurrently evaluating the Pupil Labs glasses and the EyeLink 1000. *PeerJ*, *7*, e7086. https://doi.org/10.7717/peerj.7086

Einhäuser, W. (2017). The pupil as marker of cognitive processes. In Zhao, Q. (Ed.) *Computational and cognitive neuroscience of vision* (pp. 141–169): Springer.

Ellermeier, W., & Westphal, W. (1995). Gender differences in pain ratings and pupil reactions to painful pressure stimuli. *Pain*, *61*(3), 435–439. https://doi.org/10.1016/0304-3959(94)00203-q

Ellis, S. R., & Stark, L. (1986). Statistical dependency in visual scanning. *Human Factors*, *28*(4), 421–438. https://doi.org/10.1177/001872088602800405.

Emhardt, S. N., Kok, E. M., Jarodzka, H., Brand-Gruwel, S., Drumm, C., & van Gog, T. (2020). How experts adapt their gaze behavior when modeling a task to novices. *Cognitive Science*, *44*(9), e12893. https://doi.org/10.1111/cogs.12893

Engbert, R., & Kliegl, R. (2003). Binocular coordination in microsaccades. In Hyönä, J., Radach, R., & Deubel, H. (Eds.) *The mind's eye: Cognitive and applied aspects of oculomotor research*. https://doi.org/10.1016/B978-044451020-4/50007-4, (pp. 103–117): Elsevier.

Engbert, R., Rothkegel, L., Backhaus, D., & Trukenbrod, H. A. (2016). *Evaluation of velocity-based saccade detection in the SMI-ETG 2W system* (Tech. Rep.). Retrieved from http://read.psych.uni-potsdam.de.

Engel, F. L. (1971). Visual conspicuity, directed attention and retinal locus. *Vision Research*, *11*(6), 563–575. https://doi.org/10.1016/0042-6989(71)90077-0.

Enright, J., & Hendriks, A. (1994). To stare or to scrutinize:: "grasping"? the eye for better vision. *Vision Research*, *34*(15), 2039–2042. https://doi.org/10.1016/0042-6989(94)90031-0

Erdmann, B., & Dodge, R. (1898). *Psychologische Untersuchungen über das Lesen auf experimenteller Grundlage*. Niemeyer.

Erens, R. G. F., Kappers, A. M. L., & Koenderink, J. J. (1993). Perception of local shape from shading. *Perception & Psychophysics*, *54*(2), 145–156. https://doi.org/10.3758/bf03211750

Ettinger, U., & Kumari, V. (2019). Effects of nicotine on inhibitory control in humans. In Preedy, V. R. (Ed.) *Neuroscience of Nicotine* (pp. 151–158): Elsevier.

Ettinger, U., Meyhöfer, I., Mehta, M. A., Kumari, V., Corr, P. J., Williams, S. C. R., & Perkins, A. M. (2018). Effects of lorazepam on saccadic eye movements: the role of sex, task characteristics and baseline traits. *Journal of Psychopharmacology*, *32*(6), 678–690. https://doi.org/10.1177/0269881118772450

Fick, A. (1854). Die Bewegungen des menschlichen Augapfels. *Zeitschrift für rationelle Medicin*, *4*, 101–128.

Fiedler, S., Schulte-Mecklenbeck, M., Renkewitz, F., & Orquin, J. L. (2019). Increasing reproducibility of eye-tracking studies. In Schulte-Mecklenbeck, M., Kühlberger, A., & Johnson, J. G. (Eds.) *A handbook of process tracing methods* (pp. 65–75): Routledge.

Fischler, M. A., & Bolles, R. C. (1981). Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, *24*(6), 381–395. https://doi.org/10.1145/358669.358692

Flom, M. C., Brown, B., Adams, A. J., & Jones, R. T. (1976). Alcohol and marijuana effects on ocular tracking. *American Journal of Optometry and Physiological Optics*, *53*(12), 764–773.

Fogarty, C., & Stern, J. A. (1989). Eye movements and blinks: their relationship to higher cognitive processes. *International Journal of Psychophysiology*, *8*(1), 35–42. https://doi.org/10.1016/0167-8760(89)90017-2.

Frank, M. C., Vul, E., & Saxe, R. (2012). Measuring the development of social attention using free-viewing. *Infancy*, *17*(4), 355–375. https://doi.org/10.1111/j.1532-7078.2011.00086.x

Fransson, P.-A., Patel, M., Magnusson, M., Berg, S., Almbladh, P., & Gomez, S. (2008). Effects of 24-hour and 36-hour sleep deprivation on smooth pursuit and saccadic eye movements. *Journal of Vestibular Research*, *18*(4), 209–222.

Frazier, L., & Rayner, K. (1982). Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences. *Cognitive Psychology*, *14*(2), 178–210. https://doi.org/10.1016/0010-0285(82)90008-1

Frazier, T. W., Strauss, M., Klingemier, E. W., Zetzer, E. E., Hardan, A. Y., Eng, C., & Youngstrom, E. A. (2017). A meta-analysis of gaze differences to social and nonsocial information between individuals with and without autism. *Journal of the American Academy of Child & Adolescent Psychiatry*, *56*(7), 546–555. https://doi.org/10.1016/j.jaac.2017.05.005

Frens, M. A., & van der Geest, J. N. (2002). Scleral search coils influence saccade dynamics. *Journal of Neurophysiology*, *88*(2), 692–698. https://doi.org/10.1152/jn.00457.2001

Friedman, L. (2020). Brief communication: Three errors and two problems in a recent paper: gazeNet: End-to-end eye-movement event detection with deep neural networks (Zemblys, Niehorster, and Holmqvist, 2019). *Behavior Research Methods*, *52*(4), 1671–1680. https://doi.org/10.3758/s13428-019-01342-x

Fuhl, W., Tonsen, M., Bulling, A., & Kasneci, E. (2016). Pupil detection for head-mounted eye tracking in the wild: an evaluation of the state of the art. *Machine Vision and Applications*, *27*(8), 1275–1288. https://doi.org/10.1007/s00138-016-0776-4

Funke, G., Greenlee, E., Carter, M., Dukes, A., Brown, R., & Menke, L. (2016). Which eye tracker is right for your research? performance evaluation of several cost variant eye trackers. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, (Vol. 60, pp. 1240–1244). Sage: SAGE Publications.

Gagl, B., Hawelka, S., & Hutzler, F. (2011). Systematic influence of gaze position on pupil size measurement: analysis and correction. *Behavior Research Methods*, *43*(4), 1171–1181. https://doi.org/10.3758/s13428-011-0109-5

Gegenfurtner, A., Lehtinen, E., & Säljö, R. (2011). Expertise differences in the comprehension of visualizations: A meta-analysis of eye-tracking research in professional domains. *Educational Psychology Review*, *23*(4), 523–552. https://doi.org/10.1007/s10648-011-9174-7

Ghasia, F. F., & Shaikh, A. G. (2015). Uncorrected myopic refractive error increases microsaccade amplitude. *Investigative Ophthalmology & Visual Science*, *56*(4), 2531–2535. https://doi.org/10.1167/iovs.14-15882

Gluckman, M., & Johnson, S. P. (2013). Attentional capture by social stimuli in young infants. *Frontiers in Psychology*, *4*, 527. https://doi.org/10.3389/fpsyg.2013.00527

Godara, P., Dubis, A. M., Roorda, A., Duncan, J. L., & Carroll, J. (2010). Adaptive optics retinal imaging: Emerging clinical applications. *Optometry and Vision Science*, *87*(12), 930–941. https://doi.org/10.1097/OPX.0b013e3181ff9a8b

Goldberg, J. H., Stimson, M. J., Lewenstein, M., Scott, N., & Wichansky, A. M. (2002). Eye tracking in web search tasks: design implications. In *Proceedings of the Symposium on Eye Tracking Research and Applications* (pp. 51–58).

Graupner, S.-T., & Pannasch, S. (2014). Continuous gaze cursor feedback in various tasks: Influence on eye movement behavior, task performance and subjective distraction. In *International Conference on Human-Computer Interaction* (pp. 323–329): Springer.

Greene, H. H., & Rayner, K. (2001). Eye movements and familiarity effects in visual search. *Vision Research*, *41*(27), 3763–3773. https://doi.org/10.1016/S0042-6989(01)00154-7

Guestrin, E. D., & Eizenman, M. (2006). General theory of remote gaze estimation using the pupil center and corneal reflections. *IEEE Transactions on Biomedical Engineering*, *53*(6), 1124–1133. https://doi.org/10.1109/tbme.2005.863952

Guillon, Q., Hadjikhani, N., Baduel, S., & Rogé, B. (2014). Visual social attention in autism spectrum disorder: Insights from eye tracking studies. *Neuroscience & Biobehavioral Reviews*, *42*, 279–297. https://doi.org/10.1016/j.neubiorev.2014.03.013

Gunzelmann, G., Gluck, K. A., Price, S., Van Dongen, H. P. A., & Dinges, D. F. (2007). Decreased arousal as a result of sleep deprivation: The unraveling of cognitive control. In Gray, W. D. (Ed.) *Integrated models of cognitive systems* (pp. 243–253): Oxford University Press.

Hansen, D. W., & Ji, Q. (2010). In the eye of the beholder: A survey of models for eyes and gaze. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *32*(3), 478–500. https://doi.org/10.1109/TPAMI.2009.30

Hansen, D. W., & Pece, A. E. C. (2005). Eye tracking in the wild. *Computer Vision and Image Understanding*, *98*(1), 155–181. https://doi.org/10.1016/j.cviu.2004.07.013

Hansen, J. P., Agustin, J. S., & Skovsgaard, H. (2011). Gaze interaction from bed. In *Proceedings of the 1st Conference on Novel Gaze-Controlled Applications* (pp. 1–4).

