A universal of human social cognition: Children from 17 communities process gaze in similar ways

Manuel Bohn (ORCID: 0000-0001-6006-1348)[1,2,*], Julia Prein (ORCID: 0000-0002-3154-6167)[1,2,*], Agnes Ayikoru[3], Florian M. Bednarski (ORCID: 0000-0003-4384-4791)[4], Ardain Dzabatou[5], Michael C. Frank (ORCID: 0000-0002-7551-4378)[6], Annette M. E. Henderson (ORCID: 0000-0003-4384-4791)[4], Joan Isabella[3], Josefine Kalbitz[2], Patricia Kanngiesser (ORCID:0000-0003-1068-3725)[7], Dilara Keşşafoğlu (ORCID: 0000-0002-7356-0733)[8], Bahar Köymen (ORCID: 0000-0001-5126-8240)[9], Maira V. Manrique-Hernandez[2], Shirley Magazi (ORCID: 0009-0006-0479-9800)[10], Lizbeth Mújica-Manrique[2], Julia Ohlendorf[2], Damilola Olaoba[2], Wesley R. Pieters (ORCID:0000-0002-6152-249X)[10], Sarah Pope-Caldwell[2], Katie Slocombe (ORCID: 0000-0002-7310-1887)[11], Robert Z. Sparks (ORCID: 0000-0001-7545-0522)[6], Jahnavi Sunderarajan[2], Wilson Vieira[2], Zhen Zhang (ORCID: 0000-0001-9300-0920)[12], Yufei Zong (ORCID: 0009-0000-5012-0244)[12], Roman Stengelin (ORCID: 0000-0003-2212-4613)[2,10,+], & Daniel B. M. Haun (ORCID: 0000-0002-3262-645X)[2,+]

[1] Institute of Psychology in Education, Leuphana University Lüneburg

[2] Department of Comparative Cultural Psychology, Max Planck Institute for Evolutionary Anthropology

[3] Budongo Conservation Field Station

[4] School of Psychology, University of Auckland

[5] Université Marien Ngouabi

[6] Department of Psychology, Stanford University

[7] School of Psychology, University of Plymouth

[8] Department of Psychology, Koç University

[9] Division of Psychology, Communication, and Human Neuroscience, University of Manchester

[10] Department of Psychology and Social Work, University of Namibia

[11] Department of Psychology, University of York

[12] CAS Key Laboratory of Behavioral Science, Institute of Psychology, Chinese Academy of Sciences

[*] joint first author

[+] joint last author

Author Note

The authors would like to thank Luke Maurits for statistical advice. Manuel Bohn was supported by a Jacobs Foundation Research Fellowship (2022-1484-00). We are gratful to thank all children and caregivers for participating in the study. We thank the Max Planck Society for the Advancement fo Science.

The authors made the following contributions. Manuel Bohn (ORCID: 0000-0001-6006-1348): Conceptualization, Methodology, Formal Analysis, Writing - Original Draft Preparation, Writing - Review & Editing; Julia Prein (ORCID: 0000-0002-3154-6167): Conceptualization, Methodology, Software, Investigation, Writing - Review & Editing; Agnes Ayikoru: Investigation; Florian M. Bednarski (ORCID: 0000-0003-4384-4791): Investigation, Writing - Review & Editing; Ardain Dzabatou: Investigation; Michael C. Frank (ORCID: 0000-0002-7551-4378): Investigation, Writing - Review & Editing; Annette M. E. Henderson (ORCID: 0000-0003-4384-4791): Investigation, Writing - Review & Editing; Joan Isabella: Investigation; Josefine Kalbitz: Investigation, Writing - Review & Editing; Patricia Kanngiesser (ORCID:0000-0003-1068-3725): Investigation, Writing - Review & Editing; Dilara Keşşafoğlu (ORCID: 0000-0002-7356-0733): Investigation, Writing - Review & Editing; Bahar Köymen (ORCID: 0000-0001-5126-8240): Investigation, Writing - Review & Editing; Maira V. Manrique-Hernandez: Investigation; Shirley Magazi (ORCID: 0009-0006-0479-9800): Investigation; Lizbeth Mújica-Manrique: Investigation, Writing - Review & Editing; Julia Ohlendorf: Investigation; Damilola Olaoba: Investigation; Wesley R. Pieters (ORCID:0000-0002-6152-249X): Investigation, Writing - Review & Editing; Sarah Pope-Caldwell: Investigation; Katie Slocombe (ORCID: 0000-0002-7310-1887): Investigation, Writing - Review & Editing; Robert Z. Sparks (ORCID: 0000-0001-7545-0522): Investigation; Jahnavi Sunderarajan: Investigation; Wilson Vieira: Investigation; Zhen Zhang (ORCID: 0000-0001-9300-0920): Investigation, Writing - Review & Editing; Yufei Zong (ORCID: 0009-0000-5012-0244): Investigation; Roman Stengelin

(ORCID: 0000-0003-2212-4613): Conceptualization, Methodology, Investigation, Writing - Review & Editing; Daniel B. M. Haun (ORCID: 0000-0002-3262-645X): Conceptualization, Funding acquisition, Writing - Review & Editing.

Correspondence concerning this article should be addressed to Manuel Bohn (ORCID: 0000-0001-6006-1348), Universitätsallee 1, 21335 Lüneburg, Germany. E-mail: manuel.bohn@leuphana.de

Abstract

Theoretical accounts assume that key features of human social cognition are universal. Here we focus on gaze-following, the bedrock of social interactions and coordinated activities, to test this claim. In this comprehensive cross-cultural study spanning five continents and 17 distinct cultural communities, we examined the development of gaze-following in early childhood. We identified key processing signatures through a computational model that assumes that participants follow an individual's gaze by estimating a vector emanating from the eye-center through the pupil. Using a single reliable touchscreen-based task, we found these signatures in all communities, suggesting that children worldwide processed gaze in highly similar ways. Absolute differences in performance between groups are accounted for by a cross-culturally consistent relationship between children's exposure to touchscreens and their performance in the task. These results provide strong evidence for a universal process underlying a foundational socio-cognitive ability in humans that can be reliably inferred even in the presence of cultural variation in overt behavior.

