Supplements for the manuscript 'Variation in gaze following across the life span: A process-level perspective'

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

[1] Department of Comparative Cultural Psychology
Max Planck Institute for Evolutionary Anthropology
Leipzig
Germany
[2] Institute of Psychology
Leuphana University Lüneburg
Germany
[3] Center for Lifespan Psychology
Max Planck Institute for Human Development
Berlin
Germany
[4] Max Planck School of Cognition
Leipzig
Germany
[*] Shared senior authorship

Supplements for the manuscript 'Variation in gaze following across the life span: A process-level perspective'

Study 1: Lifespan

Participants

For further information on the participants in Study 1, please see Supplementary Table S1.

Analysis

Model structures

In the paper, we reported the following model structures: linear model: `mean_imprecision ~ age_centered`; quadratic model in `R`: `mean_imprecision ~ 1 + age_centered + I(age_centered^2)`; cubic model: `mean_imprecision ~ 1 + age_centered + I(age_centered^2) + I(age_centered^3)`; Gaussian process model: `mean_imprecision ~ gp(age_centered, k=50, c=5/4, scale=TRUE)`. Note the additional parameters in the Gaussian process model. With the default settings, the underlying Gaussian process maths would get solved exactly. By providing the arguments `k` and `c`, we use an approximation process. The higher the value of k, the better the approximation: we have used `k=50` for faster processing speed and better diagnostics. `brms` suggests 5/4 as a value for `c`. Adding `scale=TRUE` is supposed to improve sampling speed and convergence.

Originally, we fitted the models on a trial-by-trial basis with the following structure in `R`: `performance ~ age + symmetricPosition + trialNr + (1 + symmetricPosition + trialNr | subjID)`. However, the Gaussian Process model was computationally heavy. Therefore, we simplified the model structure, aggregated data on a subject level, and included only age as an effect. We then visually compared the model predictions of the original and the simplified models with each other. As you can see in Figure S1, results of the two models did not differ notably.

Changepoint analysis

In our Bayesian changepoint analysis, we restricted the model to a constant mean (i.e., a flat line with zero degree polynomial) within each segment, and to have minimally 10 data points between two change points (i.e., corresponding to half of the data points we collected per adult decade) to avoid "overreactions" to individual outlying data points. In a supplementary analysis, we assessed how different parameter settings effected of our changepoint analysis. We changed the number of allowed change points, the minimum number of data points between change points, and the polynomial order. When the model had more explorative room, for example, by a greater number of change points, smaller minimum number of data points between change points, higher polynomial order, the model outputs showed more fine-grained change points. The exact location of the change points varied slightly. Overall, the interpretation stayed the same as the one we reported in the paper. 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. If you are interested into the details, please have a look at the file `supplements_changepoint_parameters.html`, which you can find in the GitHub repository in the `stats` folder.

Study 2: Computational cognitive model

Analysis

Gaze model prediction

Our gaze model predicts that TANGO trials vary in their difficulty, resulting in a U-shaped pattern: Participants' imprecision should increase, the further out the target lands (towards the very left/right sides). Since the task is presented on a screen, there is a natural border towards one side. Imagine the target lands to the very right side. Participants' imprecision cannot click further right because the screen ends; all their uncertainty about the target location faces the inner, left-hand side now. Therefore, the predicted U-shaped pattern should decrease again towards the screen borders. For previous reliability analyses

(Prein et al., 2023), we had increased the trial number for an adult sample (N = 70; each 30 trials). Interestingly, here we found the expected shape in the data: the U-pattern decreased again towards the screen ends (Figure S2).

Inference parameter estimates per individual

As can be seen in Figure S3 and Figure S4, the gaze model estimated the inference parameter for each individual. Across individuals, the inference parameter varied in the estimated magnitude and level of uncertainty. In general, estimates for more precise individuals (i.e., smaller inference parameter value) showed decreased levels of uncertainty.