Harrar, V., Le Trung, W., Malienko, A., & Khan, A. Z. (2018). A nonvisual eye tracker calibration method for video-based tracking. *Journal of Vision*, *18*(9), 1–11. https://doi.org/10.1167/18.9.13

Hartridge, H., & Thomson, L. C. (1948). Methods of investigating eye movements. *The British Journal of Ophthalmology*, *32*(9), 581–591. https://doi.org/10.1136/bjo.32.9.581

Haslwanter, T. (1995). Mathematics of three-dimensional eye rotations. *Vision Research*, *35*(12), 1727–1739. https://doi.org/10.1016/0042-6989(94)00257-m

Hayes, T. R., & Petrov, A. A. (2016). Mapping and correcting the influence of gaze position on pupil size measurements. *Behavior Research Methods*, *48*(2), 510–527. https://doi.org/10.3758/s13428-015-0588-x

Helo, A., Pannasch, S., Sirri, L., & Rämä, P. (2014). The maturation of eye movement behavior: Scene viewing characteristics in children and adults. *Vision Research*, *103*, 83–91. https://doi.org/10.1016/j.visres.2014.08.006

Henderson, J. M., Nuthmann, A., & Luke, S. G. (2013). Eye movement control during scene viewing: Immediate effects of scene luminance on fixation durations. *Journal of Experimental Psychology: Human Perception and Performance*, *39*(2), 318–322. https://doi.org/10.1037/a0031224

Hermens, F. (2015). Dummy eye measurements of microsaccades: Testing the influence of system noise and head movements on microsaccade detection in a popular video-based eye tracker. *Journal of Eye Movement Research*, *8*(1), 1–17. https://doi.org/10.16910/jemr.8.1.1

Hess, E. H., & Polt, J. M. (1964). Pupil size in relation to mental activity during simple problem-solving. *Science*, *143*(3611), 1190–1192. https://doi.org/10.1126/science.143.3611.1190

Hessels, R. S., Andersson, R., Hooge, I. T. C., Nyström, M., & Kemner, C. (2015). Consequences of eye color, positioning, and head movement for eye-tracking data quality in infant research. *Infancy*, *20*(6), 601–633. https://doi.org/10.1111/infa.12093

Hessels, R. S., Benjamins, J. S., van Doorn, A. J., Koenderink, J. J., Holleman, G. A., & Hooge, I. T. C. (2020). Looking behavior and potential human interactions during locomotion. *Journal of Vision*, *20*(10), 1–25. https://doi.org/10.1167/jov.20.10.5

Hessels, R. S., Cornelissen, T. H. W., Kemner, C., & Hooge, I. T. C. (2015). Qualitative tests of remote eyetracker recovery and performance during head rotation. *Behavior Research Methods*, *47*(3), 848–859. https://doi.org/10.3758/s13428-014-0507-6

Hessels, R. S., Holleman, G. A., Kingstone, A., Hooge, I. T. C., & Kemner, C. (2019). Gaze allocation in face-to-face communication is affected primarily by task structure and social context, not stimulus-driven factors. *Cognition*, *184*, 28–43. https://doi.org/10.1016/j.cognition.2018.12.005

Hessels, R. S., & Hooge, I. T. C. (2019). Eye tracking in developmental cognitive neuroscience – The good, the bad and the ugly. *Developmental Cognitive Neuroscience*, *40*, 100710. https://doi.org/10.1016/j.dcn.2019.100710

Hessels, R. S., Hooge, I. T. C., & Kemner, C. (2016). An in-depth look at saccadic search in infancy. *Journal of Vision*, *16*(8), 1–14. https://doi.org/10.1167/16.8.10

Hessels, R. S., Kemner, C., van den Boomen, C., & Hooge, I. T. C. (2016). The area-of-interest problem in eyetracking research: A noise-robust solution for face and sparse stimuli. *Behavior Research Methods*, *48*(4), 1694–1712. https://doi.org/10.3758/s13428-015-0676-y

Hessels, R. S., Niehorster, D. C., Kemner, C., & Hooge, I. T. C. (2017). Noise-robust fixation detection in eye movement data: Identification by two-means clustering (I2MC). *Behavior Research Methods*, *49*(5), 1802–1823. https://doi.org/10.3758/s13428-016-0822-1

Hessels, R. S., Niehorster, D. C., Nyström, M., Andersson, R., & Hooge, I. T. C. (2018). Is the eye-movement field confused about fixations and saccades? A survey among 124 researchers. *Royal Society Open Science*, *5*(8), 180502. https://doi.org/10.1098/rsos.180502

Hessels, R. S., van Doorn, A. J., Benjamins, J. S., Holleman, G. A., & Hooge, I. T. C. (2020). Task-related gaze control in human crowd navigation. *Attention, Perception, & Psychophysics*, *82*, 2482–2501.

Hohenstein, S., & Kliegl, R. (2014). Semantic preview benefit during reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*(1), 166–190. https://doi.org/10.1037/a0033670

Holleman, G. A., Hessels, R. S., Kemner, C., & Hooge, I. T. C. (2020). Implying social interaction and its influence on gaze behavior to the eyes. *PLOS ONE*, *15*(2), e0229203. https://doi.org/10.1371/journal.pone.0229203

Holmqvist, K. (2015). Common predictors of accuracy, precision and data loss in 12 eye-trackers. (Accepted for publication in *Behavior Research Methods*, 1–25; not published for legal reasons) https://doi.org/10.13140/RG.2.2.16805.22246.

Holmqvist, K. (2016). Optimal settings for commercial event detection algorithms based on the level of noise. (Accepted for publication in *Behavior Research Methods*, 1–21; not published for legal reasons) https://doi.org/10.13140/RG.2.2.26871.55200.

Holmqvist, K., & Andersson, R. (2017). Eye tracking: A comprehensive guide to methods, paradigms and measures. Lund: Lund Eye-Tracking Research Institute.

Holmqvist, K., & Blignaut, P. (2020). Small eye movements cannot be reliably measured by video-based P–CR eye-trackers. *Behavior Research Methods*, *52*, 2098–2121. https://doi.org/10.3758/s13428-020-01363-x

Holmqvist, K., Örbom, S. L., & Zemblys, R. (2021). Small head movements increase and colour noise in data from five video-based P–CR eye trackers. Behavior Research Methods, 1–16. https://doi.org/10.3758/s13428-021-01648-9.

Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., & Van de Weijer, J. (2011). Eye tracking: A comprehensive guide to methods and measures. OUP Oxford.

Holmqvist, K., Nyström, M., & Mulvey, F. (2012). Eye tracker data quality: What it is and how to measure it. In *Proceedings of the Symposium on Eye Tracking Research and Applications* (pp. 45–52): ACM.

Holmqvist, K., Örbom, S. L., Miller, M., Kashchenevsky, A., Shovman, M., & Greenlee, M. W. (2020). Validation of a prototype hybrid eye-tracker against the DPI and the Tobii Spectrum. In *Symposium on Eye Tracking Research and Applications* (pp. 1–9).

Holt, E. B. (1903). Eye movement and central anaesthesia. *The Psychological Review: Monograph Supplements*.

Hooge, I. T. C., & Camps, G. (2013). Scan path entropy and arrow plots: Capturing scanning behavior of multiple observers. *Frontiers in Psychology*, *4*, 996. https://doi.org/10.3389/fpsyg.2013.00996

Hooge, I. T. C., & Erkelens, C. (1996). Control of fixation duration in a simple search task. *Perception & Psychophysics*, *58*(7), 969–976. https://doi.org/10.3758/bf03206825

Hooge, I. T. C., Hessels, R. S., & Nyström, M. (2019). Do pupil-based binocular video eye trackers reliably measure vergence?. *Vision Research*, *156*, 1–9. https://doi.org/10.1016/j.visres.2019.01.004

Hooge, I. T. C., Holleman, G. A., Haukes, N. C., & Hessels, R. S. (2019). Gaze tracking accuracy in humans: One eye is sometimes better than two. *Behavior Research Methods*, *51*(6), 2712–2721. https://doi.org/10.3758/s13428-018-1135-3

Hooge, I. T. C., Holmqvist, K., & Nyström, M. (2016). The pupil is faster than the corneal reflection (CR): Are video based pupil-CR eye trackers suitable for studying detailed dynamics of eye movements?. *Vision Research*, *128*, 6–18. https://doi.org/10.1016/j.visres.2016.09.002

Hooge, I. T. C., Niehorster, D. C., Hessels, R. S., Cleveland, D., & Nyström, M. (2021). The pupil-size artefact (PSA) across time, viewing direction, and different eye trackers. *Behavior Research Methods*. https://doi.org/10.3758/s13428-020-01512-2.

Hooge, I. T. C., Niehorster, D. C., Nyström, M., Andersson, R., & Hessels, R. S. (2018). Is human classification by experienced untrained observers a gold standard in fixation detection? *Behavior Research Methods*, *50*(5), 1864–1881. https://doi.org/10.3758/s13428-017-0955-x

Hooge, I. T. C., Nyström, M., Cornelissen, T., & Holmqvist, K. (2015). The art of braking: Post saccadic oscillations in the eye tracker signal decrease with increasing saccade size. *Vision Research*, *112*, 55–67. https://doi.org/10.1016/j.visres.2015.03.015

Hooge, I. T. C., & Van den Berg, A. V. (2000). Visually evoked cyclovergence and extended listing's law. *Journal of Neurophysiology*, *83*(5), 2757–2775. https://doi.org/10.1152/jn.2000.83.5.2757

Horrey, W. J., & Wickens, C. D. (2007). In-vehicle glance duration: distributions, tails, and model of crash risk. *Transportation Research Record*, *2018*(1), 22–28. https://doi.org/10.3141/2018-04

Houben, M. M. J., Goumans, J., & van der Steen, J. (2006). Recording three-dimensional eye movements: Scleral search coils versus video oculography. *Investigative Ophthalmology & Visual Science*, *47*(1), 179–187. https://doi.org/10.1167/iovs.05-0234

Huestegge, L., Radach, R., & Kunert, H.-J. (2009). Long-term effects of cannabis on oculomotor function in humans. *Journal of Psychopharmacology*, *23*(6), 714–722. https://doi.org/10.1177/0269881108091601

Hyrskykari, A. (2006). Utilizing eye movements: Overcoming inaccuracy while tracking the focus of attention during reading. *Computers in Human Behavior*, *22*(4), 657–671. https://doi.org/10.1016/j.chb.2005.12.013

Inchingolo, P., & Spanio, M. (1985). On the identification and analysis of saccadic eye movements–A quantitative study of the processing procedures. *IEEE Transactions on Biomedical Engineering*, *32*(9), 683–695. https://doi.org/10.1109/tbme.1985.325586

Irving, E. L., Zacher, J. E., Allison, R. S., & Callender, M. G. (2003). Effects of scleral search coil wear on visual function. *Investigative Ophthalmology & Visual Science*, *44*(5), 1933–1938. https://doi.org/10.1167/iovs.01-0926

Jarodzka, H., Holmqvist, K., & Nyström, M. (2010). A vector-based, multidimensional scanpath similarity measure. In *Proceedings of the Symposium on Eye Tracking Research and Applications*, (pp. 211–218).

