

Measuring Variation in Gaze Following Across Communities, Ages, and Individuals: A Showcase of TANGO-CC (Task for Assessing iNdividual differences in Gaze understanding-Open-Cross-Cultural)



Julia Christin Prein^{1,2}, Florian M. Bednarski³, Ardain Dzabatou⁴,
Michael C. Frank⁵, Annette M. E. Henderson³, Josefine Kalbitz²,
Patricia Kanngiesser⁶, Dilara Keşşafoglu⁷, Bahar Köymen⁸,
Maira V. Manrique-Hernandez², Shirley Magazi⁹, Lizbeth Mújica-Manrique²,
Julia Ohlendorf², Damilola Olaoba², Wesley R. Pieters⁹,
Sarah Pope-Caldwell^{2,10}, Umay Sen¹¹, Katie Slocombe¹²,
Robert Z. Sparks⁵, Roman Stengelin^{2,9}, Jahnavi Sunderarajan²,
Kirsten Sutherland², Florence Tusiime¹³, Wilson Vieira²,
Zhen Zhang¹⁴, Yufei Zong^{14,15}, Daniel B. M. Haun², and
Manuel Bohn^{1,2}

¹Institute of Psychology in Education, Leuphana University Lüneburg, Lüneburg, Germany; ²Department of Comparative Cultural Psychology, Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany; ³School of Psychology, University of Auckland, Auckland, New Zealand; ⁴Université Marien Ngouabi, Brazzaville, Republic of the Congo; ⁵Department of Psychology, Stanford University, Stanford, California; ⁶School of Psychology, University of Plymouth, Plymouth, England; ⁷Department of Psychology, Koç University, Istanbul, Türkiye; ⁸Division of Psychology, Communication, and Human Neuroscience, University of Manchester, Manchester, England; ⁹Department of Psychology and Social Work, University of Namibia, Windhoek, Namibia; ¹⁰Department of Psychology, Georgia State University, Atlanta, Georgia; ¹¹Department of Psychology, Developmental Psychology, Uppsala University, Uppsala, Sweden; ¹²Department of Psychology, University of York, York, England; ¹³Budongo Conservation Field Station, Masindi, Uganda; ¹⁴CAS Key Laboratory of Behavioral Science, Institute of Psychology, Chinese Academy of Sciences, Beijing, China; and ¹⁵Department of Cognitive Sciences, University of California Irvine, Irvine, California

Abstract

Cross-cultural studies are crucial for investigating the cultural variability and universality of cognitive developmental processes. However, cross-cultural assessment tools in cognition across languages and communities are limited. In this article, we describe a gaze-following task designed to measure basic social cognition across individuals, ages, and communities (the Task for Assessing iNdividual differences in Gaze understanding-Open-Cross-Cultural; TANGO-CC). The task was developed and psychometrically assessed in one cultural setting and, with input of local collaborators, adapted for cross-cultural data collection. Minimal language demands and the web-app implementation allow fast and easy contextual adaptations to each community. TANGO-CC captures individual- and community-level variation and shows

Corresponding Author:

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



good internal consistency in a data set of 2.5- to 11-year-old children from 17 diverse communities. Within-communities variation outweighed between-communities variation. We provide an open-source website for researchers to customize and use the task (<https://ccp-odc.eva.mpg.de/tango-cc>). TANGO-CC can be used to assess basic social cognition in diverse communities and provides a roadmap for researching community-level and individual-level differences across cultures.

Keywords

cross-cultural psychology, social cognition, gaze following, individual differences, reliability, open data, open materials

Received 6/19/24; Revision accepted 11/26/24

For decades, researchers have advocated for more diverse samples in psychological research and cautioned against relying solely on convenience samples from the Global North (Arnett, 2008; Henrich et al., 2010; Lillard, 1998). Despite numerous calls for change, the samples reported in major psychology journals still lack diversity (Apicella et al., 2020; Gutchess & Rajaram, 2023; Thalmayer et al., 2021). This hinders progress in theory building and testing: One cannot draw inferences about universal and variable aspects of the human cognitive system from data collected exclusively in one single community (Krys et al., 2024). Although this sampling bias is often discussed in adult psychology, it is equally relevant to developmental psychology (Nielsen et al., 2017). Early experiences shape the way children think about and interact with the world, and an ontogenetic perspective is needed to explore the foundational aspects of human behavioral diversity (Amir & McAuliffe, 2020; Broesch et al., 2023; Liebal & Haun, 2018; Torr ns et al., 2023).

There are numerous challenges with collecting cross-cultural, developmental data (Amir & McAuliffe, 2020; Broesch et al., 2023). Cross-cultural studies need reliable and valid measures to capture variation between communities and/or individuals systematically. Even though this applies to all areas of cognitive development, we focus on social cognition in this article.

Social cognition refers to how individuals process information in social situations that allows them to understand and predict others' behavior (Adolphs, 1999; Decety, 2020; Frith & Frith, 2007; Zeigler-Hill et al., 2015). If, in theory, stimuli used in social-cognition tasks should relate to people's everyday experiences, then tasks themselves should be tuned to the features of specific communities. Indeed, task performance can be diminished when stimuli do not reflect the characteristics of the participants' communities (Pe a, 2007). For example, Elfenbein and Ambady (2002) found better emotion recognition for members of the same national, ethnic, or regional group. Selcuk et al. (2023) concluded that children often attribute mental states more accurately

and more frequently to individuals from the same community.

From a psychometric perspective, the situation looks dire. Studies on social cognition with U.S.-American and European samples rarely report psychometric information (for a review, see Beaudoin et al., 2020), and the picture further deteriorates when one looks at cross-cultural social-cognition tasks (Bourdage et al., 2023; Hajd k et al., 2020; Waschl & Chen, 2022). Thus, it is already challenging to find reliable and valid tasks that have measurement sensitivity to detect individual differences in one community, let alone tasks that do so across different communities. In this article, we describe the construction and psychometric evaluation of a cross-cultural measure of basic social cognition (gaze following) in children as a concrete example of how to address this problem.

The approaches that researchers can take to collect cross-cultural data lie on a continuum: The decision for a specific method partly depends on whether researchers aim to increase the depth (culture specificity) or breadth (standardization across multiple communities) of their work (Amir & McAuliffe, 2020). At one extreme, researchers translate the psychological construct into a separate design or task for each community (termed "assembly"; He & van de Vijver, 2012; Waschl & Chen, 2022). Although this approach allows greater flexibility and sensitivity to cultural differences, it might not be feasible to study a multitude of communities because it becomes too demanding and time-consuming. Furthermore, the results are limited to each community, and absolute task scores might not be comparable across communities. A study that followed this approach is Wefers et al. (2023), who investigated how cultural variations in parenting styles modulated infants' responses to disruptions in social interactions. Although studies in the Global North often apply the still-face paradigm to assess infants' reactions to unresponsive partners, Wefers et al. reasoned that this paradigm might not capture infants' everyday interaction routines in communities

with proximal (i.e., emphasis on body stimulation) parenting styles. By developing the novel no-touch paradigm, they found that indeed, infants' responses to unresponsive partners were modulated by the cultural context in which they grew up: Kichwa infants from rural Ecuador showed stronger reactions to unresponsive partners in the no-touch paradigm compared with the still-face paradigm, whereas reactions of urban German infants differed less in both paradigms.

At the other extreme, researchers use the same standardized procedure across diverse communities, potentially providing a simple translation or modification of stimuli to ensure they are culturally appropriate (termed "adoption" and "adaptation," respectively; He & van de Vijver, 2012; Waschl & Chen, 2022). This approach is less sensitive to each community's unique characteristics but renders quantitative comparisons of data more feasible. An example of this approach is the Multilingual Assessment Instrument for Narratives (MAIN; Gagarina et al., 2012), which assesses narrative abilities in monolingual and multilingual children. Extensive piloting and adaptation of MAIN materials ensured that the instrument is culturally appropriate, robust, and suitable for cross-linguistic comparisons (Gagarina et al., 2012), and new and revised language versions are continuously added to the MAIN database (Gagarina & Lindgren, 2020).

In the present article, we aim to describe the development and psychometric properties of a social-cognition task that can be adapted to diverse communities. On the continuum described above, our task lies more toward standardized approaches but allows for some customization of the stimuli to each local community. The task focuses on one of the most fundamental social-cognitive abilities: gaze following, that is, the ability to identify the attentional focus of another agent. Gaze following develops early in infancy (Del Bianco et al., 2019; Tang et al., 2024) and contributes to social learning, communication, and collaboration (Bohn & Köymen, 2018; Hernik & Broesch, 2019; Shepherd, 2010; Tomasello et al., 2007). Identifying the attentional focus of a conversational partner facilitates language development (Brooks & Meltzoff, 2015; Carpenter et al., 1998; Mundy et al., 2007) given that children might use gaze to identify the referent of a new word (Brooks & Meltzoff, 2005).