Keywords: keywords

Word count: X

A universal of human social cognition: Children from 17 communities process gaze in similar ways

Introduction

Human socio-cognitive skills enable unique forms of communication and cooperation that provide a bedrock for cumulative culture and the formation of complex societies[1–7]. The eyes are the proverbial "window to the mind" and eye gaze is essential for many social reasoning processes[8–10]. Others' eye gaze is used to infer their focus of visual attention, which is a critical aspect of coordinated activities, including communication and cooperation[11–16].

The ability to follow gaze emerges early in development[17–20]. The earliest signs of gaze-following have been found in infants as young as four months[21,22]. Initially, infants rely more on head direction than actual gaze direction[23,24]. Throughout the first two years of life, children refine their abilities: they interpret gaze in mentalistic terms, for example, they follow gaze to locations outside their own visual field by moving around barriers[25]. Importantly, individual differences in children's gaze-following abilities predict later life outcomes, most notably communicative abilities[26]. For example, gaze-following at 10 months predicts language scores at 18 months of age[27]. Difficulties with gaze-following have been linked to developmental disorders, including Autism[28–30]. This work highlights the importance of gaze-following as a foundational building block of human social interaction and its central place in theorizing.

A central assumption in the theoretical and empirical work discussed above is that, despite substantial variation in developmental contexts, gaze-following works and develops in the same way across human societies[31]. This assumption – despite being central to many developmental theories – is currently not supported by evidence. On the contrary, cross-cultural studies have revealed substantial diversity in socio-cognitive development[3,32–35]. One of the very few cross-cultural studies also found differences in the likelihood to follow gaze between communities[36].

One potential source for this paradox lies in the reliance on aggregated measures in cross-cultural studies. Absolute differences in mean performance across communities are interpreted as a signal of different underlying cognitive processes. Here, we resolve this paradox by instead focusing on processing signatures that can be investigated independently of absolute community-level differences. This allows us to directly evaluate the empirical foundation of claims about universal features of human social cognition. To this end, we conducted a pre-registered, large-scale, cross-cultural study on the development of gaze-following abilities to study potentially universal processing signatures.

The processing signatures were derived from a simple computational model that assumes that participants follow gaze by estimating a vector emanating from the eye center through the pupil[37]. The key innovation of the model is that it explains how individuals may use the same cognitive process but still differ in their measured abilities. The process always involves estimating a vector but also involves a degree of uncertainty because the eye center is not directly observable. Individuals are assumed to differ in their level of uncertainty with which they estimate the vector which causes differences in their observable behavior. Importantly, the assumed process leaves a key signature in the data that is observable independent of the absolute level of performance. In the present study, we therefor focus on this signature instead of absolute levels of performance when evaluating the claim whether there is evidence for a universal cognitive mechanism underlying gaze-following.

The 1377 participants who took part in the study lived in 17 different communities across 14 countries and five continents (Fig. 1A, Tab. 1). These countries represent ~46% of the world's population. Communities covered a broad spectrum of geographical locations, social and political systems, languages, and subsistence styles (see Supplemental Materials). This diversity allowed us to overcome the common pitfall of cross-cultural studies that compare urban communities from the global north to rural communities from the global south[38].

