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

Abstract

Abstract max 250 words

*Keywords:* social-cognitive development, theory of mind, gaze cues, individual differences, cognitive modeling, lifespan

*Word count:* xxx max 8000

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

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

**Authors:** Julia Christin Prein1, Luke Maurits1, Annika Werwach1, Daniel B. M. Haun1,\*, Manuel Bohn1,2,\*

**Affiliations:** 1 Department of Comparative Cultural Psychology, Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany. 2 Institute of Psychology, Leuphana University Lüneburg, Germany. \* shared senior authorship

**ORCiD**: *Julia Christin Prein* <https://orcid.org/0000-0002-3154-6167>

**Conflicts of interest:** The authors declare that they have no conflict of interest.

**Data availability statement:** The web application (<https://ccp-odc.eva.mpg.de/tango-demo/>) described here is open source (<https://github.com/ccp-eva/tango-demo>). The data sets generated during and/or analysed during the current study are available in the [gazecues-modeling] repository (<https://github.com/ccp-eva/gazecues-modeling>). All experiments were pre-registered (<https://osf.io/zjhsc/>).

**Acknowledgements:** We thank Jana Jurkat for her help with data collection and participant recruitment. We would also like to thank Steven Kalinke for his technical programming support.

**Funding:** This study was funded by the Max Planck Society for the Advancement of Science, a noncommercial, publicly financed scientific organization (no grant number). Manuel Bohn was supported by a Jacobs Foundation Research Fellowship. We thank all the children, caregivers, and adults who participated in the study.

**Ethical statement:** The studies were approved by an internal ethics committee at the Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany.Informed consent was obtained from all individual participants or their legal guardians.

# Research highlights

* up to four bulleted points outlining the key contributions to research the paper makes.
* The Research Highlights should be placed before the abstract.
* Each research highlight should not be longer than 25 words.

# Introduction

* why do we care about developmental trajectory? ref to stat learning paper
* variation

## Why do we need gaze understanding?

How do humans learn about their environment and navigate through their social surroundings? One possibility to extract information from the environment is through following others’ focus of attention. Building a common ground is considered especially important in communicative interactions and shared activities (Tomasello, Hare, Lehmann, & Call, 2007).

## How does gaze following emerge?

Existing studies operationalize gaze following as the ability to follow another agent’s line of sight. As one of the most fundamental social-cognitive abilities, it has been extensively studied in infancy and early childhood. Infants as young as six months can attune their gaze to that of another agent (D’Entremont, Hains, & Muir, 1997). 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 (Moll & Tomasello, 2004).

While the emergence of gaze following has been well established, less is known about the developmental trajectory throughout childhood and adolescence. One possibility is that our social-cognitive ability in question is fully developed once emerged in infancy. However, many cognitive abilities develop with age (e.g., working memory, Gathercole, Pickering, Ambridge, & Wearing, 2004). Similarly, visual processing appears to improve with age. Therefore, children could potentially improve in gaze following, fine-tuning the performance of the already existing skill.

## The scope of infants’ gaze following ability

Though these studies suggest that young infants can align their visual attention to another’s line of sight, it does not necessarily include understanding the intentions of the other agent. Infants could simply attune their orientation or be attracted by others’ gaze without processing what exactly the other is seeing (cf. Butterworth & Jarrett’s ecological and geometric mechanism, Butterworth and Jarrett (1991)]. Therefore, it is crucial to study children’s intentional understanding of gaze.

Moore, Angelopoulos, and Bennett (1997) showed that 9-month-olds followed an agent’s gaze more, when it was accompanied by a dynamic head turn in comparison to a static head turn.

In a hiding game with two search locations, Povinelli, Reaux, Bierschwale, Allain, and Simon (1997) found that three-year-olds used gaze as a cue to locate the reward, while two-year-olds performed at chance level.

In a similar object choice paradigm with two containers, Behne, Carpenter, and Tomasello (2005) investigated whether infants understand the communicative intent behind pointing and gaze cues. In contrast to Povinelli et al. (1997), they found that already 14-month-olds used the agent’s cues to select an object. In conditions with absent-minded ‘cues’, infants performed around chance. This could be interpreted as infants recognizing the nature of this joint activity: namely, that the adult’s behavior was beneficial and relevant for their object choice.

### Head vs eye direction.

It is important to note that in many existing gaze conditions, the experimenter shifted their eyes and head in synchrony (e.g., Behne et al. (2005)). Instead of pointing towards gaze understanding, a critic could claim that the results can be explained by face direction alone.

A handful of studies approached this potential confound by separately manipulating head and eye movement. Brooks and Meltzoff (2002) implemented a comparison between eye and head orientation and found that 14-month-olds were sensitive to open versus closed eyes.

Investigating the ‘cooperative eye hypothesis’, Tomasello et al. (2007) implemented six conditions, in which an experimenter oriented towards the ceiling with their eyes only, head only (eyes closed), both head and eyes, or neither. They found that human infants relied more on the eye movement, while chimpanzees paid more attention to the head movement.

Importantly, the subjects were not presented with an object choice but their attention orientation was measured.

* (Raviv & Arnon, 2018)
* (Astor & Gredebäck, 2022)
* (Colombo, 2001)
* (Scaife & Bruner, 1975)
* (Itakura & Tanaka, 1998)
* (Carpenter, Nagell, & Tomasello, 1998) “Several other studies have attempted to determine more precisely the cue that infants are using when they follow the gaze direction of others, that is, whether they use adults’ head or eye orientation. In tasks comparing infants’ responses when the experimenters turned their head and eyes together to targets with their responses when the experimenters directed their eyes to the targets but their head remained facing forward, Corkum and Moore (1995), Lempers (1979), and Lempers, Flavell, and Flavell (1977) all found that only infants age 12 months and older responded correctly when eyes and head were oriented in the same direction and that infants at all ages (i.e., through 19 months) performed poorly when eye and head direction diverged” (p.10-11) object choice.
* (Silverstein, Feng, Westermann, Parise, & Twomey, 2021) for vertical plane
* (Zhang, Zhang, Zhang, Tang, & Liu, 2019)
* (Frischen, Bayliss, & Tipper, 2007)
* (Lee, Eskritt, Symons, & Muir, 1998)
* (Coelho, George, Conty, Hugueville, & Tijus, 2006)

## Aim of the current project

### Developmental trajectory, measuring & modeling individual differences.

In this study, we were interested in the developmental trajectory of gaze understanding. While we expect the younger children to be able to follow gaze, we aimed at assessing the differentiation of their social-cognitive ability. Our goal was *not* to establish the youngest age at which children understand gaze cues. Rather, we wanted to examine how that ability changes with age.

