



Implementing mobile eye tracking in psychological research: A practical guide

Xiaoxue Fu¹ · John M. Franchak² · Leigha A. MacNeill^{3,4} · Kelley E. Gunther⁵ · Jeremy I. Borjon^{6,7,8} · Julia Yurkovic-Harding¹ · Samuel Harding¹ · Jessica Bradshaw¹ · Koraly E. Pérez-Edgar⁹

Accepted: 20 June 2024
© The Author(s) 2024

Abstract

Eye tracking provides direct, temporally and spatially sensitive measures of eye gaze. It can capture visual attention patterns from infancy through adulthood. However, commonly used screen-based eye tracking (SET) paradigms are limited in their depiction of how individuals process information as they interact with the environment in “real life”. Mobile eye tracking (MET) records participant-perspective gaze in the context of active behavior. Recent technological developments in MET hardware enable researchers to capture egocentric vision as early as infancy and across the lifespan. However, challenges remain in MET data collection, processing, and analysis. The present paper aims to provide an introduction and practical guide to starting researchers in the field to facilitate the use of MET in psychological research with a wide range of age groups. First, we provide a general introduction to MET. Next, we briefly review MET studies in adults and children that provide new insights into attention and its roles in cognitive and socioemotional functioning. We then discuss technical issues relating to MET data collection and provide guidelines for data quality inspection, gaze annotations, data visualization, and statistical analyses. Lastly, we conclude by discussing the future directions of MET implementation. Open-source programs for MET data quality inspection, data visualization, and analysis are shared publicly.

Keywords Mobile eye tracking · Head-mounted eye tracking · Method · Technology

✉ Xiaoxue Fu
xiaoxuef@mailbox.sc.edu

- ¹ Department of Psychology, University of South Carolina, Columbia, SC, USA
- ² Department of Psychology, University of California Riverside, Riverside, CA, USA
- ³ Department of Medical Social Sciences, Northwestern University, Feinberg School of Medicine, Northwestern University, Chicago, IL, USA
- ⁴ Institute for Innovations in Developmental Sciences, Northwestern University, Evanston, IL, USA
- ⁵ Neuroscience and Cognitive Science Program, University of Maryland, College Park, MD, USA
- ⁶ Department of Psychology, University of Houston, Houston, TX, USA
- ⁷ Texas Institute for Measurement, Evaluation, and Statistics, University of Houston, Houston, TX, USA
- ⁸ Texas Center for Learning Disorders, University of Houston, Houston, TX, USA
- ⁹ Department of Psychology, The Pennsylvania State University, University Park, PA, USA

Introduction

Eye movements provide a window into one’s perception, cognition, and visually guided behavior. Eye movements can indicate the deployment of visual attention (Henderson, 2003). Attention, in turn, acts as a processing mechanism that filters out excessive information from the environment by biasing selection based on the individual’s current goals (Desimone & Duncan, 1995) and affective states (Todd et al., 2012). Visual experiences influence downstream cognition, learning, action, and affect (Crick & Dodge, 1994). Controlled laboratory experiments often study visual attention in isolation. However, in everyday life, visual attention is closely linked to the individual’s ongoing behavior and experiences of the physical and social environments (Franchak, 2020a; Hayhoe & Rothkopf, 2011). While we have gained tremendous insights from screen-based tasks, without studying attention in situ, we can only approximate how attention, action, and social information dynamically influence each other in real-time and in real-life environments.

Head-mounted, or mobile eye tracking (MET), records eye movements embedded in an individual's free-flowing behaviors as they interact with the environment. The technology has been pioneered in adults since the early 1900s (Land, 2006). MET systems have become more portable and robust with technological advancement. This recent development facilitates research into attention during active visual exploration (Ballard et al., 1997), especially in infants and young children (Franchak, 2017; Franchak, 2019). MET typically consists of a scene camera that captures the wearer's first-person view and one or two eye cameras that support monocular or binocular eye tracking, respectively. The MET system records the wearer's gaze direction and maps the three-dimensional gaze point to the two-dimensional space of the scene camera, allowing researchers to visualize the point of gaze overlaid on the scene camera recording (MacInnes et al., 2018).

The use of MET yields several key advantages. Compared to video recordings, MET provides a more proximal, temporally and spatially sensitive measure of attention from the first-person perspective (Franchak, 2019; Franchak, 2020b; Fu & Pérez-Edgar, 2019; Pérez-Edgar et al., 2020). MET captures rich micro-longitudinal data by sampling looking locations within self-generated behavior for extended periods of time (see [Data visualization](#) section). MET data can then be used to probe within-person changes of attention over time and capture the moment-to-moment dynamics between the environmental inputs, individuals' attention, and behavior (see [Data analysis](#) section). Hence, MET studies may provide new understandings of human cognition operating within the individual's active motor and social behaviors (Ballard et al., 1997; Gibson, 1979; Yoshida & Burling, 2011).

The present paper provides an introduction and practical guide for MET data collection, processing, and analytic methods to new researchers in the field. Existing literature

has highlighted the utility and advantages of MET (Franchak, 2019, 2020a; Pérez-Edgar et al., 2020; Yoshida & Burling, 2011), technical challenges (Hessels, Niehorster et al., 2020b; Niehorster et al., 2020; Valtakari et al., 2021), and provided practical guides in MET data collection and data quality inspection (Franchak & Yu, 2022; Hooge et al., 2023; Niehorster et al., 2023; Slone et al., 2018). The present paper complements and extends existing method papers by providing a review of current MET methodologies and practical guidance that are applicable to MET research that covers a wide age span from infancy to adulthood. We will briefly review studies that illustrate the utility of MET as an integral tool for understanding attentional processes in locomotion, learning, and social interactions in adults, children, and infants in the "[The utility of MET technology](#)" section. This is followed by recommendations on MET data collection in the "[MET data collection considerations](#)" section, data quality assessment in the "[MET data quality inspection](#)" section, gaze annotation methods in the "[Gaze annotations](#)" section, visualization of looking events in the "[Data visualization](#)" section, and data analysis approaches in the "[Data analysis](#)" section. Methods introduced in these sections are applicable to MET research with adults, children, and infants. Finally, we will discuss remaining challenges and future directions in the "[Future directions](#)" section. In addition to reviewing existing tools, the present paper also provides computer programs and example data for demonstrating methods for data quality assessment, data visualization, and data analysis (https://github.com/xiaoxuefu/MET_methods). The example MET data were collected from two research projects: the iTRAC study that enrolled 5- to 7-year-olds and the ACTION study that involves infants at 4 and 8 months of age. Descriptions of the two projects are provided in the GitHub repository. Table 1 also lists information on open-access MET data and data analytic tools provided by studies cited in the present paper.

Table 1 Open-access data and tools cited

Name	Authors	URL
Mobile eye tracking data		
Gaze-in-Wild	Franchak et al. (2018) Kothari et al. (2020) Matthis et al. (2018)	https://nyu.databrary.org/volume/135 https://www.cis.rit.edu/~rsk3900/gaze-in-wild/ https://doi.org/10.6084/m9.figshare.6130850
Glasses Test	Niehorster et al. (2020)	https://github.com/dnieho/GlassesTestCodeData
Data Quality Assessment		
GlassesValidator	Niehorster et al. (2023)	https://github.com/dnieho/glassesValidator
Gaze Annotations		
GazeCode	Benjamins et al. (2018)	https://github.com/jsbenjamins/gazecode
ROI Coder	Franchak (unpublished)	https://github.com/JohnFranchak/roi_coder
Eye Tracker Analysis	Jongierius et al. (2021)	https://osf.io/4uy35/?view_only=785a011774cf4c4f8c5e4608b34a2a38

The utility of MET technology

MET as a tool to examine cognition embodied in individuals' sensorimotor systems

MET research has provided empirical evidence for embodied cognition (Ballard et al., 1997; Yoshida & Burling, 2011). The ecological approach suggests that visual attention operates in conjunction with a whole-body locomotor system (Gibson, 1979). Research has historically studied attention and active locomotor behavior as two separate, encapsulated systems. MET opens the opportunity for examining the “what” and “when” of visual attention during sequences of actions carried out during everyday activities (Hayhoe, 2017, 2018; Hayhoe & Rothkopf, 2011) or other fieldworks (e.g., fly a plane: Socha et al., 2022; perform a clinical procedure: Wright et al., 2022). MET research in adults reveals the tight spatial and temporal coupling between attention, action, and task demands (Hayhoe et al., 2003; Land et al., 1999). For example, when adults walk on complex terrain, they gaze at the point at which they would place their foot two steps ahead (Domínguez-Zamora & Marigold, 2019; Marigold & Patla, 2007; Matthis & Fajen, 2014), and adjust the timing of fixations to match the difficulty of foot placement. When walking over flat terrain, adults can navigate obstacles without the need to fixate (Franchak & Adolph, 2010). When adults navigate crowds, participants avoid eye contact as instructed by orienting both their heads and eyes towards the floor (Hessels et al., 2022). Hence, the synergistic eye-body coordination is constantly adjusted in real time based on in-the-moment behavioral goals.

MET recordings in infants reveal that the development of gross motor (e.g., posture) and fine motor (e.g., manual object manipulation) skills shape infant attentional behavior. For example, crawling infants (13-month-olds) look mostly at the floor, whereas age-matched infants who walk can see more distal objects and people (Kretch et al., 2014; Luo & Franchak, 2020). Moreover, infants (12-month-olds) look more at their caregivers' faces when upright and sitting, compared to when in a prone position (Franchak et al., 2018). As infants' fine motor skills mature, manual object explorations generate more salient and variable object images in the visual field (15–25 months). Visual inputs in the real world are maintained in infants (12–24 months) by consistently aligning the head and eyes while looking at objects (Borjon et al., 2021). The visual inputs facilitate learning of word-object associations (Bambach et al., 2018; Slone et al., 2019; Yu & Smith, 2012). These studies collectively underscore the importance of studying visual attention within the developing sensorimotor system.

MET as a tool to capture social attention embedded in naturalistic interactions

The “second-person” or “person-centered” perspective emphasizes that social attention needs to be examined in the context of the individuals' interaction with social partners (Fu & Pérez-Edgar, 2019; Pérez-Edgar et al., 2020; Redcay & Schilbach, 2019; Risko et al., 2012). In real-life social interactions, eye gaze serves the dual function of both collecting and communicating information (Gobel et al., 2015; Nasiopoulos et al., 2015). By recording visual attention during real-time social behaviors, MET is a unique tool for understanding the dual function of eye gaze. For example, MET studies found that adults tend to avoid directly looking at strangers when engaging in first-person social interactions, compared to being passive observers (Foulsham et al., 2011; Freeth et al., 2013; Laidlaw et al., 2011). The behavior might be driven by the implicit understanding of social customs and the effect of gaze in delivering social information. However, there are cultural differences in social attention, such that East Asians engaged in more mutual gaze than Western Caucasians during face-to-face conversations (Haensel et al., 2022). Adults also utilize eye gaze as social communicative cues. For example, when verbal instructions for a task activity are ambiguous, participants are more likely to follow the gaze of their social partners compared to when given unambiguous verbal instructions (Macdonald & Tatler, 2013). Hence, social attention is context-driven and goal-directed.

MET is also an indispensable tool for understanding the coupling between attention and affective behavior. Vallorani et al. (2022) showed that among 5- to 7-year-olds, a child's expression of positive affect predicts a greater likelihood of looking at peers during dyadic free play. Social attention, in turn, is linked to a greater likelihood of the child expressing positive affect when the peer is expressing neutral affect. Existing MET studies in adults and children underscore the importance of studying social attention nested in the individuals' affect and social experiences (Fu & Pérez-Edgar, 2019; Pérez-Edgar et al., 2020). One application of measuring social attention embedded in real-life interactions is to study threat-related attention bias linked to risk for internalizing symptoms (Fu & Pérez-Edgar, 2019). Behavioral inhibition, a temperament profile characterized by heightened vigilance and reactivity to novelty in infancy and social reticence in childhood, is a robust risk factor for anxiety disorders (Chronis-Tuscano et al., 2009; Clauss & Blackford, 2012). During a relatively benign social encounter, children (partially overlapping sample as Vallorani et al., 2022) with high behavioral inhibition show greater attention avoidance towards an adult stranger (Fu et al., 2019). Moreover, children with an attention profile characterized by avoidance

to an adult stranger exhibit greater internalizing symptoms even when controlling for their behavioral inhibition level (Gunther et al., 2022). When encountering higher social threat (i.e., an adult wearing a “scary” mask), children with high behavioral inhibition showed more attention toward the stranger (Gunther et al., 2021). Together, these findings highlight the importance of studying threat-related attention in the context of naturalistic interactions, as the nature of threat context can influence attention patterns.