One of the main open questions is the impact of social-environmental and cultural factors on gaze following (Astor & Gredebäck, 2022), but studies on variation in gaze following across communities are rare. Recently, Brooks et al. (2020) found that Deaf infants show increased gaze following compared with hearing infants, underlining the importance of visual-communicative signals for Deaf children. Callaghan et al. (2011) investigated gaze following behind barriers in 12- and 17-month-olds from rural Canada ($n = 35$), Peru ($n = 38$),

and India ($n = 65$). In their setup, an agent looked at a toy behind a barrier (experimental condition) or a sticker in front of a barrier (control condition), and children's crawling toward the barrier to follow the agent's gaze was assessed. Although the absolute crawling rates differed across communities, children in all three communities crawled more often to gain visual access when the agent looked at an object behind the barrier than in front of it. Hernik and Broesch (2019) studied 22 infants between 5 and 7 months of age from Vanuatu and used an eye-tracking procedure displaying a local actor looking at one of two objects. Even though face-to-face interactions are less common in parent-child interactions in Vanuatu than in Western communities, the resulting patterns of gaze following in Ni-Vanuatu infants resembled those of their Western counterparts: Infants from Vanuatu followed the agent's gaze to the target object only when preceded by infant-directed speech and not adult-directed speech. Astor et al. (2022) analyzed the effects of maternal postpartum depression on gaze following in 9-month-old infants from Bhutan and Sweden. Although infants showed similar gaze-following rates in both countries, maternal postpartum depression was associated with reduced gaze following only in the Swedish sample. Beyond gaze following, some studies conducted in the Global South have investigated opportunities for mutual eye gaze and factors influencing joint attention (see Akhtar & Gernsbacher, 2008). For example, Kaluli infants in Papua New Guinea are often carried facing away from their mothers, so infants and caregivers share mutual gaze less frequently (Ochs & Schieffelin, 1984). Childers et al. (2007) studied Ngas-speaking infants in Nigeria and found no effect of the caregivers' carrying style (usually on the back) on children's joint-attention tendencies.

Although these studies point toward potential cross-cultural stability of gaze following, the lack of psychometrically evaluated tasks and the small number of communities studied limit the generalizability of these findings and—more importantly—do not allow for studying individual differences. A more comprehensive cross-cultural study on gaze following was recently conducted by Bohn, Prein, et al. (2024). The researchers tested the universality of gaze following by studying 17 different communities on five continents and found evidence for a similar processing mechanism across communities.

The task presented in this article was developed for the study by Bohn, Prein, et al. (2024) and is based on a previously established gaze-following task called "TANGO" (Task for Assessing iNdividual differences in Gaze understanding-Open) by Prein, Kalinke, et al. (2024). TANGO measures how precisely participants locate an agent's attentional focus. It reliably measured individual differences in a German child sample and an

English-speaking remote adult sample (Prein, Kalinke, et al., 2024). However, we cannot claim the task’s generalizability and reliability based on a monocultural sample. In this article, we showcase TANGO-Cross-Cultural (TANGO-CC), a standardized gaze-following task that has been adapted to 13 languages and even more communities, and evaluate its psychometric properties by leveraging a large and diverse data set of 2.5- to 11-year-olds from 17 diverse communities (Bohn, Prein, et al., 2024). We describe the task’s development and provide a tutorial for the open-source website (<https://ccp-odc.eva.mpg.de/tango-cc/>).

Task Development

Approach

TANGO-CC was implemented in Leipzig, Germany, and thoroughly assessed in terms of reliability and validity (Prein, Kalinke, et al., 2024). During this process, the cross-cultural adaptation of the task was prepared by a team of cross-cultural psychologists and cognitive scientists. In this article, we assess TANGO-CC’s measurement quality (i.e.,

variability and reliability) across 17 diverse communities by analyzing the data set from Bohn, Prein, et al. (2024). In the following, we describe the different steps in detail.

TANGO-CC is a screen-based, interactive picture book with cartoon-like line drawings (see Fig. 1). Previous gaze studies have successfully implemented tasks with (schematic) line drawings for both children and adults (e.g., Anderson & Doherty, 1997; Doherty & Anderson, 1999; Doherty et al., 2009; Kingstone et al., 2000; Lee et al., 1998). TANGO-CC measures the imprecision with which participants locate a balloon by following an agent’s gaze. Participants click or touch the location on the screen where they believe the balloon to be. Precision is measured as the distance between the participant’s click on the screen and the balloon’s real position.

During the task development, we decided to implement the task’s main functionality independently of the task’s appearance. We programmed a function that calculates the x and y coordinates of where the agent’s pupil and iris should move to follow the balloon given the eyes’ and balloon’s original positions and measures. Because the measures of the eyes and balloon are read out dynamically from the image on screen, stimuli can

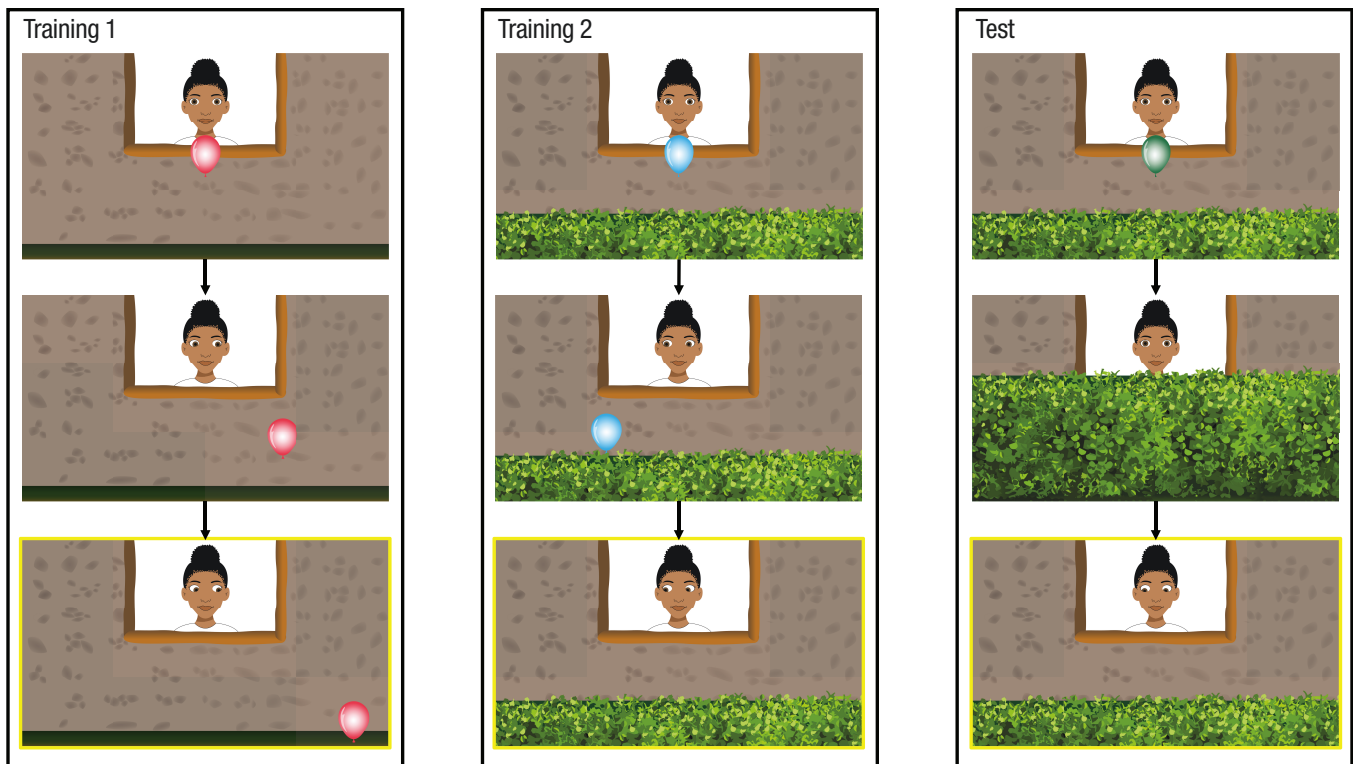


Fig. 1. Screenshots of TANGO-CC trials. In Training 1, an agent looks at a balloon that falls to the ground, and participants have to respond by clicking/touching the balloon. In Training 2, the balloon falls behind the hedge while its flight is still visible. Participants respond by clicking the hedge where they think the balloon is. In test trials, the balloon’s movement and final position are covered by a hedge, and participants respond by clicking the hedge. In the task, all movements are smoothly animated (no still pictures). Yellow frames indicate the time point when participants respond (only illustrative, not shown during the task). TANGO-CC = Task for Assessing individual differences in Gaze understanding-Open-Cross-Cultural.

be easily adapted and exchanged (i.e., no coordination values for animation are hard-coded into the task's source code). After having programmed this "backbone" functionality of the task (i.e., animate the eyes so that they follow the balloon), we added the task's audio instructions and superficial appearance (e.g., background scene, hedge, agent faces).

This basic version of TANGO was psychometrically evaluated in a German child sample and an English-speaking remote adult sample and was found to be highly reliable and valid (Prein, Kalinke, et al., 2024). Although children became more precise in locating the agent's attentional focus with age, individuals differed across all age groups and showed no floor or ceiling effects. In the Leipzig sample, performance on TANGO was weakly related to factors of children's daily social environment and could predict children's receptive vocabulary 6 months later. In a computational cognitive model, Prein, Maurits, et al. (2024) described gaze following as a form of social vector following and empirically found that performance on TANGO was related to children's nonsocial vector following and visual perspective-taking abilities. These connections to related constructs indicate the task's convergent validity in the German child sample.

To adapt the task for cross-cultural data collection, we generated a set of human cartoon faces and background scenes with input from local researchers and research assistants. The stimulus pool was adjusted and expanded until the researchers and research assistants from each target community judged the selected stimuli to be representative of the local population and typical accommodation (see Fig. 2). Audio instructions were translated from English or German into the corresponding local language(s). By back-translating these instructions, we ensured the original meaning did not change. Sometimes, specific words were slightly modified in the target language (e.g., "bush" instead of "hedge") to ensure that all participants understood the instructions well. Based on these adaptations, TANGO-CC could be applied in 17 communities and 13 different languages (Bohn, Prein, et al., 2024). In the following, we describe how researchers can use and customize TANGO-CC in more detail.