In our study, we focused on the communicative intents of gaze: we asked children to locate a target by following an agent’s gaze. While language demands were kept low, the participants had to actively respond and, therefore, make use of the presented gaze cue.

A unique contribution of this study is the richness of the data set. Methodological challenges arise when trying to compare data across ages from qualitatively and quantitatively different study tasks. We could circumvent these issues by applying the exact same task for the entire life span.

# Study 1: Gaze understanding across the lifespan

We aimed to assess the developmental trajectory of gaze understanding across the lifespan. First, we were interested in how this ability changes with age: Is the ability to understand gaze fully developed once emerged in infancy and stays stable across the lifetime? Or do we fine-tune our already existing ability in early adulthood? Do we then potentially even notice an age decay later in adulthood? Second, we were interested in individual differences across the age groups. Does the variation between individuals decrease, the older they get? Do all adults reach the same level of precision in gaze understanding?

Due to the sheer number of participants, we restricted this analysis to a remote sample. Pre-registrations can be found here: <https://osf.io/snju6> (child sample) and <https://osf.io/6yjz3> (adult sample). The study obtained ethical clearance by the MPG Ethics commission Munich, Germany, falling under an umbrella ethics application (Appl. No. 2021\_45), and was approved by an internal ethics committee at the Max Planck Institute for Evolutionary Anthropology. The research adheres to the legal requirements of psychological research with children in Germany. Data was collected between May 2021 and April 2023.

<https://osf.io/6yjz3>

## Participants

We collected data from a remote child, teenager and adult sample, reaching from 3 to 80 years of age (see Supplements for further details).

The child and teenager sample consisted of 471 participants. We recruited participants via an internal database consisting of families living in Leipzig, Germany, who volunteered to participate in child development studies and indicated an interest in online 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 (∼600,000 inhabitants; median individual monthly net income ~ 1,600€ 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 service from the University of Oxford with a predominantly European and US-American subject pool. Participants consisted of 240 English-speaking adults that reported to have normal or corrected-to-normal vision. For completing the study, subjects were paid above the fixed minimum wage (on average £10.00 per hour). For further information on age and gender of participants, see Table 1.

## Materials

We used the continuous version of the TANGO (Prein, Kalinke, Haun, & Bohn, 2023). The task was presented as an interactive web application (see Figure 1; live demo [https://ccp-odc.eva.mpg.de/tango-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 [with reliability estimates *Pearson’s r* ranging from .7 to .8 for the continuous task version; Prein et al. (2023)].

Each trial presented an agent standing in a window, watching a balloon (*i.e.*, target) falling to the ground. The target fell behind a hedge while the agent’s gaze followed the target’s trajectory: pupil and iris moved so that their center aligned with the target center. In test trials, the target flight was covered so that participants could not see where the target landed. Participants were tasked to locate the target by tracking the agent’s gaze. They could respond by touching on the screen.

Four familiarization trials ensured that participants understood the task and felt comfortable with the response format. Then, 15 test trials followed. Completing the 19 trials took approximately 5-10 minutes.

The outcome measure was imprecision, defined as the absolute difference between the target (i.e., balloon) center and the x coordinate of the participant’s click. The full screen width was divided into ten bins. Within each bin, exact target coordinates were randomly generated during runtime. Each target bin, as well as all agents and target colors, occurred equally often and did not appear in more than two consecutive trials.

## Procedure

Children and teenagers received a personalized link to the study website. Caregivers were asked to provide technical support whenever needed, while explicitly being reminded not to help their children 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.

## Analysis

We ran all analyses in R version 4.3.1 (2023-06-16) (R Core Team, 2022). Regression models were fit as Bayesian generalized linear mixed models (GLMMs) with default priors for all analyses, using the function brm from the package brms (Bürkner, 2017, 2018).

To estimate the effect of age on gaze understanding, we fit GLMMs that make different assumptions about the developmental trajectory, modeling the relationship as linear, quadratic, or cubic. In addition, we applied a Gaussian Process model (Bürkner, 2017) and examined nonlinearity. The models predicted imprecision by age (continuous), aggregated across trials and modeled as a lognormal distribution.[[1]](#footnote-40) The unit of imprecision was counted in target width, i.e., a participant with imprecision of 1 clicked on average one target 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 using leave-one-out cross-validation (LOO), the widely applicable information criterion (WAIC), and model weights (Vehtari, Gelman, & Gabry, 2017).

Confirming that the developmental change was non-linear, we performed a Bayesian change point analysis, using the package RBeast (Zhao et al., 2019). The function beast.irreg handled the irregular nature of our series, which did not have the same number of data points for each year in age. We were interested in finding the most prominent, 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 individual outlying data points, we constrained the model to have minimally 10 data points between two change points (i.e., corresponding to half of the data points we collected per adult decade). We inspected the probability of number of change points and the locations of these change points (mean and 95% CrI).[[2]](#footnote-41)

