RESEARCH ARTICLE

Variation in gaze following across the life span: A process-level perspective

Julia Christin Prein[1,2] | Luke Maurits[1] | Annika Werwach[1,3,4] | Daniel B. M. Haun[1] | Manuel Bohn[1,2]

[1]Department of Comparative Cultural Psychology, Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany

[2]Institute of Psychology in Education, Leuphana University Lüneburg, Lüneburg, Germany

[3]Center for Lifespan Psychology, Max Planck Institute for Human Development, Berlin, Germany

[4]Max Planck School of Cognition, Leipzig, Germany

Correspondence
Julia Christin Prein, Institute of Psychology in Education, Leuphana University Lüneburg, Lüneburg, Germany.
Email: julia.prein@leuphana.de

Daniel B.M. Haun, and Manuel Bohn shared senior authorship.

Abstract

Following eye gaze is fundamental for many social-cognitive abilities, for example, when judging what another agent can or cannot know. While the emergence of gaze following has been thoroughly studied on a group level, we know little about (a) the developmental trajectory beyond infancy and (b) the sources of individual differences. In Study 1, we examined gaze following across the lifespan (N = 478 3- to 19-year-olds from Leipzig, Germany; and N = 240 20- to 80-year-old international, remotely tested adults). We found a steep performance improvement during preschool years, in which children became more precise in locating the attentional focus of an agent. Precision levels then stayed comparably stable throughout adulthood with a minor decline toward old age. In Study 2, we formalized the process of gaze following in a computational cognitive model that allowed us to conceptualize individual differences in a psychologically meaningful way (N = 60 3- to 5-year-olds, 50 adults). According to our model, participants estimate pupil angles with varying levels of precision based on observing the pupil location within the agent's eyes. In Study 3, we empirically tested how gaze following relates to vector following in non-social settings and perspective-taking abilities (N = 102 4- to 5-year-olds). We found that gaze following is associated with both of these abilities but less so with other Theory of Mind tasks. This work illustrates how the combination of reliable measurement instruments and formal theoretical models allows us to explore the in(ter)dependence of core social-cognitive processes in greater detail.

KEYWORDS
cognitive modeling, gaze following, individual differences, lifespan, social-cognitive development, Theory of Mind

Research Highlights

- Gaze following develops beyond infancy. The highest precision levels in localizing attentional foci are reached in young adulthood with a slight decrease towards old age.

- We present a computational model that describes gaze following as a process of estimating pupil angles and the corresponding gaze vectors.
- The model explains individual differences and recovers signature patterns in the data. To estimate the relation between gaze- and vector following, we designed a non-social vector following task.
- We found substantial correlations between gaze following and vector following, as well as Level 2 perspective-taking. Other Theory of Mind tasks did not correlate.

1 | INTRODUCTION

Following the gaze of others is valuable for extracting information from the environment. It guides us to "informational hotspots" (Meltzoff et al., 2010, p. 1) and can be used to identify internal states such as intentions or emotions (Corkum & Moore, 1998; Pfeiffer et al., 2013). As one of the most fundamental social-cognitive abilities, gaze following is an integral part of almost every form of social interaction, including communication, collaboration, cultural and social learning (Bohn & Frank, 2019; Emery, 2000; Frith & Frith, 2012; Hessels, 2020; Moore, 2008; Rakoczy, 2022; Tomasello & Rakoczy, 2003), and has been extensively studied in infancy (for review, see Del Bianco et al., 2019). In the most commonly used paradigm (e.g., Astor et al., 2020; Byers-Heinlein et al., 2021; Gredebäck et al., 2010; Ishikawa et al., 2022), the experimenter looks directly at the infant before shifting their head and eyes to one of two objects. Infants' looking times to the target or the proportion of choosing the target over the distractor are measured. Research in this tradition finds that infants as young as 3 to 4 months can follow the gaze of another agent (Astor et al., 2021; D'Entremont et al., 1997; D'Entremont, 2000; Del Bianco et al., 2019).

Previous research has shown a refinement of gaze following abilities in a child's first and second year of life (e.g., Astor et al., 2021; Brooks & Meltzoff, 2002; Butterworth & Jarrett, 1991). At the end of their first year of life, infants can follow gaze to locations outside their current visual field and move themselves to gain proper perceptual access (Butterworth & Jarrett, 1991; Corkum & Moore, 1995; Deák et al., 2000; Moll & Tomasello, 2004). However, we do not know much about the developmental progression beyond these qualitative milestones. One possibility is that the ability to follow gaze does not improve beyond infancy. Yet, most, if not all, cognitive abilities continue to develop throughout childhood (e.g., Gathercole et al., 2004; Gredebäck et al., 2010). It seems likely that children fine-tune their gaze following as they get older—presumably while using them in social interactions. To capture the development in gaze following beyond infancy, Study 1 included participants from preschool to old age.

Until today, only a handful of gaze following studies differentiate between manipulating head and eye movement. Michel et al. (2021) found that gaze following in 4-month-olds was likely driven by other's head- instead of eye movements. Corkum and Moore (1995), Lempers (1979), and Lempers et al. (1977) suggest that infants, at least until 19

months, struggle when eye and head direction diverge. From farther distance, body or face orientation can act as more salient cues to determine another's area of attention. However, eye direction indicates a more precise location of focus (Emery, 2000; Stiefelhagen & Zhu, 2002; however, see Loomis et al., 2008 for peripheral vision), and allows us to anticipate likely future actions (Friesen & Rao, 2011; Zohary et al., 2022). The three studies included in this work, therefore, focused on subtle gaze cues and isolated eye movement alone.

Group-level analyses of gaze following abilities (e.g., the average age at which children as a group reach an above-chance performance) may mask individual differences between children. Measuring individual differences in basic aspects of social cognition is important to understand the underlying processes and to quantify the impact of environmental influences and other cognitive abilities (Birch et al., 2017; Del Bianco et al., 2019). Across the three studies, we measured gaze following continuously by using a task that is designed to capture individual-level variation (Prein et al., 2023): the TANGO (Task for Assessing iNdividual differences in Gaze understanding—Open) avoids floor and ceiling effects in children and adults and is thus particularly suited to examine how gaze following changes with age.

A promising approach to interpreting individual differences in cognitive abilities is computational cognitive modeling. Existing computational models have described gaze following via reinforcement learning (Ishikawa et al., 2020), as a consequence of goal interference and mapping self-experience onto other agents (Friesen & Rao, 2011) or an interplay of object distances/saliencies and head poses (Jasso & Triesch, 2006; Lau & Triesch, 2004; Recasens et al., 2015). As such, these models focus on the motivation behind gaze following or on situations in which the context (e.g., head orientation or surrounding objects) offers cues to the gaze direction. Our goal, however, was to formulate a theory that models how people estimate gaze direction based on the eyes alone (i.e., "literal" gaze following) when no target objects are present[1] and the other's head is frontally oriented.

To our knowledge, there are three views that focus on the eyes alone that conceptualize gaze as (1) a beam, (2) a cone, or (3) a line (alternatively called a vector, ray, or line-of-sight). Guterstam et al. (2019) proposed the idea of gaze as a force-carrying beam. This theory, however, focuses on how people's implicit assumptions about the physical properties of an object change when someone looks at it. The idea of gaze as a cone is mostly concerned with the question of

how people determine whether someone else looks at them (Gamer & Hecht, 2007; Horstmann & Linke, 2021). The conception most relevant to the present study sees eye gaze as a line: Yaniv and Shatz (1990) proposed that children extrapolate an imaginary trajectory between the agent and the object to identify the focus of attention (similar to "geometrical" gaze following; [Butterworth & Jarrett, 1991]). Anecdotal evidence was already reported by Walker and Gollin (1977), who observed two children pointing their fingers into the air, drawing a line between an agent and an object, and saying, "He sees that" (p. 354). Michelon and Zacks (2006) found that response time in Level 1 visual perspective-taking tasks depends on the distance between the agent and the object (i.e., the length of the line-of-sight). Some researchers additionally highlight iris eccentricity, that is, the ratio of the visible sclera on each side of the pupil (Anstis et al., 1969; Symons et al., 2004; Todorović, 2006). Todorović (2006) defined gaze direction as "the vector positioned along the visual axis, pointing from the fovea of the looker through the center of the pupil to the gazed-at spot" (p. 3550). Symons et al. (2004) furthermore reasoned that "the perceiver must use the asymmetrical configuration of the dark-white contrast of another individual's eyes, and trace along two invisible sight-lines to their convergent point, that is, the third part of the triad (e.g., an object or a person)" (p. 452). Based on their finding that adults' gaze direction sensitivity decreases when only one eye is shown, Symons et al. (2004) conclude that information from both eyes must be integrated.

While the previously mentioned conceptualizations focus on the direction of eye gaze, they (A) cannot explain how people differ in their abilities to precisely estimate gaze direction, and (B) are not clearly expressed as formal, mathematical models with explicit assumptions and testable predictions. In the words of (Gamer & Hecht, 2007): "Given the social relevance of determining gaze direction, the psychophysics of gaze is underdeveloped" (p. 705).

Here, we propose a cognitive model of gaze following, which builds upon the notion of eye gaze as line-of-sight tracing and extends this by explicitly modeling individual differences. Our gaze model assumes participants infer the locus of someone's attention to be where two estimated gaze vectors meet. Each of these gaze vectors results from connecting the center of the agent's eyeball and the center of the pupil. Because the center of the eyeball is not directly observable at that moment, vector estimation happens with a degree of uncertainty. Development of gaze following corresponds to a decrease in uncertainty. Individual differences correspond to systematic differences in uncertainty.

By focusing on individual differences, we can further address the relationship between gaze following and other cognitive abilities. A longstanding question has been whether gaze following is related to Theory of Mind (ToM) (e.g., Brooks & Meltzoff, 2015). Moll and Meltzoff (2011) have suggested that joint attention (including gaze following) might be seen as "Level 0 perspective-taking," which provides the foundation for later-emerging, more complex perspective-taking abilities. On the other hand, an alignment of infants' visual attention to another's gaze does not necessarily indicate understanding the intentions of the agent (Aslin, 2007). Infants could simply align their orientation

without processing what exactly the other is seeing (Butterworth & Jarrett, 1991). In fact, one might question if such an alignment reflects an understanding of visual perspectives at all because the "target" or "object of representation" is not necessarily specified (Perner et al., 2003, p. 358). Consequently, Astor and Gredebäck (2022) have listed the relationship between gaze following and perspective-taking as one of their five big open questions in gaze following research. Therefore, in Study 3, we assessed how gaze following relates to ToM abilities, especially visual perspective-taking.

Taken together, the present study had three main goals: first, we studied the development of gaze following beyond infancy (Study 1). Instead of capturing the youngest age at which children follow the gaze, we examined how this ability changes with age. Our second goal was to provide a process-level theory of gaze following—and, most importantly, individual differences therein. We proposed a computational cognitive model, which formalized gaze following as a form of vector following, and tested whether our model explained empirical data (Study 2). Third, we examined which (social-)cognitive components comprise gaze following (Study 3). Based on our model, we predicted that gaze following should be related to non-social vector following. Additionally, we assessed the link between gaze following and ToM measures, with a particular focus on visual perspective-taking.

2 | STUDY 1: GAZE FOLLOWING ACROSS THE LIFESPAN

The study was pre-registered prior to data collection: https://osf.io/snju6 (child sample) and https://osf.io/6yjz3 (adult sample). The study obtained ethical clearance from the MPG Ethics Commission in Munich, Germany, falling under an umbrella ethics application (Appl. No. 2021_45). Data was collected between May 2021 and April 2023.

2.1 | Participants

We collected data online from 3- to 80- year-olds (see Supplementary information for further details). The child sample consisted of 478 3- to 19-year-olds and was recruited via an internal database of families in Leipzig, Germany, who volunteered to participate in child development studies. Participants came from ethnically homogeneous, mixed socioeconomic backgrounds with mid to high parental education levels. They lived in an industrialized, urban Central-European context in a mid-size German city (approx. 600,000 inhabitants; median individual monthly net income approx. 1600€ as of 2021). Most were raised monolingually in a nuclear two-generational family setting. Information on demographics and socioeconomic status was not formally recorded on a participant level.