Features of TANGO-CC

Trials. The task consists of three different trial types: Training 1, Training 2, and test trials (see Fig. 1). In every trial, participants see an agent (boy or girl) looking out of a house with a balloon (red, blue, green, or yellow) in front of the agent. The balloon falls down to the ground. The eyes of the agent follow the movement of the balloon in a way that the balloon center and the pupil center always align. Depending on the trial type, participants' visual

access to the balloon's position varies. This experimental design builds on existing gaze-following paradigms that have used barriers or obstacles (e.g., Butler et al., 2000; Dunphy-Lelii & Wellman, 2004; Franco & Gagliano, 2001; Moll & Tomasello, 2004). In Training 1, participants see the full trajectory of the balloon and directly have to click on the balloon itself. In Training 2, participants see most of the balloon's movement, but a hedge covers the final location. In test trials, a hedge grows at the beginning of the trial, and participants see neither the movement nor the final position of the balloon.

The first trial of each type contains an audio description of the presented events (see the Supplemental Material available online). Note that the instructions explicitly state that the agent is looking at the balloon. This ensures all participants understand the purpose of the task and minimizes learning effects (increasing comparability between earlier and later trials, avoiding that participants notice the gaze cue only after, e.g., half of the trials). TANGO-CC measures how precisely participants use the gaze cue rather than if they notice it in the first place.

The outcome variable is the distance between the participant's click and the balloon's center. Trials can be completed quickly and efficiently so that children can complete 15 trials within 10 min, and few children fail to complete the task. By using self-explanatory animations, language demands are kept to a minimum. The task uses simple audio instructions, which makes the task accessible to children from different age groups, and no reading skills are required. There is no feedback during the task to prevent learning effects across trials.

Randomization. The order of the agents, balloon colors (red, yellow, green, blue), and balloon positions are each randomized independently. For the balloon positions, the entire width of the screen (1,920 in scalable-vector-graphics [SVG] units) is divided into 10 bins. Exact coordinates (value between 0 for the far left and 1,920 for the far right) in each bin are then randomly generated. The number of repetitions for each agent, balloon color, and balloon bin is calculated based on the total number of trials and the number of unique agents, balloon colors, and bins, respectively. All agents, balloon colors, and bins appear equally often and are not repeated in more than two consecutive trials. If the total number of trials is not divisible by the number of unique elements, some elements (i.e., some agents, balloon colors, bins) are randomly repeated to make up for the remainder.

Cross-cultural customization. When researchers visit TANGO-CC website (<https://ccp-odc.eva.mpg.de/tango-cc/>), they can select the language for audio instructions that are currently available for 13 different languages and five more dialects (see Table 1). To add a new language, researchers have two options. (a) For using their own

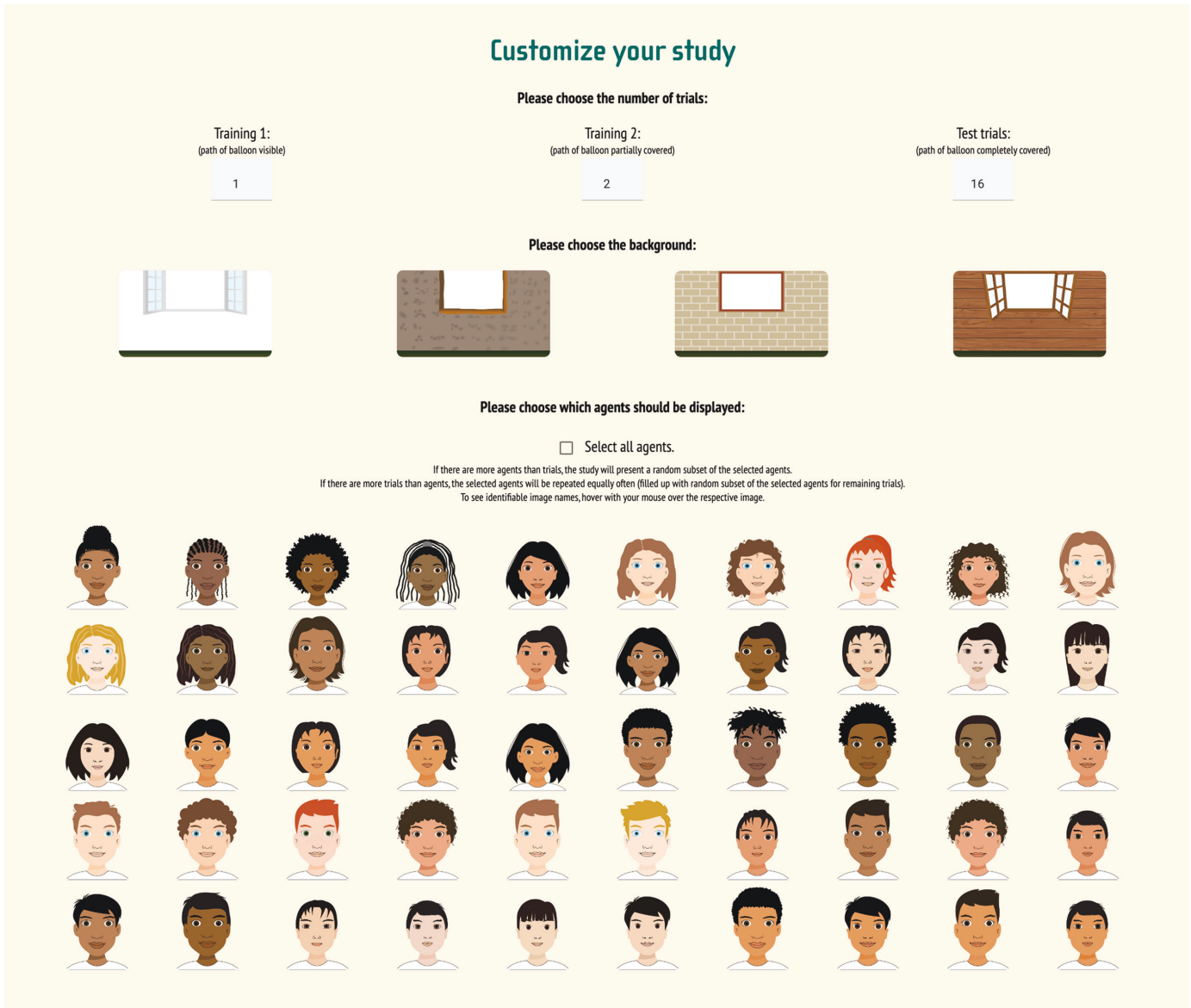


Fig. 2. Screenshot of the customizable components of TANGO-CC. Researchers can select the language of the audio instructions (see Table 1), the number of trials per trial type, the background, and the agent’s face. TANGO-CC = Task for Assessing iIndividual differences in Gaze understanding-Open-Cross-Cultural.

audio instructions in the offline version of the task, researchers can download the task, exchange the audio instructions in the “dist” folder (in the folder sounds>custom), and select “Custom” in the language drop-down menu. For detailed instructions, see TANGO-CC manual (<https://ccp-odc.eva.mpg.de/tango-cc/manual.html>). (b) For adding a new language in the online version of the task, researchers can contact J. C. Prein. Note that this option requires new audio recordings by the interested researchers, which will then be openly available for all users of the task. All written instructions in the task are

solely for the research assistant to help them guide participants through the task; these instructions are solely available in English. The task can either be started with the default settings or further customized by adapting the number of trials, agents, and background scenes. The default settings use the version applied in Bohn, Prein, et al. (2024) based on the selected language (see the Supplemental Material).

If researchers choose to customize the task (see Fig. 2), they can adjust the number of trials for each trial type but not their sequence. Specifically, trial types build

Table 1. Current Language Options Available for the Audio Instructions in TANGO-CC

Languages	Language family	Speaker's country of origin
Bemba	Bantu	Zambia
Chinese	Sino-Tibetan	China
English	Indo-European	United States, United Kingdom, India, Nigeria, New Zealand
German	Indo-European	Germany
(≠Akhoe) Hai om	Khoe-Kwadi	Namibia
Khwedam	Khoe-Kwadi	Namibia
Lingala	Bantu	Republic of the Congo
Marathi	Indo-European	India
Shona ^a	Bantu	Zimbabwe
Spanish	Indo-European	Argentina (Rioplatense Spanish), Mexico (Mexiquense Spanish)
Kiswahili	Bantu	Uganda
Turkish	Turkic	Türkiye
Yaka	Bantu	Republic of the Congo

Note: In cases in which more than one speaker's country of origin is listed, the audio instructions were recorded multiple times by different speakers. For example, the English instructions are available in five different versions.

^aAudio instructions are available in Shona, but no data of this version are included in the present data set.

on each other, and participants need to complete each trial type (without skipping any) to understand the structure of the task. The minimum number of trials per type is one; the maximum is 100. Furthermore, researchers can customize backgrounds by selecting one of four different backgrounds. Finally, researchers can choose from 50 diverse cartoon-like human faces (50% female, 50% male) and freely select how many different faces to include (minimum = 1, maximum = 50). Once all the settings are selected, the customized task is compiled. To save the selected settings, researchers can bookmark the URL to easily access the customized task.

In the last step, researchers can enter an alphanumeric participant identifier (one to eight characters) and enable a webcam recording of the participant if needed. A webcam recording might prove especially helpful for unsupervised online data collection to ensure that the participant is alone during the task and no help is provided. The participant identifier and webcam choice have to be provided every time the task is run.

The source code of the task is openly available on GitHub (<https://github.com/ccp-eva/tango-cc>). By directly editing the HTML and JavaScript code, researchers can further modify the task as needed. To ensure the

accessibility of TANGO-CC long-term, we have additionally stored the source code on Zenodo (<https://zenodo.org/doi/10.5281/zenodo.13643836>).