## Results

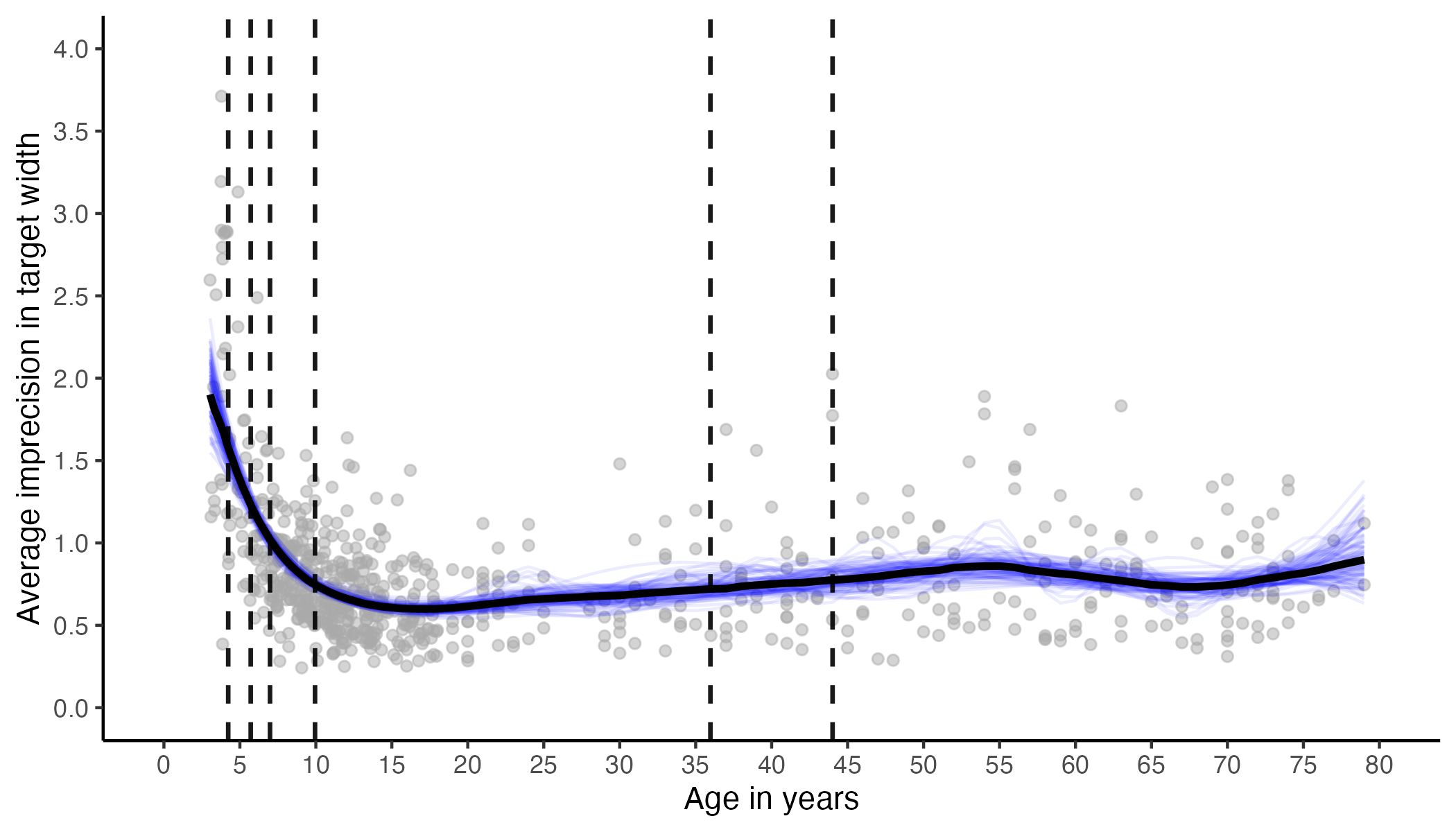


Figure 1: **Gaze model** (A) Visualization of the gaze model. is defined as the pupil angle, i.e., the angle between a line connecting the eye center to the pupil and a line extended vertically downward from the pupil. is the variance around the Normal distribution centered on the true pupil angle. This component is expected to vary between participants. (B) Developmental trajectory of the estimated model parameter. Grey dots show individual level parameter values. The black line shows the maximum a posteriori (MAP) estimate; blue lines show 1000 draws from it. (C) Correlation between estimated mode of the model parameter and data mean per individual, color-coded by age. The grey regression line with 95% CI shows smooth conditional mean based on a linear model, with *Pearson*’s correlation coefficient *r*. (D) Geometrical features of the gaze model. As the pupil location varies, a fixed amount of uncertainty around the gaze vector corresponds to a varying degree of uncertainty in the estimated target location. Top: Agent gazes centrally to the ground. Bottom: Agent gazes toward the side. The distribution around the gaze vector from which participants sample is wider compared to when the agent 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. (E) Pattern recovery. Imprecision in target width for each target bin by age group. Model predictions in blue; data in grey. (F) Correlation between the observed data and the predictions of the three models by target position (across age and individuals).

High levels of variation pointed to individual differences in all age groups (overall imprecision mean = 0.81, sd = 0.82, range = [0 - 10.73]). For example, there were some children that were more accurate than the average adult.

In our model comparison, we found clear evidence for a non-linear development in gaze understanding across the lifespan. Compared against a linear, quadratic, and cubic model, the Gaussian Process model showed the best model fit, demonstrated by the greatest model weight and smallest WAIC values. For the imprecision in gaze understanding, the standard deviation (*b* = 1.52, 95% CrI [0.25; 5.19]) and visual inspection indicated nonlinearity (see Figure 2).

Going one step further, we investigated the most prominent change points in the data. The Bayesian change point analysis revealed 6.10 (with 23.71% probability) major shifts in gaze understanding during the lifespan. The change points occurred at the following ages: 4.23 years (95% CrI [4.13; 4.33]); followed by 5.71 years (95% CrI [5.37; 5.90]); followed by 9.94 years (95% CrI [9.33; 12.47]); and finally at 6.98 years (95% CrI [6.61; 7.52]). In other words: we found a very rapid initial improvement in early childhood (three change points in rapid succession), followed by a long period of minor, very slow change with slightly increasing levels of imprecision toward the eldest in our sample.

## Discussion

We investigated the shape of change in gaze understanding across the lifespan. By applying the exact same task for the entire life span, we could directly compare gaze understanding in all ages. Therefore, we circumvented methodological challenges that often arise when trying to compare data across ages from qualitatively and quantitatively different study tasks.

We found a non-linear developmental trajectory in gaze understanding: Early in childhood, children quickly enhanced their level of proficiency. Performance peaked (i.e., imprecision was lowest) around early adulthood, while there was a minor decay in later adulthood. This is consistent with the view that we fine-tune our existing gaze understanding ability after the first emergence in early childhood. Furthermore, we observed individual differences in all age groups. While variation was highest in the three- and four-year-olds, it remained relatively stable across the lifespan.

Previous studies found that children start to follow gaze in the second half of their first year of life (Moll & Tomasello, 2004). In our sample, three-year-olds were still rather imprecise in their gaze understanding ability. How can we explain this divergence? First of all, 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), therefore establishing a confound with greater (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). Furthermore, our study required participants to (1) precisely follow an agent’s gaze, (2) interpret this as a cue, and then (3) make use of this cue to guide their own behavior (i.e., touching the screen at the cued location). We measured an active location choice instead of an attention orientation. It is conceivable that three-year-olds followed the agent’s gaze but were still learning to translate this understanding into precise, active behavior.

Regarding the sample of elderly adults, we expect a sampling bias to be present (Bethlehem, 2010; Gosling, Vazire, Srivastava, & John, 2004; Remillard, Mazor, Cutrona, Gurwitz, & Tjia, 2014). First, certainly not all older people have working a high-speed Internet connection or are knowledgeable and trained in its use. Second, it takes a certain amount of independence and motivation to participate in *Prolific* studies. The elderly who know how to participate in *Prolific* studies certainly might show greater cognitive fitness and flexibility compared to their offline counterparts. Therefore, a representative sample might show a greater age decline in gaze understanding compared to our reported sample. In addition, it is noteworthy that 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.

