Measuring variation in gaze following across communities, ages, and individuals — a showcase of the TANGO–CC

Julia Christin Prein (ORCID: 0000-0002-3154-6167)[1,2], Florian M. Bednarski (ORCID: 0000-0003-4384-4791)[3], Ardain Dzabatou[4], Michael C. Frank (ORCID: 0000-0002-7551-4378)[5], Annette M. E. Henderson (ORCID: 0000-0003-4384-4791)[3], Josefine Kalbitz[2], Patricia Kanngiesser (ORCID:0000-0003-1068-3725)[6], Dilara Keşşafoğlu (ORCID: 0000-0002-7356-0733)[7], Bahar Köymen (ORCID: 0000-0001-5126-8240)[8], Maira V. Manrique-Hernandez[2], Shirley Magazi (ORCID: 0009-0006-0479-9800)[9], Lizbeth Mújica-Manrique[2], Julia Ohlendorf[2], Damilola Olaoba[2], Wesley R. Pieters (ORCID:0000-0002-6152-249X)[9], Sarah Pope-Caldwell[2], Umay Sen (ORCID: 0000-0001-9488-0851)[10], Katie Slocombe (ORCID: 0000-0002-7310-1887)[11], Robert Z. Sparks (ORCID: 0000-0001-7545-0522)[5], Roman Stengelin (ORCID: 0000-0003-2212-4613)[2,9], Jahnavi Sunderarajan[2], Kirsten Sutherland[2], Florence Tusiime[12], Wilson Vieira (ORCID: 0009-0001-9400-6328)[2], Zhen Zhang (ORCID: 0000-0001-9300-0920)[13], Yufei Zong (ORCID: 0009-0000-5012-0244)[13], Daniel B. M. Haun (ORCID: 0000-0002-3262-645X)[2,+], and & Manuel Bohn (ORCID: 0000-0001-6006-1348)[1,2,+]

[1] Institute of Psychology in Education
Leuphana University Lüneburg
Germany
[2] Department of Comparative Cultural Psychology
Max Planck Institute for Evolutionary Anthropology

Germany
[3] School of Psychology
University of Auckland
New Zealand
[4] Université Marien Ngouabi
Republic of the Congo
[5] Department of Psychology
Stanford University
USA
[6] School of Psychology
University of Plymouth
UK
[7] Department of Psychology
Koç University
Türkiye
[8] Division of Psychology
Communication
and Human Neuroscience
University of Manchester
UK
[9] Department of Psychology and Social Work
University of Namibia
Namibia
[10] Department of Psychology
Developmental Psychology
Uppsala University
Sweden

[11] Department of Psychology
University of York
UK
[12] Budongo Conservation Field Station
Uganda
[13] CAS Key Laboratory of Behavioral Science
Institute of Psychology
Chinese Academy of Sciences
China
[+] joint last author

Author Note

The authors made the following contributions. Julia Christin Prein (ORCID: 0000-0002-3154-6167): Conceptualization, Formal Analysis, Methodology, Project administration, Resources, Software, Visualization, Writing - Original Draft Preparation, Writing - Review & Editing; Florian M. Bednarski (ORCID: 0000-0003-4384-4791): Methodology, Resources, Writing - Review & Editing; Ardain Dzabatou: Methodology, Resources, Writing - Review & Editing; Michael C. Frank (ORCID: 0000-0002-7551-4378): Methodology, Resources, Writing - Review & Editing; Annette M. E. Henderson (ORCID: 0000-0003-4384-4791): Methodology, Resources, Writing - Review & Editing; Josefine Kalbitz: Methodology, Resources, Writing - Review & Editing; Patricia Kanngiesser (ORCID:0000-0003-1068-3725): Methodology, Resources, Writing - Review & Editing; Dilara Keşşafoğlu (ORCID: 0000-0002-7356-0733): Methodology, Resources, Writing - Review & Editing; Bahar Köymen (ORCID: 0000-0001-5126-8240): Methodology, Resources, Writing - Review & Editing; Maira V. Manrique-Hernandez: Methodology, Resources, Writing - Review & Editing; Shirley Magazi (ORCID: 0009-0006-0479-9800): 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 (ORCID:0000-0002-6152-249X): Methodology, Resources, Writing - Review & Editing; Sarah Pope-Caldwell: Methodology, Resources, Writing - Review & Editing; Umay Sen (ORCID: 0000-0001-9488-0851): Methodology, Resources, Writing - Review & Editing; Katie Slocombe (ORCID: 0000-0002-7310-1887): Methodology, Resources, Writing - Review & Editing; Robert Z. Sparks (ORCID: 0000-0001-7545-0522): Methodology, Resources, Writing - Review & Editing; Roman Stengelin (ORCID: 0000-0003-2212-4613): 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 (ORCID: 0009-0001-9400-6328): Methodology, Resources, Writing - Review & Editing; Zhen Zhang (ORCID: 0000-0001-9300-0920): Methodology, Resources, Writing - Review & Editing; Yufei Zong (ORCID: 0009-0000-5012-0244): Methodology, Resources, Writing - Review & Editing; Daniel B. M. Haun (ORCID: 0000-0002-3262-645X): Conceptualization, Funding acquisition, Methodology, Writing - Review & Editing; Manuel Bohn (ORCID: 0000-0001-6006-1348): Conceptualization, Funding acquisition, Methodology, Project administration, Supervision, Writing - Review & Editing.