Simulations

In the manuscript, we have described the gaze model and two alternatives: random guessing and a center bias. Here, we consider two more alternatives. Let us consider a model that assumes participants can infer the agent's focus without any noise (for example, by tracing the line of sight). A model like this would assume that participants follow gaze without any uncertainty. Therefore, no U curve would be predicted. Furthermore, let us assume another model in that participants still show no inferential noise, but they vary in their amount of motor noise, so how accurately participants then click at the corresponding location. A model like this would assume equal variance across the target locations, so we would not expect a U-shape here.

Please note that our random guessing model acts like 100% noise: the predictions of a participant's click on a trial level range uniformly from 0 to 1920 (the whole screen range). The mean comes down to the center, namely 960. However, as you can see in Figure S5A below, the U shape is weaker compared to the other models. Most importantly, note that this is only the case when you average across all the trials. When you look at the individual trial-by-trial level, the models are defined by different data-generating processes, and therefore, their predictions differ, too (Figure S5C). In our correlational plots (Figure SS5B), we see that the gaze model is clearly favored. This highlights the benefit that we gain through the modeling approach: even though a certain (U) pattern in the data could be elicited from several different models, we can disentangle which process is most likely causing this pattern.

Study 3: Components of gaze understanding

Procedure Theory of mind battery

For the Theory of Mind battery, we followed the standardized procedure as described below. The battery was administered in German and presented in the order as stated below.

(1) Diverse Beliefs (Wellman & Liu, 2004)

Material: Girl figure, leaf with garage & bush. Experimenter (E): “Look, this is Linda. Linda wants to find her cat. The cat is either hiding behind the bush or in the garage. Where do you think the cat is hiding? Behind the bush or in the garage?” (own-belief question). Child: “Garage”. E: “That’s a good idea! But Linda thinks her cat is behind the bush. She thinks her cat is behind the bush. So, where will Linda look for her cat? Behind the bush or in the garage?” (target question)

(2) Knowledge Access (Wellman & Liu, 2004)

Material: Yellow box, pig figure, female figure. E: “Look, here’s a box. What do you think is in the box?” [Child answers] [Box opens] E: “Let’s see... oh, there’s actually a pig inside!” [Box is closed] E: “What’s in the box?” [E pulls out figure] E: “That’s Polly. Polly has never looked in the box. So, does Polly know what’s in the box? (target question) Has Polly ever looked in the box? (control question)”

(3) Contents False Belief (Wellman & Liu, 2004)

Material: Smarties box, figure of a dog, figure of a boy. E: “Here is a Smartie box. What do you think is inside?” [Smartie box is opened] E: “Let’s see... there’s actually a dog inside!” [Smartie box is closed] E: “Okay, what’s in the Smartie box?” [Man’s figure is taken out] E: “This is Peter. Peter has never looked inside the Smartie box. What does Peter think is in the box? Smarties or a dog? (target question) Has Peter ever looked in the box? (memory question)”

(4) Explicit False Belief (Wellman & Liu, 2004)

Material: Figure of a boy, sheet with cupboard & rucksack on it. E: “This is Scott. Scott wants to find his gloves. The gloves could be in his backpack or they could be in the closet. In reality, they are in the backpack. But Scott thinks his gloves are in the closet. So, where will Scott look for his gloves? In the backpack or in the closet? (target question) Where are his gloves really? In the backpack or in the closet? (reality question)”

(5) Perspective-Taking Level 2 version I (Flavell, Flavell, et al., 1981; Flavell, Everett, et al., 1981)

Material: Picture of turtle. E: "Look, here's a picture of a turtle. I'll put it between us. What does the turtle look like to you, is it standing on its feet or lying on its back? (own perspective question) And what does it look like to me? Does it stand on its feet for me or does it lie on its back? (other perspective question)."