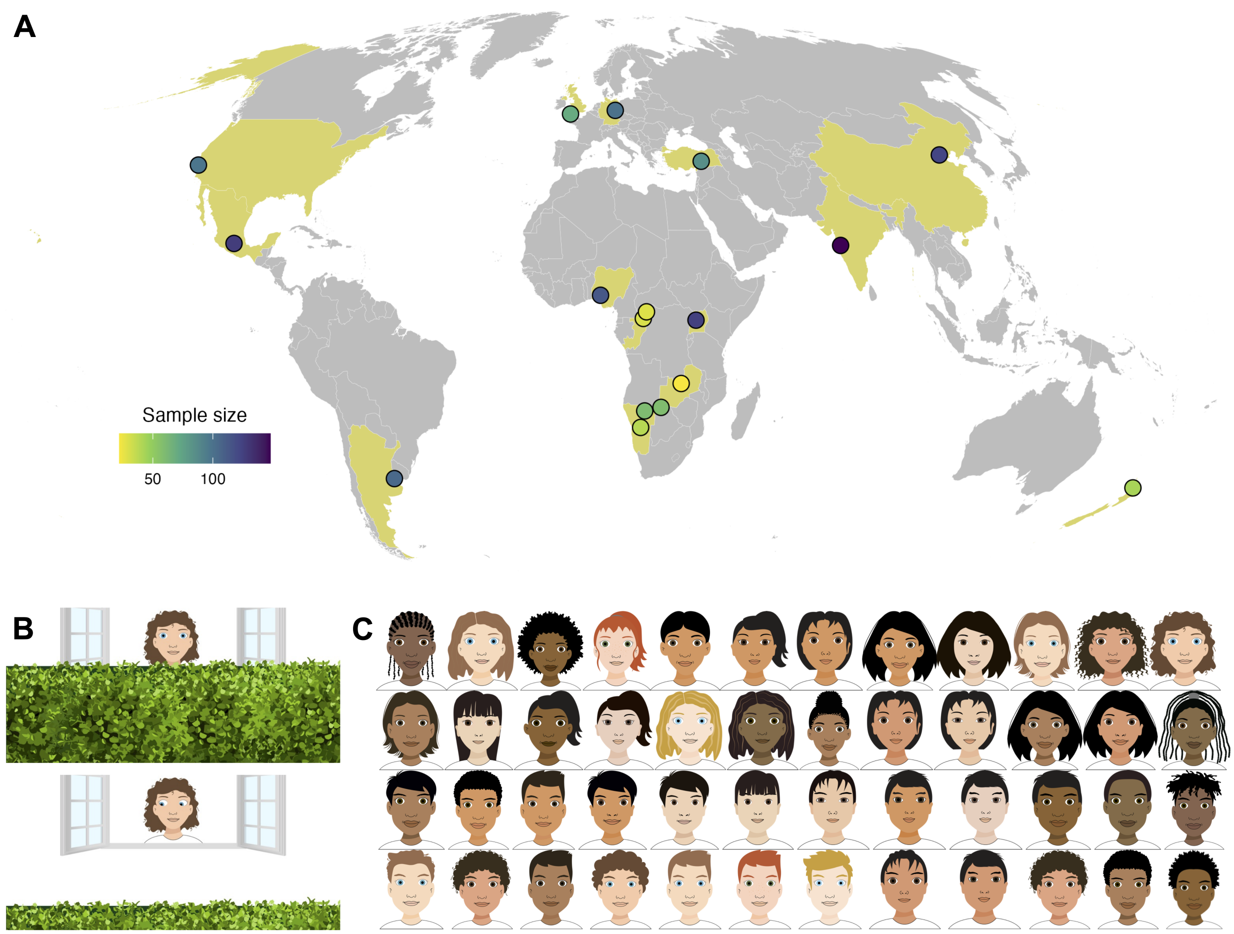

Figure 1. (A) Data collection sites. Points show the approximate geographical location of the data collection sites, coloring shows the sample sizes. (B) Screenshots from the task. Screenshots from the task. The upper scene depicts the start and the lower the choice phase in a test trial. Participants had to use the gaze of the agent to locate the balloon and touch the location on the hedge where they thought the balloon was. Agents, audio recordings and backgrounds were adapted to each cultural setting. (C) Drawings used as agents across cultural settings.

Table 1

Participant demographics.

Continent	Country	Community	N(male)	Age (range)	Language	Touchscreen exposure1
Americas	Argentina	Buenos Aires	105 (53)	4.72 (3.00 - 6.96)	Spanish (Rio-platense)	0.90
	México	Ocuilan	127 (63)	4.96 (2.57 - 6.95)	Spanish (Mexican)	0.77
	USA	Stanford	98 (54)	4.99 (2.52 - 7.90)	English (American)	0.98
Africa	Namibia	Hai||om	60 (38)	5.85 (2.74 - 8.34)	Hai||om	0.05
		Khwe	59 (24)	5.84 (3.38 - 8.63)	Khwedam	0.19
		Windhoek	39 (17)	5.69 (2.66 - 8.66)	English (Nigerian)2	0.95
	Nigeria	Akure	114 (54)	5.07 (2.57 - 7.33)	English (Nigerian)	0.91
	Rep. Congo	BaYaka	29 (13)	7.80 (3.94 - 10.56)	BaYaka	0.00
		Bandongo	30 (11)	7.45 (3.50 - 10.95)	Lingala	0.00
	Uganda	Nyabyeya	125 (62)	5.94 (2.67 - 8.92)	Kiswahili	0.34

Table 1 continued

Continent	Country	Community	N(male)	Age (range)	Language	Touchscreen exposure1
	Zambia	Chimfunshi	22 (5)	5.98 (2.88 - 8.00)	Bemba	0.14
Europe	Germany	Leipzig	100 (48)	4.88 (2.53 - 6.95)	German	0.89
	UK	Plymouth	70 (30)	6.02 (2.38 - 8.94)	English (British)	0.99
Asia	China	Beijing	123 (62)	5.47 (2.69 - 8.48)	Mandarin	0.95
	India	Pune	148 (73)	6.14 (3.06 - 8.83)	English (Indian) / Marathi	0.93
	Türkiye	Malatya	85 (40)	5.02 (2.75 - 7.12)	Turkish	1.00
Oceania	New Zealand	Auckland	43 (19)	5.14 (2.81 - 8.75)	English (New Zealand)	0.95

Note. 1 Proportion of participants who have access to touchscreens according to parental questionnaire. 2 Local collaborators and piloting suggested that Nigerian English is suitable for Windhoek as well.

We used an animated picture book tablet task in which participants had to locate a hidden object based on observing an agent's gaze. Children watched a balloon disappear behind a hedge. An agent followed the trajectory of the balloon with their eyes (Fig. 1B). The key dependent variable was the (im)precision with which children located the agent's focus of attention, that is, the deviation between where the agent looked (where the balloon was) and the child's response. We adapted visuals and audio instructions specifically for each of the 17 communities. Previous work demonstrated excellent individual-level measurement properties for this task in a German sample[39].

Results

Cross-cultural variation in development

As the first step, we investigated developmental improvements, that is, how children become more precise at estimating the target location with age. Across all 17 communities, we found a substantial increase in average levels of precision with age (fixed effect in Bayesian regression model[40]: β = -0.30, 95% Credible Interval (CrI) (-0.40 - -0.21); range of community-level (random) effects: β_{min} = -0.06, 95% CrI (-0.18 - 0.05) to β_{max} = -0.59, 95% CrI (-0.71 - -0.48)).

Nevertheless, there were also marked differences between communities (see Fig. 2A). In a six-fold cross-validation procedure, we trained a regression model on a subset of the data (training data) to later predict the held-out data (testing data)[41]. This procedure was repeated 100 times. We found that a model assuming cross-cultural variation in average performance as well as cross-cultural variation in developmental trajectories outperformed simpler models – assuming no variation in the shape of developmental trajectories or no variation between settings at all – in 98% of cases (see Supplemental Material). There are numerous ways in which communities could be grouped that would fall in line with these absolute differences (e.g., market integration, average levels of education, or average household size). However, we deliberately want to avoid any such overly simplistic explanation based on group-level data. Instead, we think these results can be best understood in methodological terms in the form of exposure to touch-screen devices; a finding we discuss in more detail below. Importantly, average differences in precision between communities were small compared to differences between individuals: communities did not form homogeneous clusters but largely overlapping distributions in that some individuals from communities with a lower average level of precision performed better compared to some individuals from a setting with a very high average level of precision. Similarly, in all communities, some 4-year-olds outperformed children two years older than them (see Fig. 2A). The lack of adequate individual-level measurement instruments in previous large-scale developmental cross-cultural studies made it impossible to contrast these perspectives. The substantial overlap between communities found here speaks against categorical differences in gaze-following and is suggestive of a universal underlying process. However, consistent developmental improvements and overlapping distributions alone cannot inform us about the cognitive processes children use when locating the agent's focus of attention.