Correspondence concerning this article should be addressed to Julia Christin Prein (ORCID: 0000-0002-3154-6167), Universitätsallee 1, 21335 Lüneburg, Germany. E-mail: julia.prein@leuphana.de

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. This paper describes a gaze following task designed to measure basic social cognition across individuals, ages, and communities (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. The TANGO–CC captures individual- and community-level variation and shows good internal consistency in a data set from 2.5- to 11-year-old children from 17 diverse communities. Within-community variation outweighed between-community variation. We provide an open-source website for researchers to customize and use the task (https://ccp-odc.eva.mpg.de/tango-cc). The 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

Word count: XXX

Measuring variation in gaze following across communities, ages, and individuals — a showcase of the TANGO–CC

Introduction

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: we 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). While this sampling bias is often discussed within 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; Torréns 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 paper.

Social cognition refers to how an individual processes information in social situations which 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 (Peña, 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 US-American and European samples rarely report psychometric information (for a review, see Beaudoin et al., 2020), and the picture further deteriorates when we look at cross-cultural social cognition tasks (Bourdage et al., 2023; Hajdúk 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 within one community, let alone tasks that do so across different communities. In this paper, 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 and Vijver (2012), Waschl and Chen (2022)). While this approach allows greater flexibility and sensitivity to cultural differences, it might not be feasible to study a multitude of communities as 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 following this approach is Wefers et al. (2023) who investigated how cultural variations in parenting styles modulated infants' responses to disruptions in social interactions. While studies in the Global North often apply the Still-Face Paradigm to assess infants' reactions to unresponsive partners, Wefers et al. (2023) 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 to the Still-Face Paradigm, while 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; Waschl & Chen (2022)). This approach is less sensitive to each community's unique characteristics but renders quantitative comparisons of data more feasible. An example following this approach is the Multilingual Assessment Instrument for Narratives (MAIN; (Gagarina et al., 2012)), which assesses narrative abilities in mono- 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).

The present paper aims 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). While the question of how social-environmental and cultural factors impact gaze following was recently posed as one of the big open questions in gaze following research (Astor & Gredebäck, 2022), studies focusing on cultural variations of gaze following are rare. 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 (experimental condition) or a sticker in front of (control condition) a barrier, and children's crawling toward the barrier to follow the agent's gaze was assessed. While 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 to 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 result 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, and not adult-directed, speech. While these two studies point towards 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 – does not allow for studying individual-differences. A more comprehensive cross-cultural study on gaze following was recently conducted by Bohn 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 manuscript was developed for the study by Bohn 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 et al. (2023). The 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 et al., 2023). However, we cannot claim the task's generalizability and reliability based on a mono-cultural sample. This paper showcases the TANGO–CC (TANGO – Cross-Cultural), a standardized gaze following task that has been adapted to 13 languages and even more communities, and evaluates its psychometric properties by leveraging a large and diverse data set of 2.5- to 11-year-olds from 17 diverse communities (Bohn 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

The TANGO-CC was implemented in Leipzig, Germany, and thoroughly assessed in terms of reliability and validity (Prein et al., 2023). During this process, the cross-cultural adaptation of the task was prepared by a team of cross-cultural psychologists and cognitive scientists. In this paper, we assess the TANGO–CC's measurement quality (*i.e.,* variability and reliability) across 17 diverse communities by analyzing the data set from Bohn et al. (2024). In the following, we describe the different steps in detail.