Moreover, developmental scientists have used MET to study how learning emerges from free-flowing interactions in infant–caregiver dyads. Joint attention (JA) is a key conduit for language learning. JA reflects children’s ability to coordinate attention with their social partners, creating a critical context for language acquisition (Suarez-Rivera et al., 2022; Tomasello & Farrar, 1986). Traditional laboratory tasks assess infants’ ability to achieve JA by focusing on visual attention patterns, encompassing face looking, gaze following, and object looking (Brooks & Meltzoff, 2005; Tomasello & Farrar, 1986). In contrast, through studying infant–parent free-flowing play behaviors, MET studies in infants and toddlers (9–48 months) show that it is the hand–eye coordination between infants and caregivers, not infants’ visual attention alone, which contributes to the formation of JA (Abney et al., 2020; Yu & Smith, 2013, 2017a, 2017b; Yurkovic-Harding et al., 2022). Parents are more likely to name and touch the toy during bouts of JA, and the multimodal behavior increases infants’ sustained attention to the objects and facilitates real-time learning of the word–referent association (Chen et al., 2021; Suarez-Rivera et al., 2019; Yu & Smith, 2012). As infants actively interact with the environment through sensorimotor (e.g., hand–eye) coordination, they create idiosyncratic inputs for learning (Smith et al., 2018). Hence, MET provides a tool for understanding the formation and characteristics of the environmental inputs from the first-person perspective, and the downstream impacts of these inputs on cognitive development (Yoshida & Burling, 2011).

MET data collection considerations

Decisions on eye-tracker hardware and MET task procedures are driven by researchers’ requirements regarding (1) participant characteristics, including age, (2) freedom of movement, and (3) data collection environment, such as in controlled laboratory settings or less controlled indoor or outdoor environments (e.g., homes and streets). While the hardware choices and study procedures may vary, a common goal for eye-tracking research is to safeguard data quality, defined as the reliability, validity, and availability of usable data (Hessels & Hooge, 2019; Niehorster et al.,

2018). Disruptions of pupil detection (due to factors such as ambient lighting, headset slippage, and eye makeup) and the alignment of the headset relative to the participant’s head (due to movement and slippage) negatively impact data quality (Hessels et al., 2022; Niehorster et al., 2020).

Calibration is a critical procedure for obtaining high data quality. The commonly used video-based pupil–corneal reflection (P-CR) eye tracker records the relative locations of the pupil and corneal reflection. Calibration involves mapping the recorded pupil and corneal reflection locations when the gaze was directed to the calibration targets to the spatial locations of the calibration stimuli (Blignaut et al., 2014). Poor calibration reduces the validity of MET data. Furthermore, care needs to be taken to ensure that experimental manipulations do not create differential impacts on MET data quality between conditions (Hessels et al., 2022). While calibration-free MET devices are commercially available (e.g., Tonsen et al., 2020), we recommend researchers evaluate different calibration options based on participant age and experiment needs. We include information on calibration here, as gaze-estimation accuracy of the calibration-free MET device is yet to be published for children and infants. This section will discuss hardware setup, calibration, and study design issues in example research scenarios based on study considerations on (1) participant characteristics, (2) freedom of movement, and (3) environment. Additional guidance on MET setups is provided in Valtakari et al. (2021) and Slone et al. (2018) for adult and child participants, respectively.

MET data collection with adults and older children in controlled laboratory environments

Collecting MET data in older participants in controlled environments allows for greater flexibility in hardware setups given the minimal customization required for “out-of-the-box” eye-trackers and participants’ better tolerance and abilities to cooperate (compared to infants and toddlers). One main consideration that determines MET setups is the participants’ freedom of movement. Published studies in participants above 5 years old commonly connect the headset directly to a computer device (e.g., laptop) for data recording and storage (e.g., 5–69 years old: Fu et al., 2019; Hessels et al., 2022; Matthis et al., 2018; Woody et al., 2019). This setup can be burdensome for the participants, and the restrained movement can affect eye–body coordination, a key construct of interest in many MET studies. Newer setups involve connecting the headset to a lightweight smartphone device, which functions as a recording and local storage device (Nasrabadi & Alonso, 2022; Tonsen et al., 2020).

MET research with adults and children who can be instructed to fixate calibration targets has greater flexibility

in calibration methods. In a typical calibration session, participants are asked to look at calibration points displayed on a screen, comparable to screen-based eye tracking (SET) calibration (e.g., Fu et al., 2019; Kothari et al., 2020) or a calibration marker fixed on a naturalistic object (e.g., Niehorster et al., 2020; Woody et al., 2019). Studies in this age range may employ online calibration, where the mapping between pupil-corneal-reflection locations and the locations of the calibration points take effect immediately. Offline calibration performs the spatial mapping after data collection. Hence, the advantage of online calibration is allowing for real-time data monitoring and providing the opportunity for just-in-time recalibration.

We recommend a few best practices for performing both online and offline calibration accuracy:

- 1) **Display calibration targets at a distance comparable to the distance between the participant and primary areas of interest (AOIs).** AOIs refer to the targets of the participants' looks that will be annotated for data analysis (also see the section "[Gaze annotations](#)"). Parallax error is a gaze estimation error introduced when the distance between the wearer and the AOI (i.e., the fixation plane) is different than the distance between the wearer and the calibration target (i.e., the calibration plane). This causes an offset between the true gaze location and the estimated gaze location on the fixation plane and the scene camera coordinate space, which bias the experimenter's identification of actual gaze location (Mardanbegi & Hansen, 2012; Valtakari et al., 2021). Hence, it is recommended that the calibration targets be presented at the same approximate location of the AOIs. If participants will engage in various viewing distances during the experiment, it is best to perform multiple calibration sessions to accommodate the distance changes or employ a mid-range distance if there is a lack of experimental control on the distance change.
- 2) **Present multiple calibration targets that cover the participant's entire field of view (FOV).** We use FOV here to refer to the view of the participants captured by the scene camera. The FOV tends to be smaller than the participant's visual field, and it is not necessarily equivalent to the reported FOV specifications of a given MET model, depending on factors such as viewing distance, participant's posture, and the camera angle. Five or more calibration targets can be presented across the participant's FOV, comparable to SET. This step is to ensure that calibration accuracy is maintained from the center of the FOV to peripheral locations. The experiment should verify that participants do not turn their heads to orient toward peripheral targets, which will result in target clustering in the center of the FOV.
- 3) **Perform a validation procedure (i.e., calibration check) at the beginning and end of the experiment session, and after any MET headset movement.** A validation procedure is conducted by directing participants to look at specific target locations. Similarly, the targets should be presented in a location that is comparable to the location of the AOIs. Conducting multiple calibration checks during the experiment helps to ensure that the data quality is maintained throughout the experiment. The headset slippage issue can be effectively prevented through monitoring online calibration accuracy and recalibrating to correct accuracy drift (Niehorster et al., 2020). The validation procedure also provides additional calibration points for corrections in offline calibrations. For example, if the eye gaze capture is perturbed by headset movement, the experimenter needs to adjust the eye camera and perform calibration checks. The points of gaze obtained post-adjustment can be used to update the spatial mapping in offline calibration. Finally, performing per-participant validation checks allows reporting of the accuracy metric (see the section "[Accuracy](#)"), which can also be used as a control variable in analyses (Franchak & Yu, 2022).

MET data collection in infants and toddlers

Existing MET studies with infants and toddlers (4–26 months old) in laboratory (e.g., Franchak et al., 2011; Schroer & Yu, 2023; Yu & Smith, 2012, 2017a) and home settings (Bradshaw et al., 2023) have largely followed a common set of equipment setup and calibration procedures. The headset needs to be stably placed on the head to minimize the negative effect of slippage on data quality (Niehorster et al., 2020). Researchers may customize the "out-of-the-box" eye tracker by affixing it on a tailored headband, cap, or beanie for secure placement. For young infants (< 8 months), we recommend utilizing a series of headsets that can accommodate different head sizes, head shapes, and hair textures. Some headsets can be connected to a smartphone to increase children's mobility (Schroer & Yu, 2023).

It is challenging to instruct infants and toddlers to follow calibration points. Hence, MET studies in this age group commonly implement offline calibration. The calibration procedure can be integrated into a child-experimenter play session during which the experimenter presents engaging calibration targets (e.g., toys and/or laser points) at various locations across the child's FOV. The calibration target distance from the child and the child's posture should match the specifications for the formal data collection. Researchers should closely monitor the eye image recording throughout data collection. Additional calibration(s) are required if the eye image capture is perturbed by headset movement. If the study involves interactions with an adult partner, such as a

caregiver, the social partner can be trained in calibration target presentations to minimize disruptions during naturalistic interactions. After data collection, a trained researcher marks the calibration target locations on the scene camera recording where the child's point of gaze is clearly identifiable and directed to the calibration target. An algorithm is then applied to map the pupil and corneal reflection locations with the specified calibration target locations (e.g., Hassoumi et al., 2019). The manual identification of points of gaze and automated mapping procedure are run iteratively to establish satisfactory calibration (Slone et al., 2018).

MET data collection outside controlled laboratory environments

MET studies have been conducted in naturalistic outdoor (adults: Foulsham et al., 2011; Matthis et al., 2018) and indoor environments (infants at homes: Bradshaw et al., 2023; adults in an event hall: Hessels et al., 2022; a child in a museum: Jung et al., 2018; adults in a clinical setting: Wright et al., 2022). Factors that can then compromise MET data quality include a lack of control over ambient lighting, locations of the target objects (i.e., AOIs), and insufficient calibration procedures (Evans et al., 2012; Hessels et al., 2022). For example, infrared light from the sun when outdoors interferes with pupil and corneal reflection tracking. A remedy is to provide participants with an infrared-blocking visor (Matthis et al., 2018). The distance between the participant and different AOIs can greatly vary. This is both an advantage (greater visual selection) and a disadvantage (greater analytic complexity) of MET. In tasks where participants tilt their heads down, AOIs that are close to the participant and lower than eye level are captured in the lower part of the scene camera view, while farther objects are captured in the higher part of the scene camera view (Slone et al., 2018). Hence, appropriate scene camera positioning needs to be determined to ensure it can capture all AOIs in the study.

Offline calibration can be advantageous in less-controlled environments, given real-time data monitoring might not be possible. Offline calibration offers researchers the opportunity to update the spatial mapping between pupil and corneal reflection locations and the points of gaze directed to the calibration targets after pupil capture is altered by slippage, posture changes, or lighting. The emerging calibration-free MET technology is also promising for maintaining acceptable accuracy in outdoor settings (Tonsen et al., 2020).

MET data quality inspection

Eye-tracking data quality is quantified by accuracy, precision, and availability of usable data (or data loss) (Hessels & Hooge, 2019). Accuracy is operationalized as the distance

(spatial offset) between the gaze location detected by the eye tracker and the actual gaze location measured in degree of visual angle. Precision indexes the level of noise in the eye-tracking data that produces spatial variability between gaze samples. Accuracy and precision provide an index of validity and reliability of the eye-tracking data (Hessels & Hooge, 2019). Data loss can be calculated using the number of valid gaze data points recorded and the expected number of samples based on the specified sampling frequency (Hessels & Hooge, 2019; Hooge et al., 2023; Niehorster et al., 2020).

As illustrated in the "MET data collection considerations" section, there are MET-specific data quality concerns relative to SET studies due to the less constrained nature of MET data collection. After the initial calibration, changes in illumination conditions, the distance between the participant and AOIs, and disruptions in the detection of pupil locations and corneal reflection can introduce errors (Franchak & Yu, 2022; Niehorster et al., 2020). Indeed, the accuracy and precision achieved from the initial calibration may not be maintained at the end of the experiment after unconstrained movement and headset slippage occurred (Niehorster et al., 2020; Santini et al., 2018). Compromised data quality could bias eye-tracking measurements and lead to false conclusions (Wass et al., 2014). Thus, it is critical to examine data quality before data analyses. This section will provide strategies for assessing MET data accuracy, precision, and data loss based on published definitions (Franchak & Yu, 2022; Hessels & Hooge, 2019; Niehorster et al., 2020). We will also discuss strategies to make informed decisions on data analyses based on data quality assessments. An additional pipeline for computing these data quality indices is provided in (Hooge et al., 2023). The authors underscored the importance of inspecting synchronization between the eye and scene camera recordings before computing the data quality indices. While some MET systems provide a built-in function for synchronization, there could still be intermittent periods of asynchronization between the recordings of gaze and target location, which will bias the accuracy index (Hooge et al., 2023).