# Study 2: Computational cognitive model

Our lifespan study showed that gaze understanding develops throughout childhood, and variation between individuals appears in all age groups. The TANGO has previously been shown to be reliably capture inter-individual differences in gaze understanding (Prein et al., 2023). The variation between participants is thus likely genuine and not due to random noise. Now, we aim to understand what explains the developmental change and the variation across participants on a process level. We present a theory of gaze understanding that explains how children process the available information (i.e., the agent’s eyes) to make inferences about the agent’s gaze and its attentional focus, which leads them to identify the target position. We formalize this inference process in a computational cognitive model.

Computational modeling frameworks allow researchers to establish mechanistic explanations of psychological phenomena (Grahek, Schaller, & Tackett, 2021). As formal, mathematical accounts of the psychological process in question, they force researchers to accurately and comprehensively state all their underlying assumptions (Simmering, Triesch, Deák, & Spencer, 2010). In addition, models can be used to simulate behavior and form testable predictions. The expected patterns can then, in turn, be compared to the empirically observed behavior.

We seek to explain how participants solved the TANGO. Our gaze model assumes that participants use available gaze information to infer the target location.

Task design, data collection, and sample sizes were pre-registered: <https://osf.io/r3bhn>. The study design and procedure obtained ethical clearance by the MPG Ethics commission Munich, Germany, falling under a packaged ethics application (Appl. No. 2021\_45), and was approved by an internal ethics committee at the Max Planck Institute for Evolutionary Anthropology. The research adheres to the legal requirements of psychological research with children in Germany. Data were collected between May and August 2021.

## Participants

The sample consisted of 60 children, including 20 three-year-olds (mean age = 3.47 years, SD = 0.34, range = 3.07 - 3.97, 11 girls), 20 four-year-olds (mean age = 4.61 years, SD = 0.26, range = 4.09 - 4.98, 10 girls), 20 five-year-olds (mean age = 5.66 years, SD = 0.24, range = 5.01 - 5.96, 12 girls). Data of children was collected in kindergartens located in Leipzig, Germany. The children within each kindergarten were recruited via an internal database, where each parent priorly consented to child development studies.

In addition, we included 50 adults from our Lifespan study (mean age = 31.92 years, SD = 12.15, range = 18 - 63, 36 female). Adults were recruited over *Prolific* (Palan & Schitter, 2018). Since developmental change was minimal in our adult sample (see Lifespan study) and the cognitive models were computationally heavy, we decided to only include the first 50 adults that had completed the study.

## Procedure

We applied exactly the same procedure as in the first study, employing the continuous version of the TANGO (Prein et al., 2023). 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.

## 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:

where is an individual’s cognitive ability to locate the focus of the agent’s attention, is the coordinate the participant clicked, and and are the pupil angles for the left and right eye, respectively. The pupil angle 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 pupil.[[3]](#footnote-53)

Based on our verbal task instructions, we assume that participants (1) expect the agent’s looks to be directed at the target, and (2) to click on the coordinate they estimate the agent to look at. Consequently, we do not assume that participant’s clicks are noisy in any way but that they click on the screen location where they genuinely think the target is (and that the agent is looking at). However, the true eye angles ( and ) cannot be directly observed. These have to be estimated based on the position of the pupils within the eyes, resulting in approximate values ( and ). We presume this estimation to be a noisy process. Thus, the development of the cognitive ability to follow gaze corresponds to a reduction in the magnitude of the noise in the estimates (i.e., an increased certainty about the eye angles).

Any clicked value of implies a “matched pair” of the estimated pupil angles and , with the property that lines extended along those two angles meet at the precise location of the target. As a consequence, we can rewrite the likelihood function of the model above:

The second term of the right-hand side equation above, , is a prior over potential target locations, which we assume to be skewed towards the screen center: We anticipate that participants have an a priori expectation that the target will land close to the middle, partly because the target was last visible in the screen center before disappearing behind the hedge, and because the agent is located centrally on the screen. We estimate the strength of this center bias (i.e., the standard deviation of a Normal distribution around the center of the screen) based on the data.

The main inferential task for the participant lies in estimating the pupil angles, i.e., sampling from the first term of the right-hand side equation above, . For this, we assume that the pair of estimated angles are sampled from a probability distribution which is the product of two Normal distributions of equal variance centered on the true pupil angles:

Broadly summarizing: Participants are assumed to observe the pupil locations and estimate the center of the agent’s eyes. 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. Taking the angle between this vector and a line pointing vertically to the ground yields the pupil angle. Participants are assumed to sample from Normal distributions centered around the true pupil angle. They are assumed to do this independently for the left and right eye and then integrate the information to estimate the target’s location.

As determines the level of accuracy with which the participant estimates the pupil angles, it is the component that defines . When 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 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 expect to vary between individuals. Consequently, individuals differ in the level of precision with which they can locate the target based on observing the agent’s eyes.

The shape of the distribution leads to an interesting, 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. When the agent directs their gaze toward the very left or right side, the distribution around the pupil angle from which participants sample is comparatively wider than when the agent gazes centrally to the ground in front of them. 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 , 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 decreased performance in the task. As decreases and the cone narrows, the extent to which performance varies at different angles decreases. Therefore, our gaze model predicts that our trials vary in difficulty: participants should be more imprecise in locating the target the further out it lands.[[4]](#footnote-54) If our data matches the pattern of this model prediction, this can act as evidence for the gaze model. Therefore, our gaze model provides a quantitative theory of gaze understanding, with testable model predictions.

## Analysis

Our goal was to describe the inferential process of gaze understanding. We quantified how well our gaze model explained the gaze understanding process by comparing it to two alternative models that make different assumptions of which information participants use and where they consequently click to locate the target. Our modeling framework consisted of three mutually exclusive models: (1) a gaze model, (2) a random guessing model, and (3) a center bias model. We gauged which model can best explain our data by conducting model comparisons. All cognitive models were implemented in WebPPL (Goodman & Stuhlmüller, 2014).

To gauge the plausibility of our gaze model, we implemented two models that represent alternatives about how participants solve the TANGO. Participants who were overall very imprecise in locating the target might be less likely to use the agent’s gaze as a cue at all. The alternative models, therefore, do not assume that participants made use of the gaze cue. The first alternative model assumed participants were randomly guessing. This was implemented as sampling from a uniform distribution over all possible coordinates (i.e., 0 - 1920). The second alternative model assumed the participants always want to click at the screen center: Participants could be drawn toward the screen center since the agent and the starting point of the balloon were located there. This was implemented as sampling from a Normal distribution, with the center of the screen as the mean and one target width as the standard deviation.