(6) Perspective-Taking Level 2 version II (Flavell, Flavell, et al., 1981; Flavell, Everett, et al., 1981)

Material: Picture of worm between two pillows. E: "Now we have a worm lying between two pillows. I'll put it between us again. How does it look to you, is the worm lying on the red or the blue cushion? (own perspective question) And what does it look like to me - is the worm lying on the red or the blue cushion? (other perspective question)"

Analysis

Animal vs. human faces

In Study 1 and Study 2, we presented the TANGO (Prein et al., 2023) with animal characters. For Study 3, we exchanged the animals with human faces, modelled in appearance after the local population. We decided to do so in order to enhance the social context of this task and to make it more comparable to the Theory of Mind task battery (where there is live interaction with the experimenter). To ensure the change from animal to human faces did not notably change children's responses, we conducted an exploratory analysis. We conducted a GLMM analysis with the following model structure in `R`: `click ~ age_scaled + stimuli + symmetric_position + trial_nr + (1 + symmetric_position + trial_nr | subj_id)`; where `stimuli` denoted either human or animal faces. The estimate for the fixed effect of `stimuli` was small and the 95% CrI included zero: $\beta = 0.16$; 95% CrI [-0.06; 0.37]). Therefore, we concluded that the animal vs. human version of the TANGO did not differ substantially.

Model comparisons

To identify which (social-)cognitive components were needed to best explain the TANGO score, we compared GLMMs that predicted the mean imprecision in gaze understanding by age + the respective task score: imprecision in non-social vector estimation, the ToM aggregate score, and/ or the aggregate of the two perspective-taking tasks (subset of ToM battery). For example, the model notation in `R: tango_mean ~ age_centered + magnet_scaled + perspective_scaled`). The model including the non-social vector estimation task (magnet) and the two perspective-taking tasks won, as indicated by the model comparison results shown in Supplementary Table S2.

References

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*, 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

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

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

Supplementary Tables

Table S1

Participants in Study 1. Age is given in years.

Age group	n	Age mean	Age range	Age SD
3	19 (7 female)	3.62	3.04 - 3.99	0.31
4	17 (9 female)	4.45	4.05 - 4.91	0.30
5	22 (13 female)	5.56	5.08 - 5.99	0.31
6	24 (16 female)	6.50	6.1 - 6.99	0.28
7	39 (20 female)	7.48	7.04 - 7.95	0.25
8	41 (20 female)	8.46	8.03 - 8.98	0.27
9	56 (29 female)	9.46	9.01 - 9.96	0.28
10	35 (22 female)	10.49	10.01 - 11	0.28
11	54 (26 female)	11.43	11.01 - 11.96	0.28
12	43 (19 female)	12.41	12.01 - 12.99	0.30
13	42 (19 female)	13.50	13.09 - 13.99	0.27
14	20 (14 female)	14.37	14.05 - 14.98	0.23
15	21 (11 female)	15.56	15.05 - 15.98	0.30
16	19 (10 female)	16.51	16.17 - 16.97	0.24
17	19 (10 female)	17.53	17.01 - 17.95	0.28
18	2 (0 female)	18.00	18 - 18	0.00
19	5 (4 female)	19.00	19 - 19	0.00
20	40 (25 female)	23.02	20 - 29	2.77
30	40 (21 female)	34.42	30 - 39	3.00
40	40 (24 female)	44.17	40 - 49	2.92
50	40 (21 female)	54.38	50 - 59	3.04
60	40 (21 female)	63.73	60 - 69	2.56
70	40 (20 female)	72.75	70 - 79	2.44

Table S2

Model comparisons ToM tasks.

Model	WAIC	SE_WAIC	Weight	ELPD_DIFF	SE_ELPD_DIFF
Magnet mean (scaled) + Perspective-taking aggregate (scaled)	200.83	16.16	0.92	0.00	0.00
Magnet mean (scaled)	206.51	16.92	0.05	-2.84	2.64
Magnet mean (scaled) + ToM aggregate (scaled)	208.51	16.79	0.02	-3.84	2.38
Perspective-taking aggregate (scaled)	212.21	15.42	0.00	-5.69	2.48
Null model with Age (scaled)	218.72	15.96	0.00	-8.95	3.35
ToM aggregate (scaled)	220.52	15.83	0.00	-9.85	3.09