Accuracy

One off-the-shelf tool for calculating accuracy is *Glasses-Validator* (Niehorster et al., 2023; Table 1). *GlassesValidator* is suited for data collection with adults and older children, as participants are required to look at fixation targets displayed on the poster that is included with the tool. The computation is automated and does not require manual annotations. Briefly, the poster contains arrays of ArUco markers (i.e., barcodes) that allow automated estimation of the participant's viewing distance and gaze location. A fixation classifier is applied to determine the valid fixation (> 50-ms duration) towards each fixation target. Accuracy

is calculated as the deviation between the fixation target and the estimated gaze location in degree of visual angle (i.e., the angle between the line from the eye to the fixation target and the line from the eye to the gaze location).

We provide an additional tool for calculating the spatial offset (in degree of visual angle) between the gaze location and the validation target that the participant is directed to look at. The tool can be applied to validation recordings

obtained from adults or older children in the "MET data collection with adults and older children in controlled laboratory environments" section and calibrated recordings using offline calibration when online calibration is not possible in the "MET data collection in infants and toddlers" and "MET data collection outside controlled laboratory environments" sections. We provide both a MATLAB (https://github.com/xiaoxuefu/MET_methods/tree/main/1.%20Accuracy) and an

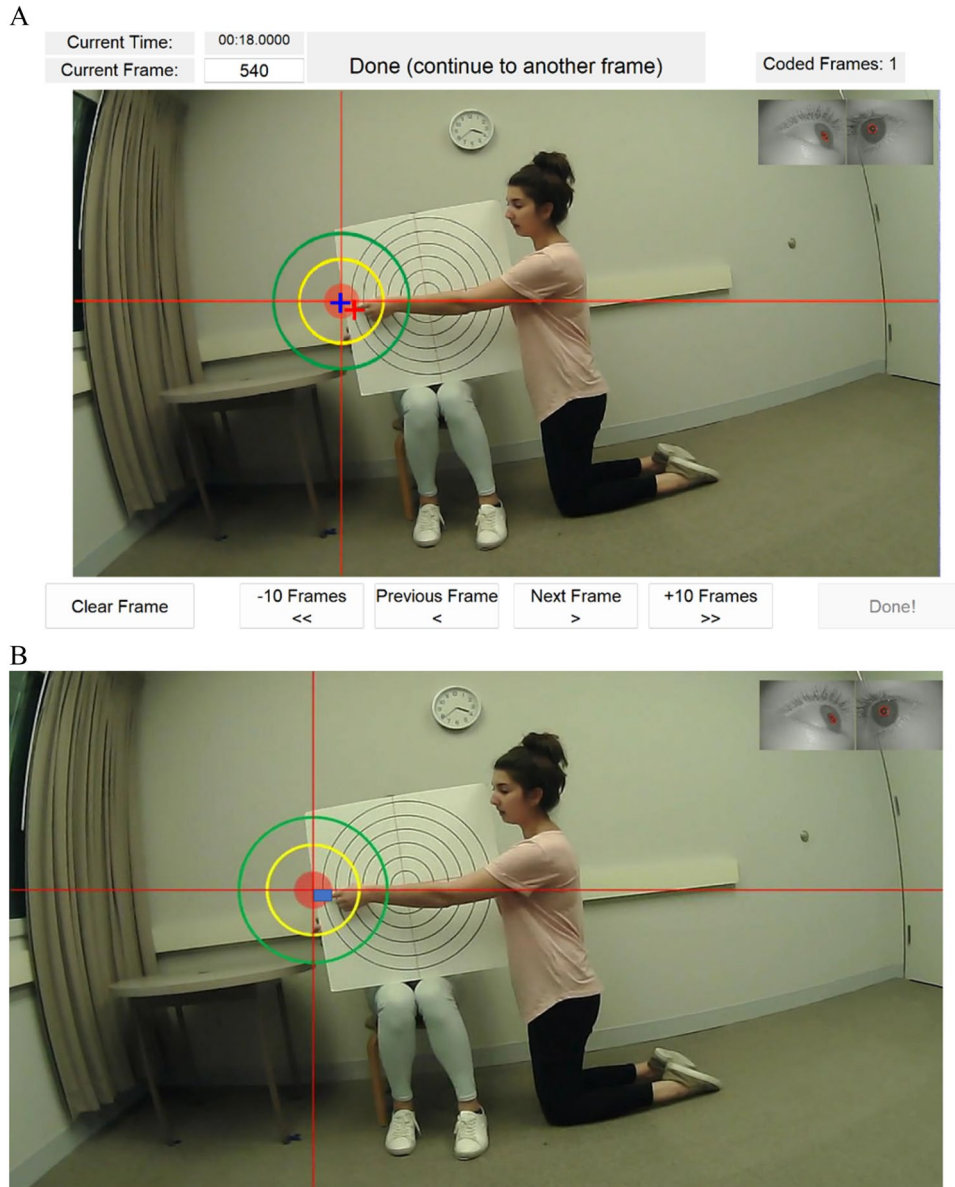


Fig. 1 Screenshots of the graphical user interfaces (GUIs) for accuracy calculation. Both the MATLAB tool (**A**) and R Shiny app (**B**) enable users to annotate the point of gaze (i.e., crosshair) and the target location (i.e., where the experimenter is pointing). The data are then used to compute the spatial offset (in degree of visual angle) between the gaze location and the target location. **A** The MATLAB tool takes video inputs. In this example, 25 frames from the validation session were used for the calculation of accuracy. Based on the

specifications of Pupil Core eye tracker: fov_x (horizontal field of view in degrees) = 82.1, fov_y (vertical field of view in degrees) = 52.2, fov_res_x (horizontal resolution in pixels) = 1280, fov_res_y (vertical resolution in pixels) = 720. The average spatial offset across these frames is 0.766°. **B** The R Shiny app uses frames extracted from the video recordings. The blue box (edited) represents the region between the gaze and the target location. Detailed instructions are provided on https://github.com/JohnFranchak/et_accuracy

R Shiny app version of the tool (<https://john-franchak.shinyapps.io/Eye-Tracking-Accuracy-Calculator/>). The spatial offset computation method is based on the definition described in Franchak and Yu (2022). Figure 1 displays the MATLAB graphical user interface (GUI) and R Shiny app for obtaining the spatial offset. The user can annotate target and gaze locations in the MATLAB GUI or the R Shiny app. Both versions of the tool will compute the spatial offset for each frame based on user-specified target and gaze locations, the scene camera FOV and resolution specifications provided by the manufacturer. A lower spatial offset indicates better accuracy.

Participant-specific accuracy values are recommended for use in scientific reports and making analytical decisions. Participant-specific accuracy is likely to be worse (i.e., larger spatial offset) than manufacturer-reported values (Franchak & Yu, 2022; Niehorster et al., 2020; Santini et al., 2018). Participant-specific accuracy is used to evaluate whether it is valid to determine looking towards AOIs specified for the study. The experimenter should determine the accuracy required to distinguish looking between AOIs. When viewing AOIs at a comparable distance as the validation target, the radius of the AOI, or the distance between two AOIs, should not be smaller than the distance between the target location and the actual point of gaze measured during validation. If the participant-specific accuracy is lower than required, data from the participant may be excluded or the AOI(s) may be adjusted for the participant. For example, the experimenter will determine looking to the person rather than the person's face for participants with lower accuracy.

Precision

Precision can also be operationalized as the root mean squared error (RMSE) of sample-to-sample deviation when a participant is assumed to fixate at the same location (Hessels & Hooge, 2019; Niehorster et al., 2020). A larger RMSE value indicates higher sample-to-sample deviation, thus lower precision (Hessels & Hooge, 2019; Niehorster et al., 2020). *GlassesValidator* (Niehorster et al., 2023) provides precision indices using gaze points that are directed at the fixation targets on the poster provided with the tool. We provide a MATLAB program (https://github.com/xiaoxuefu/MET_methods/tree/main/2.%20Precision) for calculating sample-to-sample RMSE based on the published definition (Hessels & Hooge, 2019; Niehorster et al., 2020). The expected input data are x - and y -coordinates of gaze points when the participants were instructed to look at the same location (i.e., a target object), such as during a calibration procedure. User-input parameters are the scene camera specifications, the size of the target object, and the distance between the participant and the target object.

Less precise gaze data bias the parcellation of fixations and saccades, as they may erroneously suggest shifts in gaze locations when in fact the gaze remains stable (Wass et al., 2013, 2014). To minimize the impact of low precision, larger AOIs can be defined to allow for more error margins. In addition, data analysis can be less dependent on fixation or saccade categorization by computing the duration of continuous looking toward an AOI (further discussed in the "Gaze annotations").

Data loss

Data loss can be computed the proportion of data loss (the total amount of valid data points expected to be sampled based on the sampling frequency of the MET device minus the number of valid data points collected) over the total amount of expected data points (Niehorster et al., 2020). Data loss can occur when the eye tracker fails to detect the corneal reflection or pupil (Wass et al., 2014). This can be caused by blinking, lighting, eye camera being moved out of alignment, or other eye tracker technical errors. Hence, with more data loss, shorter durations of AOI looking could be caused by MET failing to detect eye gazes, rather than the participant not looking at the AOI (Wass et al., 2014). Hence, to accurately quantify AOI looking, it is important to measure the amount of both valid and invalid MET data. The amount of AOI looking can then be indexed as the proportion of time looking at the AOI over the total amount of valid MET data recorded (rather than the total recording duration).

Gaze annotations

Automated annotations

One challenge in processing MET data is fixation classification. During MET data collection, the participant moves, the AOI moves, or both move in a three-dimensional space. Hence, it is challenging to classify different types of gaze events, including fixations, saccade, and gaze pursuit. For example, during a bout of fixation, the AOI being foveated moves when the participant's head moves. Classifier algorithms are available for automatic fixation detection (GazeCode: Benjamins et al., 2018; Kothari et al., 2020; Table 1). While the classifiers yielded substantial agreement with human coders, it remains challenging to accurately classify gaze pursuit, defined as the tracking of an AOI moving across the scene camera view (Kothari et al., 2020). However, differentiating fixations from other gaze events or counting the number of fixations might not be the key aims of most MET studies. Depending on the research questions, it might be sufficient to measure the proportion

of time, or frames during which, the gaze was directed to an AOI (Franchak & Yu, 2022). This dependent variable can be computed using either manual annotations by human coders or automated classifiers.

Another challenge in MET gaze annotation is identifying the AOI being foveated (Brône et al., 2011). The AOI coordinates need to be defined in the participant-specific egocentric space. They also need to be defined frame-by-frame as the AOI's appearance can change due to motion, viewing perspectives, and occlusion. Researchers have traditionally conducted manual AOI annotations (e.g., Franchak & Adolph, 2010; Franchak et al., 2011). The development of open-source deep learning algorithms has made it possible to automate AOI identification. Off-the-shelf computer vision algorithms enable automated detection of human faces and bodies in the scene camera view. Once the AOIs (e.g., bounding boxes for faces) are specified, an additional procedure is applied to map the gaze locations (synchronized with the scene camera recordings) to the AOIs (Duchowski et al., 2019; Gehrer et al., 2020; Hessels et al., 2022; Hessels, Benjamins et al., 2020a; Jongerius et al., 2021). Jongerius et al. (2021) found high agreement (Cohen's kappa \geq .89) between automated annotation of face looking using *OpenPose* (Cao et al., 2017) and manual annotations by trained coders.

However, computer vision AOI detection can be more challenging for addressing certain research goals than others. A common application has been detecting faces in MET recordings collected during laboratory-controlled face-to-face interactions (Duchowski et al., 2019; Gehrer et al., 2020; Haensel et al., 2022; Hessels, Benjamins et al., 2020a; Jongerius et al., 2021). In contrast, Long et al. (2022) applied *OpenPose* to head-mounted camera recordings obtained from infants during parent–infant free play of toys to detect parents' wrists as an index of hand presence, as hands are often occluded by the toys. They found more misses in detecting the presence of hands than faces. *OpenPose* detection of human figures is more challenging during unrestrained locomotion when the distance between the wearer and the AOIs varies moment-to-moment (Hessels, Benjamins et al., 2020a). Furthermore, additional training on deep learning models using manually annotated data is required when the AOIs are novel and/or complex objects (e.g., toys; Bambach et al., 2016). Together, automated AOI annotations are faster and can be more objective than manual annotations (Jongerius et al., 2021). The increased data processing capacity can advance our knowledge about the characteristics of visual inputs in the natural environment (Smith & Slone, 2017). However, off-the-shelf computer vision algorithms might not be applicable to all detection tasks. They are also not error-free. Depending on the task requirement and error tolerance, manual annotations might still be necessary

for providing training datasets (Bambach et al., 2018) or to complement the automated detection (Haensel et al., 2022).