Adults were recruited via *Prolific* (Palan & Schitter, 2018). *Prolific* is an online participant recruitment tool from the University of Oxford with predominantly European and US-American subjects. Participants consisted of 240 English-speaking adults between 20 and 80 years of age who reported to have a normal or corrected-to-normal vision.

For completing the study, subjects were paid above the fixed minimum wage (~£10.00/h).

2.2 | Materials

We used the continuous version of the TANGO (Prein et al., 2023). The task was presented as a web application (demo https://ccp-odc.eva.mpg.de/tango-demo/; source code https://github.com/ccp-eva/tango-demo). The TANGO showed satisfactory internal consistency and retest reliability (*Pearson's r* from 0.7 to 0.8; Prein et al. [2023]) and no floor or ceiling effects for children and adults.

2.3 | Procedure

Children and teenagers received a personalized link to the study website. Caregivers were asked to provide technical support, while explicitly being reminded not to help in responding. Webcam videos were recorded whenever consented and technically feasible in order to monitor whether children and teenagers responded on their own. Adults completed the online study unsupervised.

Each trial presented an agent standing in a window, watching a balloon (i.e., target) falling to the ground (see Figure 4a; however, Study 1 presented animal agents). The target fell behind a hedge while the agent's gaze followed the target's trajectory. In test trials, a hedge covered the target's position. Participants were asked to touch or click where they estimated the target to be based on the agent's gaze. Four familiarization trials ensured participants understood the task and felt comfortable with the response format. Then, 15 test trials followed. Completing 19 trials took 5–10 min.

We measured imprecision, defined as the absolute difference between the target center and the x coordinate of the participant's click. The screen width was divided into ten bins. Within each bin, exact target coordinates were randomly generated. Each target bin, agent, and target color occurred equally often and did not appear in more than two consecutive trials.

2.4 | Analysis

We ran all analyses in R version 4.4.0 (2024-04-24) (R Core Team, 2022). Regression models were fit as Bayesian generalized linear mixed models (GLMMs) with default priors using the function brm from the package brms (Bürkner, 2017, 2018).

We fit GLMMs that make different assumptions about the developmental trajectory, modeling the relationship between age and performance as linear, quadratic, or cubic. In addition, we fit a Gaussian Process model (Bürkner, 2017), which assumes a smooth relationship but avoids enforcing a particular shape. Per individual, imprecision was aggregated across trials and modeled as a lognormal distribution.[2] The unit of imprecision was counted in target widths, that is, an imprecision of 1 meant clicking one balloon width to the left or right of the true target center. We inspected the posterior distributions (mean and 95% credible interval [CrI]) for the age estimates and compared models via model weights and the difference in expected log pointwise predictive density (ELPD) estimated using leave-one-out cross-validation (LOO) (Vehtari et al., 2017).

To obtain a concise but principled characterization of the developmental trajectory, we additionally performed a Bayesian change point analysis, using the package RBeast (Zhao et al., 2019). We sought the most likely change points in our data, assuming a constant mean (i.e., a flat line, zero-degree polynomial) within each segment. To avoid "overreactions" to outlying data points, we constrained the model to have minimally 10 data points between consecutive change points (= half of the data points collected per adult decade). We inspected the posterior probability of different numbers of change points and the locations of these change points (mean and 95% CrI).[3]

2.5 | Results

High levels of variation pointed to substantial individual differences in all age groups (overall imprecision mean = 0.81, SD = 0.82, range = [0–10.73]). We found strong evidence for a non-linear development in gaze following across the lifespan (see Figure 1). The Gaussian process model was clearly preferred over the polynomial models due to the highest predictive accuracy according to the LOO ELPD estimates: elpd_diff between Gaussian process and cubic model = −33.75 (SE = 8.83); elpd_diff between Gaussian process and quadratic model = −95.07 (SE = 15.19); elpd_diff between Gaussian process and linear model = −127.17 (SE = 18.18); all in favor of the Gaussian process model. Moreover, the Gaussian Process model showed the greatest model weight (approximating 1). For the imprecision in gaze following, the standard deviation of the Gaussian process (SD = 1.52, 95% CrI [0.25; 5.19]) indicated nonlinearity.

The Bayesian change point analysis revealed 6 major shifts in gaze following during the lifespan (MAP estimate with 23.49% probability). The change points occurred at 4.23 years (95% CrI [4.13; 4.33], mean imprecision until change point = 2.15, SD = 0.85); 5.71 years (95% CrI [5.38; 5.90], mean = 1.36, SD = 0.53); 6.98 years (95% CrI [6.64; 8.04], mean = 1.06, SD = 0.40); 9.94 years (95% CrI [8.66; 12.55], mean = 0.82, SD = 0.23); 35.97 years (95% CrI [27.30; 39.69], mean = 0.65, SD = 0.23); and 44.01 years (95% CrI [40.37; 44.43], mean = 0.79, SD = 0.40). In short: we found a rapid initial improvement in gaze following in early childhood, followed by a long period of minor, very slow change with slightly increasing levels of imprecision toward old age.

2.6 | Discussion

We investigated the shape of change in gaze following across the lifespan and found a non-linear developmental trajectory, in which young children quickly enhanced their level of proficiency. Performance peaked around early adulthood, while there was a minor decay
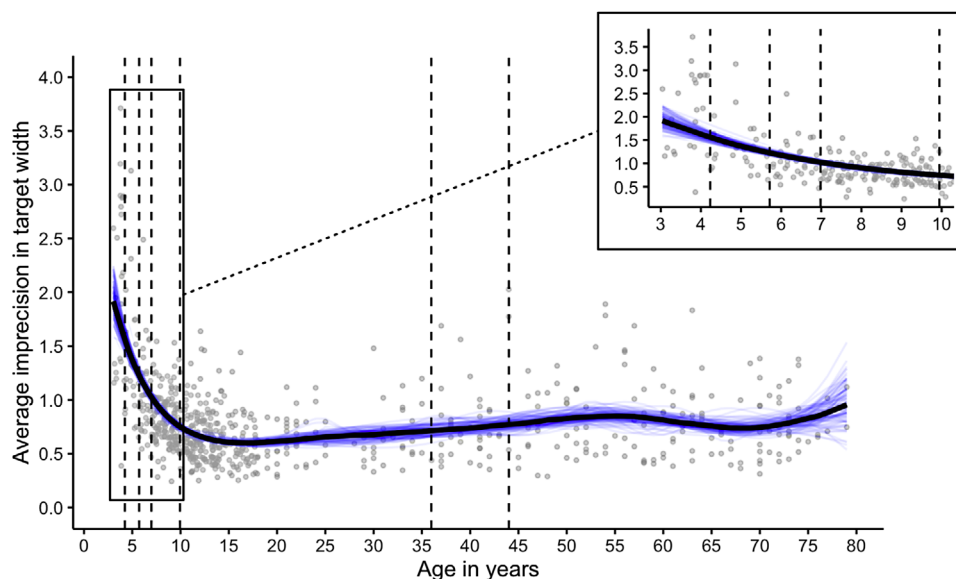
FIGURE 1 Developmental trajectory of gaze following across the lifespan. Gray dots show the mean level of imprecision (i.e., absolute distance between the target's center and the participant's click) for each subject averaged across trials. The unit of imprecision is counted in target width, that is, a participant with imprecision of 1 clicked on average one target width to the left or right of the true target center. Blue lines show 100 draws from the expectation of the posterior predictive distribution of the Gaussian Process model, with its mean predicted developmental trajectory as a solid black line. Vertical, black, dashed lines show the locations of the changes with the highest posterior probability according to our Bayesian change point analysis.

in later adulthood. These results support the idea that humans fine-tune their existing gaze following ability after the first emergence in infancy. Furthermore, we observed substantial individual differences in all age groups. While variation was highest in the 3- and 4-year-olds, it remained relatively stable across the lifespan.

Previous studies found that already 4-month-olds demonstrate basic gaze following abilities (Del Bianco et al., 2019). Since we measured an active location choice on a touchscreen, we could not collect data from infants. In our sample, 3-year-olds were still rather imprecise in their gaze following ability (average imprecision was approx. two target widths). How can we explain this divergence? First, we used subtle eye movements as cues. Many existing studies let the agents move eye and head in parallel (Behne et al., 2005; Povinelli et al., 1997), establishing a confound with the more salient head movement. Relying exclusively on eye movements might be more difficult for children than presenting them with a combined eye and head orientation (Carpenter et al., 1998). Silverstein et al. (2021) used a similar manipulation of gaze cues without head rotation and found that 6- to 18-month-olds were around or just above chance for gaze following. The authors argue that infants might fixate on another's face most of the time, while eye movement alone might not be strong enough to guide their attention. Furthermore, our study required participants to (1) precisely follow an agent's gaze, (2) interpret this as a cue, and (3) use the cue to guide their own behavior. It is conceivable that 3-year-olds followed the agent's gaze but did not translate this into precise, active behavior. Moll and Kadipasaoglu (2013) argue that social forms of perspective-taking evolve prior to visual perspective-taking, which only emerges within the third year of life. Young children might simply not be interested in a differential, spatial representation of the surrounding objects. Taken

together, this might explain why our younger participants located the agent's gaze rather imprecisely.

Regarding our sample of elderly adults, we expect a sampling bias (Bethlehem, 2010; Gosling et al., 2004; Remillard et al., 2014). First, certainly, not all older people have a high-speed internet connection or are knowledgeable in its use. Second, the elderly adults participating in *Prolific* studies might show greater cognitive flexibility compared to their offline counterparts. Therefore, a representative sample may show a greater age decline in gaze following compared to our reported sample. In addition, older people might be more likely to suffer from visual impairments. Even though we filtered participants to only include normal- to correct-to-normal vision, we cannot guarantee that our participants showed no symptoms of reduced vision.

3 | STUDY 2: COMPUTATIONAL COGNITIVE MODEL

Our lifespan study showed that gaze following develops throughout childhood, and variation between individuals appears in all age groups. The TANGO has previously been shown to reliably capture inter-individual differences in gaze following (Prein et al., 2023). The variation between participants was thus likely genuine and not due to random noise. In Study 2, we aimed to understand the developmental change and individual differences on a process level. We present a theory of gaze following that explains how participants process the available gaze information and trace a line-of-sight to identify the agent's focus. We formalized this inference process in a computational cognitive model that replicates a schematic representation of how
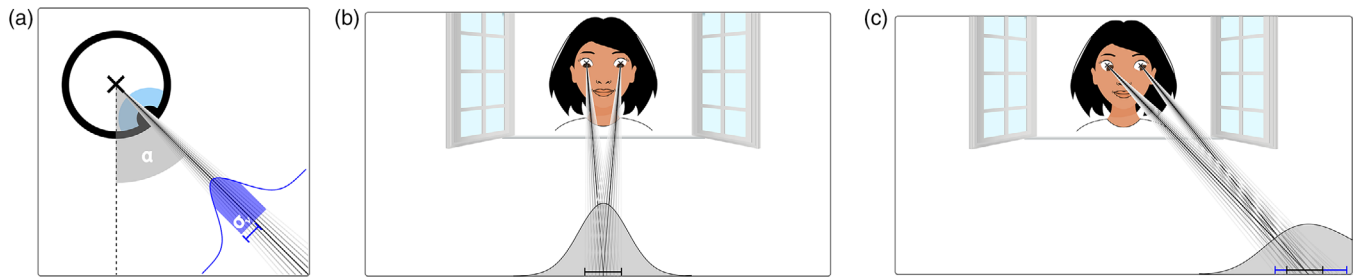
FIGURE 2 Gaze model. (a) Visualization of the gaze model (simplified, for one eye). Participants are assumed to observe the pupil location and estimate the center of the agent's eye. Connecting these two-point estimates as a line yields the unique vector that extends from the center of the agent's eyeball through the center of the pupil to the attentional focal point (line-of-sight). The angle between this vector and a line pointing vertically to the ground (black dashed line) is the pupil angle (α). Participants are assumed to sample (gray lines) from Normal distributions (blue line) centered around the true pupil angle. The variance of the Normal distribution (σ_v) is expected to vary between participants. (b, c) Geometrical features of the gaze model. As the pupil angle varies, a fixed amount of uncertainty in the angle corresponds to a varying degree of uncertainty in the estimated target location. The distribution around the target location from which participants sample is wider when the agent gazes toward the side than when she gazes centrally. The blue line on the ground shows the added level of uncertainty in the estimated target position for the target location further outward (c).

participants make inferences in the task's context (i.e., a model of the task and not the data).