We created a public OSF page (<https://doi.org/10.17605/OSF.IO/P2EGU>) on which we plan to collect data sets that used TANGO-CC. Researchers who have collected data using TANGO-CC can share their data with the community by contacting J. C. Prein or visiting the OSF repository.

Task implementation

The task was implemented in JavaScript, HTML, and CSS and is presented as a web app. It can be accessed on any modern web browser on any device (e.g., computer or tablet) and does not require prior installation (although note that configurations of browsers and JavaScript may change in the future). Participants' responses can be recorded on a touchscreen or with a mouse or trackpad. The online version of the task can be used for unsupervised data collection (e.g., using online platforms such as Prolific; see Prein, Kalinke, et al., 2024). The task can be shared easily internationally by providing the URL. The web-app implementation does not require a working WIFI connection: An offline version of the task can be downloaded and quickly set up for devices that support Node.js (<https://nodejs.org/en>). This is an especially useful feature for researchers working in locations with limited internet access.

The stimuli are embedded as SVGs (an XML format that describes elements in mathematical formulas based on points and lines on a grid). SVGs ensure that the picture quality, aspect ratio, and relative object positioning are constant. Furthermore, stimuli are added as individual components to the image scene, which allows for an easy adaptation of the task's elements (in contrast to other image formats that consist of only one combined layer that would need entire replacement). The task is programmed so that responses are registered only when the participant clicks on the relevant part of the screen (i.e., in test trials, when they click on the hedge). Furthermore, clicks are registered only after the voice recordings stop playing. An audio reminder is played again if no click is registered within 5 s.

The website does not use cookies or upload any data to servers; that is, the data are stored only locally on the device. The output of the task is a CSV file (and WEBM file if a webcam recording was selected) that contains the participants' responses and can be easily imported into statistical software for further analysis. The file will be stored in the device's downloads folder and is named after the following pattern: "tangoCC-participantID-YYYY-MM-DD_hh_mm_ss." To modify the storage location on the device, researchers can change the designated downloads folder in their browser settings.

Psychometric Evaluation

Data set

We used the data set from Bohn, Prein, et al. (2024) for the psychometric evaluation of TANGO-CC. The data set contains a sample of 1,377 children, ages 2.5 to 11 years (see the Supplemental Material). Participants came from 17 communities on five continents, in rural and urban settings, with varying degrees of market integration and technology exposure. Bohn, Prein, et al. carried out 19 trials (one Training 1, two Training 2, and 16 test trials, of which the first of each type had audio instructions) on a touchscreen device. Faces, backgrounds, and languages were chosen by researchers and assistants with experience in the specific community. For further details on the communities, participant information, and data-collection procedures, see the supplements of Bohn, Prein, et al. (2024).

Individual differences

All analyses and the data set are available on GitHub (<https://github.com/ccp-eva/tango-cc-methods/>). First, we inspected the mean and standard deviations by community and compared performance in each trial type (Training 1, Training 2, test trials). Performance was defined as the absolute distance between the target center and the x coordinate of the participant's click (measured in balloon widths). Across communities, children performed best in Training 1 ($M = 0.19$, $SD = 0.63$), followed by Training 2 ($M = 0.79$, $SD = 1.44$) and test trials ($M = 2.21$, $SD = 2.03$; see Fig. 3).

To formally estimate the effect of trial type on performance in TANGO-CC, we fitted a generalized linear mixed model (GLMM) predicting the task performance by trial type (reference category: test trials). All analyses were run in R (Version 4.4.1; R Core Team, 2024). GLMMs were fitted with default priors using the function *brm* from the package *brms* (Bürkner, 2017, 2018). The model included random effects for trial type by community (model notation in R: $\text{imprecision} \sim \text{trialtype} + (\text{trialtype} | \text{community})$), and imprecision was modeled by a log-normal distribution. We inspected the posterior distribution (mean and 95% credible interval [CrI]) for the trial-type estimates.

Our GLMM analysis supported the visual inspection of the data: The fixed-effect estimates for Training 1 ($\beta = -3.26$, 95% CrI = $[-3.41, -3.10]$) and Training 2 ($\beta = -1.47$, 95% CrI = $[-1.58, -1.35]$) were negative and reliably different from zero.¹ This effect was found across all communities (random effects of trial type within community: minimum estimate for Training 1 = -2.87 , 95% CrI = $[-3.11, -2.60]$; minimum estimate for Training 2 = -1.27 , 95% CrI = $[-1.51, -0.98]$). The almost perfect performance in training trials indicated that children understood the

task and were able to correctly indicate the location of the balloon when its path was (mostly) visible. In test trials, children's imprecision was higher, indicating that the task was more challenging. All communities showed substantial individual variation and overlapped in their imprecision levels (see Fig. 3).

To identify the sources of variation, we computed intraclass correlations (ICCs). The variation in children's imprecision within communities was substantially larger than the variation between the communities. The mean within-communities variance was 1.28, ranging from 0.24 (in Pune, India) to 3.46 (in Chimfunshi, Zambia). Between-communities variance was 0.34. The ICC, representing the proportion of between-communities variance relative to the total variance (sum of within- and between-communities variance), was 0.02. This indicated that only 2% of the total variability in the data could be attributed to differences between communities, whereas the remaining 98% were attributed to differences within communities (Kusano et al., 2024).

Reliability

To assess reliability, we estimated internal consistency in each community in three different ways. First, data of each participant were split into odd and even trials, and a Pearson correlation was calculated between the aggregated scores of the two halves. Second, using the function *by_split* from the *splithalfr* package (Pronk et al., 2022), data were stratified by target centrality, and a Pearson correlation was calculated between the matched halves. Target centrality measured how closely the target landed to the center of the screen (i.e., distance from the balloon center to the screen center), which served as a proxy for trial difficulty and was previously found to be a relevant component in predicting participants' performance in a cognitive model (Prein, Maurits, et al., 2024). Third, a data set was generated with stratified test halves by target centrality, and we applied the GLMM approach introduced by Rouder and Haaf (2019). A GLMM was fitted with the mean imprecision as the outcome, age as the predictor, and test half and participant ID as random effects (model notation: $\text{imprecision} \sim \text{age} + (0 + \text{half} | \text{subjid})$). The model estimates correlations between participant-specific estimates for each test half. The hierarchical shrinkage of the model enables accurate person-specific estimates. By incorporating age as a fixed effect, the correlation between the two person-specific estimates represents the age-independent estimate for internal consistency. This removes the possibility that a good internal-consistency estimate results from general cognitive development rather than task-specific interindividual differences. Because the process of generating stratified data sets is partly random, the model was fitted 50 times for each community. The posterior

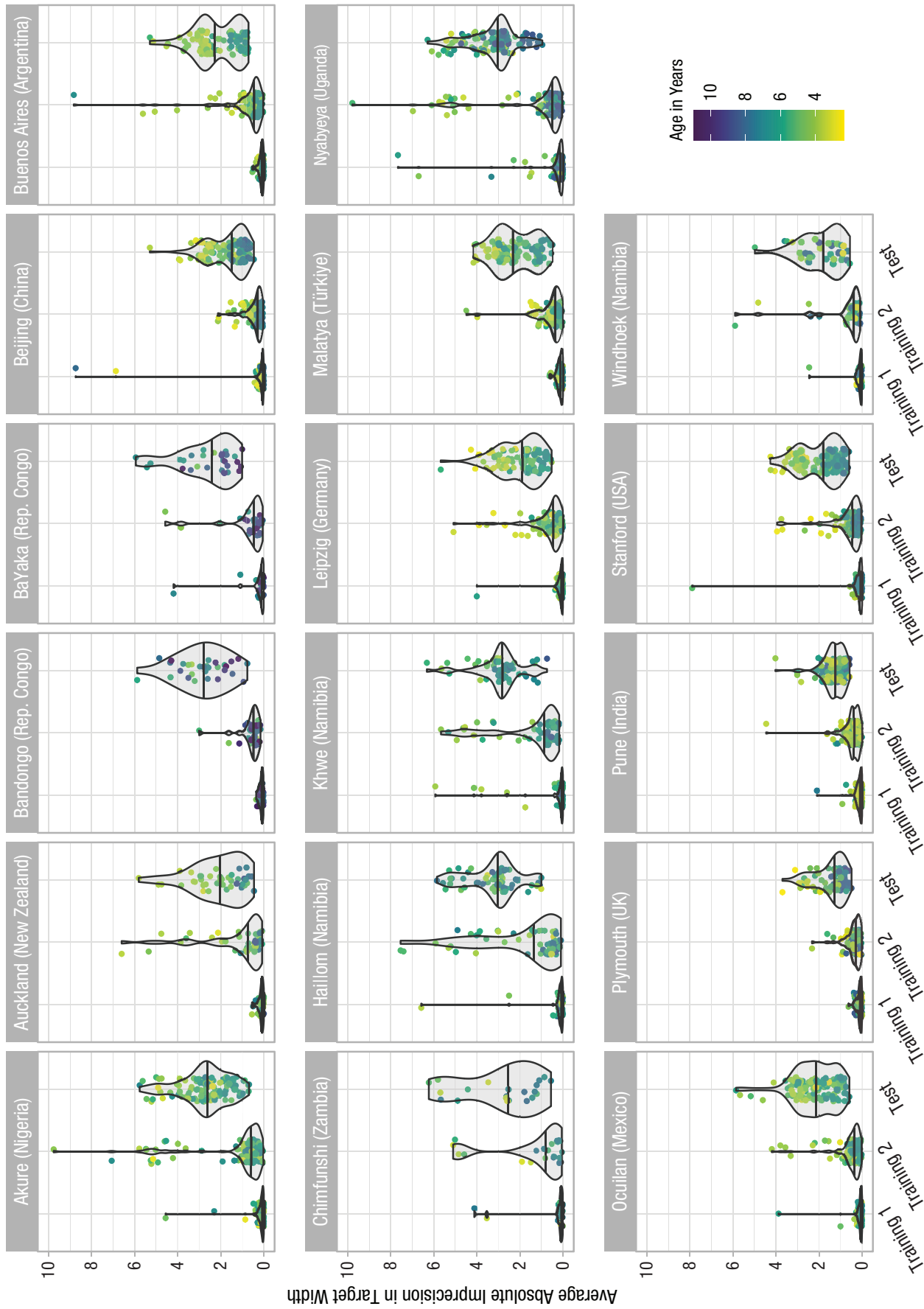


Fig. 3. Variability measured by TANGO-CC: mean imprecision in locating the agent's attentional focus by community (alphabetically) and trial type. Imprecision is defined as the distance between the participant's click and the balloon's center in units of balloon width. For a depiction of each trial's procedure, see Figure 1. TANGO-CC = Task for Assessing Individual Differences in Gaze understanding-Open-Cross-Cultural.