Universal processing signatures

Recent computational work modeled gaze-following as social vector estimation[37]. When following gaze, onlookers observer the location of the pupil within the eye and estimate a vector emanating from the center of the eye through the pupil. The focus of attention is the location where the estimated vectors from both eyes hit a surface (Fig. 3). It is assumed that this estimation process has some uncertainty becasue the center of the eye is not directly observable and that individuals vary in their level of uncertainty. As a consequence, even though individuals use the same general process, they might differ in their absolute levels of precision. Crucially, this process model predicts a clear performance signature in our gaze-following task: Trials in which the agent looks further away from the center should result in lower levels of precision compared to trials in which the agent looks closer to the center. This prediction is best understood by considering a similar

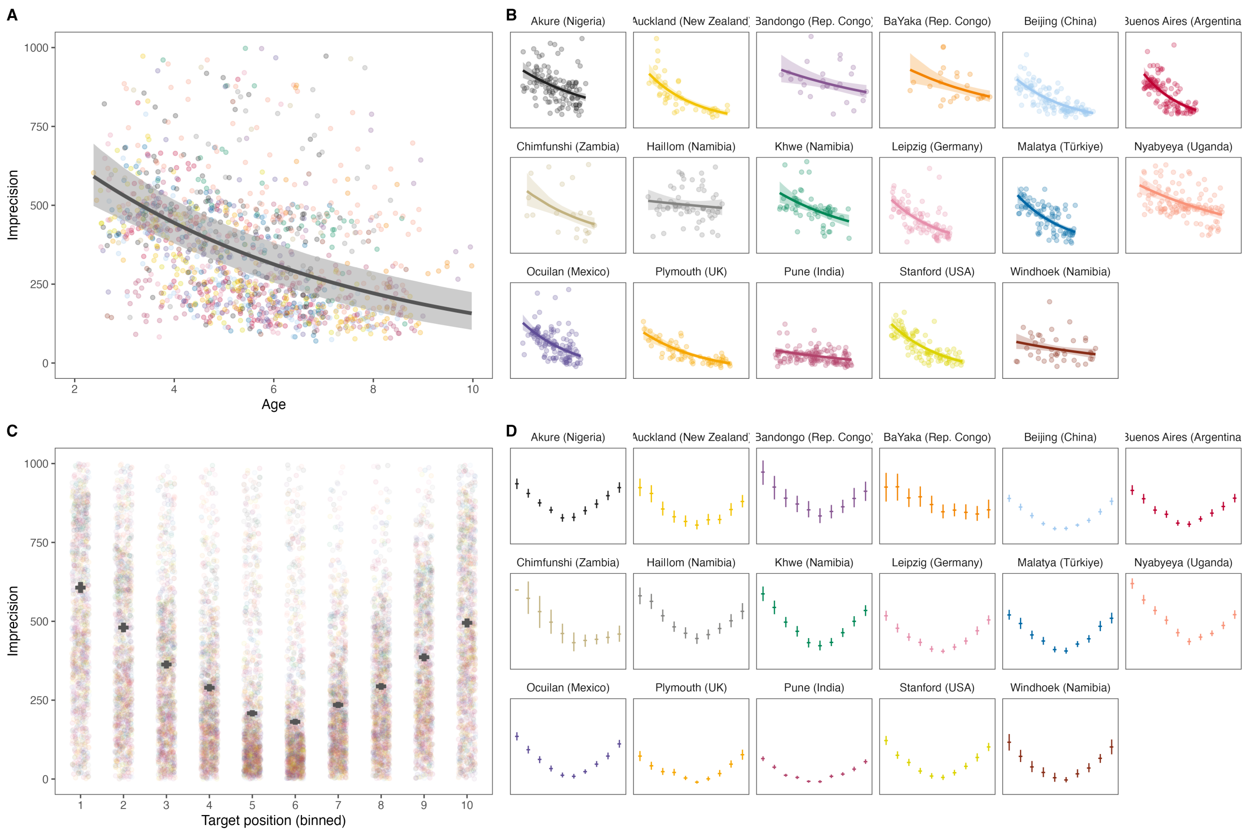

Figure 2. A) Developmental trajectory across and B) by community. The developmental trajectories are predicted based on a model of the data aggregated for each participant. C) Performance by target location on the screen across, and D) by community. Each bin covers 1/10th of the screen. Points show means, and error bars 95% confidence intervals for the data within that bin aggregated across participants. Transparent dots in A) and C) show aggregated data for each individual.

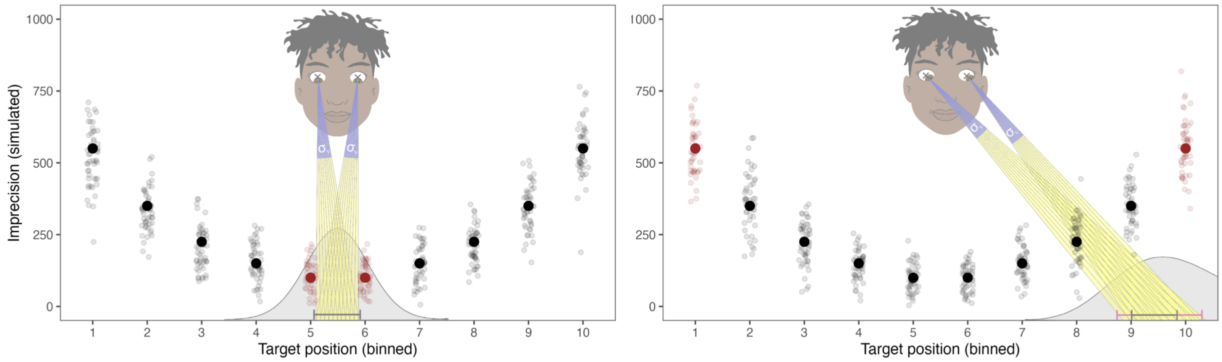

Figure 3. Graphical illustration of the cognitive model. Individuals infer the target of an agent's attention by estimating a vector based on the position of the pupils within the eyes. This process is noisy, illustrated by the different vectors (transparent lines). Individuals differ in their level of precision (indicated by sigma). For a given level of precision, the further the target lands from the centre of the screen, the less precise the model predicts individuals to be. Solid and transparent dots show simulated means and individual data points to illustrate the predicted effect of target position.

phenomenon: pointing a torch light to a flat surface. The width of the light beam represents each individual's level of uncertainty in vector estimation. When the torch is directed straight down, the light beam is concentrated in a relatively small area. When the torch is rotated to the side, the light from one half of the cone must travel further than the light from the other half to reach the surface. As a consequence, the light is spread over a wider area (see Fig. 3).