The study design and procedure obtained ethical clearance in the same way as Study 1. The study and the model were pre-registered prior to data collection: https://osf.io/r3bhn. Data were collected between May and August 2021.

3.1 | Computational model

Our model quantifies a participant's cognitive ability to follow gaze by inverting a probabilistic process that generates the participant's clicks from observing the eyes of the agent. It is formally defined as:

$$P(\theta|x_c, \alpha_l, \alpha_r) \propto P(x_c|\alpha_l, \alpha_r, \theta) P(\theta)$$

where θ is an individual's cognitive ability to locate the focus of the agent's attention, x_c is the coordinate the participant clicked, and α_l and α_r are the pupil angles for the left and right eye, respectively. The pupil angle α is defined as the angle between a line connecting the center of the eye to the pupil and a line extended vertically downward from the center of the eye (see Figure 2a)[4]. Please note that in our case, the center of the pupil is simultaneously the center of the iris (see Anstis (2018) for the influence of moving irises, pupils, and corneal reflexes).

Based on our verbal task instructions, we assumed that participants (1) expected the agent's looks to be directed at the target, and (2) to click on the coordinate they estimated the agent to look at. Consequently, we did not assume that participants' clicks were noisy in any way but that they clicked on the screen location where they genuinely thought the target was (and that the agent was looking at).

The true eye angles (α_l and α_r) cannot be directly observed and have to be estimated based on the position of the pupils within the eyes, resulting in approximate values ($\hat{\alpha}_l$ and $\hat{\alpha}_r$). We presumed this estimation to be a noisy process. Thus, we conceptualized the development

of the cognitive ability to follow gaze as a reduction of noise in the estimates (i.e., an increased certainty about the pupil angles).

Any clicked value of x_c implied a "matched pair" of the estimated pupil angles $\hat{\alpha}_l$ and $\hat{\alpha}_r$, with the property that lines extended along those two angles met at the precise location of the target. As a consequence, we can rewrite the likelihood function of the model above:

$$P(x_c|\alpha_l, \alpha_r, \theta) \propto P(\hat{\alpha}_l, \hat{\alpha}_r|\alpha_l, \alpha_r, \theta) P(x_c)$$

$P(x_c)$ is a prior over potential target locations, which we assumed to be skewed towards the screen center: We anticipated that participants have an a priori expectation that the target will land close to the middle, because the target was last visible in the screen center before disappearing behind the hedge and because the agent was located centrally on the screen. We estimated the strength of this center bias (i.e., the standard deviation of a Normal distribution around the screen center) based on the data: $P(x_c) \sim \mathcal{N}(960, \sigma^p)$.

The width of this distribution is defined by σ^p. For children, we assumed that the center bias changed with age and estimated σ^p via a linear regression as a function of the child's age (age_i): $\sigma^p = \beta_0^{\sigma^p} + age_i \cdot \beta_1^{\sigma^p}$. Therefore, the participant-specific distribution for $P(x_c)$ was constrained by the performance in the TANGO and the child's age. For the adults, σ^p was not age-specific.

The main inferential task for the participant lay in estimating the pupil angles, that is, sampling from the first term of the right-hand side equation above, $P(\hat{\alpha}_l, \hat{\alpha}_r|\alpha_l, \alpha_r, \theta)$. For this, we assumed that the pair of estimated pupil angles was sampled from a probability distribution which is the product of two Normal distributions of equal variance, σ_v, centered on the true pupil angles:

$$P(\hat{\alpha}_l, \hat{\alpha}_r|\alpha_l, \alpha_r, \theta) \propto \phi(\hat{\alpha}_l; \alpha_l, \sigma_v) \phi(\hat{\alpha}_r; \alpha_r, \sigma_v),$$

As σ_v determined the level of accuracy with which participants estimated the pupil angles, it is the component of the model that defines θ. When σ_v is very small (i.e., the distribution around the pupil

angle is narrow), clicks far away from the target are unlikely, as these would require estimated pupil angles very different from the true pupil angles. When σ_v is very large (i.e., the distribution around the pupil angle is wide), almost any pupil angles may be sampled, corresponding to a roughly uniform distribution over click coordinates. We expected σ_v to vary between individuals. Consequently, individuals differed in the level of precision with which they can locate the target based on observing the agent's eyes.

The shape of the $P(x_c|\alpha_l,\alpha_r,\theta)$ distribution leads to a testable group-level prediction. As the pupil location varies, a fixed amount of uncertainty around the pupil angle corresponds to a varying degree of uncertainty in the estimated target location (see Figure 2b,c). When the agent directs their gaze toward the very left or right side, the distribution around the target location from which participants sample is comparatively wider than when the agent gazes centrally to the ground. For illustrative purposes, imagine a similar phenomenon: pointing a torch light to a flat surface on the ground. When one points the light cone directly at the surface, the light beam is concentrated in a clearly defined, small, symmetric area. When one points the light cone further away from oneself (shining at an angle), the light from one half of the cone must travel further to reach the surface than the light from the other half, resulting in an asymmetric light pattern. As the angle increases, the light is spread over a wider area, and the surface is illuminated less evenly. Consequently, for the same σ_v, the further out a target coordinate lies, the wider and less symmetric the distribution. This increases both the variance and the bias in a participant's estimate of the agent's attentional focus, resulting in a decreased performance in the task. As σ_v decreases and the cone narrows, the extent to which performance varies at different angles decreases.

Our gaze model consequently predicted that TANGO trials vary in difficulty (see Figure 2b,c): participants should be more imprecise in locating the target the further out it lands, resulting in a U-shaped pattern[5]. If our data matched the pattern of this model prediction, this could act as evidence for the gaze model. Therefore, our gaze model provides a quantitative theory of gaze following. In the following, we tested these predictions in children and adults.

3.2 | Participants

The sample consisted of 60 children, including 20 3-year-olds (mean age = 3.47 years, SD = 0.34, range = 3.07–3.97, 11 girls), 20 4-year-olds (mean age = 4.61 years, SD = 0.26, range = 4.09–4.98, 10 girls), 20 5-year-olds (mean age = 5.66 years, SD = 0.24, range = 5.01–5.96, 12 girls). Children were recruited via an internal database, where parent previously consented to child development studies, and data was collected in kindergartens in Leipzig, Germany.

In addition, we included 50 adults from Study 1 (mean age = 31.92 years, SD = 12.15, range = 18–63, 36 female). Since developmental change was minimal in our adult sample (see Study 1) and the cognitive models were computationally heavy, we decided to only include the first 50 adults who had completed the study.

3.3 | Procedure

We applied the same procedure as in Study 1. Children were tested in a quiet room in their kindergarten, while an experimenter guided the child through the study on a tablet. Adults participated online.

3.4 | Analysis

We quantified how well our gaze model explained the gaze following process in two ways. First, we aggregated the model predictions and data for each target bin and age group (3-, 4-, 5-years-olds, adults), and computed correlation to quantify how well the model was able to recover the data. Second, we compared the predictions of our gaze model to two simple alternative models that assume participants do not rely on the agent's gaze at all: a random guessing model and a center bias model. The random guessing model assumed participants randomly clicked on the screen and was implemented as sampling from a uniform distribution over all possible coordinates, $\mathcal{U}(0, 1920)$. The center bias model assumed participants always clicked near the screen center and was implemented as sampling from a Normal distribution with the screen center as the mean, and one balloon width as the variance, $\mathcal{N}(960, 160)$. Note that the center bias model also predicted imprecision should be higher for targets further out on the screen. However, compared to the gaze model, it predicted a steep effect towards the sides, resulting in a V-shaped pattern (imprecision as the distance between the target location and the screen center). All cognitive models were implemented in WebPPL (Goodman & Stuhlmüller, 2014).[6]

We compared models via the marginal likelihood of the data under each model. The pairwise ratio of marginal likelihoods for two models is also known as the Bayes Factor, which quantifies the quality of a model's predictions by averaging over the possible values of the model's parameters weighted by the prior probabilities of those parameter values. It can be used to estimate how much more likely the data under one model are compared to the other. Bayes Factors implicitly consider model complexity: models with more parameters often have broader prior distributions over parameters, which might weaken potential gains in predictive accuracy.

3.5 | Results

We found very clear support for our gaze model, both in children as well as adults. A strong correlation between the data mean and the gaze model estimate (σ_v) showed that the data mean is suitable to quantify individual differences: child sample $r(60) = 0.95$, $p < 0.001$, 95%CI [0.92, 0.97]; adult sample $r(50) = 0.96$, $p < 0.001$, 95%CI [0.93, 0.98]. The gaze model predicted a U-shaped pattern which we also observed in our data (see Figure 3c). The model comparison strongly favored our gaze model over the center bias model (child sample $logBF_{10} = 1015.33$; adult sample $logBF_{10} = 2575.75$)
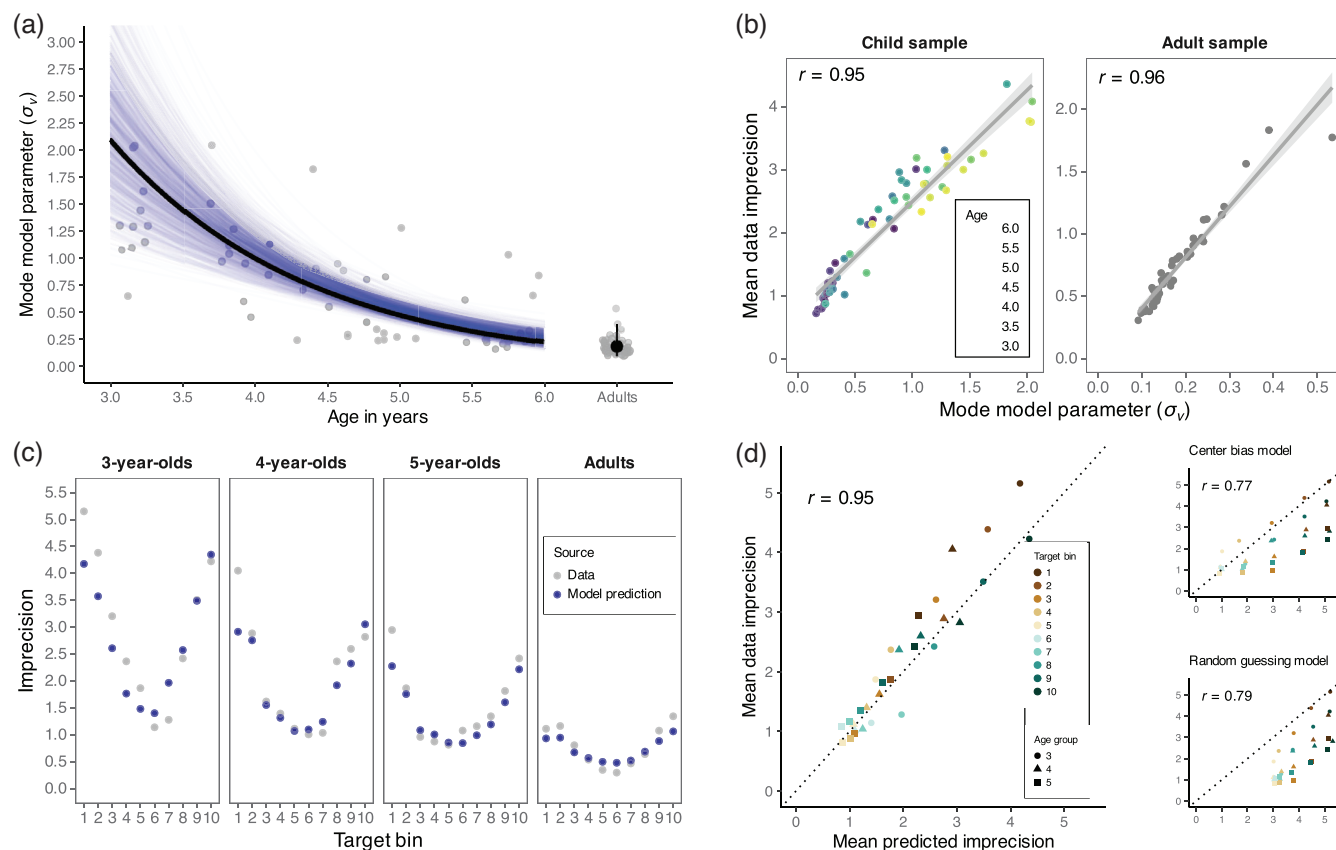
FIGURE 3 Gaze model results. (a) Developmental trajectory of the estimated model parameter. Gray dots show individual level parameter values. The black line shows the maximum a posteriori (MAP) estimate; blue lines show 1000 draws from it. The large black point with 95% HDI shows the mean of the model parameter for the adult sample. We added minimal horizontal and vertical noise to the adult individual level parameter values to avoid overplotting. (b) Correlation between estimated mode of the model parameter and data mean per individual, faceted by age group. In the child sample, color denotes age in years. Gray regression lines with 95% CI show smooth conditional means based on linear models, with Pearson's correlation coefficient *r*. (c) Pattern recovery. Imprecision in target width for each target bin by age group. Model predictions in blue; data in gray. (d) Correlation between the observed data and the predictions of the three models by target position and age group (across individuals; only for the child sample).