The TANGO–CC is a screen-based task that measures the imprecision with which participants locate a balloon by following an agent's gaze (see Figure 1). 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. As the measures of the eyes and balloon are read out dynamically from the image on screen, stimuli can 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 the 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 et al., 2023). While 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 in the 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 et al. (2024) described gaze following as a form of social vector following and empirically found that performance in the TANGO was related to children's non-social 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 Figure 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, the TANGO–CC could be applied in 17 communities and 13 different languages (Bohn et al., 2024). In the following, we describe how researchers can use and customize the TANGO–CC in more detail.

Features of the TANGO–CC

Trials

The task consists of three different trial types: training 1, training 2, and test trials (see Figure 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 them. 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

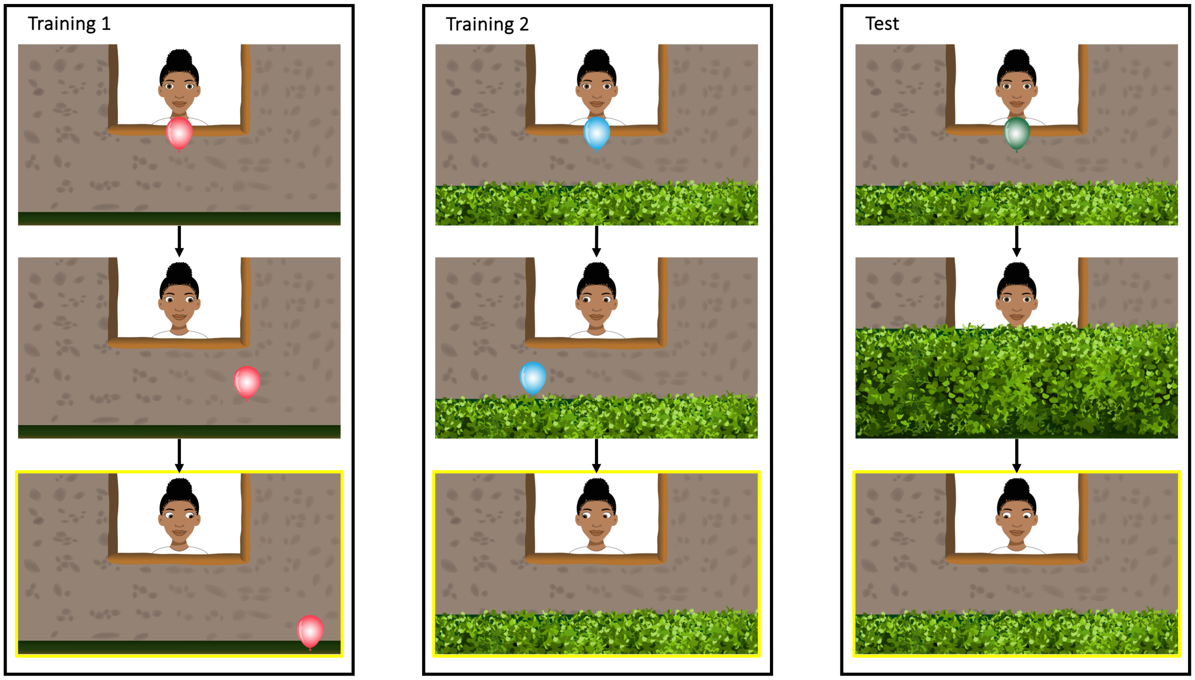

Figure 1

Screenshots of the 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).*

pupil center always align. Depending on the trial type, participants have different visual access to the balloon's position. In training 1, participants see the full trajectory of the balloon and directly have to click 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 Supplements). Notably, the instructions explicitly state that the agent is looking at the balloon.