Jaschinski, W. (2016). Pupil size affects measures of eye position in video eye tracking: implications for recording vergence accuracy. *Journal of Eye Movement Research*, *9*(4), 1–14. https://doi.org/10.16910/jemr.9.4.2.

Jaschinski, W., Jainta, S., & Kloke, W. B. (2010). Methods for objective measurement of fixation disparity; appendix (online only). *Ophthalmic and Physiological Optics*, *30*(4), 1–8.

Jones, P. R., Kalwarowsky, S., Atkinson, J., Braddick, O. J., & Nardini, M. (2014). Automated measurement of resolution acuity in infants using remote eye-tracking. *Investigative Ophthalmology & Visual Science*, *55*(12), 8102–8110. https://doi.org/10.1167/iovs.14-15108

Jones, W., Carr, K., & Klin, A. (2008). Absence of preferential looking to the eyes of approaching adults predicts level of social disability in 2-year-old toddlers with autism spectrum disorder. *Archives of General Psychiatry*, *65*(8), 946–954. https://doi.org/10.1001/archpsyc.65.8.946

Jones, W., & Klin, A. (2013). Attention to eyes is present but in decline in 2–6-month-old infants later diagnosed with autism. *Nature*, *504*(7480), 427–431. https://doi.org/10.1038/nature12715

Juhola, M., Jäntti, V., & Pyykkö, I. (1985). Effect of sampling frequencies on computation of the maximum velocity of saccadic eye movements. *Biological Cybernetics*, *53*(2), 67–72. https://doi.org/10.1007/bf00337023

Kahneman, D., & Beatty, J. (1966). Pupil diameter and load on memory. *Science*, *154*(3756), 1583–1585. https://doi.org/10.1126/science.154.3756.1583

Kaneko, D., Hogervorst, M., Toet, A., van Erp, J. B. F., Kallen, V., & Brouwer, A.-M. (2019). Explicit and implicit responses to tasting drinks associated with different tasting experiences. *Sensors*, *19*(20), 4397. https://doi.org/10.3390/s19204397

Kangas, J., Koskinen, O., & Raisamo, R. (2020). Gaze tracker accuracy and precision measurements in virtual reality headsets. In *Proceedings of the 2020 International Conference on Multimodal Interaction* (pp. 640–644).

Kapoula, Z., Yang, Q., Otero-Millan, J., Xiao, S., Macknik, S. L., Lang, A., ..., Martinez-Conde, S. (2014). Distinctive features of microsaccades in Alzheimers disease and in mild cognitive impairment. *Age*, *36*(2), 535–543. https://doi.org/10.1007/s11357-013-9582-3

Karolus, J., Wozniak, P. W., Chuang, L. L., & Schmidt, A. (2017). Robust gaze features for enabling language proficiency awareness. In *Proceedings of the 2017 CHI conference on human factors in computing systems* (pp. 2998–3010).

Karpouzian, T., Petrovsky, N., Ettinger, U., & Reilly, J. (2019). Eye movements as biomarkers to evaluate pharmacological effects on brain systems. In C. Klein, & U. Ettinger (Eds.) *Eye Movement Research* (pp. 775–816). Cham: Springer.

Kasneci, E., Sippel, K., Aehling, K., Heister, M., Rosenstiel, W., Schiefer, U., & Papageorgiou, E. (2014). Driving with binocular visual field loss? A study on a supervised on-road parcours with simultaneous eye and head tracking. *PLOS ONE*, *9*(2), e87470. https://doi.org/10.1371/journal.pone.0087470

Katsanis, J., Kortenkamp, S., Iacono, W. G., & Grove, W. M. (1997). Antisaccade performance in patients with schizophrenia and affective disorder. *Journal of Abnormal Psychology*, *106*(3), 468–472. https://doi.org/10.1037/0021-843X.106.3.468

Keehn, B., & Joseph, R. M. (2016). Exploring what's missing: What do target absent trials reveal about autism search superiority?. *Journal of Autism and Developmental Disorders*, *46*(5), 1686–1698. https://doi.org/10.1007/s10803-016-2700-1

Keehn, B., Müller, R.-A., & Townsend, J. (2013). Atypical attentional networks and the emergence of autism. *Neuroscience & Biobehavioral Reviews*, *37*(2), 164–183. https://doi.org/10.1016/j.neubiorev.2012.11.014

Kemner, C., Van Ewijk, L., Van Engeland, H., & Hooge, I. T. C. (2008). Eye movements during visual search tasks indicate enhanced stimulus discriminability in subjects with PDD. *Journal of Autism and Developmental Disorders*, *38*(3), 553–557. https://doi.org/10.1007/s10803-007-0406-0

Kenward, B., Koch, F.-S., Forssman, L., Brehm, J., Tidemann, I., Sundqvist, A., ..., Gredebäck, G. (2017). Saccadic reaction times in infants and adults: Spatiotemporal factors, gender, and interlaboratory variation. *Developmental Psychology*, *53*(9), 1750. https://doi.org/10.1037/dev0000338

Kevic, K., Walters, B. M., Shaffer, T. R., Sharif, B., Shepherd, D. C., & Fritz, T. (2015). Tracing software developers' eyes and interactions for change tasks. In *Proceedings of the 2015 10th Joint Meeting on Foundations of Software Engineering* (pp. 202–213).

Kimmel, D. L., Mammo, D., & Newsome, W. T. (2012). Tracking the eye non-invasively: simultaneous comparison of the scleral search coil and optical tracking techniques in the macaque monkey. *Frontiers in Behavioral Neuroscience*, *6*, 49. https://doi.org/10.3389/fnbeh.2012.00049

Kirkby, J. A., Blythe, H. I., Drieghe, D., Benson, V., & Liversedge, S. P. (2013). Investigating eye movement acquisition and analysis technologies as a causal factor in differential prevalence of crossed and uncrossed fixation disparity during reading and dot scanning. *Behavior Research Methods*, *45*(3), 664–678. https://doi.org/10.3758/s13428-012-0301-2

Klin, A., Jones, W., Schultz, R., Volkmar, F., & Cohen, D. (2002). Visual fixation patterns during viewing of naturalistic social situations as predictors of social competence in individuals with autism. *Archives of General Psychiatry*, *59*(9), 809–816. https://doi.org/10.1001/archpsyc.59.9.809

Ko, H.-k., Snodderly, D. M., & Poletti, M. (2016). Eye movements between saccades: Measuring ocular drift and tremor. *Vision Research*, *122*, 93–104. https://doi.org/10.1016/j.visres.2016.03.006

Kok, E. M., Aizenman, A. M., Võ, M. L.-H., & Wolfe, J. M. (2017). Even if I showed you where you looked, remembering where you just looked is hard. *Journal of Vision*, *17*(12), 1–11. https://doi.org/10.1167/17.12.2

Kok, E. M., Jarodzka, H., de Bruin, A. B. H., BinAmir, H. A. N., Robben, S. G. F., & van Merriënboer, J. J. G. (2016). Systematic viewing in radiology: seeing more, missing less?. *Advances in Health Sciences Education*, *21*(1), 189–205. https://doi.org/10.1007/s10459-015-9624-y

Kolakowski, S. M., & Pelz, J. B. (2006). Compensating for eye tracker camera movement. In *Proceedings of the Symposium on Eye Tracking Research and Applications* (pp. 79–85). New York: ACM.

Komínková, B., Pedersen, M., Hardeberg, J. Y., & Kaplanová, M. (2008). Comparison of eye tracking devices used on printed images. In Rogowitz, B. E., & Pappas, T. N. (Eds.) *Human Vision and Electronic Imaging XIII*, (Vol. 6806, pp. 545–556): SPIE.

Komogortsev, O. V., Jayarathna, S., Koh, D. H., & Gowda, S. M. (2010). Qualitative and quantitative scoring and evaluation of the eye movement classification algorithms. In *Proceedings of the Symposium on Eye Tracking Research and Applications* (pp. 65–68).

Kooiker, M. J. G., Pel, J. J. M., van der Steen-Kant, S. P., & van der Steen, J. (2016). A method to quantify visual information processing in children using eye tracking. *Journal of Visualized Experiments*, *113*, e54031. https://doi.org/10.3791/54031

Kothari, R., Yang, Z., Kanan, C., Bailey, R., Pelz, J. B., & Diaz, G. J. (2020). Gaze-in-wild: A dataset for studying eye and head coordination in everyday activities. *Scientific Reports*, *10*(1), 1–18. https://doi.org/10.1038/s41598-020-59251-5

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

Krejtz, K., Duchowski, A., Szmidt, T., Krejtz, I., González Perilli, F., Pires, A., ..., Villalobos, N. (2015). Gaze transition entropy. *ACM Transactions on Applied Perception*, *13*(1), 1–20. https://doi.org/10.1145/2834121

Krejtz, K., Szmidt, T., Duchowski, A. T., & Krejtz, I. (2014). Entropy-based statistical analysis of eye movement transitions. In *Proceedings of the Symposium on Eye Tracking Research and Applications* (pp. 159–166).

Kübler, T. C., Kasneci, E., & Rosenstiel, W. (2014). Subsmatch: Scanpath similarity in dynamic scenes based on subsequence frequencies. In *Proceedings of the Symposium on Eye Tracking Research and Applications* (pp. 319–322).

Kulke, L. V. (2015). *Cortical mechanisms of visual attention in typically developing infants and adults*. Ph.D. Thesis, UCL (University College London). https://discovery.ucl.ac.uk/id/eprint/1471117.

Lamare, M. (1892). Des mouvements des yeux dans la lecture. *Bulletins et Mémoires de la Société Française d'Ophthalmologie*, *10*, 354–364.

Lappe-Osthege, M., Talamo, S., Helmchen, C., & Sprenger, A. (2010). Overestimation of saccadic peak velocity recorded by electro-oculography compared to video-oculography and scleral search coil. *Clinical Neurophysiology*, *121*(10), 1786–1787. https://doi.org/10.1016/j.clinph.2010.03.051

Lappi, O. (2016). Eye movements in the wild: Oculomotor control, gaze behavior & frames of reference. *Neuroscience & Biobehavioral Reviews*, *69*, 49–68. https://doi.org/10.1016/j.neubiorev.2016.06.006

Larsson, L. (2010). *Event detection in eye-tracking data*. Master's Thesis, Lund University, Lund, Sweden.