This processing signature was clearly visible across all 17 communities. Precision decreased when the agent looked at locations further away from the center (fixed effect: β = 0.47, 95% CrI (0.40 - 0.54); range of community-level effects: β_{min} = 0.58, 95% CrI (0.51 - 0.66) to β_{max} = 0.16, 95% CrI (-0.01 - 0.33)). Visualization of the data showed the predicted u-shaped pattern in all communities (see Fig. 2B). These results indicate a universal cognitive process used by children in all communities. There are, however, alternative ways in which the u-shaped pattern might arise: if participants ignored the agent's gaze and instead always selected the middle of the screen (center bias) or randomly selected locations (random guessing), precision would also decrease when the balloon lands further away from the center. To rule out these alternative explanations, we directly compared three models that made different assumptions about how participants' responses were generated: the focal vector-based gaze estimation model described above, a center-bias model where participants always select the center, and a random guessing model where participants select random locations. For every community, we found overwhelming support for the gaze estimation model (min $BF_{10} > 100\,000$ for comparisons with both alternative models). Taken together, children from all 17 communities processed gaze in similar ways.

Predictors of variation

Next, we looked at factors that could explain community- and individual-level variation. In addition to the gaze-following task, caregivers responded to a short questionnaire about children's access to screen-based technology and household composition. On an individual level, we found that children with access to touchscreen devices had higher levels of precision (β = -0.14, SE = 0.04, 95% CrI = -0.21 - -0.07). This effect was consistent across communities in that allowing the effect of access to touchscreens to vary across communities did not improve model fit (see Supplemental Materials). On a community level, we also saw that average performance was lowest in communities in which touchscreen devices were the least frequent (community-level correlation between age-corrected imprecision and proportion of children with access to touchscreens: r = -0.90, 95% CI = -0.96 - -0.74). Thus, familiarity with the device used for data collection likely explains variation between communities. Children with more touchscreen experience were probably better at task handling and thus more likely to precisely touch the location they inferred the agent to look at.

However, there was substantial variation between individuals that could not be explained by differential exposures to touchscreens alone. For example, in Malatya (Türkyie) where 100% of children had access to touchscreens there was still substantial variation between individuals (see Fig.1B). This strongly indicates that other factors likely contributed to individual differences. Social interaction has been highlighted as an important driver of social-cognitive development[e.g., 31,42–45] and thus we hypothesized (and pre-registered) that more opportunities for social interaction – approximated by living in larger households with more children – would be associated with higher levels of precision. When predicting performance by relative opportunities for social interactions within a community – while accounting for absolute differences and the prevalence of touchscreens – we found no strong associations between any of the demographic indicators and performance (see Supplemental Material). Whilst household size was a useful proxy for regular social interaction opportunities, the measure does not directly measure the factors that previous work has suggested to be related to the development of gaze-following in younger children, such as attachment quality or the use of gaze in early communicative interactions[46–48].

Discussion

Following and understanding gaze is a foundational building block of human social cognition[11–16]. A substantial body of work has explored the developmental onset of gaze-following in a few selected cultural communities[17–19,49]. The data reported here provides strong evidence that children from a large and diverse set of communities process others' gaze in similar ways. We found key performance signatures predicted by a model treating gaze-following as a form of social vector estimation across all 17 communities. With the focus on individual-level processing signatures, the study goes beyond previous studies on gaze-following – focused on the onset of gaze-following in infancy[36,50] – as well as comprehensive cross-cultural studies that compared average developmental trajectories[51–54].

The cognitive processes underlying gaze-following might be rooted in humans' evolved cognitive architecture, which is – presumably – later refined during social interaction[46–48]. The phylogenetic roots of these processes might possibly lie much deeper as primates from a wide range of species follow gaze[55–58]. Yet, similarities in overt behavior do not imply the same underlying cognitive processes. The present study defines clear performance signatures that can be explored in other species to test such evolutionary hypotheses.

Our study combined precise individual-level cognitive measurement and individual-level assessment of experience (here: touchscreen exposure) in a large and diverse sample to directly investigate the impact of specific cultural experiences on developmental outcomes. Instead of establishing universality by maximizing the cultural distance between two or three tested communities[59], this large-scale cross-cultural approach treats children's cultural experience at scale, shedding light on the big "middle ground" of children's cultural experience[38].

The study has important limitations. The fact that performance in the task was correlated with exposure to touchscreens might have overshadowed other sources of variation. However, we think it is in an important innovation that we were able to account for this effect. Most developmental cross-cultural studies do not even question the portability of their measurement instruments. Importantly, the key result that the processing signatures were seen in all cultural settings, is immune to this finding. The potential that lies in the otherwise precise individual-level measurement that our task achieves is largely unexploited. The questionnaire items only offer a very coarse picture into children's actual lived experiences. Future work could increase the resolution with which everyday experiences in children from diverse communities are recorded to compare the drivers behind social-cognitive development as we observe it. Recent work in the field of language acquisition has shown how technological innovations allowed for direct recording of social interactions across communities which can be used to close this explanatory gap[60,61].

In sum, our work pioneers an approach that introduces computational modeling and precise individual-level measurement to the cross-cultural study of cognitive development. This approach allowed us to test for universals in the human cognitive architecture rather than just overt behavior. As such, it can serve as a blueprint for future research on a broad spectrum of cognitive abilities and offers a much-needed empirical foundation for theories on the nature of the human mind. Children from diverse cultures deploy similar cognitive processes in interpreting gaze, pointing to a universal foundation of basic social cognition, which is refined during development.

Methods

A more detailed description of the materials and models can be found in the supplemental material. The experimental procedure and analysis plan were pre-registered (https://osf.io/tdsvc). We report on deviations from the pre-registered plan in the supplementary material. The task itself, including all the versions used in the study, can be accessed via the following website: https://ccp-odc.eva.mpg.de/tango-cc/. Data, model and analysis scripts can be found in an online repository (https://github.com/ccp-eva/gafo-cc-analysis).

Participants

A total of 1377 children between 2.38 and 10.95 provided data for the study. Children lived in 17 different communities, located in 14 different countries. Table 1 gives the sample size per community together with some basic demographic information. The recruitment strategy for each community is reported in the respective site descriptions. For some children, the exact birthday was unknown. In such cases, we set the birthday to the 30th of June of the year that would make them fall into the reported age category.

Data from children was only included in the study when they contributed at least four valid test trials. We also excluded the data from children with a diagnosed developmental disorder. In sum, in addition to the sample size reported above, 74 additional children participated in the study but did not contribute data. The main reasons for exclusion were: contribution of less than four valid test trials, technical failures, and missing or implausible demographic information (e.g., when the number of children living in the household was reported to be larger than the household itself or when the number of children reported to live in the household equaled the number of children younger than the child being tested). We did not exclude any participants for performance reasons. A detailed description of each data collection site and the way children were recruited can be found in the supplemental material.