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 minutes, 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 (1920 in "SVG units") is divided into ten bins. Exact coordinates (value between 0 for the far left and 1920 for the far right) within 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 the TANGO–CC's website (https://ccp-odc.eva.mpg.de/tango-cc/), they can select the language for audio instructions which are currently available for 13 different languages and five more dialects (see Table 1). To add a new language, researchers have two options: (1) for using their own 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 the TANGO–CC's manual (https://ccp-odc.eva.mpg.de/tango-cc/manual.html). (2) For adding a new language in the online version of the task, researchers can contact the first author of this paper. Please 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 et al. (2024) based on the selected language (see Supplements).

If researchers choose to customize the task (see Figure 2), they can adjust the number of trials for each trial type, but not their sequence. Specifically, trial types build 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 1; 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 (min 1, max 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 (1 - 8 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.

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

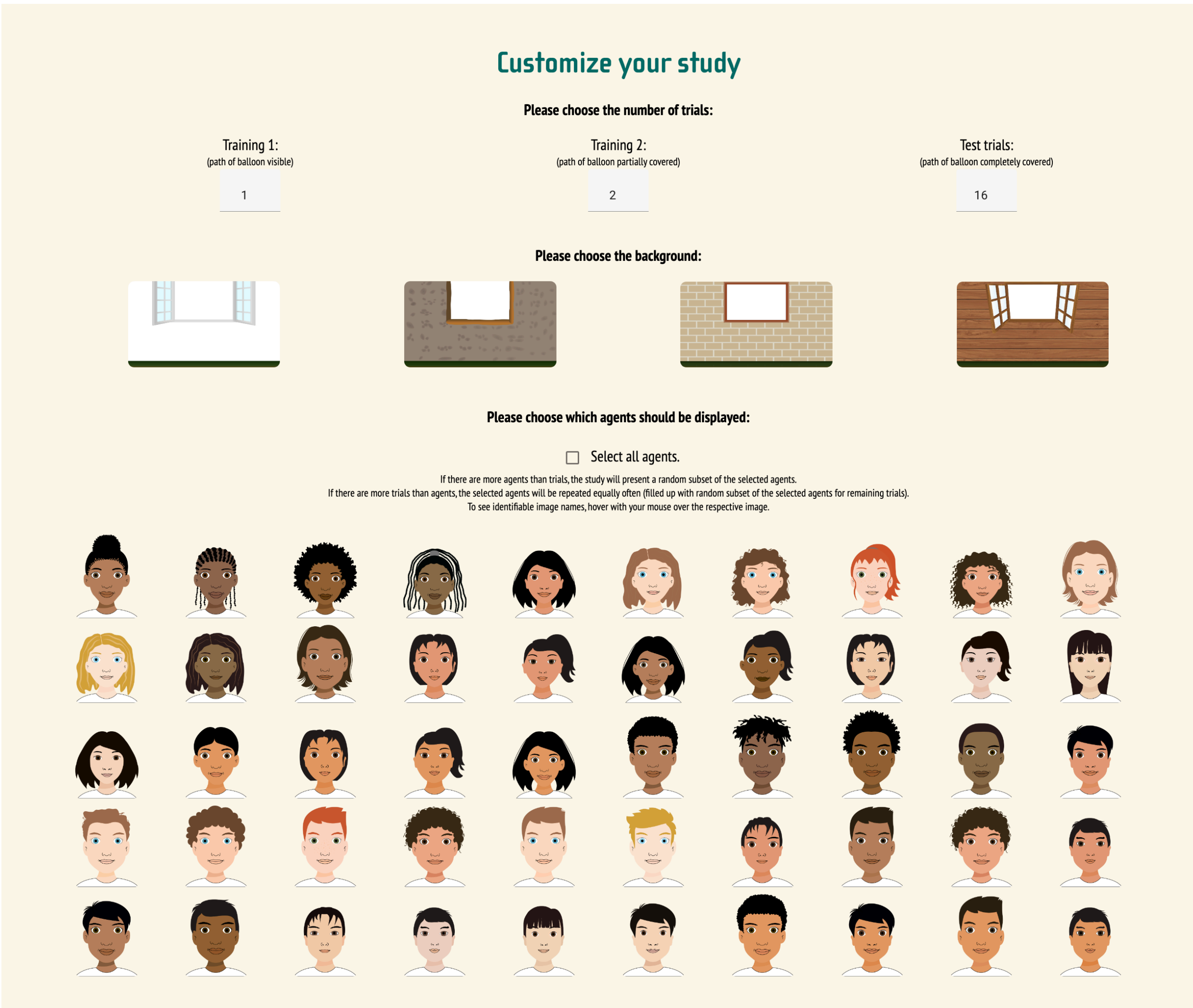

Figure 2

Screenshot of the customizable components of the 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.*