Larsson, L., Nyström, M., & Stridh, M. (2013). Detection of saccades and postsaccadic oscillations in the presence of smooth pursuit. *IEEE Transactions on Biomedical Engineering*, *60*(9), 2484–2493. https://doi.org/10.1109/tbme.2013.2258918

Larsson, L., Schwaller, A., Holmqvist, K., Nyström, M., & Stridh, M. (2014). Compensation of head movements in mobile eye-tracking data using an inertial measurement unit. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication* (pp. 1161–1167).

Lee, J. Y., Donkers, J., Jarodzka, H., Sellenraad, G., & van Merriënboer, J. J. G. (2020). Different effects of pausing on cognitive load in a medical simulation game. *Computers in Human Behavior*, *110*, 106385. https://doi.org/10.1016/j.chb.2020.106385

Lee, J. Y., Donkers, J., Jarodzka, H., & Van Merriënboer, J. J. G. (2019). How prior knowledge affects problem-solving performance in a medical simulation game: Using game-logs and eye-tracking. *Computers in Human Behavior*, *99*, 268–277. https://doi.org/10.1016/j.chb.2019.05.035

Legge, G. E., Ahn, S. J., Klitz, T. S., & Luebker, A. (1997). Psychophysics of reading—XVI. The visual span in normal and low vision. *Vision Research*, *37*(14), 1999–2010. https://doi.org/10.1016/s0042-6989(97)00017-5

Lehtinen, I., Lang, A. H., Jäntti, V., & Keskinen, E. (1979). Acute effects of alcohol on saccadic eye movements. *Psychopharmacology*, *63*(1), 17–23. https://doi.org/10.1007/BF00426915

Leigh, R. J., & Zee, D. S. (2006). *The neurology of eye movements*. Oxford: Oxford University Press.

Leppänen, J. M., Forssman, L., Kaatiala, J., Yrttiaho, S., & Wass, S. V. (2015). Widely applicable MATLAB routines for automated analysis of saccadic reaction times. *Behavior Research Methods*, *47*(2), 538–548. https://doi.org/10.3758/s13428-014-0473-z

Liu, A., Xia, L., Duchowski, A., Bailey, R., Holmqvist, K., & Jain, E. (2019). Differential privacy for eye-tracking data. In *Proceedings of the Symposium on Eye Tracking Research and Applications* (pp. 1–10).

Liversedge, S., White, S. J., Findlay, J. M., & Rayner, K. (2006). Binocular coordination of eye movements during reading. *Vision Research*, *46*(15), 2363–2374. https://doi.org/10.1016/j.visres.2006.01.013

Liversedge, S., Gilchrist, I., & Everling, S. (2011). *The Oxford handbook of eye movements*. Oxford University Press.

Loftus, G. R., & Mackworth, N. H. (1978). Cognitive determinants of fixation location during picture viewing. *Journal of Experimental Psychology: Human Perception and Performance*, *4*(4), 565–572. https://doi.org/10.1037//0096-1523.4.4.565

Loschky, L. C., & Wolverton, G. S. (2007). How late can you update gaze-contingent multiresolutional displays without detection?. *ACM Transactions on Multimedia Computing, Communications, and Applications*, *3*(4), 1–10. https://doi.org/10.1145/1314303.1314310

Luna, B., & Velanova, K. (2011). Development from reflexive to controlled eye movements. In Liversedge, S., Gilchrist, I. D., & Everling, S. (Eds.) *The Oxford handbook of eye movements*, (pp. 621–642): Oxford University Press.

Mack, D. J., Belfanti, S., & Schwarz, U. (2017). The effect of sampling rate and lowpass filters on saccades – A modeling approach. *Behavior Research Methods*, *49*(6), 2146–2162. https://doi.org/10.3758/s13428-016-0848-4

Mackworth, N. H., & Bruner, J. S. (1970). How adults and children search and recognize pictures. *Human Development*, *13*(3), 149–177. https://doi.org/10.1159/000270887

Majaranta, P., Bates, R., & Donegan, M. (2009). Eye tracking. In Stephanidis, C. (Ed.) *The universal access handbook* (pp. 36:1–20): CRC Press.

Majaranta, P. (2011). *Gaze interaction and applications of eye tracking: Advances in assistive technologies: Advances in assistive technologies*. IGI Global.

Malpeli, J. G. (1998). Measuring eye position with the double magnetic induction method. *Journal of Neuroscience Methods*, *86*(1), 55–61. https://doi.org/10.1016/s0165-0270(98)00145-9.

Manor, B. R., & Gordon, E. (2003). Defining the temporal threshold for ocular fixation in free-viewing visuocognitive tasks. *Journal of Neuroscience Methods*, *128*(1-2), 85–93. https://doi.org/10.1016/s0165-0270(03)00151-1

Marandi, R. Z., & Gazerani, P. (2019). Aging and eye tracking: in the quest for objective biomarkers. *Future Neurology*, *14*(4), FNL33. https://doi.org/10.2217/fnl-2019-0012

Mardanbegi, D., & Hansen, D. W. (2012). Parallax error in the monocular head-mounted eye trackers. In *Proceedings of the 2012 acm conference on ubiquitous computing* (pp. 689–694).

Mathur, A., Gehrmann, J., & Atchison, D. A. (2013). Pupil shape as viewed along the horizontal visual field. *Journal of Vision*, *13*(6), 1–8. https://doi.org/10.1167/13.6.3

McCamy, M. B., Collins, N., Otero-Millan, J., Al-Kalbani, M., Macknik, S. L., Coakley, D., ..., Martinez-Conde, S. (2013). Simultaneous recordings of ocular microtremor and microsaccades with a piezoelectric sensor and a video-oculography system. *PeerJ*, *1*, e14. https://doi.org/10.7717/peerj.14

McCamy, M. B., Otero-Millan, J., Leigh, R. J., King, S. A., Schneider, R. M., Macknik, S. L., & Martinez-Conde, S. (2015). Simultaneous recordings of human microsaccades and drifts with a contemporary video eye tracker and the search coil technique. *PLOS ONE*, *10*(6), e0128428. https://doi.org/10.1371/journal.pone.0128428

McConkie, G. (1981). Evaluating and reporting data quality in eye movement research. *Behavior Research Methods*, *13*(2), 97–106. https://doi.org/10.3758/bf03207916

McConkie, G. (1997). Eye movement contingent display control: Personal reflections and comments. *Scientific Studies of Reading*, *1*(4), 303–316. https://doi.org/10.1207/s1532799xssr0104_1

McConkie, G., & Rayner, K. (1975). The span of the effective stimulus during a fixation in reading. *Perception & Psychophysics*, *17*(6), 578–586. https://doi.org/10.3758/BF03203972

McFarland, R., Roebuck, H., Yan, Y., Majolo, B., Li, W., & Guo, K. (2013). Social interactions through the eyes of macaques and humans. *PLOS ONE*, *8*(2), e56437. https://doi.org/10.1371/journal.pone.0056437

McLaughlin, S. C. (1967). Parametric adjustment in saccadic eye movements. *Perception & Psychophysics*, *2*(8), 359–362. https://doi.org/10.3758/bf03210071

McPartland, J. C., Webb, S. J., Keehn, B., & Dawson, G. (2011). Patterns of visual attention to faces and objects in autism spectrum disorder. *Journal of Autism and Developmental Disorders*, *41*(2), 148–157. https://doi.org/10.1007/s10803-010-1033-8

Merchant, J. (1967). *The oculometer* (Tech. Rep. No. CR-805). Washington: NASA.

Meyers, I. (1929). Electronystagmography: A graphic study of the action currents in nystagmus. *Archives of Neurology and*

Psychiatry, *21*(4), 901–918. https://doi.org/10.1001/archneurpsyc.1929.02210220172009.

Meyhöfer, I., Kasparbauer, A.-M., Steffens, M., & Ettinger, U. (2019). Effects of nicotine on smooth pursuit eye movements in healthy non-smokers. *Psychopharmacology*, *236*(7), 2259–2271. https://doi.org/10.1007/s00213-019-05223-1

Meyhöfer, I., Kumari, V., Hill, A., Petrovsky, N., & Ettinger, U. (2017). Sleep deprivation as an experimental model system for psychosis: effects on smooth pursuit, prosaccades, and antisaccades. *Journal of Psychopharmacology*, *31*(4), 418–433. https://doi.org/10.1177/0269881116675511

Monty, R. A. (1975). An advanced eye-movement measuring and recording system. *American Psychologist*, *30*(3), 331–335. https://doi.org/10.1037/0003-066X.30.3.331

Morgante, J. D., Zolfaghari, R., & Johnson, S. P. (2012). A critical test of temporal and spatial accuracy of the Tobii T60XL eye tracker. *Infancy*, *17*(1), 9–32. https://doi.org/10.1111/j.1532-7078.2011.00089.x

Moschner, C., & Baloh, R. W. (1994). Age-related changes in visual tracking. *Journal of Gerontology*, *49*(5), M235–M238. https://doi.org/10.1093/geronj/49.5.M235

Moser, A., Heide, W., & Kömpf, D. (1998). The effect of oral ethanol consumption on eye movements in healthy volunteers. *Journal of Neurology*, *245*(8), 542–550. https://doi.org/10.1007/s004150050240

Mould, M. S., Foster, D. H., Amano, K., & Oakley, J. P. (2012). A simple nonparametric method for classifying eye fixations. *Vision Research*, *57*, 18–25. https://doi.org/10.1016/j.visres.2011.12.006

Mulligan, J. B. (1997). Recovery of motion parameters from distortions in scanned images. In *Proceedings of the NASA Image Registration Workshop (IRW97)*: Goddard space flight center, Maryland.

Murphy, P. J., Duncan, A. L., Glennie, A. J., & Knox, P. C. (2001). The effect of scleral search coil lens wear on the eye. *British Journal of Ophthalmology*, *85*(3), 332–335. https://doi.org/10.1136/bjo.85.3.332

Narcizo, F. B., Ahmed, Z., & Hansen, D. W. (2017). Using priors to compensate geometrical problems in head-mounted eye trackers. In *19th European Conference on Eye Movements*.

Nasiopoulos, E., Risko, E. F., Foulsham, T., & Kingstone, A. (2015). Wearable computing: Will it make people prosocial?. *British Journal of Psychology*, *106*(2), 209–216. https://doi.org/10.1111/bjop.12080

Nevalainen, S., & Sajaniemi, J. (2004). Comparison of three eye tracking devices in psychology of programming research. In *Proceedings of the 16th Annual Workshop of the Psychology of Programming Interest Group, Carlow, Ireland* (pp. 151–158).