and the random guessing model (child sample $logBF_{10}$ = 388.98; adult sample $logBF_{10}$ = 919.03). For the child sample, we went one step further into the analysis. When correlating the observed data across all target positions and age groups with the predictions of the three models, we found a high similarity for the gaze model: $r(30)$ = 0.95, $p < 0.001$, 95%CI [0.90, 0.98], while the correlations with the alternative models were smaller (center bias model: $r(30)$ = 0.77, $p < 0.001$, 95%CI [0.57, 0.89]; random guessing model: $r(30)$ = 0.79, $p < 0.001$, 95%CI [0.59, 0.89]). The gaze model's prior over potential target locations assumed that participants' clicks would be skewed towards the screen center. For children, we estimated the width of the prior's distribution based on the participants' age. Interestingly, we found that the differential age effects on the center bias prior were minor (intercept = 300.14, 95% highest-density interval (HDI) [277.19; 321.28]; slope = 1.18, 95%HDI [0.00; 18.78]). The slope of the prior indicated that older children were only marginally less drawn towards the screen center compared to the younger children.

3.6 | Discussion

We presented a formal cognitive model of gaze following to describe how gaze following develops with age and varies between individuals. We modeled gaze following as a process in which participants estimate pupil angles based on the pupil location within the eye. By following the resulting gaze vector, they consequently arrive at the attentional focus of the agent. Individual differences can be explained as varying levels of imprecision in the pupil angle estimation. We assume the basic process of gaze following to be the same across the lifespan, though individuals become increasingly precise with age. While we did not include an infancy sample, we believe the process of gaze following should operate similarly—if not other cues such as the object saliencies or head directions are followed instead. By conducting model comparisons, we ruled out simpler explanations of the data.

In addition, we observed differences in performance depending on gaze location: participant data showed that precision levels dropped as the agent's gaze moved further away from the center.

Our gaze model predictions recovered this "signature pattern" in the data. Future research could use this signature in the data as evidence of whether diverse communities employ the same inferential mechanism to solve the task, speaking for a shared cognitive architecture.

Interestingly, the U-shaped pattern in the TANGO can be conceptually compared to the result patterns of Michelon and Zacks (2006): in their Level 1 perspective-taking task, an increased distance between the agent and the target decreased performance in adults (i.e., reaction times). Targets closer to the midline were more easily traced than ones further away from the agent. The authors concluded that visual acuity is generally higher for locations on the vertical than the diagonal axis ("oblique effect"; [Appelle, 1972; Heeley et al., 1997; Mikellidou et al., 2015]). Similarly, Symons et al. (2004) have found that adults' acuity for gaze direction depends on the target location. Not only is the increased imprecision for TANGO trials in which the target lands further out consistent with this finding, but our gaze model poses a viable explanation for this effect.

A limitation of our model is that we cannot disentangle how much of the participants' uncertainty comes from a noisy estimate of the agent's attentional focus and how much is due to imprecise clicking (e.g., experiencing motor issues, adding random noise to the click). However, we believe imprecise clicking to be of minor concern, since the children in our sample seemed determined in where they clicked and to not have issues with aiming or motor control (see precision in training trials in Supplementary information).

A critical feature of our model is that it assumes gaze following to rely on vector following: subjects are modeled to calculate pupil angles which serve as gaze vectors to point to the attentional focus point of an agent. Even though this vector following component is a geometrical calculation, one must first interpret the agent's eyes as a relevant social stimulus. Therefore, our model describes gaze following as a particular form of vector following in a social context.

Previous models in the social- and non-social cognitive domain exist that also make use of the concept of vector estimation (see Madl et al. (2015) for a review of computational models for spatial memory). For example, Brom et al. (2012) defined egocentric self-to-object vectors and allocentric object-to-object vectors where corresponding uncertainties due to memory and motor errors were modeled by vector weights and Gaussian noise. The counterfactual simulation model (Gerstenberg et al., 2014, 2021; Gerstenberg & Lagnado, 2012; Gerstenberg & Stephan, 2021) used vector calculations to model causal reasoning; for example, in estimating billiard ball trajectories and their collisions. Gaussian noise was added to the ball's movement vector to represent people's prior expectations, uncertainties, and expertise (the stronger the expectations, the less noise is added to the vector). Therefore, the way these models capture individual differences resembles the approach taken in the here-proposed gaze model.

The idea of gaze as vector following has recently been introduced to the study of pointing gestures (O'Madagain et al. (2019), see Li et al. (2023) for computational modeling). According to this hypothesis, when humans use pointing gestures, the index finger does not consti-

tute an "arrow" pointing at the target, but is used to touch the object within one's visual field. Consequently, the touch line is a vector that connects the eye and the pointed fingertip, and marks the object that the person's fingertip seems closest to touching from her perspective (O'Madagain et al., 2019). This supports the idea that vector estimation is relevant for social-cognitive abilities.

4 | STUDY 3: COMPONENTS OF GAZE FOLLOWING

Study 3 examined the components of gaze following and whether it can be fully reduced to physical vector following. The positive link between TANGO and social-environmental factors like age of childcare entry (Prein et al., 2023) underline how social interaction is integral to gaze following. Nevertheless, it is unclear how gaze following relates to other forms of perspective-taking (Astor & Gredebäck, 2022). Therefore, we investigated associations between gaze following and other measures of social-cognitive abilities.

First, we experimentally isolated the vector following component of the TANGO. We designed a new non-social vector following task that shared all crucial design features of the TANGO. Second, we assessed children's social-cognitive abilities by administering a ToM task battery, comprising four tasks from the ToM scale by Wellman and Liu (Wellman & Liu, 2004) and two additional perspective-taking tasks (Flavell, Flavell, et al., 1981; Flavell, Everett, et al., 1981). We reasoned that the TANGO shares task demands with the non-social vector following task while it shares its social context with the ToM tasks. We could, therefore, imagine both an absence or presence of relationship between gaze following and ToM. As stated in our pre-registration, we further assessed whether the two perspective-taking tasks related to gaze following. Our reasoning was that similar underlying mechanisms might be needed to solve these tasks since they require participants to take into account another person's point of view.

The study design and procedure obtained ethical clearance in the same way as Study 1. The study was pre-registered prior to data collection: (https://osf.io/xsqkt). Data collection took place in Leipzig, Germany, between February and March 2023.

4.1 | Participants

The sample consisted of 102 children (mean age = 4.54 years, SD = 0.31, range = 3.99–5.03, 54 girls). Information on individual socio-economic status was not formally recorded.

4.2 | Procedure

Children were tested in a quiet room in their kindergarten. An experimenter guided the child through the study. For maximum control of extraneous participant variables, we employed a within-subjects study

design. Participants performed the tasks in this order: (1) non-social vector following task, (2) ToM task battery, (3) TANGO. We decided on a fixed order to compare participants' performance straight-forwardly with each other. To increase engagement and decrease fatigue or fuzziness, we switched between tablet tasks and tasks with personal interaction. We presented the non-social vector following task before the TANGO so that participants would not be biased to interpret the stimuli as "agent-like."

4.2.1 | Non-social vector following

Modeling the structure of the TANGO, we designed a non-social vector following task. This task was presented on a tablet and used the concept of magnetism. On the upper part of the screen, there was a tube with a circular window, containing a gearwheel. On the floor, there was a magnet. The magnet got switched on (with a cartoon-like sound), whereupon the gearwheel moved towards the magnet. The gearwheel moved so that its center aligned with the magnet center while staying inside the circular window. Participants were then asked to locate the magnet. Access to the magnet's true location was manipulated by a wooden wall: participants either had full, partial, or no visual access to the true magnet location. Compared to the TANGO, the circular window acted functionally similar to the agent's eyeball, while the gearwheel acted similar to the pupil. Participants were expected to estimate a vector from the center of the circular window to the gearwheel and extend this as a line toward the ground to locate the magnet. We deliberately decided against displaying an arrow: we aimed to keep the mechanistic functions of the TANGO and magnet stimuli as similar as possible. In both cases, the starting point of the vector needs to be estimated by the participant. With an arrow, we would have drastically reduced the level of uncertainty, since the arrow already displays all information (arrow tip as the "gaze" direction). Furthermore, we wanted to avoid referential or iconic stimuli.

Children received 19 trials with one full visual access trial, two partial visual access trials, and 16 test trials. The first trial of each type comprised a voice-over description of the presented events. We conducted our analysis with 15 test trials (excluding the voice-over trial). The outcome variable was imprecision, defined as the absolute distance between the magnet's x coordinate and the x coordinate of the participant's click. Magnet coordinates were randomized: The full width of the screen was divided into ten bins; each bin occurred equally often, while the same bin could occur in two consecutive trials; and exact coordinates within each bin were randomly generated.

4.2.2 | ToM task battery

We administered four tasks from the Wellman and Liu (2004) ToM scale (see Supplementary information for further detail). We excluded three tasks: the Diverse Desires task to avoid ceiling effects; and

both tasks involving emotions (Belief Emotion and Real-Apparent Emotion), as we aimed at assessing the "cold, cognitive" (vs. "emotional") aspects of social cognition. We added two perspective-taking level-2 tasks ((Flavell, Flavell, et al., 1981; Flavell, Everett, et al., 1981); where children were asked whether a turtle appeared to be on its back or feet/a worm lay on a red or blue blanket from the experimenter's point of view) with the aim of increasing the variability we can capture between individuals, and since we hypothesized that perspective-taking would rely on similar mechanisms than gaze following, both relying on another's person egocentric frame of reference.

4.2.3 | Gaze following

As in Study 1 and 2, we presented children with the TANGO (Prein et al., 2023). To accentuate the social aspect of the TANGO, we exchanged the animal agents (used in the previous two studies) with human faces, which were modeled after the local population in appearance (already created for another project on cross-cultural similarities in gaze following [https://osf.io/tdsvc]). This further highlighted the contrast (i.e., social vs. non-social context) to the non-social vector following task.[7]

4.3 | Analysis

By design, the TANGO and the non-social vector following task involved vector following. Based on our computational model, we expected children's performance in both tasks to correlate with each other. For each task, we calculated the mean level of imprecision for each subject and correlated them using *Pearson's* correlation coefficient.