Supplementary Figures

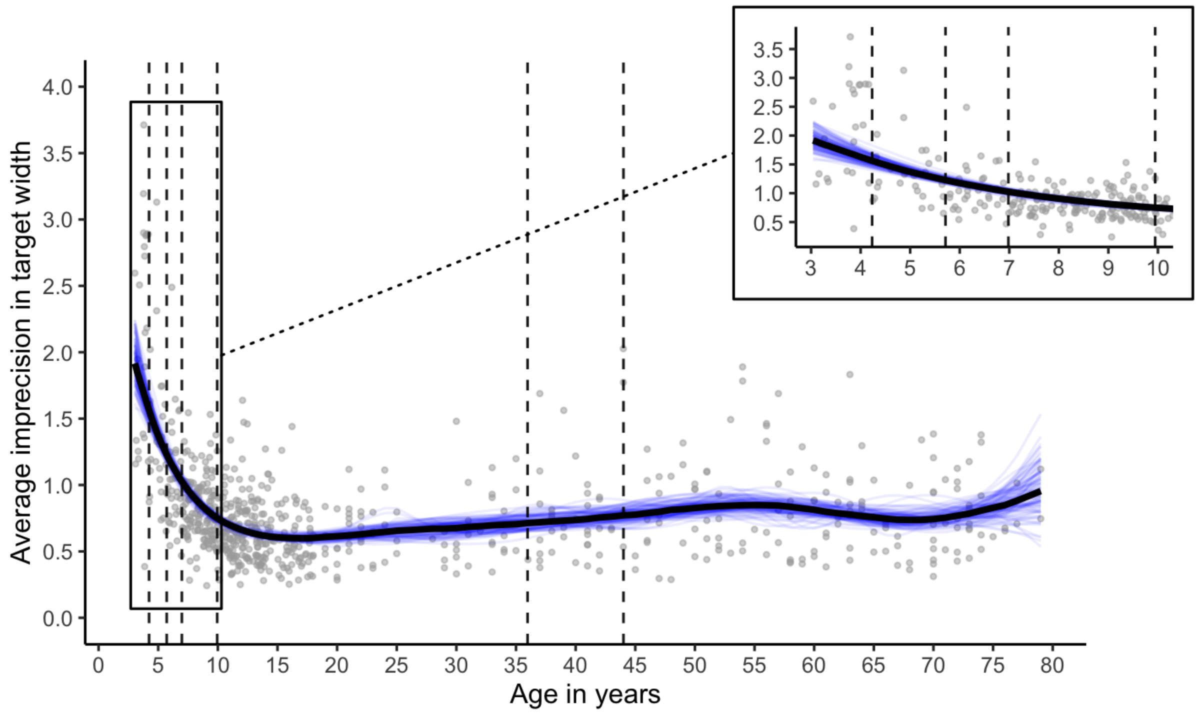

Figure S1

Comparison between models on a trial- or subject-level*. Grey dots show data of each trial. Solid lines show the mean predicted developmental trajectory for both models. Line color denotes model structure (yellow: trial-level; black: aggregated on subject-level).*