Newmeyer, M. N., Swortwood, M. J., Taylor, M. E., Abulseoud, O. A., Woodward, T. H., & Huestis, M. A. (2017). Evaluation of divided attention psychophysical task performance and effects on pupil sizes following smoked, vaporized and oral cannabis administration. *Journal of Applied Toxicology*, *37*(8), 922–932. https://doi.org/10.1002/jat.3440

Niehorster, D. C., Cornelissen, T., Holmqvist, K., & Hooge, I. T. C. (2019). Searching with and against each other: Spatiotemporal coordination of visual search behavior in collaborative and competitive settings. *Attention, Perception, & Psychophysics*, *81*(3), 666–683. https://doi.org/10.3758/s13414-018-01640-0

Niehorster, D. C., Cornelissen, T. H. W., Holmqvist, K., Hooge, I. T. C., & Hessels, R. S. (2018). What to expect from your remote eye-tracker when participants are unrestrained. *Behavior Research Methods*, *50*(1), 213–227. https://doi.org/10.3758/s13428-017-0863-0

Niehorster, D. C., Hessels, R. S., & Benjamins, J. S. (2020a). GlassesViewer: Open-source software for viewing and analyzing data from the Tobii Pro Glasses 2 eye tracker. *Behavior Research Methods*, *52*(3), 1244–1253. https://doi.org/10.3758/s13428-019-01314-1.

Niehorster, D. C., Santini, T., Hessels, R. S., Hooge, I. T. C., Kasneci, E., & Nyström, M. (2020b). The impact of slippage on the data quality of head-worn eye trackers. *Behavior Research Methods*, *52*(3), 1140–1160. https://doi.org/10.3758/s13428-019-01307-0

Niehorster, D. C., Zemblys, R., Beelders, T., & Holmqvist, K. (2020c). Characterizing gaze position signals and synthesizing noise during fixations in eye-tracking data. *Behavior Research Methods*, *52*(6), 2515–2534. https://doi.org/10.3758/s13428-020-01400-9

Niehorster, D. C., Zemblys, R., & Holmqvist, K. (2021). Is apparent fixational drift in eye-tracking data due to filters or eyeball rotation?. *Behavior Research Methods*, *53*(1), 311–324. https://doi.org/10.3758/s13428-020-01414-3

Nodine, C. F., Mello-Thoms, C., Kundel, H. L., & Weinstein, S. P. (2002). Time course of perception and decision making during mammographic interpretation. *American Journal of Roentgenology*, *179*(4), 917–923. https://doi.org/10.2214/ajr.179.4.1790917

Noton, D., & Stark, L. (1971). Scanpaths in saccadic eye movements while viewing and recognizing patterns. *Vision Research*, *11*(9), 929–942. https://doi.org/10.1016/0042-6989(71)90213-6

Nuthmann, A. (2013). On the visual span during object search in real-world scenes. *Visual Cognition*, *21*(7), 803–837. https://doi.org/10.1080/13506285.2013.832449

Nuthmann, A., & Kliegl, R. (2009). An examination of binocular reading fixations based on sentence corpus data. *Journal of Vision*, *9*(5), 1–28. https://doi.org/10.1167/9.5.31

Nyström, M., Hooge, I. T. C., & Holmqvist, K. (2013). Post-saccadic oscillations in eye movement data recorded with pupil-based eye trackers reflect motion of the pupil inside the iris. *Vision Research*, *92*, 59–66. https://doi.org/10.1016/j.visres.2013.09.009

Nyström, M., Andersson, R., Holmqvist, K., & van de Weijer, J. (2013). The influence of calibration method and eye physiology on eyetracking data quality. *Behavior Research Methods*, *45*(1), 272–288. https://doi.org/10.3758/s13428-012-0247-4

Nyström, M., & Holmqvist, K. (2010). An adaptive algorithm for fixation, saccade, and glissade detection in eye-tracking data. *Behavior Research Methods*, *42*(1), 188–204. https://doi.org/10.3758/BRM.42.1.188

Nyström, M., Hooge, I. T. C., & Andersson, R. (2016). Pupil size influences the eye-tracker signal during saccades. *Vision Research*, *121*, 95–103. https://doi.org/10.1016/j.visres.2016.01.009

Nyström, M., Niehorster, D. C., Andersson, R., & Hooge, I. T. C. (2021). The Tobii Pro Spectrum: A useful tool for studying microsaccades?. *Behavior Research Methods*, *53*(1), 335–353. https://doi.org/10.3758/s13428-020-01430-3

Oakes, L. M. (2010). Infancy guidelines for publishing eye-tracking data. *Infancy*, *15*(1), 1–5. https://doi.org/10.1111/j.1532-7078.2010.00030.x

O'Driscoll, G. A., & Callahan, B. L. (2008). Smooth pursuit in schizophrenia: A meta-analytic review of research since 1993. *Brain and Cognition*, *68*(3), 359–370. https://doi.org/10.1016/j.bandc.2008.08.023

Oliva, M., Niehorster, D. C., Jarodzka, H., & Holmqvist, K. (2017). Influence of coactors on saccadic and manual responses. *i-Perception*, *8*(1), 1–23. https://doi.org/10.1177/2041669517692814

Ooms, K., Dupont, L., Lapon, L., & Popelka, S. (2015). Accuracy and precision of fixation locations recorded with the low-cost eye tribe tracker in different experimental set-ups. *Journal of Eye Movement Research*, *8*(1), 1–20. https://doi.org/10.16910/jemr.8.1.5

Open Science Collaboration (2015). Estimating the reproducibility of psychological science. *Science, 349*(6251), aac4716. https://doi.org/10.1126/science.aac4716.

Orquin, J. L., Ashby, N. J. S., & Clarke, A. D. F. (2016). Areas of interest as a signal detection problem in behavioral eye-tracking research. *Journal of Behavioral Decision Making*, *29*(2-3), 103–115. https://doi.org/10.1002/bdm.1867

Orquin, J. L., & Holmqvist, K. (2018). Threats to the validity of eye-movement research in psychology. *Behavior Research Methods*, *50*(4), 1645–1656. https://doi.org/10.3758/s13428-017-0998-z

Oster, P. J., & Stern, J. A. (1980). Measurement of eye movement. In Martin, J., & Venables, P. H. (Eds.) *Techniques of Psychophysiology* (pp. 275–308). New York: Wiley.

Otero-Millan, J., Castro, J. L. A., Macknik, S. L., & Martinez-Conde, S. (2014). Unsupervised clustering method to detect microsaccades. *Journal of Vision*, *14*(2), 1–17. https://doi.org/10.1167/14.2.18

Otero-Millan, J., Optican, L. M., Macknik, S. L., & Martinez-Conde, S. (2018). Modeling the triggering of saccades, microsaccades, and saccadic intrusions. *Frontiers in Neurology*, *9*, 346. https://doi.org/10.3389/fneur.2018.00346

Over, E. A. B., Hooge, I. T. C., Vlaskamp, B. N. S., & Erkelens, C. J. (2007). Coarse-to-fine eye movement strategy in visual search. *Vision Research*, *47*(17), 2272–2280. https://doi.org/10.1016/j.visres.2007.05.002

Paffen, C. L. E., van der Smagt, M. J., te Pas, S. F., & Verstraten, F. A. J. (2005). Center-surround inhibition and facilitation as a function of size and contrast at multiple levels of visual motion processing. *Journal of Vision*, *5*(6), 571–578. https://doi.org/10.1167/5.6.8

Pajak, M., & Nuthmann, A. (2013). Object-based saccadic selection during scene perception: Evidence from viewing position effects. *Journal of Vision*, *13*(5), 1–21. https://doi.org/10.1167/13.5.2

Park, S. Y., Bacelar, C. E., & Holmqvist, K. (2020). Dog eye movements are slower than human eye movements. *Journal of Eye Movement Research*, *12*(8), 1–2[illegible]. https://doi.org/10.16910/jemr.12.8.4

Pastel, S., Chen, C.-H., Martin, L., Naujoks, M., Petri, K., & Witte, K. (2021). Comparison of gaze accuracy and precision in real-world and virtual reality. *Virtual Reality*, *25*, 175–189. https://doi.org/10.1007/s10055-020-00449-3

Paterson, K. B., McGowan, V. A., Warrington, K. L., Li, L., Li, S., Xie, F., . . . , Wang, J. (2020). Effects of normative aging on eye movements during reading. *Vision*, *4*(1), 7. https://doi.org/10.3390/vision4010007

Pélisson, D., Alahyane, N., Panouillères, M., & Tilikete, C. (2010). Sensorimotor adaptation of saccadic eye movements. *Neuroscience & Biobehavioral Reviews*, *34*(8), 1103–1120. https://doi.org/10.1016/j.neubiorev.2009.12.010

Poletti, M., & Rucci, M. (2016). A compact field guide to the study of microsaccades: challenges and functions. *Vision Research*, *118*, 83–97. https://doi.org/10.1016/j.visres.2015.01.018

Popelka, S., Stachoň, Z., Šašinka, C., & Doležalová, J. (2016). EyeTribe tracker data accuracy evaluation and its interconnection with hypothesis software for cartographic purposes. *Computational Intelligence and Neuroscience*, 9172506. https://doi.org/10.1155/2016/9172506.

Pérez, D. L., Radkowska, A., Rączaszek-Leonardi, J., Tomalski, P., & The Talby Study Team (2018). Beyond fixation durations: Recurrence quantification analysis reveals spatiotemporal dynamics of infant visual scanning. *Journal of Vision*, *18*(13), 1–17. https://doi.org/10.1167/18.13.5

Putnam, N. M., Hofer, H. J., Doble, N., Chen, L., Carroll, J., & Williams, D. R. (2005). The locus of fixation and the foveal cone mosaic. *Journal of Vision*, *5*(7), 632–639. https://doi.org/10.1167/5.7.3

Räihä, K.-J. (2015). Life in the fast lane: Effect of language and calibration accuracy on the speed of text entry by gaze. In *IFIP Conference on Human-Computer Interaction* (pp. 402–417): Springer.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, *124*(3), 372–422. https://doi.org/10.1037//0033-2909.124.3.372

Rayner, K., & Pollatsek, A. (1989). *The psychology of reading*. Prentice Hall.