For the ToM battery, we aggregated the score of all solved tasks. Please note that the ToM score acted as an umbrella term and included the two perspective-taking tasks. Regarding the relationship between the two vector following tasks and the ToM measures, we could imagine two possible scenarios: (A) If gaze following recruited a social-cognitive ability beyond geometric vector following, we expected that ToM measures would correlate more strongly with the gaze following task than with the non-social vector following task. (B) If gaze following relied purely on task-specific geometric processes, then the correlation between gaze following and ToM measures would be comparable to the correlation between non-social vector following and the ToM measures. For the association between the aggregate ToM scores and the gaze following/non-social vector following tasks, we used *Spearman's* rank correlation coefficients.

We compared the correlation between gaze following and ToM measures and the correlation between non-social vector following and ToM measures by using the Williams' test from the function cocor.dep.groups.overlap (designed for two dependent overlapping correlations) from the package cocor (Diedenhofen & Musch, 2015).
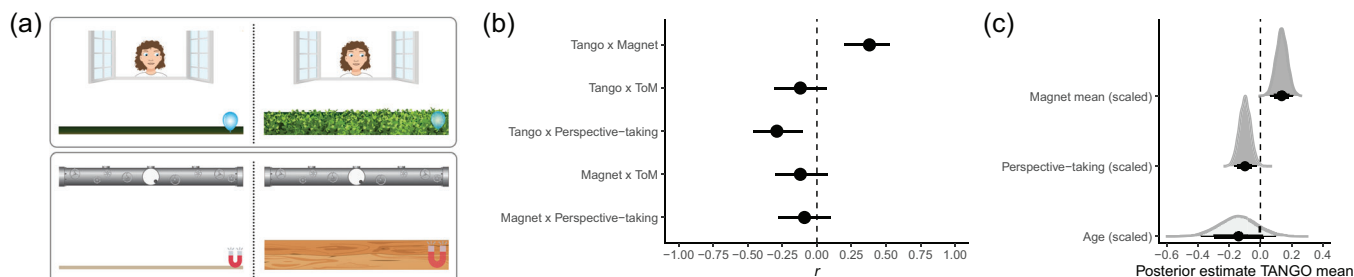
FIGURE 4 Components of gaze following. (a) Study procedures. Top: TANGO (i.e., gaze following; social vector following). Bottom: Magnet (i.e., non-social vector following). The left-hand side shows screenshots of the familiarization phase; right-hand side shows screenshots of the test phase. For illustrative purposes, translucent targets are shown in the test phase. Participants cannot see them. (b) Correlations between gaze following, non-social vector following, ToM, and perspective-taking. Dots show the correlation coefficients, error bars represent 95% CIs. Please note that the ToM score acts as an umbrella term and also includes two perspective-taking tasks. (c) Influence of age, perspective-taking and non-social vector following on gaze following. The graph shows the posterior distributions for the respective predictor. Black dots represent means, thicker black lines 80% CrI, and thinner black lines 95% CrI.

To estimate which components best explain the gaze following score, we conducted a model comparison with GLMMs predicting the mean imprecision in gaze following by age, imprecision in non-social vector following, the ToM aggregate score, or the aggregate of the two perspective-taking tasks (subset of ToM battery; example of model notation in R: tango_mean ~ age_centered + magnet_scaled + perspective_scaled). The outcome variable was modeled by a lognormal distribution. We wanted to assess whether the ToM aggregate score or the singled-out perspective-taking score added additional explanatory value when predicting the gaze following score. We hypothesized that perspective-taking seemed most closely theoretically related to gaze following as in both cases the participant was asked to judge another person's point of view.

4.4 | Results

Children performed similarly well in the social and non-social vector following tasks (gaze following mean = 1.92, SD = 0.77, range = [0.57–4.49]; vector following mean = 1.90, SD = 1.01, range = [0.57–5.83]). The mean aggregate ToM score was 3.46 (SD = 1.53, range = [0–6]), while the mean aggregate perspective-taking score was 1.38 (SD = 0.81, range = [0–2]).

Gaze following substantially correlated with the non-social vector following task, $r(102) = 0.38$, $p < 0.001$, 95%CI [0.20, 0.53] (see Figure 4b). While the tasks highly overlap in task demands, their measures shared some, but not all of their variance.

ToM abilities did not correlate with gaze following ($\rho(102) = -0.12$, $p = 0.22$, 95%CI [−0.31, 0.07]) or non-social vector following ($\rho(102) = -0.12$, $p = 0.25$, 95%CI [−0.30, 0.08]), and the correlations did not differ from each other, Williams' test $t(99) = 0$, $p = 1$.

Interestingly, gaze following and perspective-taking correlated with each other, $\rho(102) = -0.29$, $p < 0.01$, 95%CI [−0.46, −0.10]. Please note that the TANGO quantifies imprecision in gaze following. Therefore, a negative correlation suggests that imprecision in gaze following corresponds to less perspective-taking. Non-social vector following and

perspective-taking did not correlate, $\rho(102) = -0.09$, $p = 0.36$, 95%CI [−0.28, 0.10]. However, according to the Williams' test, the two correlations did not differ significantly from each other, $t(99) = -1.86$, $p = 0.06$.

Our model comparison revealed that gaze following was best predicted by a model including non-social vector following ($\beta = 0.14$, 95% CrI [0.06; 0.21]) and perspective-taking ($\beta = -0.10$; 95% CrI [−0.17, −0.03]), even when controlling for age ($\beta = -0.14$, 95% CrI [−0.38, 0.10]) (see Figure 4c and Supplementary information for the model comparison).

4.5 | Discussion

Our gaze model assumed vector calculations on a process level. By isolating non-social vector following experimentally, we could show that gaze following does indeed, to a certain degree, rely on this component. However, non-social vector following alone did not suffice to explain variation in gaze following. Additionally, perspective-taking proved to be a relevant social-cognitive ability when predicting the performance in the TANGO. The overall ToM aggregate score did not add explanatory power.

The TANGO and the two perspective-taking tasks can be seen as instances of visuospatial perspective-taking. Often, researchers distinguish between Level 1 and Level 2 perspective-taking tasks. Level 1 perspective-taking is concerned with the visibility of objects from a particular viewpoint (Surtees et al., 2013), and can usually be solved independently from another's agent frame of reference and without mental rotation [not matching the "mentalizing criterion"; Quesque and Rossetti (2020)]. In contrast, Level 2 perspective-taking tasks ask participants to judge how (visually or conceptually) the world looks for another person. Solving these tasks presumably relies on a simulation of mentally transforming one's own body schema into the space of another agent (Erle & Topolinski, 2017). Participants must understand that different viewpoints lead to different perspectives (Flavell, Everett, et al., 1981): two agents can see the same object "in different, incompatible, fine-grained ways" (Rakoczy, 2022). In short,

Level 1 perspective-taking addresses *what* an agent can see, while Level 2 perspective-taking addresses *how or where* they see it (Moll & Tomasello, 2006). The TANGO falls into the first category—however, in contrast to existing research, on a continuous scale—while the perspective-taking tasks presented within the ToM scale fall into the second category.

In the TANGO, participants locate the target by assuming that another agent perceives and looks at it. The here-applied Level 2 perspective-taking tasks add another layer by asking how one's perspective differs from another's and how exactly the other person sees an object. Michelon and Zacks (2006) assessed the different processes that participants used to master Level 1 and Level 2 perspective-taking tasks and concluded that participants rely on line-of-sight tracing in the former and perspective transformation in the latter case. Line-of-sight tracing consists of locating the avatar and the target, and drawing a line between them. Note how our gaze model in Study 2 shares the underlying idea of connecting points in space via a line. Level 2 perspective-taking requires an additional step in which participants "perform a perspective transformation so that one's imagined position matches the position of the avatar" (Michelon & Zacks, 2006). Therefore, the processes to solve Level 1 perspective-taking tasks might be computationally lighter since there is no need to adapt a new reference frame and there is no potential conflict between one's own and the other person's reference frame. Still, the assumed processes overlap, which could explain the correlation between the TANGO and the administered perspective-taking tasks.

Surtees et al. (2013) further differentiated between visual and spatial perspective-taking. While the former helps to judge if and how an agent sees an object, the latter involves judging the relative spatial locations of an agent and the object. Spatial perspective-taking does not necessitate mental states since computing a line of sight does not demand the presence of another agent (Michelon & Zacks, 2006) and can, therefore, be applied to non-agentive objects with a front (Surtees et al., 2013). This could explain how our participants solved the non-social vector following task.

Interestingly, we found weaker correlations between gaze following and the other ToM tasks. Similarly, in a longitudinal study, Brooks and Meltzoff (2015) found no direct association between gaze following at 10.5 months and explicit ToM at 4.5 years. While the ToM tasks and the TANGO shared the social context, the cognitive processes needed to solve each task might vary. As Rakoczy (2022) reflected, perception-goal psychology (which includes gaze following) comprises understanding that others see different objects or pursue different goals. However, this ability does not necessarily entail understanding more complex meta-representational aspects; for example, understanding that mental states can be false or involve aspectual information.

In previous work, we established that the TANGO is suited to capture meaningful variability across individuals (Prein et al., 2023), which is a crucial task feature when we are interested in revealing the relationship between different cognitive abilities. Importantly, the tasks we used to measure ToM abilities were not designed to capture individual differences: the aggregate score of few dichotomous items is of limited use when it comes to quantifying genuine differences between individuals. However, since these tasks are the gold standard in the social-cognitive literature (Białecka-Pikul et al., 2021; Byom & Mutlu, 2013; Poulin-Dubois et al., 2023; Rakoczy, 2022; Wellman, 2018), and measures with satisfying psychometric properties are, to the best of our knowledge, still scarce (e.g., Beaudoin et al., 2020; Mayes et al., 1996), we nonetheless relied on them in this study. Thus, lower correlations between ToM abilities and gaze following may reflect poor measurement characteristics on the side of ToM tasks rather than a genuine absence of association. We would like to point out that we already stated this concern in our pre-registration (https://osf.io/xsqkt).

If the reliability of the tasks at hand is known, one can estimate the "true" correlation between the latent constructs by applying an attenuation formula or structural equation models (Metsämuuronen, 2022; Trafimow, 2016). Adjusting for the measurement error would increase the so-called true correlation. While we can estimate the split-half reliability for the TANGO (Prein et al., 2023) and the non-social vector following task, we do not have reliability estimates of the ToM measures and cannot apply said approaches. This, in turn, underlines the importance of reporting the psychometric properties of a task. The development of new measures to capture individual differences in social-cognitive abilities seems essential to move this research further.

5 | GENERAL DISCUSSION

In three studies, we shed light on the cognitive process underlying gaze following and its developmental trajectory across the lifespan. Study 1 focused on how gaze following changes with age. We found a steep performance improvement in the preschool years in which children became more precise in locating the attentional focus of an agent. During teenage years and early adulthood, participants reached their peak performance. Precision levels then stayed comparably stable, with a minor decay toward older adulthood. Beyond these aggregated developmental patterns, we found that individual differences exist throughout the lifespan. In Study 2, we proposed a computational cognitive model that described gaze following at a process level. We modeled gaze following as a process in which participants use the pupil location within the eyes to estimate pupil angles. To locate the attentional focus of an agent (and find a target), they extend the resulting gaze vector towards the ground. Individuals vary in their levels of uncertainty around these pupil angles. Our gaze model outperformed two alternative models, which assumed participants solved the task via a center bias or random guessing. Knowing the TANGO to be a reliable individual differences measure, we investigated potential components of gaze following in Study 3. A fundamental assumption of our computational model was that gaze following relies on a vector following process. We experimentally isolated this component by designing a non-social vector following task. Furthermore, we assessed the relationship between gaze following and traditional ToM tasks. We found that gaze following does, indeed, share a substantial part of its variance with the non-social counterpart of vector following. Additionally,

perspective-taking correlated with gaze following, whereas the other ToM measures (focusing on diverse desires, knowledge access, and false beliefs) did not.