Manual annotations

Manual annotations of AOI looking remains the most accessible and robust method for data generation especially for developmental MET applications (Franchak & Yu, 2022). Manual AOI annotations are flexible. As discussed above, it might be a necessary procedure for annotating complex and irregular AOIs. Manual annotations can also be applied to additional events and behaviors that take place simultaneously. Manual annotations are accessible, as it can be carried out in any open-source annotation software, including *Datavyu* (Datavyu, 2014), *ELAN* (ELAN, 2018), and *BORIS* (Friard & Gamba, 2016). Indeed, manual annotations have been widely implemented in studies using a variety of MET systems with both adult (e.g., Laidlaw et al., 2011; Rogers et al., 2018) and child samples (e.g., Franchak et al., 2011; Fu et al., 2019; Woody et al., 2019).

There are two approaches to manual annotations of AOI looking events. One method is manually annotating AOI looking events based on the gaze overlay video entered into annotation software. Researchers may implement a cut-off duration to exclude short looks. For example, an event of continuous looking is conventionally defined as looking at the AOI for two or three successive video frames at 30 Hz, a duration of 66.7 to 99.9 ms (e.g., Franchak & Adolph, 2010; Franchak et al., 2011). The second approach is to apply fixation classifier algorithms to segment the gaze overlay videos into frames based on the detection of stable gazes. Then trained human coders annotate the AOI(s) being looked at in the video segments (e.g., Yurkovic-Harding et al., 2022; Yurkovic et al., 2021). We provide a MATLAB-based *ROI coder* program (https://github.com/JohnFranchak/roi_coder) that aids manual AOI annotations. The computer-program-guided approach can help reduce coders' cognitive effort and thus reduce human error.

A well-designed gaze annotation manual helps to ease the burden of manual annotations, reduce human errors and biases, and enhance inter-coder reliability. The manual contains descriptions of each code and instructions for the coders on how to score the looking behavior and any additional event of interest. For the looking behavior code, the manual defines the AOI codes (e.g., "b" = body looking) and provides instructions for annotating the onset and offset time of a bout of continuous look to the AOI. Adding to existing guidance (Franchak & Yu, 2022; Slone et al., 2018), we provide best practices for manual annotation of AOI looking events. Best practices for annotating general behavioral data can be accessed at <https://datavyu.org/user-guide/best-practices.html>.

- 1) **Create visual aids for manual annotations.** Researchers can superimpose a bullseye on the gaze overlay video to indicate gaze location in addition to the crosshair that presents the point of gaze. The size of the circle can be set based on the tolerance of accuracy for manual annotations. We provide a MATLAB program (https://github.com/xiaoxuefu/MET_methods/tree/main/3.%20Gaze%20Coding%20Error%20Tolerance) for estimating the visual angle of circles in the bullseye (i.e., error tolerance). An example of error margin setting is provided in Fig. 2 (also see Franchak & Yu, 2022 Fig. 4B). Additionally, the gaze overlay video must be synchronized with additional sources of video recordings, such as room cameras. The composite video displays the participant's behavior from multiple angles and perspectives, thus allowing coders to use contextual information to determine gaze shifts and locations (Slone et al., 2018).
- 2) **Downsize data to code based on research questions.** Manual annotations can be selective, given the large volume of MET data collected and the time-intensive processing of manual annotations. AOI looking events can be coded only during *events of interest*, instead of the entire recording (e.g., Franchak & Adolph, 2010). Another data-reduction method is to down-sample video frames. The typical sampling frequency of the scene camera is from 30 to 120Hz. For example, a 5-min recording could provide a range of 9000 to 36,000 frames to code. Based on initial inspections of the data, researchers can choose to resample the recordings to a lower frequency if AOIs are relatively big and sparse and gaze shifts within AOIs are not a primary interest of annotations. Researchers can code short segments from several participants at both resampled and the original frequency to make sure that the reduced frame rate does not bias the percentage scores of looking durations.
- 3) **Annotate valid and invalid AOI looking events and data.** Researchers should compute proportion scores of accumulated AOI looking durations, with the total valid AOI looking duration as the denominator. This strategy is to reduce biases produced by data loss (discussed in the "Data loss" section) and allow for comparisons of accumulated looking durations across AOIs. Thus, in addition to annotating valid AOI looking events (i.e., continuous looking exceeding a threshold duration), coders should annotate frames with invalid AOI looking (e.g., looking duration below the threshold) and data loss (i.e., no visible point of gaze).
- 4) **Code-check-revise-check.** Manual annotations are an iterative process. After a preliminary annotation plan is conceived, researchers should conduct test annotations of representative recording segments from different participants. This is to make sure that the data generated can address research questions and that satisfactory inter-rater reliability can be easily achieved. Researchers can then go back to revise the annotation methods before annotating the entire recordings. After formal annotation

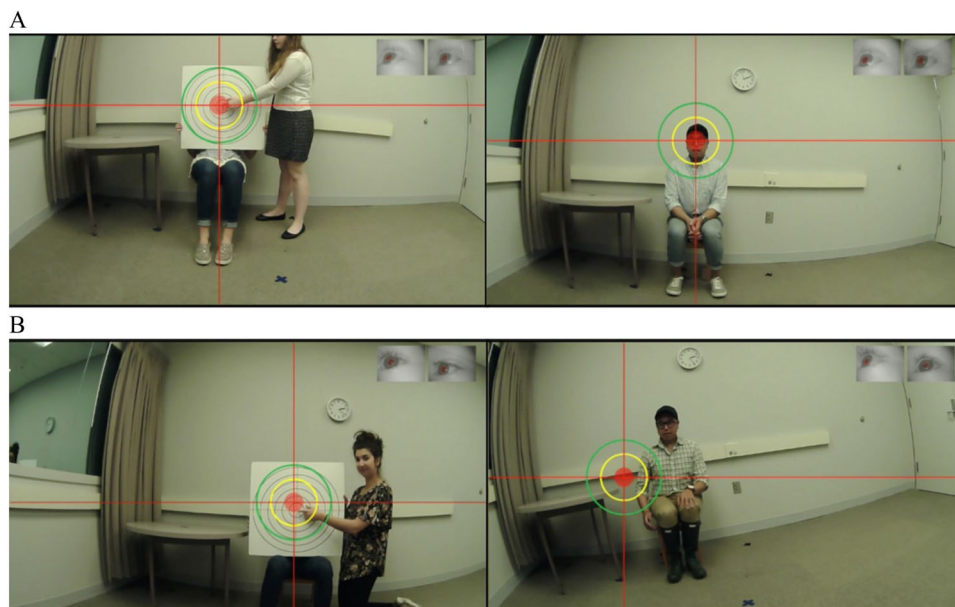


Fig. 2 Video frames taken from the validation (*left*) and task (*right*) procedures. Calibration accuracy needs to be estimated before gaze annotations. Gaze annotations that are based on the *red circle* allow reliable determination of the area of interest (AOI) for error within

2.6°. For example, looking to the researcher would be identified for **A** but not **B**. The *yellow circle* allows for an error tolerance of 6.6°. In such case, looking to the researcher would be annotated for **B**

protocol is launched, researchers should periodically check reliability to detect and resolve significant discrepancies between coders. Percentages of inter-coder agreement and kappa values need to be calculated for reliability assessment and scientific reports.

Data visualization

The gaze annotation step produces a long-format dataset that contains time series of annotated looking events. For example, each row may contain information, such as onset time, offset time, and the AOI. In addition to AOI looking events, Researchers may have coded other events, such as motor activities recorded from the room camera. The multiple types of events, which may be generated from the same individual but different modalities (e.g., looking events and motor behavior), and/or from the same modality but different individuals (e.g., looking events from two individuals during a dyadic interaction). Data visualization is a critical step for exploring the temporal characteristics of AOI looking events from an individual and/or the temporal relations between two or more data streams (e.g., time series of AOI looking and motor behavior). This section will demonstrate the methods and utility of data visualization. The example data and the programs for producing the visualization are shared with the paper.

Example 1 (iTRAC): Visualize individuals' looking behavior nested in dyadic interactions

Visualizations inform the temporal dynamics of a child's looking behavior nested in dyadic interactions. Figure 3 (https://github.com/xiaoxuefu/MET_methods/tree/main/4.%20Visualization/Figure3) presents data collected from a parent–child dyad as they completed a series of challenging tangram puzzles (MacNeill et al., 2022). The visualization explores the child's gaze patterns as the parent displays various types of parenting behavior (characterized as positive reinforcement, teaching, directives, and intrusion). Figure 3A is plotted using the MATLAB toolbox *timevp* (<https://github.com/xiaoxuefu/timevp>; Yu et al., 2012) to show how an individual child's AOI looking events and parenting behavior co-evolve during the task. It shows that there are more teaching behaviors at the beginning of the puzzle task for the dyad. As the time pressure increased as part of the task design, there are more directives and positive reinforcement towards the end of the task. Bouts of looking to the parent become shorter in the second half of the task.

State space grids (SSGs) provide a tool to display how dyadic behaviors vary over time by plotting how members

of the dyad move within a figurative space (Hollenstein, 2013; Lewis et al., 1999). Tutorials for using GridWare (<https://www.queensu.ca/psychology/adolescent-dynamics-lab/state-space-grids>; Lamey et al., 2004) are provided in Hollenstein (2013). Figure 3B demonstrates the utility of SSGs for depicting the temporal dynamics between child looking behavior and parenting behavior. The AOI categories of the child looking events and the types of parenting behaviors form a 5×4 grid (i.e., 20 possible dyadic states). We examined the dyad attractor patterns, or states that pull the dyadic system from other states under particular conditions (Thelen & Smith, 1998). GridWare can be used to identify attractors by calculating the average mean duration for a predefined grid sequence, or the average of individual cell means of interest. We characterized attractor strength in parent-focused/controlling parenting states (i.e., the child is looking at the parent while the parent is engaging in directive and intrusive behaviors). In the example, the dyad spent 26.6% of the time, for a total of 37.49 s, in the parent-focused/controlling-parenting states (highlighted in yellow). The average mean duration in these states is 0.85 seconds. Additionally, SSGs help visualize and quantify the patterns of temporal sequence and transition across states (Hollenstein et al., 2004). The level of transition across states, or dyadic flexibility in this example, is indexed by the number of cells visited, the number of transitions, dispersion (0 to 1), and transitional entropy (Lewis et al., 1999), with higher values indicating higher flexibility. The example dyad visited nine cells, made 186 transitions across cells, had a dispersion of 0.83, and an entropy value of 42.22 from looking at the puzzle as the parent engaged in teaching to looking to the parent's reference as the parent engaged in directive and intrusive behaviors.