Raynowska, J., Rizzo, J.-R., Rucker, J. C., Dai, W., Birkemeier, J., Hershowitz, J., . . . , Hudson, T. (2018). Validity of low-resolution eye-tracking to assess eye movements during a rapid number naming task: performance of the eyetribe eye tracker. *Brain Injury*, *32*(2), 200–208. https://doi.org/10.1080/02699052.2017.1374469

Recarte, M. A., & Nunes, L. M. (2003). Mental workload while driving: Effects on visual search, discrimination, and decision making. *Journal of Experimental Psychology: Applied*, *9*(2), 119–137. https://doi.org/10.1037/1076-898X.9.2.119

Recarte, M. A., Pérez, E., Conchillo, Á., & Nunes, L. M. (2008). Mental workload and visual impairment: Differences between pupil, blink, and subjective rating. *The Spanish Journal of Psychology*, *11*(2), 374–385. https://doi.org/10.1017/S1138741600004406

Reingold, E. M. (2014). Eye tracking research and technology: Towards objective measurement of data quality. *Visual Cognition*, *22*(3), 635–652. https://doi.org/10.1080/13506285.2013.876481

Reingold, E. M., & Sheridan, H. (2011). Eye movements and visual expertise in chess and medicine. In S. P. Liversedge, I. D. Gilchrist, & S. Everling (Eds.) *The Oxford handbook of eye movements*, (pp. 523–550): Oxford University Press.

Reulen, J. P. H., & Bakker, L. (1982). The measurement of eye movement using double magnetic induction. *IEEE Transactions on Biomedical Engineering*, *11*, 740–744. https://doi.org/10.1109/TBME.1982.325006.

Rice, K., Moriuchi, J. M., Jones, W., & Klin, A. (2012). Parsing heterogeneity in autism spectrum disorders: visual scanning of dynamic social scenes in school-aged children. *Journal of the American Academy of Child & Adolescent Psychiatry*, *51*(3), 238–248. https://doi.org/10.1016/j.jaac.2011.12.017

Risko, E. F., & Kingstone, A. (2011). Eyes wide shut: implied social presence, eye tracking and attention. *Attention, Perception, & Psychophysics*, *73*(2), 291–296. https://doi.org/10.3758/s13414-010-0042-1.

Roberts, D. C., Marcelli, V., Gillen, J. S., Carey, J. P., Della Santina, C. C., & Zee, D. S. (2011). MRI magnetic field stimulates rotational sensors of the brain. *Current Biology*, *21*(19), 1635–1640. https://doi.org/10.1016/j.cub.2011.08.029

Robinson, D. A. (1963). A method of measuring eye movement using a scleral search coil in a magnetic field. *IEEE Transactions on Bio-Medical Electronics*, *10*(4), 137–145. https://doi.org/10.1109/TBMEL.1963.4322822

Rosengren, W., Nyström, M., Hammar, B., & Stridh, M. (2020). A robust method for calibration of eye tracking data recorded during nystagmus. *Behavior Research Methods*, *52*(1), 36–50. https://doi.org/10.3758/s13428-019-01199-0

Rucci, M., Wu, R.-J., & Zhao, Z. (2020). System and method for real-time high-resolution eye-tracking. Google Patents. (US Patent App. 16/550,831).

Rupp, H. A., & Wallen, K. (2007). Sex differences in viewing sexual stimuli: An eye-tracking study in men and women. *Hormones and Behavior*, *51*(4), 524–533. https://doi.org/10.1016/j.yhbeh.2007.01.008

Salman, M. S., Sharpe, J. A., Eizenman, M., Lillakas, L., Westall, C., To, T., . . . , Steinbach, M. J. (2006). Saccades in children. *Vision Research*, *46*(8-9), 1432–1439. https://doi.org/10.1016/j.visres.2005.06.011

Salvucci, D. D., & Goldberg, J. H. (2000). Identifying fixations and saccades in eye-tracking protocols. In *Proceedings of the Symposium on Eye Tracking Research and Applications* (pp. 71–78).

Santini, T., Fuhl, W., & Kasneci, E. (2018). PuReST: Robust pupil tracking for real-time pervasive eye tracking. In *Proceedings of the Symposium on Eye Tracking Research and Applications* (pp. 1–5).

Sasson, N. J., Elison, J. T., Turner-Brown, L. M., Dichter, G. S., & Bodfish, J. W. (2011). Circumscribed attention in young children with autism. *Journal of Autism and Developmental Disorders*, *41*(2), 242–247. https://doi.org/10.1007/s10803-010-1038-3

Saunders, D. R., & Woods, R. L. (2014). Direct measurement of the system latency of gaze-contingent displays. *Behavior Research Methods*, *46*(2), 439–447. https://doi.org/10.3758/s13428-013-0375-5

Schalén, L., Pyykkö, I., Juhola, M., Magnusson, M., Jäntti, V., & Henriksson, N. (1983). Intra-individual variation in oculomotor performance in man. *Acta Oto-Laryngologica*, *96*(406), 212–217. https://doi.org/10.3109/00016488309123036

Schlegelmilch, K., & Wertz, A. E. (2019). The effects of calibration target, screen location, and movement type on infant eye-tracking data quality. *Infancy*, *24*(4), 636–662. https://doi.org/10.1111/infa.12294

Schnipke, S. K., & Todd, M. W. (2000). Trials and tribulations of using an eye-tracking system. In *CHI'00 Extended Abstracts on Human Factors in Computing Systems* (pp. 273–274). New York: ACM.

Schott, E. (1922). Über die Registrierung des Nystagmus und anderer Augenbewegungen vermittels des Saitengalvanometers. *Deutsches Archiv fur Klinische Medizin*, *140*, 79–90.

Scinto, L. F. M., Pillalamarri, R., & Karsh, R. (1986). Cognitive strategies for visual search. *Acta Psychologica*, *62*(3), 263–292. https://doi.org/10.1016/0001-6918(86)90091-0

Seemiller, E. S., Port, N. L., & Candy, T. R. (2018). The gaze stability of 4-to 10-week-old human infants. *Journal of Vision*, *18*(8), 1–10. https://doi.org/10.1167/18.8.15

Senju, A., & Johnson, M. H. (2009). Atypical eye contact in autism: Models, mechanisms and development. *Neuroscience & Biobehavioral Reviews*, *33*(8), 1204–1214. https://doi.org/10.1016/j.neubiorev.2009.06.001

Sharafi, Z., Shaffer, T., Sharif, B., & Guöhöneuc, Y. (2015). Eye-tracking metrics in software engineering. In *2015 Asia-Pacific Software Engineering Conference (APSEC)* (pp. 96–103).

Sharafi, Z., Sharif, B., Guéhéneuc, Y.-G., Begel, A., Bednarik, R., & Crosby, M. (2020). A practical guide on conducting eye tracking studies in software engineering. *Empirical Software Engineering*, *25*(5), 3128–3174. https://doi.org/10.1007/s10664-020-09829-4

Sharpe, J. A., & Sylvester, T. O. (1978). Effect of aging on horizontal smooth pursuit. *Investigative Ophthalmology & Visual Science*, *17*(5), 465–468.

Shic, F., Scassellati, B., & Chawarska, K. (2008). The incomplete fixation measure. In *Proceedings of the Symposium on Eye Tracking Research and Applications* (pp. 111–114).

Shukla, M., Wen, J., White, K. S., & Aslin, R. N. (2011). SMART-T: A system for novel fully automated anticipatory eye-tracking paradigms. *Behavior Research Methods*, *43*(2), 384–398. https://doi.org/10.3758/s13428-010-0056-6

Sibert, L. E., & Jacob, R. J. K. (2000). Evaluation of eye gaze interaction. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems* (pp. 281–288).

Siegenthaler, E., Costela, F. M., McCamy, M. B., Di Stasi, L. L., Otero-Millan, J., Sonderegger, A., ..., Martinez-Conde, S. (2014). Task difficulty in mental arithmetic affects microsaccadic rates and magnitudes. *European Journal of Neuroscience*, *39*(2), 287–294. https://doi.org/10.1111/ejn.12395

Slattery, T. J., Angele, B., & Rayner, K. (2011). Eye movements and display change detection during reading. *Journal of Experimental Psychology: Human Perception and Performance*, *37*(6), 1924–1938. https://doi.org/10.1037/a0024322

Šmideková, Z., Janik, M., Minarikova, E., & Holmqvist, K. (2020). Teachers' gaze over space and time in a real-world classroom. *Journal of Eye Movement Research*, *13*(4), 1–20. https://doi.org/10.16910/jemr.13.4.1

Smyrnis, N., Amado, I., Krebs, M.-O., & Sweeney, J. A. (2019). Eye movements in psychiatry. In C. Klein, & U. Ettinger (Eds.) *Eye Movement Research* (pp. 703–748). Cham: Springer.

Špakov, O. (2012). Comparison of eye movement filters used in HCI. In *Proceedings of the Symposium on Eye Tracking Research and Applications* (pp. 281–284).

Špakov, O., Isokoski, P., & Majaranta, P. (2014). Look and lean: accurate head-assisted eye pointing. In *Proceedings of the Symposium on Eye Tracking Research and Applications* (pp. 35–42).

Špakov, O., Istance, H., Hyrskykari, A., Siirtola, H., & Räihä, K.-J. (2019). Improving the performance of eye trackers with limited spatial accuracy and low sampling rates for reading analysis by heuristic fixation-to-word mapping. *Behavior Research Methods*, *51*(6), 2661–2687. https://doi.org/10.3758/s13428-018-1120-x

S. R. Research (2017). *EyeLink portable duo user manual*. SR Research Ltd.

Stampe, D. M. (1993). Heuristic filtering and reliable calibration methods for video-based pupil-tracking systems. *Behavior Research Methods, Instruments, & Computers*, *25*(2), 137–142. https://doi.org/10.3758/BF03204486

Startsev, M., Agtzidis, I., & Dorr, M. (2019). 1D CNN with BLSTM for automated classification of fixations, saccades, and smooth pursuits. *Behavior Research Methods*, *51*(2), 556–572. https://doi.org/10.3758/s13428-018-1144-2

Startsev, M., Göb, S., & Dorr, M. (2019). A novel gaze event detection metric that is not fooled by gaze-independent baselines. In *Proceedings of the Symposium on Eye Tracking Research and Applications* (pp. 1–9).

Steffen, M. (1990). A simple method for monotonic interpolation in one dimension. *Astronomy and Astrophysics*, *239*(1-2), 443–450.