The developmental trajectory seen in Study 1 shows how abilities that emerge in infancy can continue to develop throughout childhood. While previous research established that 4-month-olds can follow gaze toward one out of two objects, this is not the end point of development. By studying gaze following on a continuum, we assessed not just *whether*, but *how precisely* children locate others' attentional focus. In previous work, we have shown that these individual differences are meaningful (e.g., connected to theoretically related constructs, and showing high split half and retest reliability; [Prein et al., 2023]). Capturing individual variation is crucial when we study development and the improvement in social-cognitive abilities.

Preschool children increased their precision level to locate an agent's attentional focus, which then stayed comparably stable across adulthood. Older adults decreased slightly in their precision levels. This developmental trajectory of a first emergence with a rapid improvement, followed by a plateau and slight decline toward older age, might be representative of many cognitive processes.

In Study 2, we proposed a theoretical framework to interpret the development and individual differences in gaze following. Our computational cognitive model assumes that participants estimate a pupil angle (i.e., the angle between a line extending vertically downwards from the pupil center and a line connecting the pupil and eye center; line-of-sight; see Figure 2a). The model parameter estimates a participant's latent ability to follow gaze and can explain why individuals differ in their precision to locate an agent's attentional focus. The model proposes that development in gaze following equals a reduction in noise when estimating the agent's pupil angles. We found strong evidence for the proposed gaze model when comparing it against two alternative models and correlating its predictions with the observed data (see Figure 3d), both for children and adults. Notably, the model recovers signature patterns in the data (see Figure 3c).

In Study 3, we tested the relation between gaze following and non-social vector following. As implied by our gaze model, we found that gaze following relates to the ability to estimate vectors in space. Already Butterworth and Jarrett (1991) have argued that, as the scene of actions, space is the commonality between different minds. The spatial vector following ability might be helpful in several social-cognitive tasks, for example, action prediction (Friesen & Rao, 2011), and intention understanding. Predicting which object another agent likely wants to grasp or calculating their movement pathway could rely on similar vector following abilities.

Gaze as displayed in the TANGO versus as in real-life social interactions differs with regard to which information is available. In the TANGO, the agent's eyes are big and round, with uniform colors, white sclera, fully visible iris and pupils, and continuous eye movement. The start position of the pupil is the center of the agent's eyeball—which, later on, needs to be estimated. Eye gaze in natural social interactions often provides less information and is more ambiguous. Even though the TANGO and our gaze model are designed within a 2D world, we believe the mechanisms can be extrapolated into the 3D

world. The processes of understanding gaze in daily life likely rely on the same principles as proposed in this paper. In Study 3, we presented the first evidence that this might be the case. We administered Level 2 perspective-taking tasks in which participants needed to adapt another person's frame of reference in a real-world social interaction. The correlation between this task and the TANGO speaks toward a unified mechanism behind these two visual perspective-taking tasks, regardless of the testing setup or stimulus features. Clearly, our real-life environment is visually more cluttered and diverse than the one presented in our tablet task. Here, however, additional informational sources are available to infer where others are looking; for example, body or head orientation or common ground (Bohn & Köymen, 2018; Moll & Kadipasaoglu, 2013; Osborne-Crowley, 2020). From a modeling perspective, a shared interaction history or diverse desires might be represented as non-uniform prior distributions over locations in the visual scenery. This way, our gaze model could be expanded to include more complex processes like mental state reasoning.

Theories of gaze following differ in whether they illustrate why children pay attention to gaze in the first place versus how they identify the exact location of gaze. While we described the process behind children's increasing precision in gaze following, we still need to further explore the driving forces behind this development. Existing theories broadly vary in how much importance they place on (A) experience and social environment, and (B) social awareness versus domain-general learning mechanisms (see categorization by Astor and Gredebäck [2022]). For example, it has been hypothesized that infants might be equipped with an innate gaze module and special neural mechanisms to detect eyes (Baron-Cohen, 1995; Batki et al., 2000); that children identify contingencies in social interactions and, in these, get reinforced to follow gaze (Corkum & Moore, 1998; Silverstein et al., 2021; Triesch et al., 2006); or that children are intrinsically socially motivated and simulate their own experiences to understand others (Astor et al., 2020; Friesen & Rao, 2011; Ishikawa et al., 2020; Meltzoff, 2007; Tomasello, 1999). As seen in Prein et al. (2023), precision in gaze following is linked to receptive vocabulary and opportunities for social interaction (e.g., number of siblings and age when entering childcare). Which exact kind of interactions are most helpful to improve precision in gaze following remains unknown.

6 | LIMITATIONS

In this paper, we have focused on studying variation across ages and individuals. However, our findings rely on participants from Western, Educated, Industrialized, Rich, and Democratic (WEIRD) backgrounds (Henrich et al., 2010). Cultural variation has been found in many foundational aspects of cognition and socialization: for example, in parent-child interaction and communication (Nielsen et al., 2017). First evidence suggests that cultural variation in face-to-face interactions does not influence infants' gaze shifts (Hernik & Broesch, 2019). While we cannot generalize our here-reported developmental trajectory to different socio-environmental settings, we predict that our presented process-model of gaze following holds true across communities.

Analyzing cross-cultural variation and checking for predicted "signature patterns" in the data will inform our modeling work and theory building in further detail (Amir & McAuliffe, 2020).

In Study 1, we recruited older participants online which might have selected a particular subgroup of this age range. Seventy-year-olds who have working Wi-Fi connections, know how to use a computer, and are registered on *Prolific* might not be representative of their age group. We can imagine that results from a more diverse, in-person data collection might show different developmental trajectories toward old age.

Our computational model of gaze following estimates one person-specific parameter for how accurately participants locate another person's attentional focus. The model assumes no motor imprecision in this estimation. However, younger children could have located the agent's focus at one particular point but clicked somewhere slightly off for motor control reasons. This would blur the model's estimation of the inferential component. However, in the first training trial, in which children were simply asked to touch the balloon, we found nearly perfect precision levels (cf. Prein et al., 2023). Motor issues and inaccurate aiming, resulting in falsely wide estimations in the model's inferential component, seem unlikely.

In Study 3, we matched the non-social vector following task as closely as possible to the TANGO. However, the starting positions differ: the magnet never appeared in the center of the screen. The starting point of the balloon might be especially important when interpreting the U-shaped pattern in Study 2. Furthermore, in the TANGO, two eyes are presented and information of the two (matching) cues needs to be integrated to infer the target's location. In the non-social vector following task, only one circular window with a gearwheel inside is presented as a directional cue, and there is no need to integrate two different information sources. In addition, we want to mention that the gaze presented in the TANGO might be more prominent compared to real-life social interactions (e.g., perfectly round and visible sclera; see discussion above). Future research should investigate how factors like self-propelled movement, spatial layout, and number of information sources influence the mechanisms of gaze following.

7 | CONCLUSION

In three studies, we have illuminated the lifelong development of precisely estimating another's gaze direction, and the psycho-physical process behind this. We have shown that gaze following continues to develop beyond infancy, and that individuals differ in their precision levels to localize the gaze direction of an agent. Our proposed process-level theory of gaze following modeled individual differences in precision as varying levels of uncertainty in the estimated gaze vectors. Consequently, we found that imprecision in gaze following relates to non-social vector following, as proposed by the model. Additionally, gaze following was linked to visual perspective-taking but no other aspects of ToM. The present research shows how precise and reliable measures and process models jointly inform each other and lead to a more comprehensive understanding of the psychological phenomenon in question.

CONFLICT OF INTEREST STATEMENT
The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT
The gaze following task (https://ccp-odc.eva.mpg.de/tango-demo/) is open source (https://github.com/ccp-eva/tango-demo). The data sets generated during and/or analyzed during the current study are available in the following repository (https://github.com/ccp-eva/gazecues-modeling). All experiments and analyses were pre-registered prior to data collection (https://osf.io/zjhsc/).

ETHICS STATEMENT
The study obtained ethical clearance by the MPG Ethics commission Munich, Germany, falling under an umbrella ethics application (Appl. No. 2021_45). Informed consent was obtained from all individual participants or their legal guardians. The research adhered to the legal requirements of psychological research with children in Germany.

ORCID
Julia Christin Prein https://orcid.org/0000-0002-3154-6167

ENDNOTES
[1] As Jasso and Triesch (2006) put it: "There is a general agreement that tests of gaze following in the absence of targets are more stringent than with targets, because that eliminates the possibility that infants are simply following the target's saliency" (p. 2).

[2] Originally, we fit our models on a trial-by-trial basis with the following structure: performance ~ age + symmetricPosition + trialNr + (1 + symmetricPosition + trialNr | subjID). However, the Gaussian Process model was computationally heavy. Thus, we simplified the model structure, aggregated data on a subject level, and included only age as an effect. See Supplementary information for a comparison between the original and the here-reported model structures. The model predictions did not differ notably.

[3] In an additional analysis (see Supplementary information) analysis, we varied the parameters of our change point analysis and modified the number of allowed change points, the minimum number of data points between change points, and the polynomial order. When we allowed more explorative room, the models became more sensitive and added more fine-grained change points. The exact location of the change points varied slightly. However, the overall interpretation stayed the same, fitting our initial visual inspection: while early childhood was characterized by much

change, adults showed a relatively stable level of imprecision, with a minor decay toward elderly adulthood. See Supplementary information for further detail.

[4] This model mirrors the logic of the TANGO programming code. In the online experiment, we read out the center point coordinates of the target and the agent's eyeball (i.e., the SVG coordinates), and then calculated a line between these two points: this was our gaze vector (acting in the functionally same way as a pupil angle). Knowing the eyeball radius, we calculated the point of intersection at which the gaze vector met the eyeball boundary. Finally, the agent's pupil moved from the center of the eyeball along the gaze vector to the intersection point. This way, the agent was animated to "look at" the target. In the gaze model, we assumed participants go through these steps in reverse order.

[5] In our screen-based study, this effect should decrease again towards the most outward sides. Since the computer screen has a natural border, trials in which the target lands furthest out to the left/right become slightly easier again. In these cases, the uncertainty about the pupil angle faces practically only the inner side (facing the center) of the screen, since the natural border of the screen limits where participants can click. In another adult sample with more trials, we could recover this pattern. For further elaboration, see Supplementary information.

[6] In an exploratory analysis, we simulated data for two more alternatives: a line-of-sight tracing model that assumed no inferential noise in the participants' gaze following ability, and a model building up on the line-of-sight tracing with added motor noise. Please note that these did not predict a U-shape pattern, since all target locations would be influenced by the motor noise equally. For further details, see Supplementary information.

[7] In an exploratory analysis, we compared children's imprecision levels in the TANGO task with animal versus human agents. Based on a GLMM analysis, we conclude that there was no evidence of a stable effect of stimulus choice. See Supplementary information for further detail.

REFERENCES

Amir, D., & McAuliffe, K. (2020). Cross-cultural, developmental psychology: Integrating approaches and key insights. *Evolution and Human Behavior*, 41(5), 430–444. https://doi.org/10.1016/j.evolhumbehav.2020.06.006

Anstis, S. (2018). The role of the pupil, corneal reflex, and iris in determining the perceived direction of gaze. *I-Perception*, 9(4), 2041669518765852. https://doi.org/10.1177/2041669518765852

Anstis, S. M., Mayhew, J. W., & Morley, T. (1969). The perception of where a face or television 'portrait' is looking. *The American Journal of Psychology*, 82(4), 474–489. https://doi.org/10.2307/1420441

Appelle, S. (1972). Perception and discrimination as a function of stimulus orientation: The "oblique effect" in man and animals. *Psychological Bulletin*, 78(4), 266–278. https://doi.org/10.1037/h0033117

Aslin, R. N. (2007). What's in a look? *Developmental Science*, 10(1), 48–53. https://doi.org/10.1111/j.1467-7687.2007.00563.x

Astor, K., & Gredebäck, G. (2022). Gaze following in infancy: Five big questions that the field should answer. *Advances in child development and behavior*, 191-223. https://doi.org/10.1016/bs.acdb.2022.04.003

Astor, K., Lindskog, M., Forssman, L., Kenward, B., Fransson, M., Skalkidou, A., Tharner, A., Cassé, J., & Gredebäck, G. (2020). Social and emotional contexts predict the development of gaze following in early infancy. *Royal Society Open Science*, 7(9), 201178. https://doi.org/10.1098/rsos.201178

Astor, K., Thiele, M., & Gredebäck, G. (2021). Gaze following emergence relies on both perceptual cues and social awareness. *Cognitive Development*, 60, 101121. https://doi.org/10.1016/j.cogdev.2021.101121

Baron-Cohen, S. (1995). *Mindblindness: An essay on autism and theory of mind*. The MIT Press.