Our three proposed models made different predictions about how participants’ clicks would be distributed for different target locations. We visualized and evaluated these differences using correlations between the model predictions and the data. Furthermore, we evaluated these probabilistic models based on 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. This factor 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 (i.e., Bayesian Occam’s razor): models with more parameters often have a broader prior distribution over parameters, which might weaken potential gains in predictive accuracy. Details on models, including code to run the models, information about priors for parameter estimation, and Markov chain Monte Carlo settings, can be found in the associated online repository.

[TODO: add more infos on priors & parameters]

## Results

We found clear support for our gaze model, both in children as well as adults.

When we compared the gaze model to the two competitor models, we found little support for the two alternatives. Comparing the models using marginal likelihood of the data under each model, the data were XXX [TODO] more likely under the gaze model compared to the random clicking or center bias model.

When correlating the observed data across all target positions with the predictions of the three models, we found a high similarity for the gaze model: *r* = 0.95, 95%CI [0.90, 0.98], while the correlations with the alternative models were substantially smaller (center bias model: *r* = 0.77, 95%CI [0.57, 0.89]; random guessing model: *r* = 0.78, 95%CI [0.58, 0.89]). The developmental trajectory of the estimated gaze model parameter also correlated highly with the observed data: *r* = 0.95, 95%CI [0.92, 0.97].

[TODO: report prior]

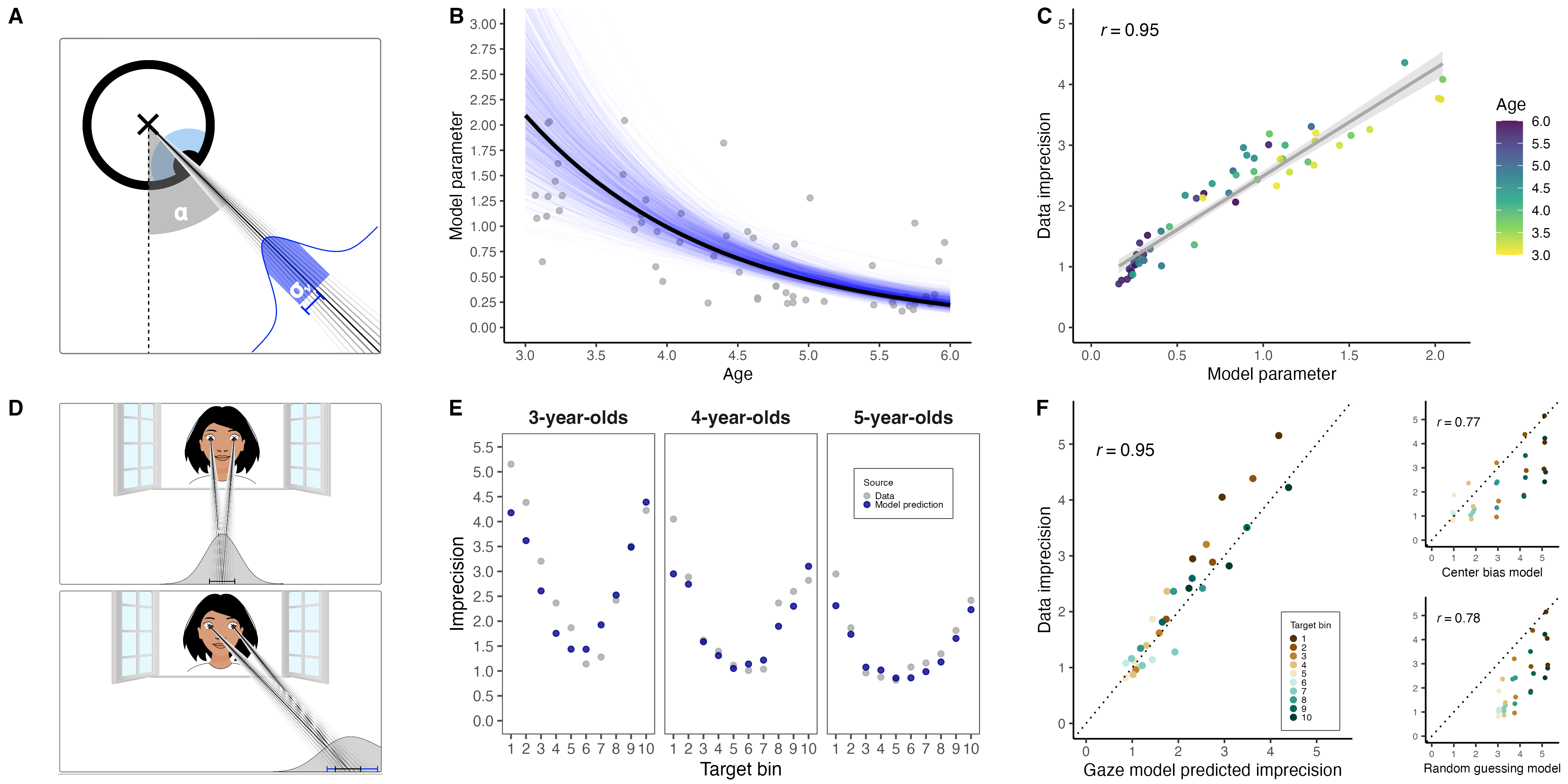


Figure 2: (ref:figlab3)

## Discussion

Our findings from Study 1 showed individual differences and a developmental change in gaze understanding across the lifespan. To answer what develops with age and how participants differ from one another, we presented a formal cognitive model of gaze understanding. Participants are modeled to observe all available gaze information, integrate it from both the agent’s eyes, and consequently arrive at the attentional focus of the agent: the target location. We assume the basic process of gaze understanding to be the same across the lifespan, though individuals become increasingly precise with age. By conducting model comparisons, we could rule out that participants’ responses can be explained by random guessing or a center bias.

In addition, we observed differences in performance depending on where the agent looks. The observed 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.

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., wanting to click somewhere but experiencing motor issues at aiming, adding random noise to the click).

A critical feature of our model is that it assumes gaze understanding to rely on vector estimations. In other words, subjects are modeled to calculate gaze vectors based on the location of an agent’s pupil. This way, they can infer the agent’s attentional focus. Even though this vector estimation component is a rather physical, geometrical calculation, it still happens in a social context. In the first place, one must interpret the agent’s eyes as a relevant social stimulus. Therefore, our computational model describes gaze understanding as a particular form of vector estimation in a social context.