Setup and Procedure

The task was implemented as a browser-based interactive picture book using `HTML` and `JavaScript`. Participants saw animated agents on a touch screen device, listened to pre-recorded audio instructions and responded by touching the screen. In all communities, a research assistant, fluent in the local language(s), guided the child through the task.

Figure 1B shows a screenshot from the task. The task was introduced verbally by the assistant as the balloon game in which the participant would play with other children to find a balloon. On each trial, participants saw an agent located in a window in the center of the screen. A balloon fell down from its starting position just below the agent. The agent's gaze followed the trajectory of the balloon. That is, the pupils and the iris were programmed to align with the center of the balloon. Once the balloon had landed on the ground, the agent was instructed to locate it, that is, to touch the location on the screen where they thought the balloon was. On each trial, we recorded the exact x-coordinate of the participant's touch.

There were two types of familiarization trials. In fam1 trials, the balloon fell down and landed in plain sight. Participants simply had to touch the visible balloon. In fam2 trials, the trajectory of the balloon was visible but it landed behind a small barrier (a hedge - see Figure 1B). Thus, participants needed to touch the hedge where they saw the balloon land. Next came test trials. Here, the barrier moved up and covered the balloon's trajectory. That is, participants only saw the agent's eyes move, but not the balloon. They had to infer the location of the balloon based on the agent's gaze direction. During fam1, fam2 and the first test trial, children heard voice overs commenting what happened on the screen. Critically, the agent was described as wanting to help the child and always looking at the balloon.

Children completed one fam1 trial, two fam2 trials and 16 test trials. We excluded the first test trial from the analysis because of the voice-over. Thus, 15 test trials were used in the analysis below.

Each child saw eight different agents, four male, four female. The agent changed from trial to trial, with alternating genders. A coin toss before the first trial decided whether the first agent was male or female. The order in which agents were shown was randomized with the constraint that all agents had to be shown once until an agent was shown again. The color of the balloon also changed from trial to trial in a random order, also with the constraint that all colors appeared once before any one was repeated.

The location (x-coordinate) where the balloon landed was determined in the following way: The screen was divided in ten equally sized bins. On each trial, one of the bins was randomly selected and the exact x-coordinate was randomly chosen within that bin. Constraints were that the balloon landed in each bin equally often and the same bin appeared no more than twice in a row.

All children were tested with a touchscreen device with a size between 11 and 13 inch equipped with a webcam. The data was either stored locally or sent to a server. In addition to the behavioral data, we stored the webcam recording of the session for verification purposes. Culture-specific adaptations were made by changing the visuals and the audio instructions (see supplementary material for details).

Analysis

We used Bayesian Regression models fit in `R`[62] using the package `brms`[63] for all analyses except the cognitive models. We used the default priors built in to `brms`. The dependent variable in all regression models was imprecision, that is, the absolute distance between the true location of the balloon (x-coordinate of its center) and the location where the participant touched the screen. We used a Log-normal distribution to model the data because the natural lower bound for imprecision is zero and the data was right skewed with a long tail. Numeric predictors that entered the models were scaled to have a mean of zero and a standard deviation of 1.

To analyse cross-cultural variation in performance, we used a cross-validation procedure[see e.g., 41]. In the supplemental material we give a detailed justification of this approach. For each cultural setting, we randomly sampled a data set that was 5/6 the size of the full data set (training data). Then, we fit the model to this training data and used the estimated model parameters to predict the remaining 1/6 of the data (testing data). We then compared the model predictions from the different models by computing the mean difference between the true and predicted imprecision, over all trials in the testing data set. We repeated the cross-validation procedure 100 times and computed the percentage of cases in which one model outperformed the other. We compared three models: a null model assuming no systematic community-level variation, a model assuming variation between communities and a model assuming variation between communities and in developmental trajectories (see supplemental material for model equations).

To evaluate the processing signature that trials in which the balloon lands further away from the center lead to larger imprecision, we fit a model predicting imprecision by age and target centrality (distance of the landing position form the center in pixel) with random intercepts for participant and cultural setting and random slopes for target centrality within participant and cultural setting (`brms` notation: `age + target_centrality + (target_centrality | participant) + (age + target_centrality | culture)`).

Cognitive model

The focal vector-based gaze estimation model has been described in detail in[37]. In brief, it inversely models the process generating touches on the screen based on observed eye movements. Formally, the model is defined as:

$$P(\theta|x_c, \alpha_l, \alpha_r) \propto P(x_c|\alpha_l, \alpha_r, \theta)P(\theta) \tag{1}$$

Here, θ represents an individual's cognitive ability to locate the focus of the agent's attention, x_c represents the touched coordinate, and α_l and α_r correspond to the left and right pupil angles (each 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).

The basic assumption in this model is that participants touch on the screen location where they think the agent is looking. The true eye angles (α_l and α_r) are not directly observable and are estimated with noise, yielding $\hat{\alpha}_l$ and $\hat{\alpha}_r$.

Each touch x_c implies a "matched pair" of estimated pupil angles $\hat{\alpha}_l$ and $\hat{\alpha}_r$, with the constraint that the lines extended along those two angles meet at the precise location of where the target is believed to be. As a consequence, we can rewrite the likelihood function of the model as:

$$P(x_c|\alpha_l, \alpha_r, \theta) \propto P(\hat{\alpha}_l, \hat{\alpha}_r|\alpha_l, \alpha_r, \theta)P(x_c) \tag{2}$$

$P(x_c)$ is a prior over potential target locations. Because the target was last visible in the screen and because the agent was located in the center, we assumed that participants have an a priori expectation that the target will land close to the middle. 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 primary inferential task for participants is therefore to estimate the pupil angles ($\hat{\alpha_l}$ and $\hat{\alpha_r}$), i.e., to sample from the term $P(\hat{\alpha_l}, \hat{\alpha_r}|\alpha_l, \alpha_r, \theta)$. Here, we assumed that the pair of estimated pupil angles were sampled from a probability distribution which is the product of two Normal distributions of equal variance, σ_v, centered on the true pupil angles:

$$P(\hat{\alpha_l}, \hat{\alpha_r}|\alpha_l, \alpha_r, \theta) \propto \phi(\hat{\alpha_l}; \alpha_l, \sigma_v)\phi(\hat{\alpha_r}; \alpha_r, \sigma_v), \tag{3}$$

Here, σ_v determines the level of accuracy with which participants estimated the pupil angles, and it is thus the component of the model that defines θ. Smaller values of σ_v result in a narrow distribution around the pupil angle, making touches far away from the target less likely. Conversely, larger values for σ_v lead to a wider distribution, making touches far away from the target more likely. To circle back to the analogy introduced above, σ_v corresponds to the width of the light beam. Thus, the goal of the model was to estimate participant-specific values for σ_v: σ_{v_i}. For more details on how σ_{v_i} was estimated, see the supplementary material.