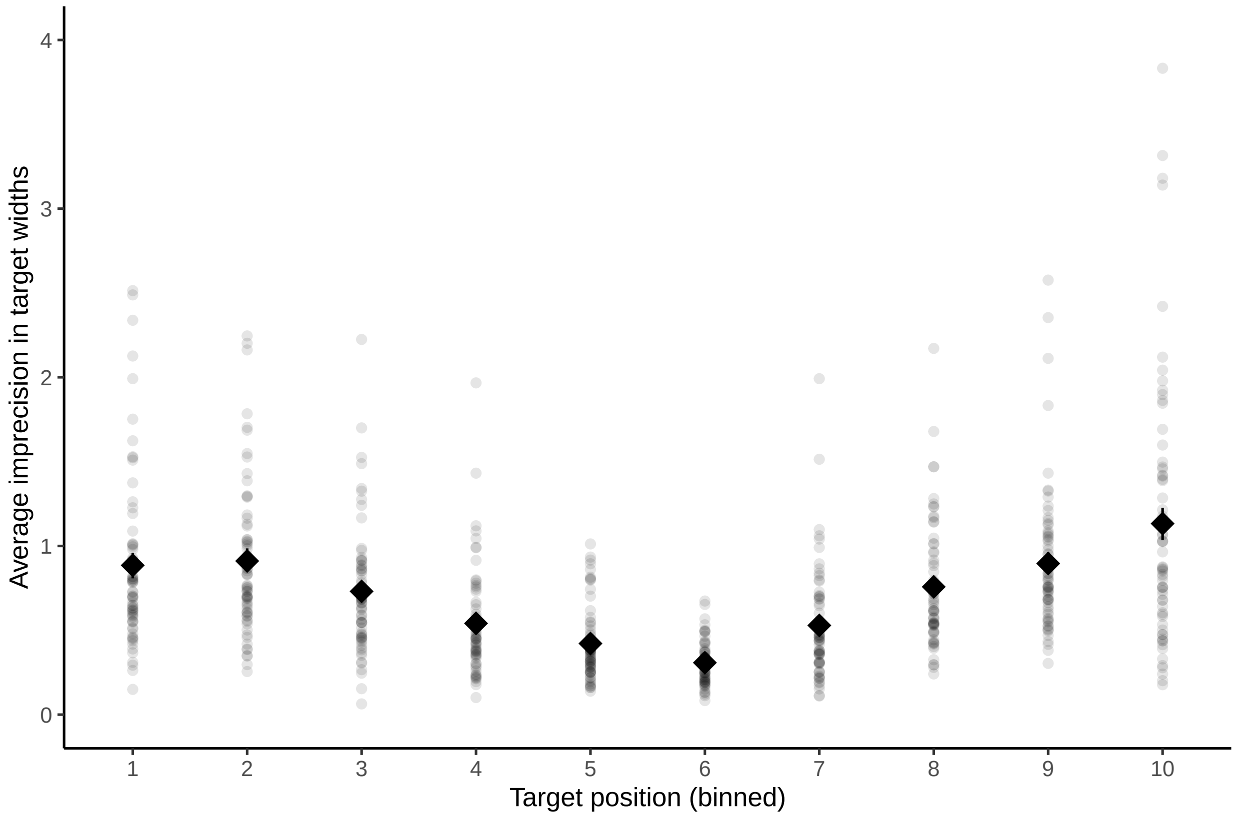

Figure S2

Gaze funnel for adult sample with higher trial number (30 trials)*. The x-axis shows the target position, binned into 10 sections. The y-axis shows participants' imprecision in target widths.*