Example 2 (ACTION): Visualize the coordination of multimodal behaviors in triadic parent–infant–object interactions

Visualization helps generate higher-order constructs that are defined based on the temporal relations of two or more event types. An example of such construct is joint attention (JA), the ability to coordinate attention with a social partner to an object or event of interest (Tomasello & Farrar, 1986). JA can be measured as the temporal alignment when two individuals are looking at the same object during triadic interactions (i.e., child–parent toy play). Visualizing moment-to-moment temporal relations between looking and bodily behaviors in the dyads over the course of interaction helps (1) identify the occurrence of JA and (2) inform the emergence and impact of JA in real time as the interaction unfolds (Yu & Smith, 2013, 2016, 2017a, 2017b; Yu et al., 2019). Figure 4 (https://github.com/xiaoxuefu/MET_methods/tree/main/4.%20Visualization/Figure4) displays

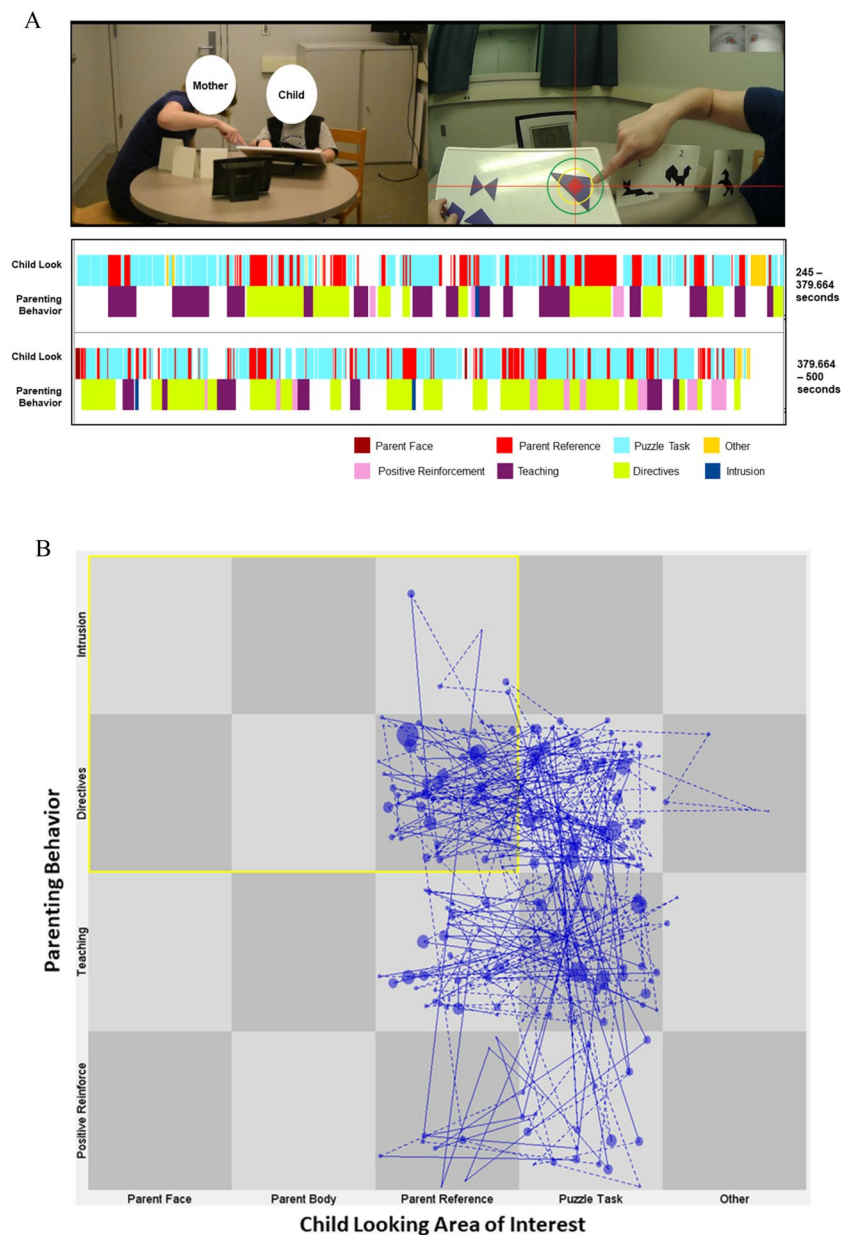


Fig. 3 Child–mother dyadic looking behavior nested in parenting behavior. Data are collected from a mother–child dyad as they completed a challenging puzzle task. **A** Screenshot of the composite video used for gaze annotation is displayed in the top panel. The child’s area of interest (AOI) looking events and parent behavior are plotted using the *timevp* MATLAB toolbox. *White gaps* in behavior represent missing child looking behavior (e.g., indeterminate looking or data loss) or parenting behavior (e.g., comforting) that occurred but are

not of interest. **B** State space grids (SSGs) depicting the child’s AOI looking event and parent behavior. Each node represents the event when both AOI looking and target parent behavior are co-occurring. The size of the nodes represents the length of time spent in each state. *Lines* between nodes denote changes from one dyadic state to the next. *Dotted lines* connect event nodes prior to missing events to nodes that follow the missing events. The *yellow box* labels the parent-focused/controlling-parenting state

a representative segment of the MET data stream from an 8-month-old. The *timevp* toolbox is used to plot events of interest. The top two rows display raw gaze annotation data of AOI looking events from the infant and his mother. Consistent with existing findings, the infant rarely looked at the social partner (and did not look at the face) during

toy play compared to the parent (e.g., Abney et al., 2020; Yu & Smith, 2017a). The third row presents bouts of JA of the toys. For data exploration, we include shorter bouts of JA (between 0.3 and 0.5 s) than Yu and Smith (2017a) considering that the dyad is given a larger variety of toys to play with than the more controlled laboratory setting. The



Fig. 4 Coordination of looking behavior and manual manipulation of toys between an infant–mother dyad during a segment of free-flowing toy play session. A screenshot of the composite video frame used for gaze annotation is displayed in the top panel. Data were plotted using the *timevp* MATLAB toolbox. The areas of interest (AOIs) for looking behavior are body, face, infant, and mother self-looking and four toy objects. The AOIs for manual manipulation are the toys. *White spaces* denote events that do not involve the AOIs. The first two rows

depict raw data from gaze annotation of AOI looking events in the infant and his mother. The third row displays bouts of joint attention (JA), defined as a period (> 0.3 s) when the dyad is looking at the same toy. When the individual is looking at the same toy as the social partner, gaps that last fewer than 0.3 s in consecutive looks of the same AOI are disregarded (e.g., the first bout of toy looking displayed in teal). The last four rows show the four types of attention-manual coordination within each individual and across the dyad

last four rows represent the four combinations of with- and across-individual attention-motor coordination. The three vertical boxes highlight example CVA bouts that emerged when either the partner was holding the toy or both. Consistent with published MET findings (Yu & Smith, 2013, 2017a, 2017b), the figure shows that JA bouts emerge in the context of infant–parent attention-motor coordination.

Data analysis

Statistical analysis with aggregated scores

Informed by data visualization, looking event data can be aggregated by AOI and task condition for each participant for subsequent data analysis. Examples of these aggregated,

high-order measures include those computed based on the individual’s AOI looking events, such as the number of AOI looks (e.g., Fu et al., 2019; Woody et al., 2019), the temporal characteristics of an individual’s AOI looking events, such as sustained attention, defined as AOI looks that are longer than 3 s (e.g., Yu et al., 2019), and the temporal relations of two individuals’ AOI looking events, such as JA (e.g., Yu et al., 2019). Statistical analysis methods, such as Pearson correlation, linear regression, analysis of variance (ANOVA), and linear mixed effects modeling, can be applied to the aggregated measures. The distribution of the looking behavior measure needs to be carefully inspected so that appropriate data transformation and statistical modeling methods can be selected for non-normally distributed outcome variables.

Several published studies have used analytical strategies using aggregated measures computed based on data

visualization (e.g., Abney et al., 2020; Fu et al., 2019; MacNeill et al., 2022; Suarez-Rivera et al., 2019; Woody et al., 2019; Yu & Smith, 2016). For example, using SSGs, MacNeill et al. (2022) examined dyadic states based on co-occurrence of child looking at specific AOIs and designated parenting behavior types (e.g., Fig. 4). Dyadic states were combined to generate two types of attractors: task-focused/positive parenting states and parent-focused/controlling parenting states. Attractor strength, the average mean duration that the child–parent dyad visited each of the two states, is computed. To account for the positive skewness of the attractor strength measures, a generalized linear model with gamma distribution and a log link (Breen, 1996) is fitted to test whether child age, behavioral inhibition, and parent anxiety symptoms predicted the attractor strengths. The results reveal that child age and parent anxiety levels jointly predicted parent-focused/controlling parenting attractor strength.

In another example, Suarez-Rivera et al. (2019) examined the impacts of parent speech and parent manual object manipulation during bouts of JA on infant sustained attention towards the objects during toy play. For each infant, mean proportion scores are computed on infant looking events that fall into five categories defined based on the temporal alignment of multimodal measures: infant looking to the toy without JA, JA with no additional parent behaviors, JA with parent touch of the toy, JA with parent speech, and JA with both parent touch and talk. The resulting mean proportion scores of infant looks are log-transformed to account for the positive skewness. A linear mixed effect model is fitted with the transformed scores as the outcome, event categories as the fixed effect, and random intercepts specified to account for individual differences in the durations of the looking events. The results indicate that infants' looking to the objects is longest (i.e., greater sustained attention) during bouts of JA that include both parent touch and speech. The aggregated summary scores effectively characterize important behaviors in individuals or dyads. However, analyses with aggregated measures may obscure within-subjects temporal effects that describe how looking behavior changes over time.

Statistical analysis to model the temporal dynamics of looking events

MET produces a high-density repeated sampling of gaze locations over a prolonged period of data collection, providing a unique opportunity for examining the temporal dynamics of looking behavior. The location and duration of looking behavior change over time within an individual in response to internal and/or external influences. The intensive longitudinal data analysis (Bolger & Laurenceau, 2013) and dynamic systems modeling (Ram & Gerstorff, 2009)

approaches provide statistical tools for understanding the patterns and dynamics of intraindividual changes in micro- (e.g., in seconds) and macro-timescales (e.g., in years), investigating factors that modulate the temporal dynamics, and characterize groups of individuals based on the trajectories of changes. The modeling methods have been widely implemented using behavioral observation and self-report data (e.g., Benson et al., 2019; Cole et al., 2020; Morales et al., 2018; Shewark et al., 2020), whereas applications to MET data are limited. However, there is increasing emphasis on a spline-based approach (e.g., Li et al., 2015) to model moment-to-moment nonlinear time-varying effects on AOI looking events (Yamashiro et al., 2019).

Emerging MET studies have modeled interindividual differences in within-subjects temporal trajectories of AOI looking behavior. For example, Gunther et al. (2021) modeled second-by-second changes in looking behavior towards a stranger wearing a gorilla mask in 5- to 7-year-olds. Figure 5 (https://github.com/xiaoxuefu/MET_methods/tree/main/5.%20Data%20Analysis%20-%20Growth%20Model) shows that the looking behavior is characterized by a quadratic trajectory (i.e., inverted U-shape) over the period of exposure. Moreover, Gunther et al. (2021) found a main effect of child behavioral inhibition. As time elapsed while the stranger had the mask on, higher levels of behavioral inhibition were related to a greater proportion of looking toward the stranger. Hence, individual differences in the temperament type shape how looking behavior unfolds over time. Furthermore, in an overlapping sample, Gunther et al. (2022) characterize latent profiles of children based on time-varying trajectories of looking behavior towards a stranger. The stranger pretended to do paperwork without initiating interaction with the child, while also holding the marbles that the child needed to play a game. Similarly, children's looking behavior exhibits quadratic trajectories over time. Group-based trajectory models (GBTM, Nagin & Odgers, 2010) are fitted to identify latent profiles underlying individual quadratic trajectories. The results indicate that 30.2% of children belong to the “orienting” group, characterized by high initial orienting to the stranger and gradual decay. The rest of the sample is categorized as the “avoidant” group who displays low initial orienting to the stranger and continued low attention. Importantly, individuals' probability of being characterized with the “avoidant” trajectory predicts variance of internalizing symptoms over and above the aggregated measure of looking towards the stranger. Together, modeling temporal dynamics of looking events may reveal important insights about underlying mechanisms (Cole et al., 2020) and enables better characterization of individual differences (Gunther et al., 2022; Shewark et al., 2020).

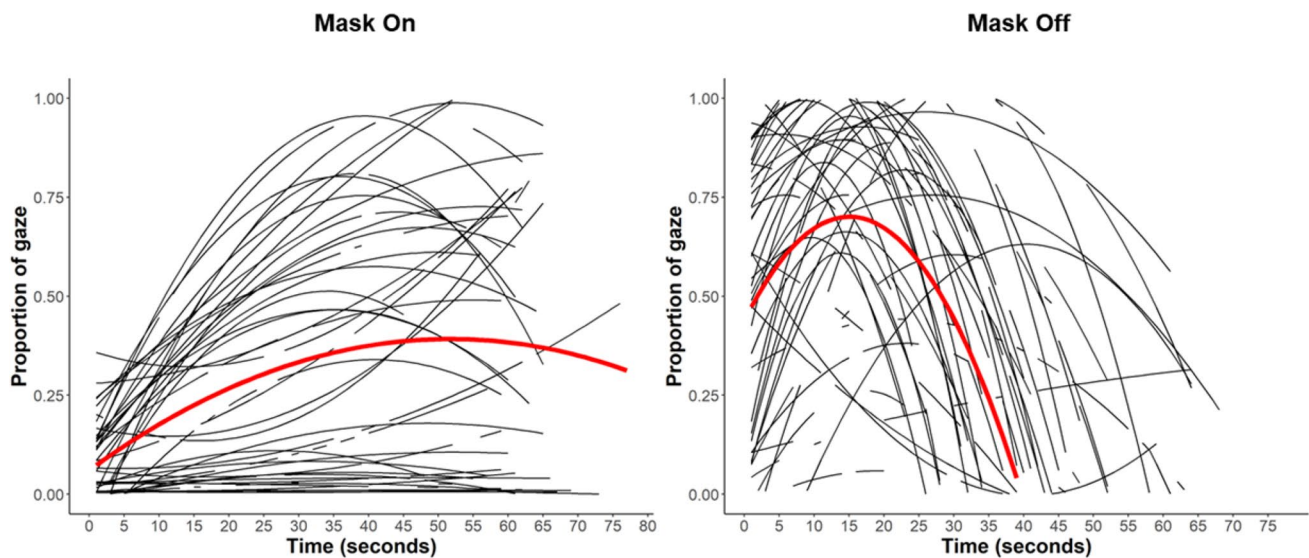


Fig. 5 Visualization of the growth curve modeling examining attention to a stranger during the period when the stranger was wearing a scary mask (*left panel*) and when the stranger took off the mask. The visualizations are presented in Figs. 2 and 4 of Gunther et al. (2021). The quadratic trajectory yielded a better fit than a linear tra-

jectory (BIC for a linear fit was 1395.65; BIC for a quadratic fit was 1324.34). The *black lines* show the model-estimated quadratic trajectory for individual participants. The *red line* displays the average quadratic trajectory

Future directions

We expect to see continuous development in MET hardware that enables MET applications to more diverse samples and data collection environments. Next-generation eye trackers (e.g., Tonsen et al., 2020) are being designed to be calibration-free and more robust to factors that reduce data quality (Niehorster et al., 2020; Valtakari et al., 2021), including participant movement, headset slippage, and changes in ambient lighting. This hardware improvement enables data collection outside the laboratory with participants who have difficulties with online calibration and lower tolerance for the headset. For example, published work has successfully collected MET data for over an hour per session (equivalent to the battery life of the smartphone used for data recording) in toddlers (27- to 31-month-old) as they go about their daily lives at home (Schroer et al., 2022). Future hardware development would benefit from data quality evaluations (e.g., Niehorster et al., 2020) in wider age ranges, clinical populations, and both indoor and outdoor environments.