# Study 3: Components of gaze understanding

## Perspective-taking:

(Erle & Topolinski, 2017)

* differentiation between visuospatial perspective-taking and psychological (i.e., cognitive and affective) perspective-taking
* unclear boundaries between ToM & perspective-taking, many correlational studies, not so many causation “terminological confusion”
* persepctive-taking: stepping into somebody else’s shoes. recognizing the other’s thoughts, feelings, mental states (=> ToM!!)
  + 1. attesting mental states to another social agent
    2. other’s states can be ≠ to own
    3. put aside egocentric bias
* little is known about underlying psychological mechanisms of perspective-taking
* visuospatial perspective taking: mentally transforming yourself and your body into the physical location of another agent. judging differences between one’s own egocentric point of view an that of another person
  + level 1: visibility of objects from certain view point. can operate independently from another’s frame of reference
  + level 2: imagine how world looks like for another person “climbing into the skin of another person”. literally rotating body schema (embodied) “transposing body schema into the target’s position”
* clinical populations show deficit in empathy & visuospatial perspective-taking
* idea that psychological & visuospatial perspective-taking share common simulation-based mechanism
* studies show that: “Taking another’s perspective led participants to adopt the thoughts of the target person more strongly (Experiments 1–3) and increased the perceived similarity of that person to the self (Experiment 4) and participants’ liking of that person (Experiment 5). These effects were independent of task difficulty (Experiment 2), and only present during trials where an embodied transformation happened (i.e., at high angular disparities; Experiment 3).” “not only affects social-cognitive outcomes, but also empathic perspective-taking” => “shared mechanism of all kinds of perspective-taking”
* non-social perspective-taking task: display empty chair. spatial perspective-taking. lead to “diametrically different results”

(Birch et al., 2017)

* “Specifically, we contend that these processes (a) are partially innate, (b) develop over time and can be honed through experience, and (c) lie on a continuum with some individuals being better than others.”
* Apperly & Buuterfill 2009: System 1 process social information relatively quickly & effortlessly, clear limitations, inflexible. System 2 explicit reasoning, involves language, inhibitory control, working memory, cognitive flexibility, reasoning, planning etc.
* chapter assumes partial innateness, develop over time, can be honed through experience, lie on a continuum with some individuals being “Better” than others
* children use gaze as indicators of others’ mental states from an early age
* individual differences:
  + children differ in onset of milestones, degree to which they can and do use ToM in daily life. vary in frequency / propensity (connected to motivation?) and accuracy
  + ind diff to “illustrate important function ToM serves in navigating the social world”. for which processes involved, contribution of other skills & environmental scaffolding
* curse of knowledge: U shaped curve. hindsight bias & tom correlated, but not inhibitory control as mediator

We previously presented a computational cognitive model of gaze understanding. Our model relies on the perhaps unexpected assumption that vector estimation is a crucial component of gaze understanding. In model comparisons, we found overwhelming support for this model in children and adults. Now, we wanted to test this assumption experimentally. We were interested in the degree to which vector estimation is a part of gaze understanding. Additionally, we investigated whether there is more to gaze understanding than the physical vector estimation component. To answer this question, we assessed the relationship between gaze understanding and other social-cognitive abilities.

First, we aimed to experimentally isolate the vector estimation component of the TANGO. We designed a new non-social vector estimation 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 (@wellman2004scaling) and two additional perspective-taking tasks [@flavell1981development; @flavell1981younga].

We aimed to assess whether there are exclusively task-specific processes at hand or whether gaze understanding recruits a general social-cognitive ability that is shared among other social-cognitive tasks.We reasoned that the TANGO shares task demands with the non-social vector estimation task while it shares its social context with the ToM tasks. This way, we aimed to disentangle what components comprise gaze understanding.

[TODO: welche sozialen prozesse überlappen hier. das hilft dann auch dabei zu sagen warum wir nicht unbedingt erwarten, dass alle ToM komponenten einen einfluss haben.]

Task design, data collection, and sample sizes were pre-registered: <https://osf.io/xsqkt>. Data were collected between February and March 2023.

## Participants

Testing took place in kindergartens in Leipzig, Germany. 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.

## Procedure

Children were tested in a quiet room in their kindergarten. An experimenter guided the child through the study. Since our research questions related to individual differences and we wanted maximum control of extraneous participant variables, we employed a within-subjects study design. All participants performed the following tasks in a fixed order: (1) non-social vector estimation task, (2) ToM task battery, (3) TANGO. Several reasons motivated this decision. First, we decided on a fixed order to be able to compare participants’ performance straight-forwardly with each other. Second, to increase participant engagement and decrease fatigue or fuzziness, we switched from a tablet task to tasks with personal interaction back to a tablet task. Third, we showed the non-social vector estimation task before the TANGO so that participants would not be biased to interpret the presented stimuli as “eye- /”agent-like”.

### Non-social vector estimation.

Modeling the setup and structure of the previously applied TANGO, we designed a non-social vector estimation task. This task was also presented as a web application on a tablet and made use of the concept of magnetism. The setup looked as follows. On the upper part of the screen, there was a tube with a gearwheel located in a circular window. On the floor, there was a magnet. The magnet then got switched on (making a cartoon-like sound), whereupon the gearwheel moved towards the magnet. The gearwheel moved in a way that its center aligned with the center of the magnet, 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. When no information about the magnet location was accessible, participants were expected to use the gearwheel inside the window as a non-social cue to locate the magnet.

As in the TANGO, there were three different trial types depending on the visual access to the true magnet location. In full visual access trials, the magnet’s location was presented without impediment (i.e., no wooden wall). In partial visual access trials, the wooden wall was moved in front of the target after the magnet’s location had already been visible. In test trials, participants had no visual access to the magnet’s location because the wall covered the magnet from the beginning of the trial.

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 difference between the magnet’s x coordinate and the x coordinate of the participant’s click. Magnet coordinates were generated as follows. 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. Exact coordinates within each bin were randomly generated.

### Theory of Mind task battery.

We administered four tasks from the Wellman and Liu (2004) Theory of Mind scale. We excluded three tasks: the Diverse Desires task in order to avoid ceiling effects; and both tasks involving emotions (Belief Emotion and Real-Apparent Emotion), as we aimed at assessing the “cold, cognitive” (as compared to the “emotional”) aspects of social cognition. Instead, we added two perspective-taking level-2 tasks (Flavell, Everett, Croft, & Flavell, 1981; Flavell, Flavell, Green, & Wilcox, 1981). We added the perspective-taking tasks (1) with the aim of increasing the variablility we can capture between individuals, and (2) since we hypothesized that perspective-taking would rely on similar mechanisms than gaze understanding. The dependent variable was the aggregate score of all solved ToM tasks (see Supplements for further detail). Additionally, we investigated if gaze understanding was more strongly associated with the two perspective-taking tasks compared to the other ToM tasks, as perspective-taking seems most closely theoretically related to gaze understanding (i.e., in both cases the participant is asked to judge another person’s point of view).