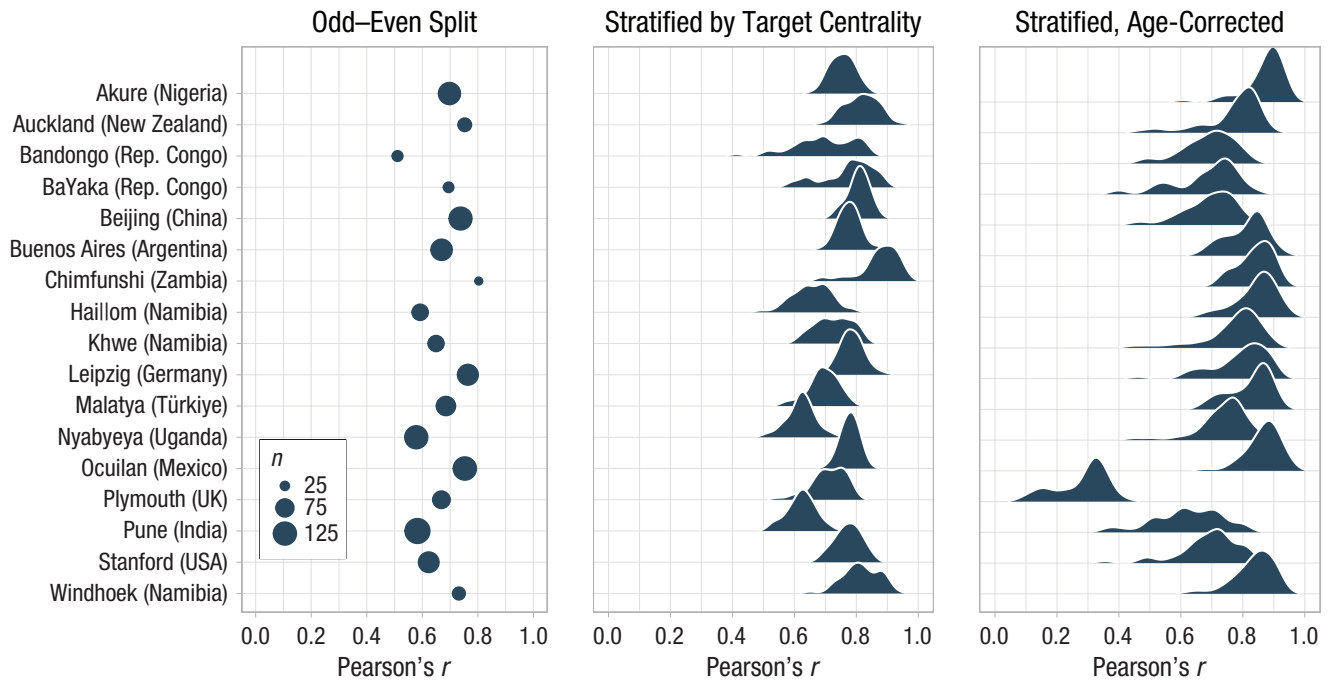


Fig. 4. Reliability of TANGO-CC by community: internal consistency estimates by community, following three different approaches. In the odd–even split, the size of points reflects the sample size in each community. In the stratified approach with and without age correction, density curves show the posterior distributions of the generalized linear mixed model. TANGO-CC = Task for Assessing Individual Differences in Gaze understanding-Open-Cross-Cultural.

estimate of the correlation between the two person-specific estimates was taken as the age-independent estimate for internal consistency.

For results, see Figure 4. Across communities, internal consistency estimates ranged from 0.51 to 0.80 for the odd–even split, 0.62 to 0.89 for the stratified internal consistency, and 0.62 to 0.87 for the age-corrected approach (Plymouth, England, was an outlier with 0.28). Following Cohen’s (1988, 1992) suggestions, these correlations constitute large effects ($r > .50$) and indicate good internal consistency.² The results are comparable with the internal-consistency estimates found in the original TANGO study (Prein, Kalinke, et al., 2024) and also resemble reliability estimates of classical false-belief tasks (Hughes et al., 2000).

In an exploratory analysis, we found that communities with larger individual variation showed higher internal consistency estimates (Pearson’s $r = .46$, 95% CI = $[-.03, .77]$). This suggests that the less variation a task can capture within a community, the lower the reliability is. However, this correlation could be influenced by outliers, and the sample size here ($N = 17$ communities) is too small to make substantial claims.

Discussion

TANGO-CC measures imprecision in gaze following across individuals, ages, and communities. Children’s

imprecision in gaze following showed highly similar result patterns across communities: Children performed better in the training than the test trials, and within-communities variation greatly exceeded between-communities variation. Furthermore, the task showed satisfactory to high reliability across all communities. Therefore, TANGO-CC is a suitable task to capture individual differences in social-cognitive development in diverse communities.

TANGO-CC’s design process lays out a much-needed pragmatic approach to studying community-level and individual-level differences across cultures: While we performed a detailed psychometric evaluation of the task in a German setting, we collaborated with local researchers for the cross-cultural stimulus development and selection. We reassessed TANGO-CC’s psychometric properties in a large and diverse data set. Although we cannot generalize our findings to all communities worldwide, we found that TANGO-CC captured reliable individual variation in all 17 communities studied by Bohn, Prein, et al. (2024). We hope that not just TANGO-CC but also our pragmatic approach to constructing it will be helpful to other researchers. We recommend that researchers consider generalizability concerns and cross-cultural applications of their tasks and collaborate with local researchers at the early stages of task development (Torréns et al., 2023). Using TANGO-CC (or any other task) in a new community requires sensitivity to the

specific context, piloting, and most important, the involvement of researchers or research assistants from the specific community.

Bourdage et al. (2023) pointed out a major challenge with adapting social-cognition tasks to diverse communities: The number of world cultures is vast, and communities are constantly changing. Therefore, a promising approach might be to provide tasks with a modular system in which components can be modified (i.e., building-block structure). In the case of TANGO-CC, the task can not only be adapted to different languages, cartoon faces, and backgrounds (see Fig. 2) but also updated with new stimuli. Unlike studies that present sequential, hand-painted pictures that are difficult to adapt (Mehta et al., 2011), TANGO-CC uses SVGs that can be easily exchanged.

Compared with one of the most commonly used social-cognition measures—the change-of-location false-belief task (Baron-Cohen et al., 1985; Wimmer & Perner, 1983)—TANGO-CC has several advantages: a continuous outcome measure that can capture individual differences from 3 years to adulthood, a short task duration that allows for more trials per child, stimuli that can be easily adapted, and known psychometric properties across 17 communities. The task is presented as a web app that enables efficient data collection with large sample sizes, although it can also be used to collect data offline in locations without a reliable internet connection. TANGO-CC follows a standardized procedure, which prevents rater errors and greatly simplifies online training of research assistants. Furthermore, minimal language demands and an engaging, playful design increase the task's usability and reduce noncompletion rates.

TANGO-CC will be a useful asset for exploring a range of research questions. First, TANGO-CC could be applied to assessing relationships between gaze following and other (social-)cognitive constructs. Examples include the relationship between gaze following and theory of mind (Prein, Maurits, et al., 2024) or eye gaze in sharing behaviors and dictator games (Kelsey et al., 2018; Manesi et al., 2016; Nettle et al., 2013; Wu et al., 2018). Second, TANGO-CC could be used to examine the influence of environmental factors on gaze following. Environmental factors of interest could include household size (Bohn, Fong, et al., 2024) or maternal postpartum depression (Astor et al., 2022). Third, scores in TANGO-CC could predict performance in other tasks at later time points. This would be helpful for assessing predictive validity of gaze following, for example, for language comprehension (Prein, Kalinke, et al., 2024). Fourth, TANGO-CC could be applied to measure children's gaze following development longitudinally, potentially combined with interim interventions. The task is suitable for children, teenagers, and adults alike and is sensitive to individual differences

throughout the life span (established in a German sample; Prein, Kalinke, et al., 2024).

TANGO-CC is a screen-based task. Bohn, Prein, et al. (2024) showed that children with no prior touchscreen exposure were less precise in TANGO-CC than children with prior experience. Therefore, the mode of stimulus presentation needs to be kept in mind when administering TANGO-CC, especially in communities with little technology exposure. Additional touchscreen training (e.g., more trials of Training 1) might prove helpful in these cases. If researchers are interested in controlling for an effect of stimulus presentation, we recommend gathering information on touchscreen exposure.