The increased ease of MET data collection facilitates multimodal research that examines physiological and neural activities concurrently as participants actively attend to external stimuli (Valtakari et al., 2021). An example is to combine MET with functional near-infrared spectroscopy (fNIRS) recording (von Lümann et al., 2020). fNIRS is a noninvasive neuroimaging tool that measures event-evoked changes in cerebral blood oxygenation. As with electroencephalogram (EEG), fNIRS is well suited

for applications in a wide age range (Vanderwert & Nelson, 2014). A key advantage of fNIRS is that robust signals can be obtained even in free-moving participants (e.g., Burgess et al., 2022; Herold et al., 2017). Recent advances in wearable and portable fNIRS devices provide the opportunity to record neural activities in a variety of indoor and outdoor environments as participants actively interact with the environment (Pinti et al., 2020). However, one barrier for the multi-modal data acquisition is signal interference of fNIRS recording, as the eye tracker may emit near-infrared light at a wavelength that can be detected by fNIRS sensors. A specially designed cover for the fNIRS headset is needed to prevent interference (Katus et al., 2019).

MET facilitates research progress in understanding the moment-to-moment unfolding of behavioral and cognitive processes and how those micro-level processes dynamically interact with environmental factors at the macro-level over time. Individuals' multisensory development, including attention and motor abilities, reciprocally influence the individuals' social and physical environment throughout the course of human development (Smith et al., 2018). Perturbations in moment-to-moment looking behavior and attention-motor coordination can cast downstream impacts over time and across multiple levels of functioning. MET data collections have been largely implemented in cross-sectional studies, while we know that there are considerable changes in attention and motor functions in the lifespan (e.g., Mason et al., 2019; Reider et al., 2022; Vallesi et al., 2021). Incorporating MET measurements

in participants' naturalistic environment in longitudinal designs can deepen our understanding of how psychological functions that operate in micro-timescales develop with age and give rise to long-term impacts.

Conclusion

MET allows researchers to sample first-person gaze behavior in the context of ongoing external events, the individual's behavior, and psychological processes (Hayhoe & Rothkopf, 2011). Commercially available MET hardware allows users to collect good quality data from participants with a wider age range, in various environments, and for longer periods (Franchak & Yu, 2022; Pérez-Edgar et al., 2020). However, challenges in maintaining data quality during acquisition and the lack of standardized protocols for data processing create barriers to applying the technology (Hessels, Niehorster et al., 2020b). This paper provides a practical guide and open-source tools aimed at addressing methodological issues and challenges. This includes maximizing mobility, ensuring MET data quality, good practices in manual gaze annotations, the utility of data visualization, and possible data analytical methods. A number of tools for MET data quality assessment are readily available. This facilitates data quality reporting and data processing. There is increased implementation of automated AOI annotations with the rapid development of computer vision algorithms. However, manual inspections and annotations are indispensable for validating automated AOI annotations and ensure AOI annotation accuracy. Finally, we encourage researchers to utilize the micro-longitudinal structure of MET data to model the temporal dynamics of AOI looking events, in addition to the use of between-subjects aggregated indices (Ram & Gerstorff, 2009). We hope the practical guide can increase the accessibility of MET technology and help to enhance the reliability, standardization, and reproducibility of MET research. In particular, we believe that these methodological advances will propel our conceptual and theoretical understanding of mechanisms that shape behavior, affect, and cognition in-the-moment and cumulatively lay the foundation for long-term or larger-scale patterns of functioning.

Acknowledgments We thank the infants, children, and families who participated in the research projects.

Funding Open access funding provided by the Carolinas Consortium. Preparation of this manuscript was supported by a James S. McDonell Foundation and a National Science Foundation grant (BCS1941449) to J.M.F, National Institute of Child Health & Human Development grants (K99HD105920, R00HD105920) to J.I.B, the James S. McDonell Foundation and a National Institute of Mental Health grant to (K23MH120476) J.B., and National Institute of Mental Health grants (R21MH111980, R56MH1266349, R01MH130007) to K.P.E.

Data availability Example data used in data visualizations are available in the GitHub repository: https://github.com/xiaoxuefu/MET_methods.

Code availability Programs for data quality inspection, manual gaze annotations, data visualizations and analysis and example datasets are available in the GitHub repository: https://github.com/xiaoxuefu/MET_methods.

Declarations

Competing Interests The authors have no relevant financial or non-financial interests to disclose.

Ethics approval All procedures for the iTRAC and ACTION study were approved by the University Institutional Review Boards.

Consent to participate Families completed informed consent prior to all study procedures.

Consent for publication Not applicable.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Abney, D. H., Suanda, S. H., Smith, L. B., & Yu, C. (2020). What are the building blocks of parent–infant coordinated attention in free-flowing interaction? *Infancy*, 25(6), 871–887. <https://doi.org/10.1111/infa.12365>
- Ballard, D. H., Hayhoe, M. M., Pook, P. K., & Rao, R. P. N. (1997). Deictic codes for the embodiment of cognition. *Behavioral and Brain Sciences*, 20(4), 723–742. <https://doi.org/10.1017/S0140525X97001611>
- Bambach, S., Smith, L. B., Crandall, D. J., & Yu, C. (2016). Objects in the center: How the infant's body constrains infant scenes. *Joint IEEE International Conference on Development and Learning and Epigenetic Robotics (ICDL-EpiRob)*. IEEE.
- Bambach, S., Crandall, D., Smith, L., & Yu, C. (2018). Toddler-inspired visual object learning. *Advances in Neural Information Processing Systems*, 31.
- Benjamins, J. S., Hessels, R. S., & Hooge, I. T. C. (2018). GazeCode: Open-source software for manual mapping of mobile eye-tracking data. In: *2018 ACM Symposium on Eye Tracking Research & Applications (Etra 2018)*. <https://doi.org/10.1145/3204493.3204568>
- Benson, L., English, T., Conroy, D. E., Pincus, A. L., Gerstorff, D., & Ram, N. (2019). Age differences in emotion regulation strategy use, variability, and flexibility: An experience sampling approach. *Developmental psychology*, 55, 1951–1964. <https://doi.org/10.1037/dev0000727>

- 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 M. Horsley, M. Eliot, B. A. Knight, & R. Reilly (Eds.), *Current Trends in Eye Tracking Research* (pp. 77–100). Springer International Publishing. https://doi.org/10.1007/978-3-319-02868-2_5
- Bolger, N., & Laurenceau, J.-P. (2013). *Intensive longitudinal methods: An introduction to diary and experience sampling research*. Guilford Press.
- Borjon, J. I., Abney, D. H., Yu, C., & Smith, L. B. (2021). Head and eyes: Looking behavior in 12- to 24-month-old infants. *Journal of Vision*, 21(8), 18–18. <https://doi.org/10.1167/jov.21.8.18>
- Bradshaw, J., Fu, X., Yurkovic-Harding, J., & Abney, D. (2023). Infant embodied attention in context: Feasibility of home-based head-mounted eye tracking in early infancy. *Developmental Cognitive Neuroscience*, 64, 101299. <https://doi.org/10.1016/j.dcn.2023.101299>
- Breen, R. (1996). *Regression models: Censored, sample selected or truncated data*. SAGE Publishing.
- Brône, G., Oben, B., & Goedemé, T. (2011). Towards a More Effective Method for Analyzing Mobile Eye-tracking Data: Integrating Gaze Data with Object Recognition Algorithms. *Proceedings of the 1st International Workshop on Pervasive Eye Tracking & Mobile Eye-Based Interaction* (pp. 53–56). ACM.
- Brooks, R., & Meltzoff, A. N. (2005). The development of gaze following and its relation to language. *Developmental Science*, 8(6), 535–543. <https://doi.org/10.1111/j.1467-7687.2005.00445.x>
- Burgess, P. W., Crum, J., Pinti, P., Aichelburg, C., Oliver, D., Lind, F., ... & Hamilton, A. (2022). Prefrontal cortical activation associated with prospective memory while walking around a real-world street environment. *NeuroImage*, 258, 119392. <https://doi.org/10.1016/j.neuroimage.2022.119392>
- Cao, Z., Simon, T., Wei, S.-E., & Sheikh, Y. (2017). Realtime multi-person 2D pose estimation using part affinity fields. *Proceedings of the IEEE conference on computer vision and pattern recognition*. IEEE.
- Chen, C.-H., Houston, D. M., & Yu, C. (2021). Parent–child joint behaviors in novel object play create high-quality data for word learning. *Child Development*, 92(5), 1889–1905. <https://doi.org/10.1111/cdev.13620>
- Chronis-Tuscano, A., Degnan, K. A., Pine, D. S., Perez-Edgar, K., Henderson, H. A., Diaz, Y., ... & Fox, N. A. (2009). Stable early maternal report of behavioral inhibition predicts lifetime social anxiety disorder in adolescence. *Journal of the American Academy of Child & Adolescent Psychiatry*, 48(9), 928–935. <https://doi.org/10.1097/CHI.0b013e3181ae09df>
- Clauss, J. A., & Blackford, J. U. (2012). Behavioral inhibition and risk for developing social anxiety disorder: A meta-analytic study. *Journal of the American Academy of Child & Adolescent Psychiatry*, 51(10), 1066–1075. <https://doi.org/10.1016/j.jaac.2012.08.002>
- Cole, P. M., Loughheed, J. P., Chow, S.-M., & Ram, N. (2020). Development of emotion regulation dynamics across early childhood: A multiple time-scale approach. *Affective Science*, 1(1), 28–41. <https://doi.org/10.1007/s42761-020-00004-y>
- Crick, N. R., & Dodge, K. A. (1994). A review and reformulation of social information-processing mechanisms in children's social adjustment. *Psychological Bulletin*, 115(1), 74–101. <https://doi.org/10.1037/0033-2909.115.1.74>
- Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual Review of Neuroscience*, 18(1), 193–222.
- Domínguez-Zamora, F. J., & Marigold, D. S. (2019). Motor cost affects the decision of when to shift gaze for guiding movement. *Journal of Neurophysiology*, 122(1), 378–388. <https://doi.org/10.1152/jn.00027.2019>
- Duchowski, A. T., Gehrler, N. A., Schönenberg, M., & Krejtz, K. (2019). *Art facing science: Artistic heuristics for face detection: Tracking gaze when looking at faces* Proceedings of the 11th ACM Symposium on Eye Tracking Research & Applications, Denver, Colorado. <https://doi.org/10.1145/3317958.3319809>
- ELAN. (2018). *ELAN (5.2)* In Nijmegen: Max Planck Institute for Psycholinguistics.
- Evans, K. M., Jacobs, R. A., Tarduno, J. A., & Pelz, J. B. (2012). Collecting and analyzing eye tracking data in outdoor environments. *Journal of Eye Movement Research*, 5(2 6), 1–19.
- Foulsham, T., Walker, E., & Kingstone, A. (2011). The where, what and when of gaze allocation in the lab and the natural environment. *Vision Research*, 51(17), 1920–1931. <https://doi.org/10.1016/j.visres.2011.07.002>
- Franchak, J. M. (2017). Using head-mounted eye tracking to study development. In B. Hopkins, E. Geangu, & S. Linkenauer (Eds.), *The Cambridge encyclopedia of child development* (pp. 113–116). Cambridge University Press.
- Franchak, J. M. (2019). Looking with the head and eyes. *Perception as Information Detection* (pp. 205–221). Routledge.
- Franchak, J. M. (2020). Chapter Three - Visual exploratory behavior and its development. In K. D. Federmeier & E. R. Schotter (Eds.), *Psychology of Learning and Motivation* (73rd ed., pp. 59–94). Academic Press. <https://doi.org/10.1016/bs.plm.2020.07.001>
- Franchak, J. M. (2020). The ecology of infants' perceptual-motor exploration. *Current Opinion in Psychology*, 32, 110–114. <https://doi.org/10.1016/j.copsyc.2019.06.035>
- Franchak, J. M., & Adolph, K. E. (2010). Visually guided navigation: Head-mounted eye-tracking of natural locomotion in children and adults. *Vision Research*, 50(24), 2766–2774. <https://doi.org/10.1016/j.visres.2010.09.024>
- Franchak, J. M., & Yu, C. (2022). Beyond screen time: Using head-mounted eye tracking to study natural behavior. *Advances in Child Development and Behavior*, 62, 61–91.
- Franchak, J. M., Kretch, K. S., Soska, K. C., & Adolph, K. E. (2011). Head-mounted eye tracking: A new method to describe infant looking. *Child Development*, 82(6), 1738–1750. <https://doi.org/10.1111/j.1467-8624.2011.01670.x>
- Franchak, J. M., Kretch, K. S., & Adolph, K. E. (2018). See and be seen: Infant–caregiver social looking during locomotor free play. *Developmental Science*, 21(4), e12626. <https://doi.org/10.1111/desc.12626>
- Freeth, M., Foulsham, T., & Kingstone, A. (2013). What affects social attention? Social presence, eye contact and autistic traits. *PLOS ONE*, 8(1), e53286. <https://doi.org/10.1371/journal.pone.0053286>
- Friard, O., & Gamba, M. (2016). BORIS: A free, versatile open-source event-logging software for video/audio coding and live observations. *Methods in Ecology and Evolution*, 7(11), 1325–1330. <https://doi.org/10.1111/2041-210X.12584>
- Fu, X., & Pérez-Edgar, K. (2019). Threat-related attention bias in socio-emotional development: A critical review and methodological considerations. *Developmental Review*, 51, 31–57. <https://doi.org/10.1016/j.dr.2018.11.002>
- Fu, X., Nelson, E. E., Borge, M., Buss, K. A., & Pérez-Edgar, K. (2019). Stationary and ambulatory attention patterns are differentially associated with early temperamental risk for socio-emotional problems: Preliminary evidence from a multimodal eye-tracking investigation. *Development and Psychopathology*, 31(3), 971–988. <https://doi.org/10.1017/S0954579419000427>
- Gehrler, N. A., Duchowski, A. T., Jusyte, A., & Schönenberg, M. (2020). *Eye contact during live social interaction in incarcerated psychopathic offenders*. Educational Publishing Foundation. <https://doi.org/10.1037/per0000400>