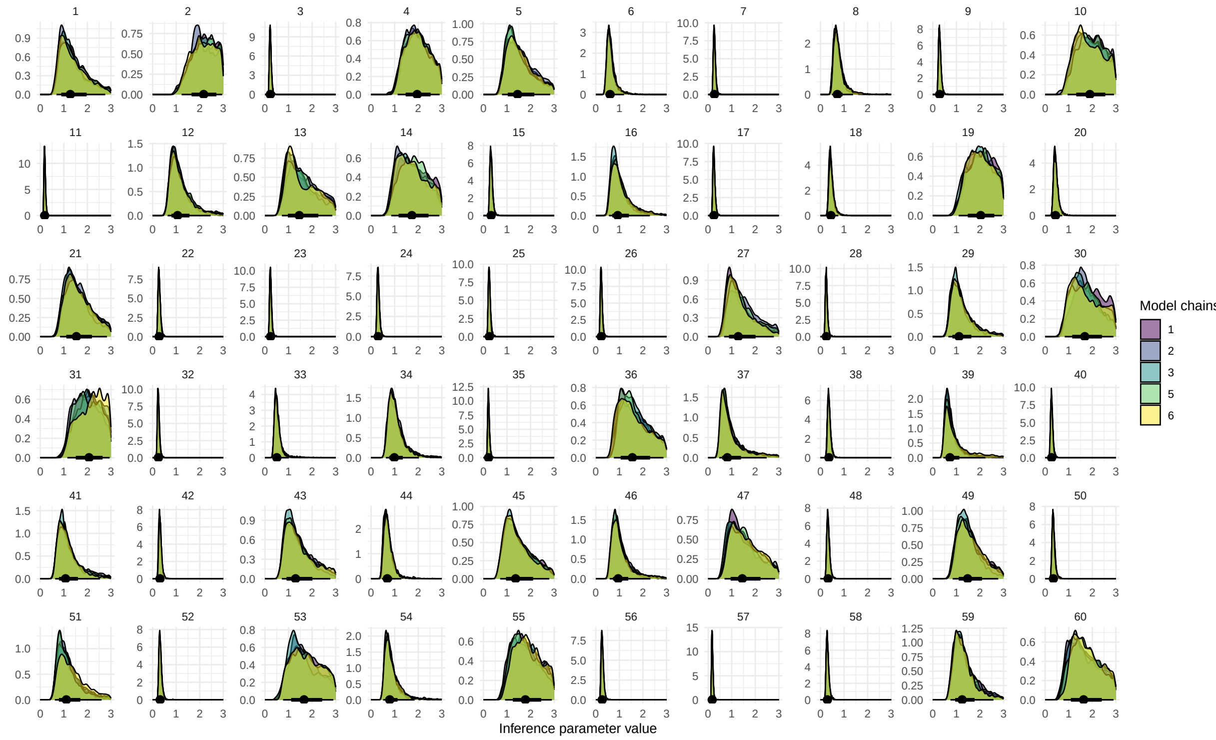

Figure S3

Gaze model estimates faceted by individual*. Density curves show the distribution of the inference parameter by individual.*