Individual differences were also present in communities with 100% touchscreen exposure, showing that this factor alone could not explain children's performance in the task (Bohn, Prein, et al., 2024). Even though the touchscreen experience caused absolute differences in task performance, all communities showed the same processing signature: Children were more precise in trials in which the agent looked to a more central position (i.e., higher target centrality) compared with a position to the far left or right of the screen. A computational cognitive model predicted this processing signature and described gaze following as a process of estimating pupil angles and the corresponding gaze vectors (Prein, Maurits, et al., 2024). Bohn, Prein, et al. (2024) found clear support for this model in every community studied, suggesting that children all over the world process gaze in a similar way. Alternative theories on gaze following exist but have not been cross-culturally validated to date. Doherty et al. (2015) proposed two separate gaze-processing systems based on luminance versus geometric cues. Other theoretical accounts assign a central role to reinforcement learning (Corkum & Moore, 1998; Ishikawa et al., 2020; Silverstein et al., 2021; Triesch et al., 2006) or self-other equivalence ("like me"; Meltzoff, 2005, 2007).

Schilbach et al. (2013) pointed out that witnessing social interactions as an observer undoubtedly differs from actively participating in social interactions. First evidence suggests that TANGO-CC indeed taps into social cognition as used in real life: Prein, Maurits, et al. (2024) found that children's perspective-taking abilities in a personal social interaction were linked to performance on TANGO but less so to a matched, non-social-vector-following task. However, this study exclusively relied on a German sample, and future research should investigate whether the relationship between TANGO-CC and perspective-taking abilities holds across communities.

We have reported reliability estimates for each community by calculating internal consistency. Ideally, we would have additionally evaluated the task's test-retest reliability in each community and checked for

relationships with theoretically related constructs to assess validity. Unfortunately, this might not always be feasible in large-scale cross-cultural studies because of organizational and financial constraints. An example of assessing TANGO's predictive validity is a study conducted in Leipzig, Germany, which used TANGO to predict children's receptive vocabulary 6 months later (Prein, Kalinke, et al., 2024). Future cross-cultural studies could investigate TANGO-CC's predictive validity and its relationship to other social-cognitive abilities (e.g., theory of mind, language development) in diverse communities.

Measurement invariance (i.e., measuring the same construct across different communities) is often seen as a requirement for a "fair" cross-cultural comparison: It is important that any group differences are not the result of the task unintentionally tapping into different underlying constructs. As Kusano et al. (2024) put it, "The research challenge is to achieve a balance between ensuring methodological 'fairness' at the individual level while also recognizing and capturing genuine sociocultural variability" (p. 34). We argue that TANGO-CC measures a fundamental social-cognitive ability that is likely similar across communities. Selcuk et al. (2023) pointed out that researchers should study both within- and between-cultures variability in the development of social cognition because sometimes within-cultures differences exceed between-cultures differences. Indeed, we found that within-groups variability was greater than between-groups variability. Although we believe that TANGO-CC can be used to compare mean differences across communities, we would recommend using it to study individual differences within communities.

For years, researchers have called for more diverse sampling and culturally valid measures of cognitive development (Matsumoto & Yoo, 2006; e.g., Mehta et al., 2011; Nielsen et al., 2017). As Hajdúk et al. (2020) said, "Using large samples and multisite approaches will align with efforts to improve reproducibility [replicability] and will clarify both the type and extent of cultural influences on social cognition" (p. 463). Likewise, Elson et al. (2023) called for standardized, psychometrically evaluated measures that can be reused by other researchers to "build a cumulative evidence base in psychology" (p. 2). This underlines how efforts to improve replicability can be combined with the goal of increasing the generalizability of psychological research findings (Li et al., 2024; Syed, 2021). Li et al. (2024) argued that replicable and generalizable results rely on stimulus sets with slight variations, more diverse samples, and data collection at a greater scale, which are indeed all steps TANGO-CC has taken. Openly sharing TANGO-CC's materials will allow other researchers to (hopefully) replicate the results and deepen the cumulative understanding of social-cognitive development across diverse communities.

Conclusion

TANGO-CC captures individual differences in social-cognitive development across diverse communities. The task's customizability, minimal language demands, and efficient data-collection method make it a valuable tool for cross-cultural research. The task showed satisfactory to high reliability (internal consistency) in a large data set including 17 diverse communities on five continents. We hope that TANGO-CC—and its pragmatic construction process—will provide a roadmap for future cross-cultural studies on cognitive development.

Transparency

Action Editor: Pamela Davis-Kean

Editor: David A. Sbarra

Author Contributions

D. B. M. Haun and M. Bohn are joint last authors.

Julia Christin Prein: Conceptualization; Formal analysis; Methodology; Project administration; Resources; Software; Visualization; Writing – original draft; Writing – review & editing.

Florian M. Bednarski: Methodology; Resources; Writing – review & editing.

Ardain Dzabatou: Methodology; Resources; Writing – review & editing.

Michael C. Frank: Methodology; Resources; Writing – review & editing.

Annette M. E. Henderson: Methodology; Resources; Writing – review & editing.

Josefine Kalbitz: Methodology; Resources; Writing – review & editing.

Patricia Kanngiesser: Methodology; Resources; Writing – review & editing.

Dilara Keşşafoglu: Methodology; Resources; Writing – review & editing.

Bahar Köymen: Methodology; Resources; Writing – review & editing.

Maira V. Manrique-Hernandez: Methodology; Resources; Writing – review & editing.

Shirley Magazi: Methodology; Resources; Writing – review & editing.

Lizbeth Mújica-Manrique: Methodology; Resources; Writing – review & editing.

Julia Ohlendorf: Methodology; Resources; Writing – review & editing.

Damilola Olaoba: Methodology; Resources; Writing – review & editing.

Wesley R. Pieters: Methodology; Resources; Writing – review & editing.

Sarah Pope-Caldwell: Methodology; Resources; Writing – review & editing.

Umay Sen: Methodology; Resources; Writing – review & editing.

Katie Slocombe: Methodology; Resources; Writing – review & editing.

Robert Z. Sparks: Methodology; Resources; Writing – review & editing.

Roman Stengelin: Conceptualization; Methodology; Resources; Writing – review & editing.

Jahnavi Sunderarajan: Methodology; Resources; Writing – review & editing.

Kirsten Sutherland: Methodology; Resources; Writing – review & editing.

Florence Tusiime: Methodology; Resources; Writing – review & editing.

Wilson Vieira: Methodology; Resources; Writing – review & editing.

Zhen Zhang: Methodology; Resources; Writing – review & editing.

Yufei Zong: Methodology; Resources; Writing – review & editing.

Daniel B. M. Haun: Conceptualization; Funding acquisition; Methodology; Writing – review & editing.

Manuel Bohn: Conceptualization; Funding acquisition; Methodology; Project administration; Supervision; Writing – review & editing.

Declaration of Conflicting Interests

The author(s) declared that there were no conflicts of interest with respect to the authorship or the publication of this article.

Funding


This publication was funded by the German Research Foundation.

Open Practices


This article has received the badges for Open Data and Open Materials. More information about the Open Practices badges can be found at <http://www.psychologicalscience.org/publications/badges>.



ORCID iDs


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


Florian M. Bednarski  <https://orcid.org/0000-0002-2459-6190>


Michael C. Frank  <https://orcid.org/0000-0002-7551-4378>


Annette M. E. Henderson  <https://orcid.org/0000-0003-4384-4791>


Patricia Kanngiesser  <https://orcid.org/0000-0003-1068-3725>


Dilara Keşşafoglu  <https://orcid.org/0000-0002-7356-0733>

Bahar Köymen  <https://orcid.org/0000-0001-5126-8240>


Umay Sen  <https://orcid.org/0000-0001-9488-0851>

Robert Z. Sparks  <https://orcid.org/0000-0001-7545-0522>

Roman Stengelin  <https://orcid.org/0000-0003-2212-4613>

Zhen Zhang  <https://orcid.org/0000-0001-9300-0920>

Yufei Zong  <https://orcid.org/0009-0000-5012-0244>

Manuel Bohn  <https://orcid.org/0000-0001-6006-1348>

Supplemental Material

Additional supporting information can be found at <http://journals.sagepub.com/doi/suppl/10.1177/25152459241308170>.

Notes

1. Please note that TANGO-CC measures imprecision in gaze following. Therefore, a negative sign indicates that children showed less imprecision (i.e., were more precise) in the training trials than in the test trials.

2. Note that for scale reliability and Cronbach's α , values of .7 to .8 have been suggested to be acceptable (Field et al., 2012; Kline, 1999). However, Kline (1999) suggested that values below .7 could be realistic for psychological constructs because of their variable nature.