- Gibson, J. J. (1979). The theory of affordances. The ecological approach to visual perception. *The People, Place and Space Reader* (pp. 56–60). Routledge.
- Gobel, M. S., Kim, H. S., & Richardson, D. C. (2015). The dual function of social gaze. *Cognition*, *136*, 359–364. <https://doi.org/10.1016/j.cognition.2014.11.040>
- Gunther, K. E., Brown, K. M., Fu, X., MacNeill, L. A., Jones, M., Ermanni, B., & Pérez-Edgar, K. (2021). Mobile eye tracking captures changes in attention over time during a naturalistic threat paradigm in behaviorally inhibited children. *Affective Science*, *2*(4), 495–505. <https://doi.org/10.1007/s42761-021-00077-3>
- Gunther, K. E., Fu, X., MacNeill, L., Vallorani, A., Ermanni, B., & Pérez-Edgar, K. (2022). Profiles of naturalistic attentional trajectories associated with internalizing behaviors in school-age children: A mobile eye tracking study. *Research on Child and Adolescent Psychopathology*, *50*(5), 637–648. <https://doi.org/10.1007/s10802-021-00881-2>
- Haensel, J. X., Smith, T. J., & Senju, A. (2022). Cultural differences in mutual gaze during face-to-face interactions: A dual head-mounted eye-tracking study. *Visual Cognition*, *30*(1–2), 100–115. <https://doi.org/10.1080/13506285.2021.1928354>
- Hassoumi, A., Peysakhovich, V., & Hurter, C. (2019). Improving eye-tracking calibration accuracy using symbolic regression. *PLOS ONE*, *14*(3), e0213675. <https://doi.org/10.1371/journal.pone.0213675>
- Hayhoe, M. M. (2017). Vision and action. *Annual Review of Vision Science*, *3*(1), 389–413. <https://doi.org/10.1146/annurev-vision-102016-061437>
- Hayhoe, M. M. (2018). Davida Teller Award Lecture 2017: What can be learned from natural behavior? *Journal of Vision*, *18*(4), 10–10. <https://doi.org/10.1167/18.4.10>
- Hayhoe, M. M., & Rothkopf, C. A. (2011). Vision in the natural world. *Wiley Interdisciplinary Reviews: Cognitive Science*, *2*(2), 158–166.
- Hayhoe, M. M., Shrivastava, A., Mruczek, R., & Pelz, J. B. (2003). Visual memory and motor planning in a natural task. *Journal of Vision*, *3*(1), 6–6. <https://doi.org/10.1167/3.1.6>
- Henderson, J. M. (2003). Human gaze control during real-world scene perception. *Trends in Cognitive Sciences*, *7*(11), 498–504. <https://doi.org/10.1016/j.tics.2003.09.006>
- Herold, F., Wiegel, P., Scholkmann, F., Thiers, A., Hamacher, D., & Schega, L. (2017). Functional near-infrared spectroscopy in movement science: A systematic review on cortical activity in postural and walking tasks. *Neurophotonics*, *4*(4), 041403. <https://doi.org/10.1117/1.NPh.4.4.041403>
- 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., 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), 5–5. <https://doi.org/10.1167/jov.20.10.5>
- Hessels, R. S., Niehorster, D. C., Holleman, G. A., Benjamins, J. S., & Hooge, I. T. C. (2020). Wearable Technology for “Real-World Research”: Realistic or Not? *Perception*, *49*(6), 611–615. <https://doi.org/10.1177/0301006620928324>
- Hessels, R. S., Benjamins, J. S., Niehorster, D. C., van Doorn, A. J., Koenderink, J. J., Holleman, G. A., ... & Hooge, I. T. C. (2022). Eye contact avoidance in crowds: A large wearable eye-tracking study. *Attention, Perception, & Psychophysics*, *84*(8), 2623–2640. <https://doi.org/10.3758/s13414-022-02541-z>
- Hollenstein, T. (2013). State Space Grids: Depicting Dynamics Across Development. In T. Hollenstein (Ed.), *State Space Grids*. Springer US. https://doi.org/10.1007/978-1-4614-5007-8_2
- Hollenstein, T., Granic, I., Stoolmiller, M., & Snyder, J. (2004). Rigidity in parent–child interactions and the development of externalizing and internalizing behavior in early childhood. *Journal of Abnormal Child Psychology*, *32*(6), 595–607. <https://doi.org/10.1023/B:JACP.0000047209.37650.41>
- Hooge, I. T. C., Niehorster, D. C., Hessels, R. S., Benjamins, J. S., & Nyström, M. (2023). How robust are wearable eye trackers to slow and fast head and body movements? *Behavior Research Methods*, *55*(8), 4128–4142. <https://doi.org/10.3758/s13428-022-02010-3>
- Jongierius, C., Callemein, T., Goedemé, T., Van Beeck, K., Romijn, J. A., Smets, E. M. A., & Hillen, M. A. (2021). Eye-tracking glasses in face-to-face interactions: Manual versus automated assessment of areas-of-interest. *Behavior Research Methods*, *53*(5), 2037–2048. <https://doi.org/10.3758/s13428-021-01544-2>
- Jung, Y. J., Zimmerman, H. T., & Pérez-Edgar, K. (2018). A methodological case study with mobile eye-tracking of child interaction in a science museum. *TechTrends*, *62*(5), 509–517. <https://doi.org/10.1007/s11528-018-0310-9>
- Katus, L., Hayes, N. J., Mason, L., Blasi, A., McCann, S., Darboe, M. K., ... & Elwell, C. E. (2019). Implementing neuroimaging and eye tracking methods to assess neurocognitive development of young infants in low- and middle-income countries. *Gates Open Research*, *3*, 1113. <https://doi.org/10.12688/gatesopenres.12951.2>
- 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), 2539. <https://doi.org/10.1038/s41598-020-59251-5>
- Kretch, K. S., Franchak, J. M., & Adolph, K. E. (2014). Crawling and walking infants see the world differently. *Child Development*, *85*(4), 1503–1518. <https://doi.org/10.1111/cdev.12206>
- Laidlaw, K. E. W., Foulsham, T., Kuhn, G., & Kingstone, A. (2011). Potential social interactions are important to social attention. *Proceedings of the National Academy of Sciences*, *108*(14), 5548–5553. <https://doi.org/10.1073/pnas.1017022108>
- Lamey, A. V., Hollenstein, T., Lewis, M. D., & Granic, I. (2004). *GridWare (Version 1.1)*. In <http://www.statespacegrids.org>
- Land, M. F. (2006). Eye movements and the control of actions in everyday life. *Progress in Retinal and Eye Research*, *25*(3), 296–324. <https://doi.org/10.1016/j.preteyeres.2006.01.002>
- Land, M., Mennie, N., & Rusted, J. (1999). The roles of vision and eye movements in the control of activities of daily living. *Perception*, *28*(11), 1311–1328. <https://doi.org/10.1068/p2935>
- Lewis, M. D., Lamey, A. V., & Douglas, L. (1999). A new dynamic systems method for the analysis of early socioemotional development. *Developmental Science*, *2*(4), 457–475. <https://doi.org/10.1111/1467-7687.00090>
- Li, R., Dziak, J. J., Tan, X., Huang, L., Wagner, A. T., & Yang, J. (2015). *TVEM (time-varying effect modeling) SAS macro users' guide (Version 3.1.1)*. The Methodology Center.
- Long, B. L., Sanchez, A., Kraus, A. M., Agrawal, K., & Frank, M. C. (2022). Automated detections reveal the social information in the changing infant view. *Child Development*, *93*(1), 101–116. <https://doi.org/10.1111/cdev.13648>
- Luo, C., & Franchak, J. M. (2020). Head and body structure infants' visual experiences during mobile, naturalistic play. *PLOS ONE*, *15*(11), e0242009. <https://doi.org/10.1371/journal.pone.0242009>
- Macdonald, R. G., & Tatler, B. W. (2013). Do as eye say: Gaze cueing and language in a real-world social interaction. *Journal of Vision*, *13*(4), 6–6. <https://doi.org/10.1167/13.4.6>
- Macinnes, J., Iqbal, S., Pearson, J., & Johnson, E. (2018). Wearable eye-tracking for research: Automated dynamic gaze mapping and accuracy/precision comparisons across devices. <https://doi.org/10.1101/299925>