To summarize, the model assumes that participant's touches are generated by a process that relies on noisy estimates of the agent's gaze direction. The precision, with which the gaze direction is estimated, varies between participants and increases with development.

The two alternative models assume that participants ignore the agent's gaze completely, instead they are assumed to follow simple heuristics. According to the center bias model, they always try to touch in the center of the screen: $P(x_c) \sim \mathcal{N}(960, 160)$. (960 is the x-coordinate of the center and 160 is the width of the balloon). According to the random guessing model, they randomly touch coordinates on the screen: $P(x_c) \sim \mathcal{U}(0, 1920)$.

All models were run separately for each cultural setting. The code to run the models can be found in the associated online repository. We also refer to this source for information on the prior distributions for all model parameters.

The cognitive models were implemented in the probabilistic programming language `webppl`[64]. We compared models based on the marginal likelihood of the data for each model, which represents the likelihood of the data while averaging over the prior distribution on parameters. The pair-wise ratio of marginal likelihoods for two models is known as the Bayes Factor. Bayes Factors are a quantitative measure of the predictive quality of a model, taking into account the possible values of the model parameters weighted by their prior probabilities. The incorporation of the prior distribution over parameters in the averaging process implicitly considers model complexity: models with more parameters typically exhibit broader prior distributions over parameter values and broader prior distribution can attenuate the potential gains in predictive accuracy that a model with more parameters might otherwise achieve[65].

References

1. Tomasello, M. The adaptive origins of uniquely human sociality. *Philosophical Transactions of the Royal Society B* **375**, 20190493 (2020).
2. Laland, K. & Seed, A. Understanding human cognitive uniqueness. *Annual Review of Psychology* **72**, 689–716 (2021).
3. Wellman, H. M. *Making minds: How theory of mind develops.* (Oxford University Press, 2014).
4. Henrich, J. *The secret of our success: How culture is driving human evolution, domesticating our species, and making us smarter.* (princeton University press, 2016).
5. Tomasello, M. & Rakoczy, H. What makes human cognition unique? From individual to shared to collective intentionality. *Mind & language* **18**, 121–147 (2003).
6. Legare, C. H. The development of cumulative cultural learning. *Annual Review of Developmental Psychology* **1**, 119–147 (2019).
7. Heyes, C. *Cognitive gadgets.* (Harvard University Press, 2018).
8. Shepherd, S. V. Following gaze: Gaze-following behavior as a window into social cognition. *Frontiers in integrative neuroscience* **4**, 5 (2010).
9. Doherty, M. J. The development of mentalistic gaze understanding. *Infant and Child Development* **15**, 179–186 (2006).
10. Emery, N. J. The eyes have it: The neuroethology, function and evolution of social gaze. *Neuroscience & biobehavioral reviews* **24**, 581–604 (2000).
11. Tomasello, M., Hare, B., Lehmann, H. & Call, J. Reliance on head versus eyes in the gaze following of great apes and human infants: The cooperative eye hypothesis. *Journal of Human Evolution* **52**, 314–320 (2007).

12. Scaife, M. & Bruner, J. S. The capacity for joint visual attention in the infant. *Nature* **253**, 265–266 (1975).

13. Langton, S. R., Watt, R. J. & Bruce, V. Do the eyes have it? Cues to the direction of social attention. *Trends in cognitive sciences* **4**, 50–59 (2000).

14. Rossano, F. Gaze in conversation. *The handbook of conversation analysis* 308–329 (2012).

15. Richardson, D. C. & Dale, R. Looking to understand: The coupling between speakers' and listeners' eye movements and its relationship to discourse comprehension. *Cognitive science* **29**, 1045–1060 (2005).

16. Sebanz, N., Bekkering, H. & Knoblich, G. Joint action: Bodies and minds moving together. *Trends in cognitive sciences* **10**, 70–76 (2006).

17. Tang, Y., Gonzalez, M. R. & Deák, G. O. The slow emergence of gaze-and point-following: A longitudinal study of infants from 4 to 12 months. *Developmental Science* e13457 (2023).

18. Gredebäck, G., Fikke, L. & Melinder, A. The development of joint visual attention: A longitudinal study of gaze following during interactions with mothers and strangers. *Developmental science* **13**, 839–848 (2010).

19. Byers-Heinlein, K. *et al.* The development of gaze following in monolingual and bilingual infants: A multi-laboratory study. *Infancy* **26**, 4–38 (2021).

20. Del Bianco, T., Falck-Ytter, T., Thorup, E. & Gredebäck, G. The developmental origins of gaze-following in human infants. *Infancy* **24**, 433–454 (2019).

21. Astor, K., Thiele, M. & Gredebäck, G. Gaze following emergence relies on both perceptual cues and social awareness. *Cognitive Development* **60**, 101121 (2021).

22. D'Entremont, B., Hains, S. M. & Muir, D. W. A demonstration of gaze following in 3-to 6-month-olds. *Infant Behavior and Development* **20**, 569–572 (1997).

23. Michel, C., Kayhan, E., Pauen, S. & Hoehl, S. Effects of reinforcement learning on gaze following of gaze and head direction in early infancy: An interactive eye-tracking study. *Child Development* **92**, e364–e382 (2021).

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

25. Moll, H. & Tomasello, M. 12-and 18-month-old infants follow gaze to spaces behind barriers. *Developmental science* **7**, F1–F9 (2004).

26. Carpenter, M., Nagell, K., Tomasello, M., Butterworth, G. & Moore, C. Social cognition, joint attention, and communicative competence from 9 to 15 months of age. *Monographs of the society for research in child development* i–174 (1998).