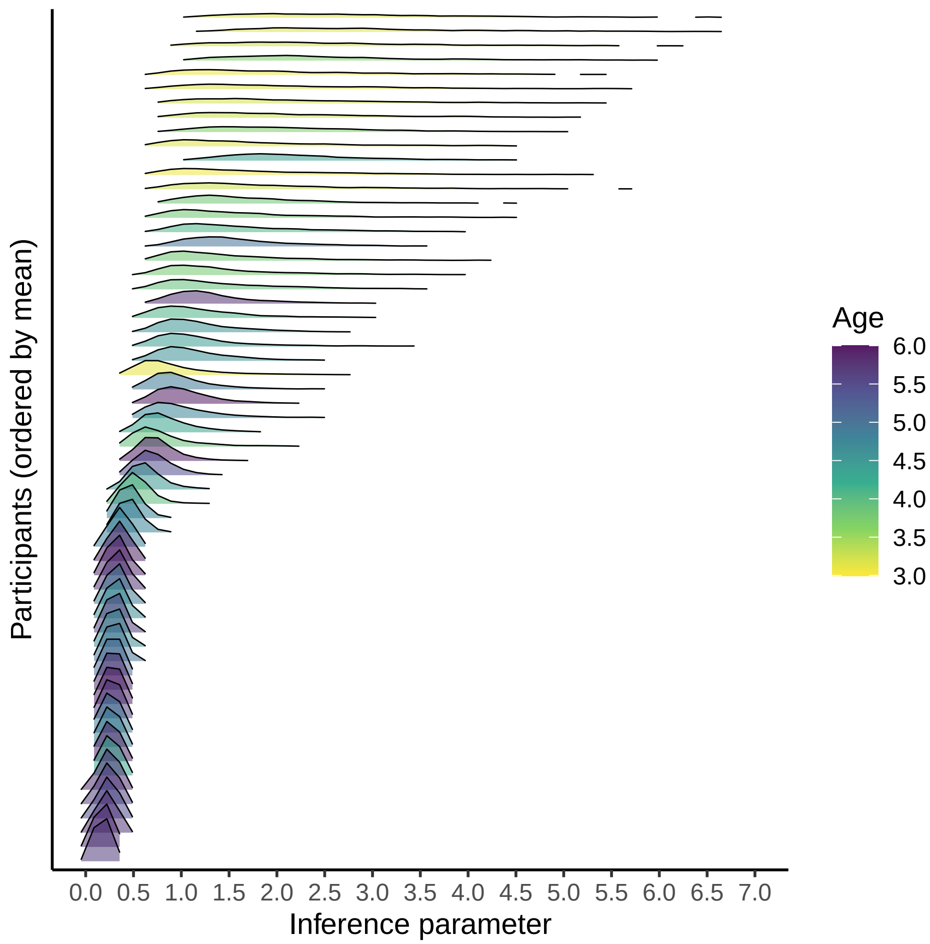

Figure S4

Gaze model estimates ordered by descending inference parameter value*. Density curves show the distribution of the inference parameter across individuals.*

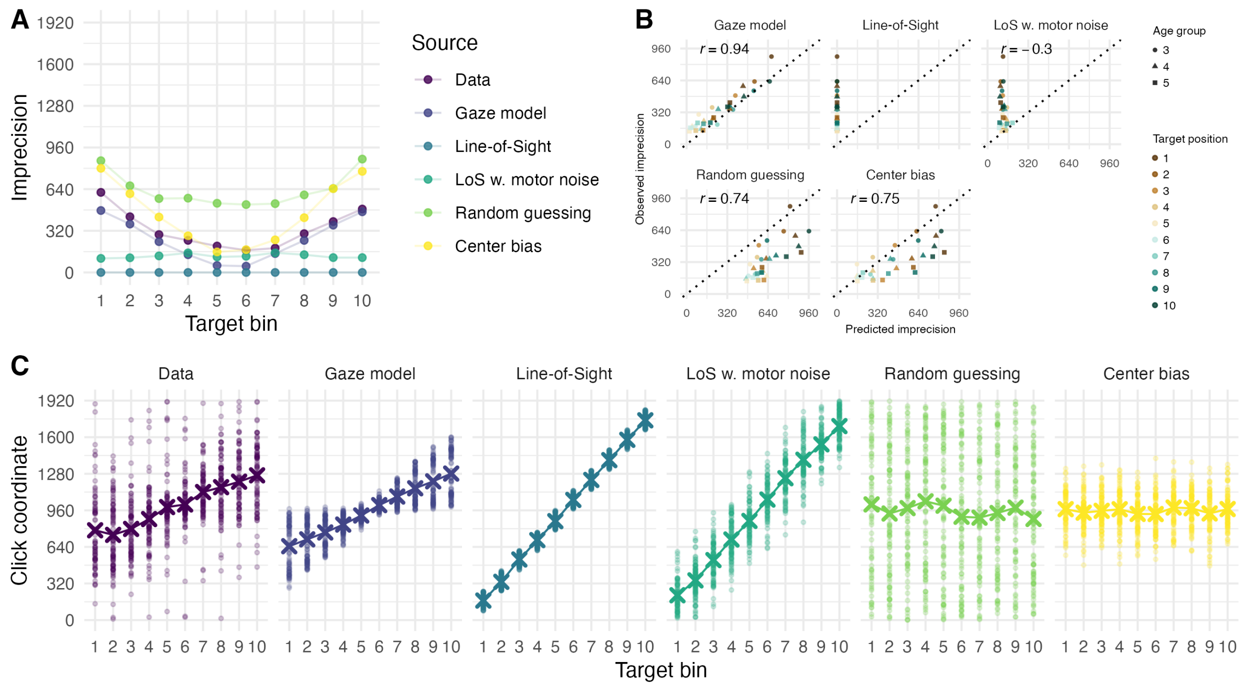

Figure S5

Predictions of alternative models across target positions. *A: Predicted U-pattern across target positions by the different models. Note that this is averaged across trials and displays imprecision, i.e., the absolute distance between the target center and the click. B: Correlation between the predicted imprecision and the observed imprecision, by target position and age group. C: Predicted click coordinates (cf. not imprecision) by target position across the different models.*