### Gaze understanding.

As in the two previously reported studies, we presented children with the continuous version of 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 understanding (<https://osf.io/tdsvc>)). This further highlighted the contrast (i.e., social vs. non-social context) to the non-social vector estimation task.[[5]](#footnote-70)

## Analysis

By design, both the TANGO as well as the non-social vector estimation task involve vector estimation. On the basis of the results from our computational cognitive model, we expected that children’s performance in both tasks correlate with each other. For each of these two tasks, we calculated the mean level of imprecision for each subject. We then correlated these two scores using *Pearson’s* correlation coefficients.

Regarding the relationship between the two vector estimation tasks and the ToM measures, we could imagine two possible scenarios: (A) If gaze understanding recruits a general social-cognitive ability beyond vector estimation, we expected that gaze understanding and ToM measures would correlate more strongly with each other than non-social vector estimation and ToM measures. (B) If gaze understanding relies purely on task-specific processes, then the correlation between gaze understanding and ToM measures would be comparable to the correlation between non-social vector estimation and the ToM measures. For the association between the aggregate ToM scores and the gaze understanding / non-social vector estimation tasks, we used *Spearman’s* rank correlation coefficients.

We compared the correlation between gaze understanding and ToM measures and the correlation between non-social vector estimation 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).

Furthermore, to estimate which components best explain the gaze understanding score, we conducted a model comparison with GLMMs predicting the mean imprecision in gaze understanding by age, imprecision in non-social vector estimation, 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). We wanted to assess whether the ToM aggregate score or the singled-out perspective-taking score added additional explanatory value when predicting the gaze understanding score. The outcome variable was modeled by a lognormal distribution.

## Results

As expected, we found that gaze understanding as a social vector estimation task correlated with the non-social vector estimation task, *r* = 0.38, 95%CI [0.20, 0.53]. Importantly, however, the two vector estimation tasks were not redundant: only a part of the variance in gaze understanding could be explained by non-social vector estimation.

Gaze understanding and perspective-taking showed a *Spearman* correlation coefficient of = -0.29, 95%CI [-0.46, -0.10], while non-social vector estimation and perspective-taking did not correlate, = -0.09, 95%CI [-0.28, 0.10]. According to the Williams’ test, these two correlations did not differ significantly from each other, *t*(99) = -1.86, *p* = 0.07.

Our model comparison revealed that gaze understanding was best predicted by a model including non-social vector estimation ( = 0.14, 95% CrI [0.06; 0.21]) and perspective-taking ( = -0.10; 95% CrI [-0.17, -0.03]), even when controlling for age ( = -0.14, 95% CrI [-0.38, 0.10]). See Supplements for further detail of the model comparison.

Taken together, this shows that the TANGO recruited social-cognitive abilities beyond vector estimation. Evidently, it shared some of its variance with other level 2 perspective-taking tasks, while the overall ToM aggregate score did not add explanatory power.

(ref:figlab3) **Components of gaze understanding.** (A) Study procedures. Top: TANGO (i.e., gaze understanding; social vector estimation). Bottom: Magnet (i.e., non-social vector estimation). Left hand side show screenshots of familiarization phase; right hand side show screenshots of the test phase. (B) Correlations between gaze understanding, physical vector estimation, ToM, and perspective-taking. Dots show the correlation coefficients, while error bars represent 95% CIs. (C) Influence of age, perspective-taking and physical vector estimation on gaze understanding. 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.

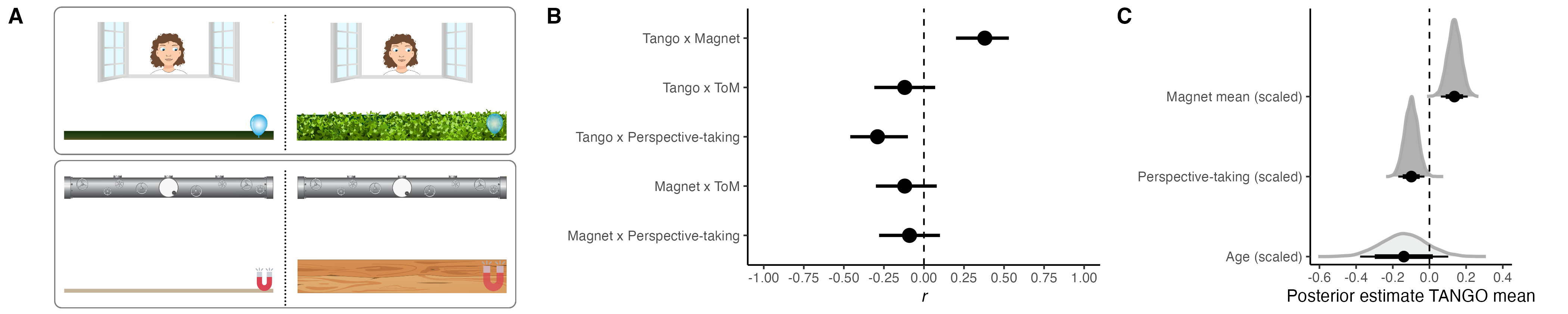


Figure 3: (ref:figlab4)

## Discussion

By carefully isolating physical vector estimation experimentally, we could show that gaze understanding does indeed, to a certain degree, rely on this component. This is in line with our computational cognitive framework that assumes vector calculations on a process-level. However, physical vector estimation alone did not suffice to explain gaze understanding. In addition, perspective-taking proved to be a relevant social-cognitive ability.

[TODO: das klingt als ob perspective taking ein prozess wäre, ist es aber nicht - versuch mal zu überlegen was hier die gemeinsamkeiten sein könnten und das zu beschreiben.]

In previous work, we could establish that the TANGO is suited as an individual differences measure (Prein et al., 2023). Capturing meaningful variability in performance 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: they relied on an aggregate score of dichotomous measures. These sum scores can only capture limited variance, which may obscure potential correlations. However, since these tasks are the gold standard in the social-cognitive literature and continuous measures with satisfying psychometric properties are, to the best of our knowledge, still scarce, we nonetheless relied on them in this study. It seems noteworthy to point out that lower correlations between ToM abilities and gaze understanding could be grounded in the design features of the applied ToM tasks. We already stated this concern in the Pre-registration (<https://osf.io/xsqkt>). The development of new measures to capture individual differences in social-cognitive abilities like false-belief understanding seems desirable and essential to move this line of research further.