References

- Adolphs, R. (1999). Social cognition and the human brain. *Trends in Cognitive Sciences*, 3(12), 469–479. [https://doi.org/10.1016/S1364-6613\(99\)01399-6](https://doi.org/10.1016/S1364-6613(99)01399-6)
- Akhtar, N., & Gernsbacher, M. A. (2008). On privileging the role of gaze in infant social cognition. *Child Development Perspectives*, 2(2), 59–65. <https://doi.org/10.1111/j.1750-8606.2008.00044.x>
- Amir, D., & McAuliffe, K. (2020). Cross-cultural, developmental psychology: Integrating approaches and key insights. *Evolution and Human Behavior*, 41(5), 430–444. <https://doi.org/10.1016/j.evolhumbehav.2020.06.006>
- Anderson, J. R., & Doherty, M. J. (1997). Preschoolers' perception of other people's looking: Photographs and drawings. *Perception*, 26(3), 333–343. <https://doi.org/10.1068/p260333>
- Apicella, C., Norenzayan, A., & Henrich, J. (2020). Beyond WEIRD: A review of the last decade and a look ahead to the global laboratory of the future. *Evolution and Human Behavior*, 41(5), 319–329. <https://doi.org/10.1016/j.evolhumbehav.2020.07.015>
- Arnett, J. J. (2008). The neglected 95%: Why American psychology needs to become less American. *American Psychologist*, 63(7), 602–614. <https://doi.org/10.1037/0003-066X.63.7.602>
- Astor, K., & Gredebäck, G. (2022). Gaze following in infancy: Five big questions that the field should answer. *Advances in Child Development and Behavior*, 63, 191–223. <https://doi.org/10.1016/bs.acdb.2022.04.003>
- Astor, K., Lindskog, M., Juvrud, J., Wangchuk Namgyel, S. C., Wangmo, T., Tshering, K., & Gredebäck, G. (2022). Maternal postpartum depression impacts infants' joint attention differentially across cultures. *Developmental Psychology*, 58(12), 2230–2238. <https://doi.org/10.1037/dev0001413>
- Baron-Cohen, S., Leslie, A. M., & Frith, U. (1985). Does the autistic child have a “theory of mind”? *Cognition*, 21, 37–46. [https://doi.org/10.1016/0010-0277\(85\)90022-8](https://doi.org/10.1016/0010-0277(85)90022-8)
- Beaudoin, C., Leblanc, É., Gagner, C., & Beauchamp, M. H. (2020). Systematic review and inventory of theory of mind measures for young children. *Frontiers in Psychology*, 10, Article 2905. <https://doi.org/10.3389/fpsyg.2019.02905>
- Bohn, M., Fong, F. T. K., Pope-Caldwell, S., Stengelin, R., & Haun, D. B. M. (2024). Understanding cultural variation in cognition one child at a time. *Nature Reviews Psychology*, 3, 641–643. <https://doi.org/10.1038/s44159-024-00351-8>
- Bohn, M., & Köymen, B. (2018). Common ground and development. *Child Development Perspectives*, 12(2), 104–108. <https://doi.org/10.1111/cdep.12269>

- Bohn, M., Prein, J. C., Ayikoru, A., Bednarski, F. M., Dzabatou, A., Frank, M. C., Henderson, A. M. E., Isabella, J., Kalbitz, J., Kanngiesser, P., Keşşafoglu, D., Koymen, B., Manrique-Hernandez, M., Magazi, S., Mújica-Manrique, L., Ohlendorf, J., Olaoba, D., Pieters, W., Pope-Caldwell, S., . . . Haun, D. (2024). *A universal of human social cognition: Children from 17 communities process gaze in similar ways*. PsyArXiv. <https://doi.org/10.31234/osf.io/z3ahv>
- Bourdage, R., Narme, P., Neeskens, R., Papma, J., & Franzen, S. (2023). An evaluation of cross-cultural adaptations of social cognition testing: A systematic review. *Neuropsychology Review*, *34*, 1048–1094. <https://doi.org/10.1007/s11065-023-09616-0>
- Broesch, T., Lew-Levy, S., Kärtner, J., Kanngiesser, P., & Kline, M. (2023). A roadmap to doing culturally grounded developmental science. *Review of Philosophy and Psychology*, *14*(2), 587–609. <https://doi.org/10.1007/s13164-022-00636-y>
- Brooks, R., & Meltzoff, A. N. (2005). The development of gaze following and its relation to language. *Developmental Science*, *8*(6), 535–543. <https://doi.org/10.1111/j.1467-7687.2005.00445.x>
- Brooks, R., & Meltzoff, A. N. (2015). Connecting the dots from infancy to childhood: A longitudinal study connecting gaze following, language, and explicit theory of mind. *Journal of Experimental Child Psychology*, *130*, 67–78. <https://doi.org/10.1016/j.jecp.2014.09.010>
- Brooks, R., Singleton, J. L., & Meltzoff, A. N. (2020). Enhanced gaze-following behavior in Deaf infants of Deaf parents. *Developmental Science*, *23*(2), Article e12900. <https://doi.org/10.1111/desc.12900>
- 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), 395–411. <https://doi.org/10.32614/RJ-2018-017>
- Butler, S. C., Caron, A. J., & Brooks, R. (2000). Infant understanding of the referential nature of looking. *Journal of Cognition and Development*, *1*(4), 359–377. https://doi.org/10.1207/S15327647JCD0104_01
- Callaghan, T., Moll, H., Rakoczy, H., Warneken, F., Liszkowski, U., Behne, T., & Tomasello, M. (2011). Early social cognition in three cultural contexts. *Monographs of the Society for Research in Child Development*, *76*(2), vii–viii, 1–142. <https://doi.org/10.1111/j.1540-5834.2011.00603.x>
- Carpenter, M., Nagell, K., & Tomasello, M. (1998). Social cognition, joint attention, and communicative competence from 9 to 15 months of age. *Monographs of the Society for Research in Child Development*, *63*(4), i–vi, 1–143.
- Childers, J. B., Vaughan, J., & Burquest, D. A. (2007). Joint attention and word learning in Ngas-speaking toddlers in Nigeria. *Journal of Child Language*, *34*(2), 199–225. <https://doi.org/10.1017/S0305000906007835>
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates.
- Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*(1), 155–159. <https://doi.org/10.1037/0033-2909.112.1.155>
- Corkum, V., & Moore, C. (1998). The origins of joint visual attention in infants. *Developmental Psychology*, *34*(1), 28–38. <https://doi.org/10.1037/0012-1649.34.1.28>
- Decety, J. (Ed.). (2020). *The social brain: A developmental perspective*. The MIT Press.
- Del Bianco, T., Falck-Ytter, T., Thorup, E., & Gredebäck, G. (2019). The developmental origins of gaze-following in human infants. *Infancy*, *24*(3), 433–454. <https://doi.org/10.1111/inf.12276>
- Doherty, M. J., & Anderson, J. R. (1999). A new look at gaze: Preschool children’s understanding of eye-direction. *Cognitive Development*, *14*(4), 549–571. [https://doi.org/10.1016/s0885-2014\(99\)00019-2](https://doi.org/10.1016/s0885-2014(99)00019-2)
- Doherty, M. J., Anderson, J. R., & Howieson, L. (2009). The rapid development of explicit gaze judgment ability at 3 years. *Journal of Experimental Child Psychology*, *104*(3), 296–312. <https://doi.org/10.1016/j.jecp.2009.06.004>
- Doherty, M. J., McIntyre, A. H., & Langton, S. R. H. (2015). Developmentally distinct gaze processing systems: Luminance versus geometric cues. *Cognition*, *137*, 72–80. <https://doi.org/10.1016/j.cognition.2015.01.001>
- Dunphy-Lelii, S., & Wellman, H. M. (2004). Infants’ understanding of occlusion of others’ line-of-sight: Implications for an emerging theory of mind. *European Journal of Developmental Psychology*, *1*(1), 49–66. <https://doi.org/10.1080/17405620444000049>
- Elfenbein, H. A., & Ambady, N. (2002). On the universality and cultural specificity of emotion recognition: A meta-analysis. *Psychological Bulletin*, *128*(2), 203–235. <https://doi.org/10.1037/0033-2909.128.2.203>
- Elson, M., Hussey, I., Alsalti, T., & Arslan, R. C. (2023). Psychological measures aren’t toothbrushes. *Communications Psychology*, *1*, Article 25. <https://doi.org/10.1038/s44271-023-00026-9>
- Field, A., Miles, J., & Field, Z. (2012). *Discovering statistics using R*. Sage.
- Franco, F., & Gagliano, A. (2001). Toddlers’ pointing when joint attention is obstructed. *First Language*, *21*(63), 289–321. <https://doi.org/10.1177/014272370102106305>
- Frith, C. D., & Frith, U. (2007). Social cognition in humans. *Current Biology*, *17*(16), R724–R732. <https://doi.org/10.1016/j.cub.2007.05.068>
- Gagarina, N., Klop, D., Kunnari, S., Tantele, K., Välimaa, T., Balčiūnienė, I., Bohnacker, U., & Walters, J. (2012). *MAIN: Multilingual assessment instrument for narratives*. Zentrum für Allgemeine Sprachwissenschaft.
- Gagarina, N., & Lindgren, J. (2020). New language versions of MAIN: Multilingual Assessment Instrument for Narratives – Revised. *ZAS Papers in Linguistics*, *64*, Article 274. <https://doi.org/10.21248/zaspil.64.2020.543>
- Gutchess, A., & Rajaram, S. (2023). Consideration of culture in cognition: How we can enrich methodology and theory. *Psychonomic Bulletin & Review*, *30*(3), 914–931. <https://doi.org/10.3758/s13423-022-02227-5>
- Hajdúk, M., Achim, A. M., Brunet-Gouet, E., Mehta, U. M., & Pinkham, A. E. (2020). How to move forward in social cognition research? Put it into an international perspective. *Schizophrenia Research*, *215*, 463–464. <https://doi.org/10.1016/j.schres.2019.10.001>
- He, J., & van de Vijver, F. (2012). Bias and equivalence in cross-cultural research. *Online Readings in Psychology and Culture*, *2*(2). <https://doi.org/10.9707/2307-0919.1111>
- Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *The Behavioral and Brain*