Batki, A., Baron-Cohen, S., Wheelwright, S., Connellan, J., & Ahluwalia, J. (2000). Is there an innate gaze module? Evidence from human neonates. *Infant Behavior and Development*, 23(2), 223–229. https://doi.org/10.1016/S0163-6383(01)00037-6

Beaudoin, C., Leblanc, É., Gagner, C., & Beauchamp, M. H. (2020). Systematic review and inventory of theory of mind measures for young children. *Frontiers in Psychology*, 10, 2905. https://doi.org/10.3389/fpsyg.2019.02905

Behne, T., Carpenter, M., & Tomasello, M. (2005). One-year-olds comprehend the communicative intentions behind gestures in a hiding game. *Developmental Science*, 8(6), 492–499. https://doi.org/10.1111/j.1467-7687.2005.00440.x

Bethlehem, J. (2010). Selection bias in web surveys. *International Statistical Review*, 78(2), 161–188. https://doi.org/10.1111/j.1751-5823.2010.00112.x

Białecka-Pikul, M., Białek, A., Kosno, M., Stępień-Nycz, M., Blukacz, M., & Zubek, J. (2021). Early mindreading scale: From joint attention to false-belief understanding. *European Journal of Developmental Psychology*, 19(2), 301–318. https://doi.org/10.1080/17405629.2021.1911799

Birch, S. A. J., Li, V., Haddock, T., Ghrear, S. E., Brosseau-Liard, P., Baimel, A., & Whyte, M. (2017). Perspectives on perspective taking. *Advances in child development and behavior* (Vol., 52, pp. 185–226). Elsevier. https://doi.org/10.1016/bs.acdb.2016.10.005

Bohn, M., & Frank, M. C. (2019). The pervasive role of pragmatics in early language. *Annual Review of Developmental Psychology*, 1(1), 223–249. https://doi.org/10.1146/annurev-devpsych-121318-085037

Bohn, M., & Köymen, B. (2018). Common ground and development. *Child Development Perspectives*, 12(2), 104–108. https://doi.org/10.1111/cdep.12269

Brom, C., Vyhnánek, J., Lukavský, J., Waller, D., & Kadlec, R. (2012). A computational model of the allocentric and egocentric spatial memory by means of virtual agents, or how simple virtual agents can help to build complex computational models. *Cognitive Systems Research*, 17–18, 1–24. https://doi.org/10.1016/j.cogsys.2011.09.001

Brooks, R., & Meltzoff, A. N. (2002). The importance of eyes: How infants interpret adult looking behavior. *Developmental Psychology*, 38(6), 958–966. https://doi.org/10.1037/0012-1649.38.6.958

Brooks, R., & Meltzoff, A. N. (2015). Connecting the dots from infancy to childhood: A longitudinal study connecting gaze following, language, and explicit theory of mind. *Journal of Experimental Child Psychology*, 130, 67–78. https://doi.org/10.1016/j.jecp.2014.09.010

Bürkner, P.-C. (2017). Brms: An R package for bayesian multilevel models using stan. *Journal of Statistical Software*, 80(1), 1–28. https://doi.org/10.18637/jss.v080.i01

Bürkner, P.-C. (2018). Advanced bayesian multilevel modeling with the R Package brms. *The R Journal*, 10(1), 395. https://doi.org/10.32614/RJ-2018-017

Butterworth, G., & Jarrett, N. (1991). What minds have in common is space: Spatial mechanisms serving joint visual attention in infancy. *British Journal of Developmental Psychology*, 9(1), 55–72. https://doi.org/10.1111/j.2044-835X.1991.tb00862.x

Byers-Heinlein, K., Tsui, R. K. Y., van Renswoude, D., Black, A. K., Barr, R., Brown, A., Colomer, M., Durrant, S., Gampe, A., Gonzalez-Gomez, N., Hay, J. F., Hernik, M., Jartó, M., Kovács, Á. M., Laoun-Rubenstein, A., Lew-Williams, C., Liszkowski, U., Liu, L., Noble, C., … Singh, L. (2021). The development of gaze following in monolingual and bilingual infants: A multi-laboratory study. *Infancy*, 26(1), 4–38. https://doi.org/10.1111/infa.12360

Byom, L., & Mutlu, B. (2013). Theory of mind: Mechanisms, methods, and new directions. *Frontiers in Human Neuroscience*, 7, 413. https://doi.org/10.3389/fnhum.2013.00413

Carpenter, M., Nagell, K., & Tomasello, M. (1998). Social cognition, joint attention, and communicative competence from 9 to 15 months of age. *Monographs of the Society for Research in Child Development*, 63(4), i–vi, 1–143. https://doi.org/10.2307/1166214

Corkum, V., & Moore, C. (1995). Development of joint visual attention in infants. *Joint attention: Its origins and role in development* (pp. 61–83). Lawrence Erlbaum Associates, Inc.

Corkum, V., & Moore, C. (1998). The origins of joint visual attention in infants. *Developmental Psychology*, 34(1), 28–38. https://doi.org/10.1037/0012-1649.34.1.28

D'Entremont, B. (2000). A perceptual–attentional explanation of gaze following in 3- and 6-month-olds. *Developmental Science*, 3(3), 302–311. https://doi.org/10.1111/1467-7687.00124

D'Entremont, B., Hains, S. M. J., & Muir, D. W. (1997). A demonstration of gaze following in 3- to 6-month-olds. *Infant Behavior and Development*, 20(4), 569–572. https://doi.org/10.1016/S0163-6383(97)90048-5

Deák, G. O., Flom, R. A., & Pick, A. D. (2000). Effects of gesture and target on 12- and 18-month-olds' joint visual attention to objects in front of or behind them. *Developmental Psychology*, 36(4), 511–523. https://doi.org/10.1037/0012-1649.36.4.511

Del Bianco, T., Falck-Ytter, T., Thorup, E., & Gredebäck, G. (2019). The developmental origins of gaze-following in human infants. *Infancy*, 24(3), 433–454. https://doi.org/10.1111/infa.12276

Diedenhofen, B., & Musch, J. (2015). Cocor: A comprehensive solution for the statistical comparison of correlations. *PLoS ONE*, 10(4), e0121945. https://doi.org/10.1371/journal.pone.0121945

Emery, N. J. (2000). The eyes have it: The neuroethology, function and evolution of social gaze. *Neuroscience & Biobehavioral Reviews*, 24(6), 581–604. https://doi.org/10.1016/S0149-7634(00)00025-7

Erle, T. M., & Topolinski, S. (2017). The grounded nature of psychological perspective-taking. *Journal of Personality and Social Psychology*, 112(5), 683–695. https://doi.org/10.1037/pspa0000081

Flavell, J. H., Everett, B. A., Croft, K., & Flavell, E. R. (1981). Young children's knowledge about visual perception: Further evidence for the Level 1–Level 2 distinction. *Developmental Psychology*, 17(1), 99–103. https://doi.org/10.1037/0012-1649.17.1.99

Flavell, J. H., Flavell, E. R., Green, F. L., & Wilcox, S. A. (1981). The development of three spatial perspective-taking rules. *Child Development*, 52(1), 356–358. https://doi.org/10.2307/1129250

Friesen, A. L., & Rao, R. P. N. (2011). Gaze following as goal inference: A bayesian model. *Proceedings of the Annual Meeting of the Cognitive Science Society*, 33, 2457–2462.

Frith, C. D., & Frith, U. (2012). Mechanisms of social cognition. *Annual Review of Psychology*, 63(1), 287–313. https://doi.org/10.1146/annurev-psych-120710-100449

Gamer, M., & Hecht, H. (2007). Are you looking at me? Measuring the cone of gaze. *Journal of Experimental Psychology: Human Perception and Performance*, 33(3), 705–715. https://doi.org/10.1037/0096-1523.33.3.705

Gathercole, S. E., Pickering, S. J., Ambridge, B., & Wearing, H. (2004). The structure of working memory from 4 to 15 years of age. *Developmental Psychology*, 40(2), 177–190. https://doi.org/10.1037/0012-1649.40.2.177

Gerstenberg, T., Goodman, N. D., Lagnado, D. A., & Tenenbaum, J. B. (2021). A counterfactual simulation model of causal judgments for physical events. *Psychological Review*, 128(5), 936–975. https://doi.org/10.1037/rev0000281

Gerstenberg, T., Goodman, N., Lagnado, D., & Tenenbaum, J. (2014). From counterfactual simulation to causal judgment. *Proceedings of the Annual Meeting of the Cognitive Science Society*, 36, 523–528.

Gerstenberg, T., & Lagnado, D. A. (2012). Noisy Newtons: Unifying process and dependency accounts of causal attribution. *Proceedings of the Annual Meeting of the Cognitive Science Society*, 34, 378-383.

Gerstenberg, T., & Stephan, S. (2021). A counterfactual simulation model of causation by omission. *Cognition*, 216, 104842. https://doi.org/10.1016/j.cognition.2021.104842

Goodman, N. D., & Stuhlmüller, A. (2014). *The Design and Implementation of Probabilistic Programming Languages*. (Computer software). http://dippl.org

Gosling, S. D., Vazire, S., Srivastava, S., & John, O. P. (2004). Should we trust web-based studies? A comparative analysis of six preconceptions about internet questionnaires. *American Psychologist*, 59(2), 93–104. https://doi.org/10.1037/0003-066X.59.2.93

Gredebäck, G., Fikke, L., & Melinder, A. (2010). The development of joint visual attention: A longitudinal study of gaze following during interactions with mothers and strangers. *Developmental Science*, 13(6), 839–848. https://doi.org/10.1111/j.1467-7687.2009.00945.x

Guterstam, A., Kean, H. H., Webb, T. W., Kean, F. S., & Graziano, M. S. A. (2019). Implicit model of other people's visual attention as an invisible, force-carrying beam projecting from the eyes. *Proceedings of the National Academy of Sciences of the United States of America*, 116(1), 328–333. https://doi.org/10.1073/pnas.1816581115

Heeley, D. W., Buchanan-Smith, H. M., Cromwell, J. A., & Wright, J. S. (1997). The oblique effect in orientation acuity. *Vision Research*, 37(2), 235–242. https://doi.org/10.1016/S0042-6989(96)00097-1

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *The Behavioral and Brain Sciences*, 33(2-3), 61–135. https://doi.org/10.1017/S0140525X0999152X

Hernik, M., & Broesch, T. (2019). Infant gaze following depends on communicative signals: An eye-tracking study of 5- to 7-month-olds in Vanuatu. *Developmental Science*, 22(4), 1–8. https://doi.org/10.1111/desc.12779

Hessels, R. S. (2020). How does gaze to faces support face-to-face interaction? A review and perspective. *Psychonomic Bulletin & Review*, 27(5), 856–881. https://doi.org/10.3758/s13423-020-01715-w

Horstmann, G., & Linke, L. (2021). Examining gaze cone shape and size. *Perception*, 50(12), 1056–1065. https://doi.org/10.1177/03010066211059930

Ishikawa, M., Senju, A., & Itakura, S. (2020). Learning process of gaze following: Computational modeling based on reinforcement learning. *Frontiers in Psychology*, 11(213), 1–10. https://doi.org/10.3389/fpsyg.2020.00213

Ishikawa, M., Senju, A., Kato, M., & Itakura, S. (2022). Physiological arousal explains infant gaze following in various social contexts. *Royal Society Open Science*, 9(8), 220592. https://doi.org/10.1098/rsos.220592

Jasso, H., & Triesch, J. (2006). Using eye direction cues for gaze following—A developmental model. *Proceedings of the 5th International Conference on Development and Learning*.