[TODO: die “wahren” Korrealtionen (die zwischen den latenten konstrukten) eigentloch stärker sind weil hier ja auch measurement error drin steckt. das kann man berechnen wenn man die reliabilität jeweils kennt. Die haben wir hier halt nciht für tango (außer vll split half??)]

# General discussion

# Limitations

# Conclusion

# Declarations

… can be found on the title page.

# References

Astor, K., & Gredebäck, G. (2022). Gaze following in infancy: Five big questions that the field should answer. In *Advances in Child Development and Behavior* (p. S0065240722000192). Elsevier. <https://doi.org/10.1016/bs.acdb.2022.04.003>

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>

Birch, S. A. J., Li, V., Haddock, T., Ghrear, S. E., Brosseau-Liard, P., Baimel, A., & Whyte, M. (2017). Perspectives on Perspective Taking. In *Advances in Child Development and Behavior* (Vol. 52, pp. 185–226). Elsevier. <https://doi.org/10.1016/bs.acdb.2016.10.005>

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>

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>

Carpenter, M., Nagell, K., & Tomasello, M. (1998). [Social cognition, joint attention, and communicative competence from 9 to 15 months of age](https://www.ncbi.nlm.nih.gov/pubmed/9835078). *Monographs of the Society for Research in Child Development*, *63*(4), i–vi, 1–143.

Coelho, E., George, N., Conty, L., Hugueville, L., & Tijus, C. (2006). Searching for asymmetries in the detection of gaze contact versus averted gaze under different head views: A behavioural study. *Spatial Vision*, *19*(6), 529–545. <https://doi.org/10.1163/156856806779194026>

Colombo, J. (2001). The development of visual attention in infancy. *Annual Review of Psychology*, *52*, 337–367. <https://doi.org/10.1146/annurev.psych.52.1.337>

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>

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>

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 2 distinction. *Developmental Psychology*, *17*, 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>

Frischen, A., Bayliss, A. P., & Tipper, S. P. (2007). Gaze cueing of attention: Visual attention, social cognition, and individual differences. *Psychological Bulletin*, *133*(4), 694–724. <https://doi.org/10.1037/0033-2909.133.4.694>

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*, 177–190. <https://doi.org/10.1037/0012-1649.40.2.177>

Goodman, N. D., & Stuhlmüller, A. (2014). *The Design and Implementation of Probabilistic Programming Languages*.

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>

Grahek, I., Schaller, M., & Tackett, J. L. (2021). Anatomy of a Psychological Theory: Integrating Construct-Validation and Computational-Modeling Methods to Advance Theorizing. *Perspectives on Psychological Science*, *16*(4), 803–815. <https://doi.org/10.1177/1745691620966794>

Itakura, S., & Tanaka, M. (1998). Use of experimenter-given cues during object-choice tasks by chimpanzees (Pan troglodytes), an orangutan (Pongo pygmaeus), and human infants (Homo sapiens). *Journal of Comparative Psychology*, *112*(2), 119–126. <https://doi.org/10.1037/0735-7036.112.2.119>

Lee, K., Eskritt, M., Symons, L. A., & Muir, D. (1998). Children’s use of triadic eye gaze information for "mind reading". *Developmental Psychology*, *34*(3), 525–539. <https://doi.org/10.1037//0012-1649.34.3.525>

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>

Moore, C., Angelopoulos, M., & Bennett, P. (1997). The role of movement in the development of joint visual attention. *Infant Behavior and Development*, *20*(1), 83–92. <https://doi.org/10.1016/S0163-6383(97)90063-1>

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

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*. <https://doi.org/10.3758/s13428-023-02159-5>

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

Raviv, L., & Arnon, I. (2018). The developmental trajectory of children’s auditory and visual statistical learning abilities: Modality-based differences in the effect of age. *Developmental Science*, *21*(4), e12593. <https://doi.org/10.1111/desc.12593>

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>

Scaife, M., & Bruner, J. S. (1975). The capacity for joint visual attention in the infant. *Nature*, *253*(5489), 265–266. <https://doi.org/10.1038/253265a0>

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>

Simmering, V. R., Triesch, J., Deák, G. O., & Spencer, J. P. (2010). A Dialogue on the Role of Computational Modeling in Developmental Science. *Child Development Perspectives*, *4*(2), 152–158. <https://doi.org/10.1111/j.1750-8606.2010.00134.x>

Tomasello, M., Hare, B., Lehmann, H., & Call, J. (2007). Reliance on head versus eyes in the gaze following of great apes and human infants: The cooperative eye hypothesis. *Journal of Human Evolution*, *52*(3), 314–320. <https://doi.org/10.1016/j.jhevol.2006.10.001>

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>

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>

Zhang, X., Zhang, Z., Zhang, Z., Tang, Y., & Liu, W. (2019). The role of the motion cue in the dynamic gaze-cueing effect: A study of the lateralized ERPs. *Neuropsychologia*, *124*, 151–160. <https://doi.org/10.1016/j.neuropsychologia.2018.12.016>

Zhao, K., Wulder, M. A., Hu, T., Bright, R., Wu, Q., Qin, H., … 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>

1. Originally, our models were fitted on a trial-by-trial basis and implemented the following structure: performance ~ age + symmetricPosition + trialNr + (1 + symmetricPosition + trialNr | subjID). However, the Gaussian Process model was computationally heavy. To circumvent issues of convergence and for practicability, we simplified the model structure, aggregated data on a subject level, and included only age as an effect. In the Supplements, you can find a comparison between the original and the here reported model structures. Essentially, the model predictions did not differ notably. [↑](#footnote-ref-40)
2. In a supplementary analysis, we varied the parameters of our changepoint analysis. We 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 (i.e., greater nr of change points, smaller minimum nr of data points between change points, higher polynomial order), 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. There was a minor change in that elderly adults became slightly more imprecise again. See Supplements for further detail. [↑](#footnote-ref-41)
3. 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). We then calculate a line between these two points: this is our gaze vector. Now, knowing the eyeball radius, we calculate the point of intersection at which the gaze vector meets the eyeball boundary. Finally, the agent’s pupil moves from the center of the eyeball along the gaze vector to the intersection point. This way, the agent is animated to “look at” the target. In the gaze model, we assume participants go through these steps in reverse order. [↑](#footnote-ref-53)
4. 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 true gaze vector 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 Supplements. [↑](#footnote-ref-54)
5. In an exploratory analysis, we compared children’s imprecision levels in the TANGO task with animal vs. human agents. Based on a GLMM analysis, we conclude that there was no evidence of a stable effect of stimulus choice (human vs. animal). See Supplements for further detail. [↑](#footnote-ref-70)