Table 1

Current language options available for the audio instructions in the TANGO–CC

Languages	Language family	Speaker's country of origin
Bemba	Bantu	Zambia
Chinese	Sino-Tibetan	China
English	Indo-European	USA / UK / India / Nigeria / New Zealand
German	Indo-European	Germany
(≠Akhoe) Hai‖om	Khoe-Kwadi	Namibia
Khwedam	Khoe-Kwadi	Namibia
Lingala	Bantu	Rep. Congo
Marathi	Indo-European	India
Shona *	Bantu	Zimbabwe
Spanish	Indo-European	Argentina (Rioplatense Spanish) / Mexico (Mexiquense Spanish)
Kiswahili	Bantu	Uganda
Turkish	Turkic	Türkiye
Yaka	Bantu	Rep. Congo

Note. In cases where 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. * Please note that audio instructions are available in Shona but no data of this version is included in the present data set.

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 (though please note that configurations of browsers and JavaScript may change in the future). Participant's 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 (for example, using online platforms like *Prolific*; see Prein et al. (2023)). The task can be shared easily internationally by providing the URL. Importantly, 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 Scalable Vector Graphics (SVG; an image format that stores image elements via 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 only registered when the participant clicks the relevant part of the screen (*i.e.,* in test trials, when they click the hedge). Furthermore, clicks are only registered after the voice recordings stop playing. An audio reminder is played again if no click is registered within 5 seconds.

The website does not use cookies, nor does it upload any data to servers; that is, the data is only stored 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 et al. (2024) for the psychometric evaluation of the TANGO–CC. The data set contains a sample of N = 1377 children, aged 2.5 to 11 years. Participants came from 17 communities on five continents, in rural and urban settings, with varying degrees of market integration and technology exposure. Bohn et al. (2024) carried out 19 trials (1 training 1, 2 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 et al. (2024).