Stein, N., Niehorster, D. C., Watson, T., Steinicke, F., Rifai, K., Wahl, S., & Lappe, M. (2021). A comparison of eye tracking latencies among several commercial head-mounted displays. *i-Perception*, *12*(1), 1–16. https://doi.org/10.1177/2041669520983338

Steinman, R. M. (1965). Effect of target size, luminance, and color on monocular fixation. *Journal of the Optical Society of America*, *55*(9), 1158–1164. https://doi.org/10.1364/josa.55.001158

Stern, J. A., Boyer, D., & Schroeder, D. (1994). Blink rate: A possible measure of fatigue. *Human Factors*, *36*(2), 285–297. https://doi.org/10.1177/001872089403600209

Stevenson, S. B., & Roorda, A. (2005). Correcting for miniature eye movements in high-resolution scanning laser ophthalmoscopy. In *Ophthalmic Technologies XV*, (Vol. 5688, pp. 145–151): International Society for Optics and Photonics.

Stevenson, S. B., Sheehy, C. K., & Roorda, A. (2016). Binocular eye tracking with the Tracking Scanning Laser Ophthalmoscope. *Vision Research*, *118*, 98–104. https://doi.org/10.1016/j.visres.2015.01.019

Strohmaier, A. R., MacKay, K. J., Obersteiner, A., & Reiss, K. M. (2020). Eye-tracking methodology in mathematics education research: A systematic literature review. *Educational Studies in Mathematics*, *104*, 147–200. https://doi.org/10.1007/s10649-020-09948-1

Stuart, S., Hickey, A., Vitorio, R., Welman, K., Foo, S., Keen, D., & Godfrey, A. (2019). Eye-tracker algorithms to detect saccades during static and dynamic tasks: a structured review. *Physiological Measurement*, *40*(2), 02TR01. https://doi.org/10.1088/1361-6579/ab02ab

Sun, Z.-K., Wang, J.-Y., & Luo, F. (2016). Experimental pain induces attentional bias that is modified by enhanced motivation: An eye tracking study. *European Journal of Pain*, *20*(8), 1266–1277. https://doi.org/10.1002/ejp.851

Švede, A., Treija, E., Jaschinski, W., & Krūmiņa, G. (2015). Monocular versus binocular calibrations in evaluating fixation disparity with a video-based eye-tracker. *Perception*, *44*(8-9), 1110–1128. https://doi.org/10.1177/0301006615596886

Talanow, T., Kasparbauer, A.-M., Lippold, J. V., Weber, B., & Ettinger, U. (2020). Neural correlates of proactive and reactive inhibition of saccadic eye movements. *Brain Imaging and Behavior*, *14*(1), 72–88. https://doi.org/10.1007/s11682-018-9972-3

Tari, B., Fadel, M. A., & Heath, M. (2019). Response suppression produces a switch-cost for spatially compatible saccades. *Experimental Brain Research*, *237*(5), 1195–1203. https://doi.org/10.1007/s00221-019-05497-z

Tatham, A. J., Murray, I. C., McTrusty, A. D., Cameron, L. A., Perperidis, A., Brash, H. M., . . . , Minns, R. A. (2020). Speed and accuracy of saccades in patients with glaucoma evaluated using an eye tracking perimeter. *BMC Ophthalmology*, *20*(1), 259. https://doi.org/10.1186/s12886-020-01528-4

Tatler, B. W., & Vincent, B. T. (2008). Systematic tendencies in scene viewing. *Journal of Eye Movement Research*, *2*(2), 1–18. https://doi.org/10.16910/jemr.2.2.5

Tatler, B. W., Hansen, D. W., & Pelz, J. B. (2019). Eye movement recordings in natural settings. In C. Klein, & U. Ettinger (Eds.) *Eye Movement Research* (pp. 703–748). Cham: Springer.

Tatler, B. W., & Hutton, S. B. (2007). Trial by trial effects in the antisaccade task. *Experimental Brain Research*, *179*(3), 387–396. https://doi.org/10.1007/s00221-006-0799-6

Thaler, L., Schütz, A. C., Goodale, M. A., & Gegenfurtner, K. R. (2013). What is the best fixation target? the effect of target shape on stability of fixational eye movements. *Vision Research*, *76*, 31–42. https://doi.org/10.1016/j.visres.2012.10.012

Thomas, M. G., Gottlob, I., McLean, R. J., Maconachie, G., Kumar, A., & Proudlock, F. A. (2011). Reading strategies in infantile nystagmus syndrome. *Investigative Ophthalmology & Visual Science*, *52*(11), 8156–8165. https://doi.org/10.1167/iovs.10-6645

Toet, A., Bijlsma, M., & Brouwer, A.-M. (2017). Stress response and facial trustworthiness judgments in civilians and military. *SAGE Open*, *7*(3), 1–11. https://doi.org/10.1177/2158244017725386

Toet, A., & Levi, D. M. (1992). The two-dimensional shape of spatial interaction zones in the parafovea. *Vision Research*, *32*(7), 1349–1357. https://doi.org/10.1016/0042-6989(92)90227-a

Tomalski, P., & Malinowska-Korczak, A. (2020). What do young infants do during eye-tracking experiments? IP-BET–a coding scheme for quantifying spontaneous infant and parent behaviour. *Frontiers in Psychology*, *11*, 764. https://doi.org/10.3389/fpsyg.2020.00764

Török, N., Guillemin, V., & Barnothy, J. M. (1951). Photoelectric nystagmography. *Annals of Otology, Rhinology & Laryngology*, *60*(4), 917–926. https://doi.org/10.1177/000348945106000402

Träisk, F., Bolzani, R., & Ygge, J. (2005). A comparison between the magnetic scleral search coil and infrared reflection methods for saccadic eye movement analysis. *Graefe's Archive for Clinical and Experimental Ophthalmology*, *243*(8), 791–797. https://doi.org/10.1007/s00417-005-1148-3

Tran, T., & Kaufman, L. M. (2003). The child's eye in systemic diseases. *Pediatric Clinics of North America*, *50*(1), 241–258. https://doi.org/10.1016/s0031-3955(02)00116-5

Uesbeck, P. M., Peterson, C. S., Sharif, B., & Stefik, A. (2020). A randomized controlled trial on the effects of embedded computer language switching. In *Proceedings of the 28th ACM Joint Meeting on European Software Engineering Conference and Symposium on the Foundations of Software Engineering* (pp. 410–420).

Unema, P. J. A., Pannasch, S., Joos, M., & Velichkovsky, B. M. (2005). Time course of information processing during scene perception: The relationship between saccade amplitude and fixation duration. *Visual Cognition*, *12*(3), 473–494. https://doi.org/10.1080/13506280444000409

Valtakari, N. V., Hooge, I. T. C., Viktorsson, C., Nyström, P., Falck-Ytter, T., & Hessels, R. S. (2021). Eye tracking in human interaction: Possibilities and limitations. *Behavior Research Methods*, 1–17. https://doi.org/10.3758/s13428-020-01517-x.

van der Geest, J. N., Kemner, C., Verbaten, M. N., & van Engeland, H. (2002). Gaze behavior of children with pervasive developmental disorder toward human faces: a fixation time study. *Journal of Child Psychology and Psychiatry*, *43*(5), 669–678. https://doi.org/10.1111/1469-7610.00055

van der Laan, L. N., Papies, E. K., Hooge, I. T. C., & Smeets, P. A. M. (2017). Goal-directed visual attention drives health goal priming: An eye-tracking experiment. *Health Psychology*, *36*(1), 82–90. https://doi.org/10.1037/hea0000410

Van der Stigchel, S., Hessels, R. S., van Elst, J. C., & Kemner, C. (2017). The disengagement of visual attention in the gap paradigm across adolescence. *Experimental Brain Research*, *235*(12), 3585–3592. https://doi.org/10.1007/s00221-017-5085-2

Van Der Werf, Y. D., Altena, E., Schoonheim, M. M., Sanz-Arigita, E. J., Vis, J. C., De Rijke, W., & Van Someren, E. J. W. (2009). Sleep benefits subsequent hippocampal functioning. *Nature Neuroscience*, *12*(2), 122–123. https://doi.org/10.1038/nn.2253

Van Dongen, H. P., & Dinges, D. F. (2005). Sleep, circadian rhythms, and psychomotor vigilance. *Clinics in Sports Medicine*, *24*(2), 237–249. https://doi.org/10.1016/j.csm.2004.12.007

Van Orden, K. F., Jung, T.-P., & Makeig, S. (2000). Combined eye activity measures accurately estimate changes in sustained visual task performance. *Biological Psychology*, *52*(3), 221–240. https://doi.org/10.1016/s0301-0511(99)00043-5

van Renswoude, D. R., Raijmakers, M. E. J., Koornneef, A., Johnson, S. P., Hunnius, S., & Visser, I. (2018). Gazepath: An eye-tracking analysis tool that accounts for individual differences and data quality. *Behavior Research Methods*, *50*(2), 834–852. https://doi.org/10.3758/s13428-017-0909-3

Villanueva, A., & Cabeza, R. (2008). Evaluation of corneal refraction in a model of a gaze tracking system. *IEEE Transactions on Biomedical Engineering*, *55*(12), 2812–2822. https://doi.org/10.1109/tbme.2008.2002152

Viviani, P. (1990). Eye movements in visual search: Cognitive, perceptual, and motor control aspects. *Eye movements and their role in visual and cognitive processes*, 353–383.

Vlaskamp, B. N. S., Over, E. A. B., & Hooge, I. T. C. (2005). Saccadic search performance: the effect of element spacing. *Experimental Brain Research*, *167*(2), 246–259. https://doi.org/10.1007/s00221-005-0032-z.

Walcher, S., Körner, C., & Benedek, M. (2017). Looking for ideas: Eye behavior during goal-directed internally focused cognition. *Consciousness and Cognition*, *53*, 165–175. https://doi.org/10.1016/j.concog.2017.06.009

Walshe, R. C., & Nuthmann, A. (2014). Asymmetrical control of fixation durations in scene viewing. *Vision Research*, *100*, 38–46. https://doi.org/10.1016/j.visres.2014.03.012

Wang, D., Mulvey, F. B., Pelz, J. B., & Holmqvist, K. (2017). A study of artificial eyes for the measurement of precision in eye-trackers. *Behavior Research Methods*, *49*(3), 947–959. https://doi.org/10.3758/s13428-016-0755-8

Wang, Q., Kim, E., Chawarska, K., Scassellati, B., Zucker, S., & Shic, F. (2014). On relationships between fixation identification algorithms and fractal box counting methods. In *Proceedings of the Symposium on Eye Tracking Research and Applications* (pp. 67–74).