Lau, B., & Triesch, J. (2004). Learning gaze following in space: A computational model. *Proceedings of the Third International Conference on Development and Learning*.

Lempers, J. D. (1979). Young children's production and comprehension of nonverbal deictic behaviors. *The Journal of Genetic Psychology*, 135(1), 93–102. https://doi.org/10.1080/00221325.1979.10533420

Lempers, J. D., Flavell, E. R., & Flavell, J. H. (1977). The development in very young children of tacit knowledge concerning visual perception. *Genetic Psychology Monographs*, 95(1), 3–53.

Li, Y., Chen, X., Zhao, H., Gong, J., Zhou, G., Rossano, F., & Zhu, Y. (2023). Understanding embodied reference with touch-line transformer. *The Eleventh International Conference on Learning Representations*.

Loomis, J. M., Kelly, J. W., Pusch, M., Bailenson, J. N., & Beall, A. C. (2008). Psychophysics of perceiving eye-gaze and head direction with peripheral vision: Implications for the dynamics of eye-gaze behavior. *Perception*, 37(9), 1443–1457. https://doi.org/10.1068/p5896

Madl, T., Chen, K., Montaldi, D., & Trappl, R. (2015). Computational cognitive models of spatial memory in navigation space: A review. *Neural Networks*, 65, 18–43. https://doi.org/10.1016/j.neunet.2015.01.002

Mayes, L. C., Klin, A., Tercyak, K. P., Cicchetti, D. V., & Cohen, D. J. (1996). Test-retest reliability for false-belief tasks. *Journal of Child Psychology and Psychiatry*, 37(3), 313–319. https://doi.org/10.1111/j.1469-7610.1996.tb01408.x

Meltzoff, A. N. (2007). "Like me": A foundation for social cognition. *Developmental Science*, 10(1), 126–134. https://doi.org/10.1111/j.1467-7687.2007.00574.x

Meltzoff, A. N., Brooks, R., Shon, A. P., & Rao, R. P. N. (2010). "Social" robots are psychological agents for infants: A test of gaze following. *Neural Networks*, 23(8), 966–972. https://doi.org/10.1016/j.neunet.2010.09.005

Metsämuuronen, J. (2022). Attenuation-corrected estimators of reliability. *Applied Psychological Measurement*, 46(8), 720–737. https://doi.org/10.1177/01466216221108131

Michel, C., Kayhan, E., Pauen, S., & Hoehl, S. (2021). Effects of reinforcement learning on gaze following of gaze and head direction in early infancy: An Interactive Eye-Tracking Study. *Child Development*, 92(4), e364–e382. https://doi.org/10.1111/cdev.13497

Michelon, P., & Zacks, J. M. (2006). Two kinds of visual perspective taking. *Perception & Psychophysics*, 68(2), 327–337. https://doi.org/10.3758/BF03193680

Mikellidou, K., Cicchini, G. M., Thompson, P. G., & Burr, D. C. (2015). The oblique effect is both allocentric and egocentric. *Journal of Vision*, 15(8), 24. https://doi.org/10.1167/15.8.24

Moll, H., & Kadipasaoglu, D. (2013). The primacy of social over visual perspective-taking. *Frontiers in Human Neuroscience*, 7(558), 1–9. https://doi.org/10.3389/fnhum.2013.00558

Moll, H., & Meltzoff, A. (2011). Perspective-taking and its foundation in joint attention. In J. Roessler, H. Lerman, & N. Eilan (Eds.), *Perception, causation, and objectivity* (pp. 286–304). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199692040.003.0016

Moll, H., & Tomasello, M. (2004). 12- and 18-month-old infants follow gaze to spaces behind barriers. *Developmental Science*, 7(1), F1–F9. https://doi.org/10.1111/j.1467-7687.2004.00315.x

Moll, H., & Tomasello, M. (2006). Level 1 perspective-taking at 24 months of age. *British Journal of Developmental Psychology*, 24(3), 603–613. https://doi.org/10.1348/026151005X55370

Moore, C. (2008). The Development of Gaze Following. *Child Development Perspectives*, 2(2), 66–70. https://doi.org/10.1111/j.1750-8606.2008.00052.x

Nielsen, M., Haun, D., Kärtner, J., & Legare, C. H. (2017). The persistent sampling bias in developmental psychology: A call to action. *Journal of Experimental Child Psychology*, 162, 31–38. https://doi.org/10.1016/j.jecp.2017.04.017

O'Madagain, C., Kachel, G., & Strickland, B. (2019). The origin of pointing: Evidence for the touch hypothesis. *Science Advances*, 5(7), eaav2558. https://doi.org/10.1126/sciadv.aav2558

Osborne-Crowley, K. (2020). Social Cognition in the real world: Reconnecting the study of social cognition with social reality. *Review of General Psychology*, 24(2), 144–158. https://doi.org/10.1177/1089268020906483

Palan, S., & Schitter, C. (2018). Prolific.ac—A subject pool for online experiments. *Journal of Behavioral and Experimental Finance*, 17, 22–27. https://doi.org/10.1016/j.jbef.2017.12.004

Perner, J., Brandl, J. L., & Garnham, A. (2003). What is a Perspective Problem? *Developmental Issues in Belief Ascription and Dual Identity. Facta Philosophica*, 5(2), 355–378. https://doi.org/10.5840/factaphil20035220

Pfeiffer, U. J., Vogeley, K., & Schilbach, L. (2013). From gaze cueing to dual eye-tracking: Novel approaches to investigate the neural correlates of gaze in social interaction. *Neuroscience & Biobehavioral Reviews*, 37(10, Part 2), 2516–2528. https://doi.org/10.1016/j.neubiorev.2013.07.017

Poulin-Dubois, D., Goldman, E. J., Meltzer, A., & Psaradellis, E. (2023). Discontinuity from implicit to explicit theory of mind from infancy to preschool age. *Cognitive Development*, 65, 101273. https://doi.org/10.1016/j.cogdev.2022.101273

Povinelli, D. J., Reaux, J. E., Bierschwale, D. T., Allain, A. D., & Simon, B. B. (1997). Exploitation of pointing as a referential gesture in young children, but not adolescent chimpanzees. *Cognitive Development*, 12(4), 423–461. https://doi.org/10.1016/S0885-2014(97)90017-4

Prein, J. C., Kalinke, S., Haun, D. B. M., & Bohn, M. (2023). TANGO: A reliable, open-source, browser-based task to assess individual differences in gaze understanding in 3 to 5-year-old children and adults. *Behavior Research Methods*, 56(3), 2469–2485. https://doi.org/10.3758/s13428-023-02159-5

Quesque, F., & Rossetti, Y. (2020). What Do Theory-of-mind tasks actually measure? Theory and practice. *Perspectives on Psychological Science*, 15(2), 384–396. https://doi.org/10.1177/1745691619896607

R Core Team. (2022). *R: A language and environment for statistical computing* [Manual]. R Foundation for Statistical Computing.

Rakoczy, H. (2022). Foundations of theory of mind and its development in early childhood. *Nature Reviews Psychology*, 1(4), 223–235. https://doi.org/10.1038/s44159-022-00037-z

Recasens, A., Khosla, A., Vondrick, C., & Torralba, A. (2015). Where are they looking? *Proceedings of the 28th International Conference on Neural Information Processing Systems Advances in Neural Information Processing Systems*, 1(2), 199–207.

Remillard, M. L., Mazor, K. M., Cutrona, S. L., Gurwitz, J. H., & Tjia, J. (2014). Systematic review of the use of online questionnaires of older adults. *Journal of the American Geriatrics Society*, 62(4), 696–705. https://doi.org/10.1111/jgs.12747

Silverstein, P., Feng, J., Westermann, G., Parise, E., & Twomey, K. E. (2021). Infants learn to follow gaze in stages: Evidence confirming a robotic prediction. *Open Mind*, 5, 174–188. https://doi.org/10.1162/opmi_a_00049

Stiefelhagen, R., & Zhu, J. (2002). Head orientation and gaze direction in meetings. *CHI '02 Extended Abstracts on Human Factors in Computing Systems*, 858–859. https://doi.org/10.1145/506443.506634

Surtees, A., Apperly, I., & Samson, D. (2013). Similarities and differences in visual and spatial perspective-taking processes. *Cognition*, 129(2), 426–438. https://doi.org/10.1016/j.cognition.2013.06.008

Symons, L. A., Lee, K., Cedrone, C. C., & Nishimura, M. (2004). What are you looking at? Acuity for triadic eye gaze. *The Journal of General Psychology*, 131(4), 451–469.

Todorović, D. (2006). Geometrical basis of perception of gaze direction. *Vision Research*, 46(21), 3549–3562. https://doi.org/10.1016/j.visres.2006.04.011

Tomasello, M. (1999). Social cognition before the revolution. In P. Rochat (Ed.), *Early social cognition: Understanding others in the first months of life*. Psychology Press (pp. 301–314). Lawrence Erlbaum Associates Publishers.

Tomasello, M., & Rakoczy, H. (2003). What makes human cognition unique? From individual to shared to collective intentionality. *Mind and Language*, 18(2), 121–147. https://doi.org/10.1111/1468-0017.00217

Trafimow, D. (2016). The attenuation of correlation coefficients: A statistical literacy issue. *Teaching Statistics*, 38(1), 25–28. https://doi.org/10.1111/test.12087

Triesch, J., Teuscher, C., Deák, G. O., & Carlson, E. (2006). Gaze following: Why (not) learn it? *Developmental Science*, 9(2), 125–147. https://doi.org/10.1111/j.1467-7687.2006.00470.x

Vehtari, A., Gelman, A., & Gabry, J. (2017). Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Statistics and Computing*, 27(5), 1413–1432. https://doi.org/10.1007/s11222-016-9696-4

Walker, L. D., & Gollin, E. S. (1977). Perspective role-taking in young children. *Journal of Experimental Child Psychology*, 24(2), 343–357. https://doi.org/10.1016/0022-0965(77)90012-1

Wellman, H. M. (2018). Theory of mind: The state of the art. *European Journal of Developmental Psychology*, 15(6), 728–755. https://doi.org/10.1080/17405629.2018.1435413

Wellman, H. M., & Liu, D. (2004). Scaling of Theory-of-Mind Tasks. *Child Development*, 75(2), 523–541. https://doi.org/10.1111/j.1467-8624.2004.00691.x

Yaniv, I., & Shatz, M. (1990). Heuristics of reasoning and analogy in children's visual perspective taking. *Child Development*, *61*(5), 1491–1501. https://doi.org/10.2307/1130758

Zhao, K., Wulder, M. A., Hu, T., Bright, R., Wu, Q., Qin, H., Li, Y., Toman, E., Mallick, B., Zhang, X., & Brown, M. (2019). Detecting change-point, trend, and seasonality in satellite time series data to track abrupt changes and nonlinear dynamics: A Bayesian ensemble algorithm. *Remote Sensing of Environment*, *232*, 111181. https://doi.org/10.1016/j.rse.2019.04.034

Zohary, E., Harari, D., Ullman, S., Ben-Zion, I., Doron, R., Attias, S., Porat, Y., Sklar, A. Y., & Mckyton, A. (2022). Gaze following requires early visual experience. *Proceedings of the National Academy of Sciences*, *119*(20), e2117184119. https://doi.org/10.1073/pnas.2117184119

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Prein, J. C., Maurits, L., Werwach, A., Haun, D. B. M., & Bohn, M. (2024). Variation in gaze following across the life span: A process-level perspective. *Developmental Science*, e13546. https://doi.org/10.1111/desc.13546