Individual differences

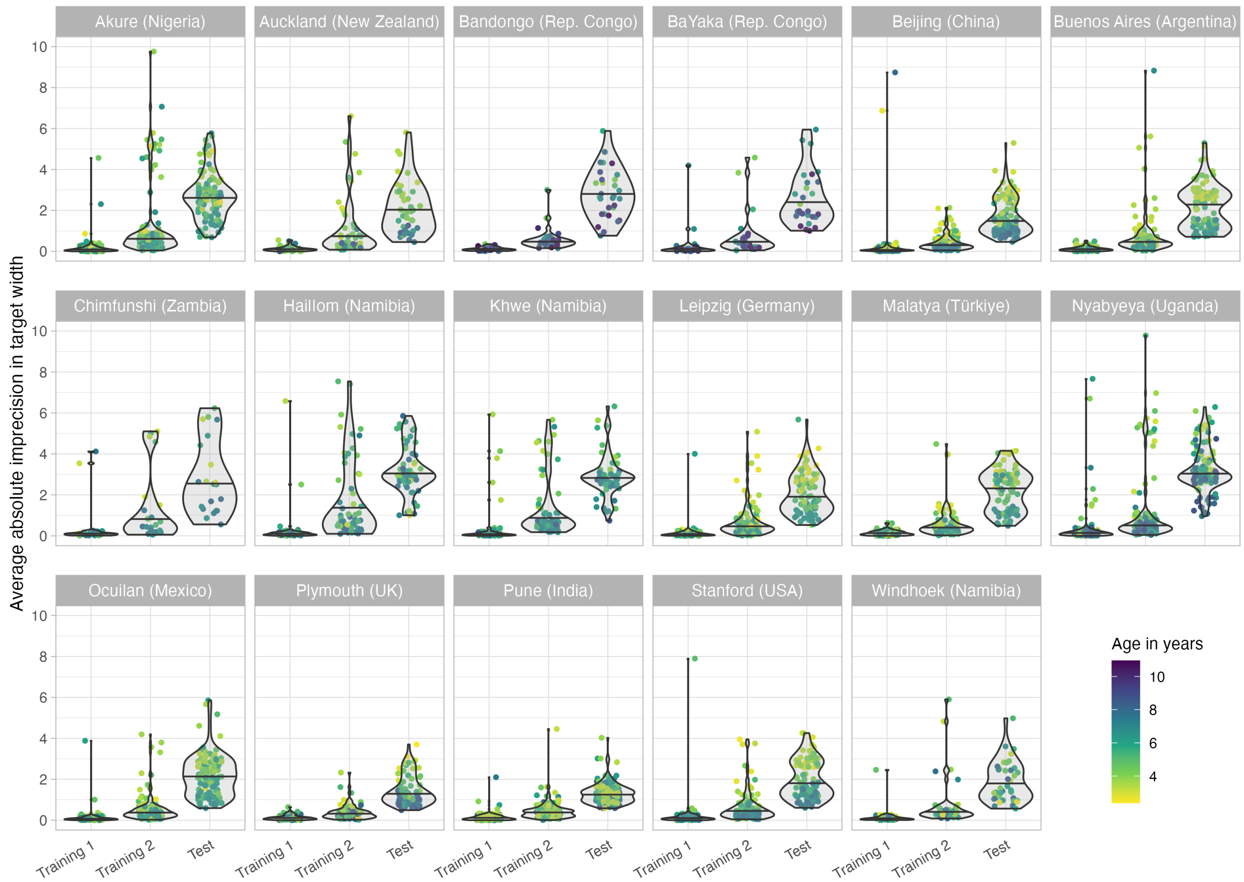

Figure 3

Variability measured by the 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.*

All analyses and the data set can be accessed 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 (mean = 0.19, sd = 0.63), followed by training 2 (mean = 0.79, sd = 1.44) and test trials (mean = 2.21, sd = 2.03; see Figure 3).

To formally estimate the effect of trial type on performance in the 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 4.4.0 (2024-04-24) (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: imprecision ~ trialtype + (trialtype | community)`), and imprecision was modeled by a `lognormal` 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 (β = -3.26; 95% CrI [-3.41; -3.10]) and training 2 (β = -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 Figure 3).

To identify the sources of variation, we computed intraclass correlations (ICC). The variation in children's imprecision within communities was substantially larger than the variation between the communities. The mean within-community variance was 1.28, ranging from 0.24 (in Pune, India) to 3.46 (in Chimfunshi, Zambia). Between-community variance was 0.34. The ICC, representing the proportion of between-community variance relative to the total variance (sum of within- and between-community variance), was 0.02. This indicated that only

[1] Please note that the 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% of the total variability in the data could be attributed to differences between communities, while the remaining 98% were attributed to differences within communities (Kusano et al., 2024).