Wang, X., Holmqvist, K., & Alexa, M. (2019). The recorded mean point of vergence is biased. *Journal of Eye Movement Research*, *12*(4), 1–27. https://doi.org/10.16910/jemr.12.4.2

Wang, Y., Mehler, B., Reimer, B., Lammers, V., D'Ambrosio, L. A., & Coughlin, J. F. (2010). The validity of driving simulation for assessing differences between in-vehicle informational interfaces: A comparison with field testing. *Ergonomics*, *53*(3), 404–420. https://doi.org/10.1080/00140130903464358

Wass, S. V., Smith, T. J., & Johnson, M. H. (2013). Parsing eye-tracking data of variable quality to provide accurate fixation duration estimates in infants and adults. *Behavior Research Methods*, *45*(1), 229–250. https://doi.org/10.3758/s13428-012-0245-6

Wass, S. V. (2014). Comparing methods for measuring peak look duration: Are individual differences observed on screen-based tasks also found in more ecologically valid contexts?. *Infant Behavior and Development*, *37*(3), 315–325. https://doi.org/10.1016/j.infbeh.2014.04.007

Wass, S. V., Forssman, L., & Leppänen, J. (2014). Robustness and precision: How data quality may influence key dependent variables in infant eye-tracker analyses. *Infancy*, *19*(5), 427–460. https://doi.org/10.1111/infa.12055

Webb, S. J., Shic, F., Murias, M., Sugar, C. A., Naples, A. J., Barney, E., ..., Autism Biomarkers Consortium for Clinical Trials (2020). Biomarker acquisition and quality control for multi-site studies: The autism biomarkers consortium for clinical trials. *Frontiers in Integrative Neuroscience*, *13*, 71. https://doi.org/10.3389/fnint.2019.00071.

Wenzlaff, F., Briken, P., & Dekker, A. (2018). If there's a penis, it's most likely a man: Investigating the social construction of gender using eye tracking. *PLOS ONE*, *13*(3), e0193616. https://doi.org/10.1371/journal.pone.0193616

Wierts, R., Janssen, M. J. A., & Kingma, H. (2008). Measuring saccade peak velocity using a low-frequency sampling rate of 50 hz. *IEEE Transactions on Biomedical Engineering*, *55*(12), 2840–2842. https://doi.org/10.1109/tbme.2008.925290

Wildenmann, U., & Schaeffel, F. (2013). Variations of pupil centration and their effects on video eye tracking. *Ophthalmic and Physiological Optics*, *33*(6), 634–641. https://doi.org/10.1111/opo.12086

Wilkinson, I. M. S., Kime, R., & Purnell, M. (1974). Alcohol and human eye movement. *Brain*, *97*(1), 785–792. https://doi.org/10.1093/brain/97.1.785

Wilson, MA., Campbell, M. C., & Simonet, P. (1992). Change of pupil centration with change of illumination and pupil size. *Optometry and Vision Science*, *69*(2), 129–136. https://doi.org/10.1097/00006324-199202000-00006

Wu, D. W.-L., Bischof, W. F., Anderson, N. C., Jakobsen, T., & Kingstone, A. (2014). The influence of personality on social attention. *Personality and Individual Differences*, *60*, 25–29. https://doi.org/10.1016/j.paid.2013.11.017

Wyatt, H. J. (1995). The form of the human pupil. *Vision Research*, *35*(14), 2021–2036. https://doi.org/10.1016/0042-6989(94)00268-q.

Wyatt, H. J. (2010). The human pupil and the use of video-based eyetrackers. *Vision Research*, *50*(19), 1982–1988. https://doi.org/10.1016/j.visres.2010.07.008

Yarbus, A. L. (1967). *Eye movements and vision*. New York: Plenum Press.

Young, L. R., & Sheena, D. (1975). Survey of eye movement recording methods. *Behavior Research Methods & Instrumentation*, *7*(5), 397–429. https://doi.org/10.3758/bf03201553

Zemblys, R., Niehorster, D. C., & Holmqvist, K. (2019). gazeNet: End-to-end eye-movement event detection with deep neural networks. *Behavior Research Methods*, *51*(2), 840–864. https://doi.org/10.3758/s13428-018-1133-5.

Zemblys, R., Niehorster, D. C., & Holmqvist, K. (2021). Evaluating three approaches to binary event-level agreement scoring. A reply to Friedman (2020). *Behavior Research Methods*, *53*(1), 325–334. https://doi.org/10.3758/s13428-020-01425-0

Zemblys, R., Niehorster, D. C., Komogortsev, O., & Holmqvist, K. (2018). Using machine learning to detect events in eye-tracking data. *Behavior Research Methods*, *50*(1), 160–181. https://doi.org/10.3758/s13428-017-0860-3.

Zingale, C. M., & Kowler, E. (1987). Planning sequences of saccades. *Vision Research*, *27*(8), 1327–1341. https://doi.org/10.1016/0042-6989(87)90210-0.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Affiliations

Kenneth Holmqvist[1,2,3] · **Saga Lee Örbom**[3] · **Ignace T. C. Hooge**[4] · **Diederick C. Niehorster**[5] · **Robert G. Alexander**[6] · **Richard Andersson**[7] · **Jeroen S. Benjamins**[4,8] · **Pieter Blignaut**[2] · **Anne-Marie Brouwer**[9] · **Lewis L. Chuang**[10,11] · **Kirsten A. Dalrymple**[12] · **Denis Drieghe**[13] · **Matt J. Dunn**[14] · **Ulrich Ettinger**[15] · **Susann Fiedler**[16] · **Tom Foulsham**[17] · **Jos N. van der Geest**[18] · **Dan Witzner Hansen**[19] · **Samuel B. Hutton**[20] · **Enkelejda Kasneci**[21] · **Alan Kingstone**[22] · **Paul C. Knox**[23] · **Ellen M. Kok**[24,25] · **Helena Lee**[26] · **Joy Yeonjoo Lee**[27] · **Jukka M. Leppänen**[28] · **Stephen Macknik**[6] · **Päivi Majaranta**[29] · **Susana Martinez-Conde**[6] · **Antje Nuthmann**[30] · **Marcus Nyström**[31] · **Jacob L. Orquin**[32,33] · **Jorge Otero-Millan**[34] · **Soon Young Park**[35] · **Stanislav Popelka**[36] · **Frank Proudlock**[37] · **Frank Renkewitz**[38] · **Austin Roorda**[34] · **Michael Schulte-Mecklenbeck**[39,40] · **Bonita Sharif**[41] · **Frederick Shic**[42,43] · **Mark Shovman**[44,45] · **Mervyn G. Thomas**[37] · **Ward Venrooij**[46] · **Raimondas Zemblys**[47] · **Roy S. Hessels**[4]

[1] Department of Psychology, Nicolaus Copernicus University, Torun, Poland

[2] Department of Computer Science and Informatics, University of the Free State, Bloemfontein, South Africa

[3] Department of Psychology, Regensburg University, Regensburg, Germany

[4] Experimental Psychology, Helmholtz Institute, Utrecht University, Utrecht, The Netherlands

[5] Lund University Humanities Lab and Department of Psychology, Lund University, Lund, Sweden

[6] Department of Ophthalmology, SUNY Downstate Health Sciences University, Brooklyn, NY, USA

[7] Tobii Pro AB, Danderyd, Sweden

[8] Social, Health and Organizational Psychology, Utrecht University, Utrecht, The Netherlands

[9] TNO, Soesterberg, The Netherlands

[10] Department of Ergonomics, Leibniz Institute for Working Environments and Human Factors, Dortmund, Germany

[11] Institute of Informatics, LMU Munich, Munich, Germany

[12] Institute of Child Development, University of Minnesota, Minneapolis, USA

[13] School of Psychology, University of Southampton, Southampton, UK

[14] School of Optometry and Vision Sciences, Cardiff University, Cardiff, UK

[15] Department of Psychology, University of Bonn, Bonn, Germany

[16] Vienna University of Economics and Business, Vienna, Austria

[17] Department of Psychology, University of Essex, Essex, UK

[18] Department of Neuroscience, Erasmus MC, Rotterdam, The Netherlands

[19] Machine Learning Group, Department of Computer Science, IT University of Copenhagen, Copenhagen, Denmark

[20] SR Research Ltd, Ottawa, Canada

[21] Human-Computer Interaction, University of Tübingen, Tübingen, Germany

[22] University of British Columbia, Columbia, Canada

[23] Department of Eye and Vision Science, Institute of Life Course and Medical Sciences, University of Liverpool, Liverpool, UK

[24] Department of Education and Pedagogy, Division Education, Faculty of Social and Behavioral Sciences, Utrecht University, Utrecht, The Netherlands

[25] Department of Online Learning and Instruction, Faculty of Educational Sciences, Open University of the Netherlands, Heerlen, The Netherlands

[26] University of Southampton, Southampton, UK

[27] School of Health Professions Education, Faculty of Health, Medicine, and Life Sciences, Maastricht University, Maastricht, The Netherlands

[28] Department of Psychology and Speed-Language Pathology, University of Turku, Turku, Finland

[29] TAUCHI Research Center, Computing Sciences, Faculty of Information Technology and Communication Sciences, Tampere University, Tampere, Finland

[30] Institute of Psychology, University of Kiel, Kiel, Germany

[31] Lund University Humanities Lab, Lund University, Lund, Sweden

[32] Department of Management, Aarhus University, Aarhus, Denmark

[33] Center for Research in Marketing and Consumer Psychology, Reykjavik University, Reykjavik, Iceland

[34] Herbert Wertheim School of Optometry and Vision Science, University of California, Berkeley, CA, USA

[35] Comparative Cognition, Messerli Research Institute, University of Veterinary Medicine Vienna, Medical University of Vienna, Vienna, Austria

[36] Department of Geoinformatics, Palacký University Olomouc, Olomouc, Czech Republic

[37] The University of Leicester Ulverscroft Eye Unit, Department of Neuroscience, Psychology and Behaviour, University of Leicester, Leicester, UK

[38] Department of Psychology, University of Erfurt, Erfurt, Germany

[39] University of Bern, Bern, Switzerland

[40] Max Planck Institute for Human Development, Berlin, Germany

[41] School of Computing, University of Nebraska-Lincoln, Lincoln, Nebraska, USA

[42] Center for Child Health, Behavior and Development, Seattle Children's Research Institute, Seattle, WA, USA

[43] Department of General Pediatrics, University of Washington School of Medicine, Seattle, WA, USA

[44] Eyeviation Systems, Herzliya, Israel

[45] Department of Industrial Design, Bezalel Academy of Arts and Design, Jerusalem, Israel

[46] Electrical Engineering, Mathematics and Computer Science (EEMCS), University of Twente, Enschede, The Netherlands

[47] Smart Eye AB, Göteborg, Sweden