- MacNeill, L. A., Fu, X., Buss, K. A., & Pérez-Edgar, K. (2022). Do you see what I mean?: Using mobile eye tracking to capture parent-child dynamics in the context of anxiety risk. *Development and Psychopathology*, *34*(3), 997–1012. <https://doi.org/10.1017/S0954579420001601>
- Mardanbegi, D., & Hansen, D. W. (2012). Parallax error in the monocular head-mounted eye trackers. In: *Ubicomp'12: Proceedings of the 2012 ACM International Conference on Ubiquitous Computing*, 689–694.
- Marigold, D. S., & Patla, A. E. (2007). Gaze fixation patterns for negotiating complex ground terrain. *Neuroscience*, *144*(1), 302–313. <https://doi.org/10.1016/j.neuroscience.2006.09.006>
- Mason, G. M., Goldstein, M. H., & Schwade, J. A. (2019). The role of multisensory development in early language learning. *Journal of Experimental Child Psychology*, *183*, 48–64. <https://doi.org/10.1016/j.jecp.2018.12.011>
- Matthis, J. S., & Fajen, B. R. (2014). Visual control of foot placement when walking over complex terrain. *Journal of Experimental Psychology: Human Perception and Performance*, *40*, 106–115. <https://doi.org/10.1037/a0033101>
- Matthis, J. S., Yates, J. L., & Hayhoe, M. M. (2018). Gaze and the control of foot placement when walking in natural terrain. *Current Biology*, *28*(8), 1224–1233.e1225. <https://doi.org/10.1016/j.cub.2018.03.008>
- Morales, S., Ram, N., Buss, K. A., Cole, P. M., Helm, J. L., & Chow, S.-M. (2018). Age-related changes in the dynamics of fear-related regulation in early childhood. *Developmental Science*, *21*(5), e12633. <https://doi.org/10.1111/desc.12633>
- Nagin, D. S., & Odgers, C. L. (2010). Group-based trajectory modeling in clinical research. *Annual Review of Clinical Psychology*, *6*(1), 109–138.
- Nasiopoulos, E., Risko, E. F., & Kingstone, A. (2015). Social attention, social presence, and the dual function of gaze. In *The many faces of social attention* (pp. 129–155). Springer.
- Nasrabadi, H. R., & Alonso, J.-M. (2022). Modular streaming pipeline of eye/head tracking data using Tobii Pro Glasses 3. *bioRxiv*, 2022.2009.2002.506255. <https://doi.org/10.1101/2022.09.02.506255>
- 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., Santini, T., Hessels, R. S., Hooge, I. T. C., Kasneci, E., & Nyström, M. (2020). 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., Hessels, R. S., Benjamins, J. S., Nyström, M., & Hooge, I. T. C. (2023). GlassesValidator: A data quality tool for eye tracking glasses. *Behavior Research Methods*, *56*(3), 1476–84. <https://doi.org/10.3758/s13428-023-02105-5>
- Pérez-Edgar, K., MacNeill, L. A., & Fu, X. (2020). Navigating through the experienced environment: Insights from mobile eye tracking. *Current Directions in Psychological Science*, *29*(3), 286–292. <https://doi.org/10.1177/0963721420915880>
- Pinti, P., Tachtsidis, I., Hamilton, A., Hirsch, J., Aichelburg, C., Gilbert, S., & Burgess, P. W. (2020). The present and future use of functional near-infrared spectroscopy (fNIRS) for cognitive neuroscience. *Annals of the New York Academy of Sciences*, *1464*(1), 5–29. <https://doi.org/10.1111/nyas.13948>
- Ram, N., & Gerstorff, D. (2009). Time-structured and net intraindividual variability: Tools for examining the development of dynamic characteristics and processes. *Psychology and Aging*, *24*(4), 778–791.
- Redcay, E., & Schilbach, L. (2019). Using second-person neuroscience to elucidate the mechanisms of social interaction. *Nature Reviews Neuroscience*, *20*(8), 495–505. <https://doi.org/10.1038/s41583-019-0179-4>
- Reider, L. B., Bierstedt, L., Burris, J. L., Vallorani, A., Gunther, Kelley E., Buss, K. A., ... & LoBue, V. (2022). Developmental patterns of affective attention across the first 2 years of life. *Child Development*, *93*(6), e607–e621. <https://doi.org/10.1111/cdev.13831>
- Risko, E., Laidlaw, K., Freeth, M., Foulsham, T., & Kingstone, A. (2012). Social attention with real versus reel stimuli: Toward an empirical approach to concerns about ecological validity [Review]. *Frontiers in Human Neuroscience*, *6*, 143. <https://doi.org/10.3389/fnhum.2012.00143>
- Rogers, S. L., Spelman, C. P., Guidetti, O., & Longmuir, M. (2018). Using dual eye tracking to uncover personal gaze patterns during social interaction. *Scientific Reports*, *8*(1), 4271. <https://doi.org/10.1038/s41598-018-22726-7>
- Santini, T., Brinkmann, H., Reitstätter, L., Leder, H., Rosenberg, R., Rosenstiel, W., & Kasneci, E. (2018). The art of pervasive eye tracking: Unconstrained eye tracking in the Austrian Gallery Belvedere. *Proceedings of the 7th workshop on pervasive eye tracking and mobile eye-based interaction*. Association for Computing Machinery Inc.
- Schroer, S. E., & Yu, C. (2023). Looking is not enough: Multimodal attention supports the real-time learning of new words. *Developmental Science*, *26*(2), e13290. <https://doi.org/10.1111/desc.13290>
- Schroer, S. E., Peters, R. E., Yarbrough, A., & Yu, C. (2022). Visual attention and language exposure during everyday activities: An at-home study of early word learning using wearable eye trackers. In: *Proceedings of the Annual Meeting of the Cognitive Science Society*.
- Shewark, E. A., Brick, T. R., & Buss, K. A. (2020). Capturing temporal dynamics of fear behaviors on a moment-to-moment basis. *Infancy*, *25*(3), 264–285. <https://doi.org/10.1111/inf.12328>
- Slone, L. K., Abney, D. H., Borjon, J. I., Chen, C.-H., Franchak, J. M., Percy, D., ... & Yu, C. (2018). Gaze in action: Head-mounted eye tracking of children's dynamic visual attention during naturalistic behavior. *JOVE: Journal of Visualized Experiments*, *141*, e58496. <https://doi.org/10.3791/58496>
- Slone, L. K., Smith, L. B., & Yu, C. (2019). Self-generated variability in object images predicts vocabulary growth. *Developmental Science*, *22*(6), e12816. <https://doi.org/10.1111/desc.12816>
- Smith, L. B., & Slone, L. K. (2017). A developmental approach to machine learning? [Hypothesis and Theory]. *Frontiers in Psychology*, *8*, 296143. <https://doi.org/10.3389/fpsyg.2017.02124>
- Smith, L. B., Jayaraman, S., Clerkin, E., & Yu, C. (2018). The developing infant creates a curriculum for statistical learning. *Trends in Cognitive Sciences*, *22*(4), 325–336. <https://doi.org/10.1016/j.tics.2018.02.004>
- Socha, V., Vidensky, J., Kasmirek, S., Hanakova, L., & Valenta, V. (2022, 26–27 Oct. 2022). Design of Wearable Eye Tracker with Automatic Cockpit Areas of Interest Recognition. 2022 New Trends in Civil Aviation (NTCA),
- Suarez-Rivera, C., Smith, L. B., & Yu, C. (2019). Multimodal parent behaviors within joint attention support sustained attention in infants. *Developmental Psychology*, *55*(1), 96–109.
- Suarez-Rivera, C., Schatz, J. L., Herzberg, O., & Tamis-LeMonda, C. S. (2022). Joint engagement in the home environment is frequent, multimodal, timely, and structured. *Infancy*, *27*(2), 232–254. <https://doi.org/10.1111/inf.12446>
- Datavyu. (2014). Datavyu: A video coding tool (Version 1.3.4). *Datavyu Project*. New York University. <http://datavyu.org>.
- Thelen, E. & Smith, L. B. (Eds.)(1998). Dynamic systems theories. In: *Handbook of child psychology: Theoretical models of human development* (vol. 1. 5th ed., pp. 563–634). John Wiley & Sons Inc.

- Todd, R. M., Cunningham, W. A., Anderson, A. K., & Thompson, E. (2012). Affect-biased attention as emotion regulation. *Trends in Cognitive Sciences*, 16(7), 365–372. <https://doi.org/10.1016/j.tics.2012.06.003>
- Tomasello, M., & Farrar, M. J. (1986). Joint attention and early language. *Child Development*, 57(6), 1454–1463. <https://doi.org/10.2307/1130423>
- Tonsen, M., Baumann, C. K., & Dierkes, K. (2020). A high-level description and performance evaluation of pupil invisible. *arXiv preprint arXiv:2009.00508*.
- Vallesi, A., Tronelli, V., Lomi, F., & Pezzetta, R. (2021). Age differences in sustained attention tasks: A meta-analysis. *Psychonomic Bulletin & Review*, 28(6), 1755–1775. <https://doi.org/10.3758/s13423-021-01908-x>
- 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*, 53(4), 1592–1608. <https://doi.org/10.3758/s13428-020-01517-x>
- Vallorani, A., Brown, K. M., Fu, X., Gunther, K. E., MacNeill, L. A., Ermanni, B., ... & Pérez-Edgar, K. (2022). Relations between social attention, expressed positive affect and behavioral inhibition during play. *Developmental Psychology*, 58(11), 2036–2048. <https://doi.org/10.1037/dev0001412>
- Vanderwert, R. E., & Nelson, C. A. (2014). The use of near-infrared spectroscopy in the study of typical and atypical development. *NeuroImage*, 85, 264–271.
- von Lüthmann, A., Zimmermann, B. B., Ortega-Martinez, A., Perkins, N., Yücel, M. A., & Boas, D. A. (2020). Towards Neuroscience in the Everyday World: Progress in wearable fNIRS instrumentation and applications. *OSA Technical Digest Biophotonics Congress: Biomedical Optics 2020 (Translational, Microscopy, OCT, OTS, BRAIN)*, Washington, DC.
- 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., 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/inf.12055>
- Woody, M. L., Rosen, D., Allen, K. B., Price, R. B., Hutchinson, E., Amole, M. C., & Silk, J. S. (2019). Looking for the negative: Depressive symptoms in adolescent girls are associated with sustained attention to a potentially critical judge during in vivo social evaluation. *Journal of Experimental Child Psychology*, 179, 90–102. <https://doi.org/10.1016/j.jecp.2018.10.011>
- Wright, G. A., Patel, R., Pérez-Edgar, K., Fu, X., Brown, K., Adhikary, S., & Zurca, A. (2022). Eye-tracking technology to determine procedural proficiency in ultrasound-guided regional anesthesia. *The Journal of Education in Perioperative Medicine: JEPM*, 24(1), E684.
- Yamashiro, A., Shrout, P. E., & Vouloumanos, A. (2019). Using spline models to analyze event-based changes in eye tracking data. *Journal of Cognition and Development*, 20(3), 299–313. <https://doi.org/10.1080/15248372.2019.1583231>
- Yoshida, H., & Burling, J. M. (2011). A new perspective on embodied social attention. *Cognition, Brain, Behavior: an Interdisciplinary Journal*, 15(4), 535–552.
- Yu, C., & Smith, L. B. (2012). Embodied attention and word learning by toddlers. *Cognition*, 125(2), 244–262. <https://doi.org/10.1016/j.cognition.2012.06.016>
- Yu, C., & Smith, L. B. (2013). Joint attention without gaze following: Human infants and their parents coordinate visual attention to objects through eye–hand coordination. *PLOS ONE*, 8(11), e79659. <https://doi.org/10.1371/journal.pone.0079659>
- Yu, C., & Smith, Linda B. (2016). The Social origins of sustained attention in one-year-old human infants. *Current Biology*, 26(9), 1235–1240. <https://doi.org/10.1016/j.cub.2016.03.026>
- Yu, C., & Smith, L. B. (2017). Hand–eye coordination predicts joint attention. *Child Development*, 88(6), 2060–2078. <https://doi.org/10.1111/cdev.12730>
- Yu, C., & Smith, L. B. (2017). Multiple sensory-motor pathways lead to coordinated visual attention. *Cognitive Science*, 41(S1), 5–31. <https://doi.org/10.1111/cogs.12366>
- Yu, C., Yurovsky, D., & Xu, T. (2012). Visual data mining: An exploratory approach to analyzing temporal patterns of eye movements. *Infancy*, 17(1), 33–60. <https://doi.org/10.1111/j.1532-7078.2011.00095.x>
- Yu, C., Suanda, S. H., & Smith, L. B. (2019). Infant sustained attention but not joint attention to objects at 9 months predicts vocabulary at 12 and 15 months. *Developmental Science*, 22(1), e12735. <https://doi.org/10.1111/desc.12735>
- Yurkovic, J. R., Lisandrelli, G., Shaffer, R. C., Dominick, K. C., Pedapati, E. V., Erickson, C. A., ... & Yu, C. (2021). Using head-mounted eye tracking to examine visual and manual exploration during naturalistic toy play in children with and without autism spectrum disorder. *Scientific Reports*, 11(1), 3578. <https://doi.org/10.1038/s41598-021-81102-0>
- Yurkovic-Harding, J., Lisandrelli, G., Shaffer, R. C., Dominick, K. C., Pedapati, E. V., Erickson, C. A., ... & Kennedy, D. P. (2022). Children with ASD establish joint attention during free-flowing toy play without face looks. *Current Biology*, 32(12), 2739–2746. e2734. <https://doi.org/10.1016/j.cub.2022.04.044>

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