- Sciences*, 33(2–3), 61–135. <https://doi.org/10.1017/S0140525X0999152X>
- Hernik, M., & Broesch, T. (2019). Infant gaze following depends on communicative signals: An eye-tracking study of 5- to 7-month-olds in Vanuatu. *Developmental Science*, 22(4), Article e12779. <https://doi.org/10.1111/desc.12779>
- Hughes, C., Adlam, A., Happé, F., Jackson, J., Taylor, A., & Caspi, A. (2000). Good test-retest reliability for standard and advanced false-belief tasks across a wide range of abilities. *Journal of Child Psychology and Psychiatry*, 41(4), 483–490. <https://doi.org/10.1111/1469-7610.00633>
- Ishikawa, M., Senju, A., & Itakura, S. (2020). Learning process of gaze following: Computational modeling based on reinforcement learning. *Frontiers in Psychology*, 11, Article 213. <https://doi.org/10.3389/fpsyg.2020.00213>
- Kelsey, C., Grossmann, T., & Vaish, A. (2018). Early reputation management: Three-year-old children are more generous following exposure to eyes. *Frontiers in Psychology*, 9, Article 968. <https://doi.org/10.3389/fpsyg.2018.00698>
- Kingstone, A., Friesen, C. K., & Gazzaniga, M. S. (2000). Reflexive joint attention depends on lateralized cortical connections. *Psychological Science*, 11(2), 159–166. <https://doi.org/10.1111/1467-9280.00232>
- Kline, P. (1999). *The handbook of psychological testing* (2nd ed.). Routledge.
- Krys, K., De Almeida, I., Wasiel, A., & Vignoles, V. L. (2024). WEIRD–Confucian comparisons: Ongoing cultural biases in psychology’s evidence base and some recommendations for improving global representation. *The American Psychologist*. Advance online publication. <https://doi.org/10.1037/amp0001298>
- Kusano, K., Napier, J., & Jost, J. (2024). *The mismeasure of culture: When measurement invariance requirements hinder cross-cultural research in psychology*. OSF. <https://doi.org/10.31234/osf.io/9qe2k>
- 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>
- Li, W., Germine, L. T., Mehr, S. A., Srinivasan, M., & Hartshorne, J. (2024). Developmental psychologists should adopt citizen science to improve generalization and reproducibility. *Infant and Child Development*, 33(1), Article e2348. <https://doi.org/10.1002/icd.2348>
- Liebal, K., & Haun, D. B. M. (2018). Why cross-cultural psychology is incomplete without comparative and developmental perspectives. *Journal of Cross-Cultural Psychology*, 49(5), 751–763. <https://doi.org/10.1177/0022022117738085>
- Lillard, A. (1998). Ethnopsychologies: Cultural variations in theories of mind. *Psychological Bulletin*, 123(1), 3–32. <https://doi.org/10.1037/0033-2909.123.1.3>
- Manesi, Z., Van Lange, P. A. M., & Pollet, T. V. (2016). Eyes wide open: Only eyes that pay attention promote prosocial behavior. *Evolutionary Psychology*, 14(2), 1–15. <https://doi.org/10.1177/1474704916640780>
- Matsumoto, D., & Yoo, S. H. (2006). Toward a new generation of cross-cultural research. *Perspectives on Psychological Science*, 1(3), 234–250. <https://doi.org/10.1111/j.1745-6916.2006.00014.x>
- Mehta, U. M., Thirthalli, J., Naveen Kumar, C., Mahadevaiah, M., Rao, K., Subbakrishna, D. K., Gangadhar, B. N., & Keshavan, M. S. (2011). Validation of Social Cognition Rating Tools in Indian Setting (SOCRATIS): A new test-battery to assess social cognition. *Asian Journal of Psychiatry*, 4(3), 203–209. <https://doi.org/10.1016/j.ajp.2011.05.014>
- Meltzoff, A. N. (2005). Imitation and other minds: The “like me” hypothesis. In S. Hurley & N. Chater (Eds.), *Perspectives on imitation: From neuroscience to social science: Vol. 2: Imitation, human development, and culture* (pp. 55–77). MIT Press.
- Meltzoff, A. N. (2007). “Like me”: A foundation for social cognition. *Developmental Science*, 10(1), 126–134. <https://doi.org/10.1111/j.1467-7687.2007.00574.x>
- 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>
- Mundy, P., Block, J., Delgado, C., Pomares, Y., Van Hecke, A. V., & Parlade, M. V. (2007). Individual differences and the development of joint attention in infancy. *Child Development*, 78(3), 938–954. <https://doi.org/10.1111/j.1467-8624.2007.01042.x>
- Nettle, D., Harper, Z., Kidson, A., Stone, R., Penton-Voak, I. S., & Bateson, M. (2013). The watching eyes effect in the Dictator Game: It’s not how much you give, it’s being seen to give something. *Evolution and Human Behavior*, 34(1), 35–40. <https://doi.org/10.1016/j.evolhumbehav.2012.08.004>
- Nielsen, M., Haun, D., Kärtner, J., & Legare, C. H. (2017). The persistent sampling bias in developmental psychology: A call to action. *Journal of Experimental Child Psychology*, 162, 31–38. <https://doi.org/10.1016/j.jecp.2017.04.017>
- Ochs, E., & Schieffelin, B. B. (1984). Language acquisition and socialization: Three developmental stories and their implications. In R. A. Shweder & R. A. LeVine (Eds.), *Culture theory: Essays on mind, self, and emotion* (pp. 276–320). Cambridge University Press.
- Peña, E. D. (2007). Lost in translation: Methodological considerations in cross-cultural research. *Child Development*, 78(4), 1255–1264. <https://www.jstor.org/stable/4620701>
- Prein, J. C., Kalinke, S., Haun, D. B. M., & Bohn, M. (2024). 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>
- Prein, J. C., Maurits, L., Werwach, A., Haun, D. B. M., & Bohn, M. (2024). Variation in gaze following across the life span: A process-level perspective. *Developmental Science*, 27(6), Article e13546. <https://doi.org/10.1111/desc.13546>
- Pronk, T., Molenaar, D., Wiers, R. W., & Murre, J. (2022). Methods to split cognitive task data for estimating split-half reliability: A comprehensive review and systematic assessment. *Psychonomic Bulletin & Review*, 29(1), 44–54. <https://doi.org/10.3758/s13423-021-01948-3>
- R Core Team. (2024). *R: A language and environment for statistical computing* [Manual]. R Foundation for Statistical Computing.

- Rouder, J. N., & Haaf, J. M. (2019). A psychometrics of individual differences in experimental tasks. *Psychonomic Bulletin & Review*, *26*(2), 452–467. <https://doi.org/10.3758/s13423-018-1558-y>
- Schillbach, L., Timmermans, B., Reddy, V., Costall, A., Bente, G., Schlicht, T., & Vokeley, K. (2013). Toward a second-person neuroscience. *The Behavioral and Brain Sciences*, *36*(4), 393–414. <https://doi.org/10.1017/S0140525X12000660>
- Selcuk, B., Gonultas, S., & Ekerim-Akbulut, M. (2023). Development and use of theory of mind in social and cultural context. *Child Development Perspectives*, *17*(1), 39–45. <https://doi.org/10.1111/cdep.12473>
- Shepherd, S. (2010). Following gaze: Gaze-following behavior as a window into social cognition. *Frontiers in Integrative Neuroscience*, *4*, Article 5. <https://doi.org/10.3389/fnint.2010.00005>
- 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
- Syed, M. (2021). *Reproducibility, diversity, and the crisis of inference in psychology*. OSF. <https://doi.org/10.31234/osf.io/89buj>
- Tang, Y., Gonzalez, M. R., & Deák, G. O. (2024). The slow emergence of gaze- and point-following: A longitudinal study of infants from 4 to 12 months. *Developmental Science*, *27*(3), Article e13457. <https://doi.org/10.1111/desc.13457>
- Thalmayer, A. G., Toscanelli, C., & Arnett, J. J. (2021). The neglected 95% revisited: Is American psychology becoming less American? *American Psychologist*, *76*(1), 116–129. <https://doi.org/10.1037/amp0000622>
- 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>
- Torréns, M. G., Wefers, H., & Kärtner, J. (2023). Der Einfluss von Kultur auf die Entwicklung in den ersten drei Lebensjahren [The influence of culture on development in the first three years of life]. *Zeitschrift für Entwicklungspsychologie Und Pädagogische Psychologie*, *55*(1), 14–18. <https://doi.org/10.1026/0049-8637/a000271>
- Triesch, J., Teuscher, C., Deák, G. O., & Carlson, E. (2006). Gaze following: Why (not) learn it? *Developmental Science*, *9*(2), 125–147. <https://doi.org/10.1111/j.1467-7687.2006.00470.x>
- Waschl, N., & Chen, M. (2022). Cross-cultural considerations for adapting valid psychoeducational assessments. In O. S. Tan, K. K. Poon, B. A. O'Brien, & A. Rifkin-Graboi (Eds.), *Early childhood development and education in Singapore* (pp. 113–140). Springer. https://doi.org/10.1007/978-981-16-7405-1_7
- Wefers, H., Schuhmacher, N., Chacón, L. H., & Kärtner, J. (2023). Universality without uniformity – Infants' reactions to unresponsive partners in urban Germany and rural Ecuador. *Memory & Cognition*, *51*(3), 807–823. <https://doi.org/10.3758/s13421-022-01318-x>
- Wimmer, H., & Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition*, *13*(1), 103–128. [https://doi.org/10.1016/0010-0277\(83\)90004-5](https://doi.org/10.1016/0010-0277(83)90004-5)
- Wu, Z., Chen, X., Gros-Louis, J., & Su, Y. (2018). “She is looking at me! Shall I share?” How Chinese and American preschoolers respond to eye gaze during sharing. *Social Development*, *27*(2), 447–460. <https://doi.org/10.1111/sode.12278>
- Zeigler-Hill, V., Welling, L. L., & Shackelford, T. K. (2015). How can an understanding of evolutionary psychology contribute to social psychology? In V. Zeigler-Hill, L. L. M. Welling, & T. K. Shackelford (Eds.), *Evolutionary perspectives on social psychology* (pp. 3–12). Springer.