27. Brooks, R. & Meltzoff, A. N. The development of gaze following and its relation to language. *Developmental science* **8**, 535–543 (2005).

28. Itier, R. J. & Batty, M. Neural bases of eye and gaze processing: The core of social cognition. *Neuroscience & Biobehavioral Reviews* **33**, 843–863 (2009).

29. Thorup, E. *et al.* Altered gaze following during live interaction in infants at risk for autism: An eye tracking study. *Molecular autism* **7**, 1–10 (2016).

30. Thorup, E. *et al.* Reduced alternating gaze during social interaction in infancy is associated with elevated symptoms of autism in toddlerhood. *Journal of Abnormal Child Psychology* **46**, 1547–1561 (2018).

31. Tomasello, M. *Becoming human: A theory of ontogeny.* (Harvard University Press, 2019).

32. Miller, J. G., Wice, M. & Goyal, N. Contributions and challenges of cultural research on the development of social cognition. *Developmental Review* **50**, 65–76 (2018).

33. Mayer, A. & Träuble, B. E. Synchrony in the onset of mental state understanding across cultures? A study among children in samoa. *International Journal of Behavioral Development* **37**, 21–28 (2013).

34. Dixson, H. G., Komugabe-Dixson, A. F., Dixson, B. J. & Low, J. Scaling theory of mind in a small-scale society: A case study from vanuatu. *Child Development* **89**, 2157–2175 (2018).

35. Taumoepeau, M., Sadeghi, S. & Nobilo, A. Cross-cultural differences in children's theory of mind in iran and new zealand: The role of caregiver mental state talk. *Cognitive Development* **51**, 32–45 (2019).

36. Callaghan, T. *et al.* Early social cognition in three cultural contexts. *Monographs of the society for research in child development* i–142 (2011).

37. Prein, J. C., Maurits, L., Werwach, A., Haun, D. B. M. & Bohn, M. Variation in gaze understanding across the life span: A process-level perspective. *PsyArXiv* (2023). doi:10.31234/osf.io/dy73a

38. Barrett, H. C. Towards a cognitive science of the human: Cross-cultural approaches and their urgency. *Trends in Cognitive Sciences* **24**, 620–638 (2020).

39. Prein, J. C., Kalinke, S., Haun, D. B. & Bohn, M. 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* 1–17 (2023).

40. Bürkner, P.-C. Bayesian item response modeling in r with brms and stan. *arXiv preprint arXiv:1905.09501* (2019).

41. Stengelin, R., Ball, R., Maurits, L., Kanngiesser, P. & Haun, D. B. Children over-imitate adults and peers more than puppets. *Developmental Science* **26**, e13303 (2023).

42. Carpendale, J. & Lewis, C. *What makes us human: How minds develop through social interactions.* (Routledge, 2020).

43. Barresi, J. & Moore, C. Intentional relations and social understanding. *Behavioral and brain sciences* **19**, 107–122 (1996).

44. Perner, J., Ruffman, T. & Leekam, S. R. Theory of mind is contagious: You catch it from your sibs. *Child Development* **65**, 1228–1238 (1994).

45. Rakoczy, H. Foundations of theory of mind and its development in early childhood. *Nature Reviews Psychology* **1**, 223–235 (2022).

46. Astor, K. *et al.* Social and emotional contexts predict the development of gaze following in early infancy. *Royal Society open science* **7**, 201178 (2020).

47. Movellan, J. R. & Watson, J. S. The development of gaze following as a bayesian systems identification problem. in *Proceedings 2nd international conference on development and learning. ICDL 2002* 34–40 (IEEE, 2002).

48. Senju, A. *et al.* Early social experience affects the development of eye gaze processing. *Current Biology* **25**, 3086–3091 (2015).

49. Moore, C. The development of gaze following. *Child Development Perspectives* **2**, 66–70 (2008).

50. Hernik, M. & Broesch, T. Infant gaze following depends on communicative signals: An eye-tracking study of 5-to 7-month-olds in vanuatu. *Developmental science* **22**, e12779 (2019).

51. House, B. R. *et al.* Universal norm psychology leads to societal diversity in prosocial behaviour and development. *Nature Human Behaviour* **4**, 36–44 (2020).

52. Van Leeuwen, E. J. *et al.* The development of human social learning across seven societies. *Nature Communications* **9**, 2076 (2018).

53. Blake, P. R. *et al.* The ontogeny of fairness in seven societies. *Nature* **528**, 258–261 (2015).

54. Kanngiesser, P. *et al.* Children across societies enforce conventional norms but in culturally variable ways. *Proceedings of the National Academy of Sciences* **119**, e2112521118 (2022).

55. Rosati, A. G. & Hare, B. Looking past the model species: Diversity in gaze-following skills across primates. *Current opinion in neurobiology* **19**, 45–51 (2009).

56. Tomasello, M., Call, J. & Hare, B. Five primate species follow the visual gaze of conspecifics. *Animal behaviour* **55**, 1063–1069 (1998).

57. Itakura, S. Gaze-following and joint visual attention in nonhuman animals. *Japanese Psychological Research* **46**, 216–226 (2004).

58. Kano, F. & Call, J. Cross-species variation in gaze following and conspecific preference among great apes, human infants and adults. *Animal Behaviour* **91**, 137–150 (2014).

59. Norenzayan, A. & Heine, S. J. Psychological universals: What are they and how can we know? *Psychological Bulletin* **131**, 763 (2005).

60. Bergelson, E. *et al.* Everyday language input and production in 1,001 children from six continents. *Proceedings of the National Academy of Sciences* **120**, e2300671120 (2023).

61. Donnelly, S. & Kidd, E. The longitudinal relationship between conversational turn-taking and vocabulary growth in early language development. *Child Development* **92**, 609–625 (2021).

62. R Core Team. *R: A language and environment for statistical computing.* (R Foundation for Statistical Computing, 2023).

63. Bürkner, P.-C. Brms: An r package for bayesian multilevel models using stan. *Journal of Statistical Software* **80**, 1–28 (2017).

64. Goodman, N. D. & Stuhlmüller, A. The design and implementation of probabilistic programming languages. (2014).

65. Lee, M. D. & Wagenmakers, E.-J. *Bayesian cognitive modeling: A practical course.* (Cambridge University Press, 2014).