Reliability

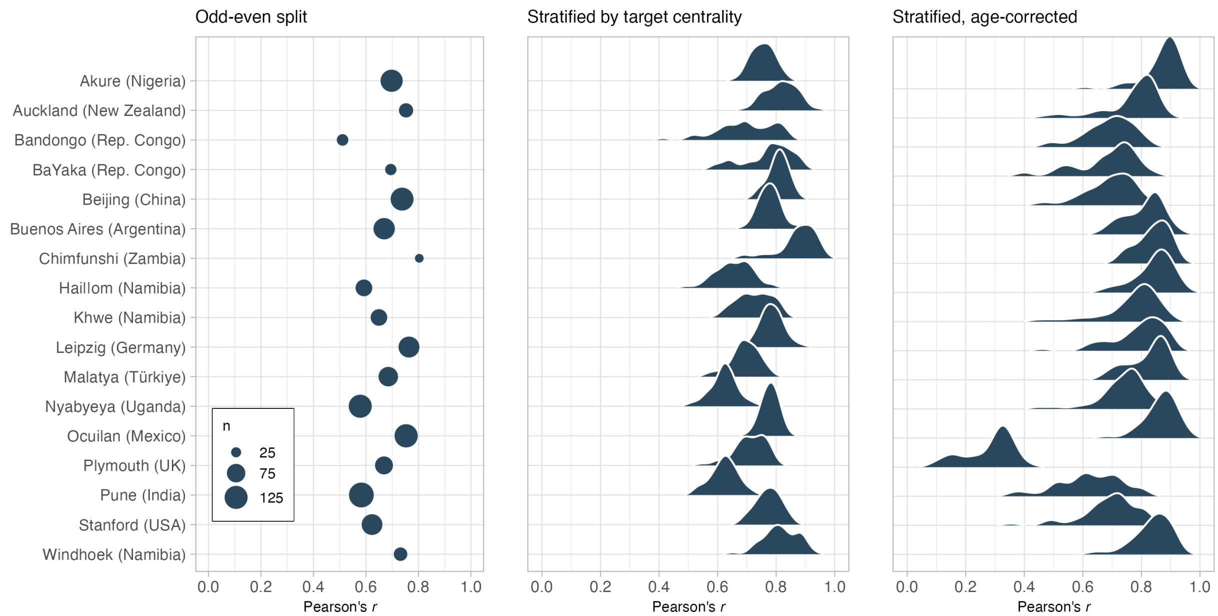

Figure 4

Reliability of the 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 GLMM.*

To assess reliability, we estimated internal consistency in each community in three different ways. First, data of each participant was 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 was stratified by target centrality (capturing trial difficulty), and a Pearson correlation was calculated between the matched halves. 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: `imprecision ~ age + (0 + half | 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 inter-individual differences. Because the process of generating stratified data sets is partly random, the model was fitted 50 times for each community. The posterior estimate of the correlation between the two person-specific estimates was taken as the age-independent estimate for internal consistency.

The results are shown in 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, UK, being an outlier with 0.28). Following Cohen's suggestions (Cohen, 1988, 1992), these correlations constitute large effects ($r > .50$), and indicate good internal consistency.[2] The results are comparable to the internal consistency estimates found in the original TANGO study (Prein et al., 2023), 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 = 0.46, 95%CI [-0.03; 0.77]). This suggests that the less variation a task can capture within a community, the lower the reliability. However, please note that this correlation could be influenced by outliers and that the sample size here (N = 17 communities) is too small to make substantial claims.

Discussion

The 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

[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 due to their variable nature.

within-community variation greatly exceeded between-community variation. Furthermore, the task showed satisfactory to high reliability across all communities. Therefore, the TANGO–CC is a suitable task to capture individual differences in social-cognitive development in diverse communities.

The 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. Importantly, we re-assessed the TANGO–CC's psychometric properties in a large and diverse data set. While we cannot generalize our findings to all communities worldwide, we found that the TANGO-CC captured reliable individual variation in all 17 communities studied by Bohn et al. (2024). We hope that not just the 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 the TANGO–CC (or any other task) in a new community requires sensitivity to the specific context, piloting, and, most importantly, 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 where components can be modified (*i.e.*, building block structure). In the case of the TANGO–CC, the task cannot only be adapted to different languages, cartoon faces, and backgrounds (see Figure 2) but also updated with new stimuli. Unlike studies that present sequential, hand-painted pictures that are difficult to adapt (Mehta et al., 2011), the TANGO–CC uses SVGs (Scalable Vector Graphics) that can be easily exchanged.

Compared to one of the most commonly used social cognition measures – the change-of-location false belief task (Baron-Cohen et al., 1985; Wimmer & Perner, 1983) – the TANGO–CC has several advantages: a continuous outcome measure that can capture individual differences from three 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, while it can also be used to collect data offline in locations without a reliable internet connection. The TANGO–CC follows a standardized procedure, which leaves no room for 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 non-completion rates.

The TANGO–CC is a screen-based task. Bohn et al. (2024) have shown that children with no prior touchscreen exposure were less precise in the TANGO–CC than children with prior experience. However, individual differences were also found in communities with 100% touchscreen exposure, showing that this factor alone could not explain children's performance in the task (Bohn et al., 2024). Notably, even though the touchscreen experience caused absolute differences in task performance, all communities showed the same processing signature. A recent computational cognitive model described gaze following as a process of estimating pupil angles and the corresponding gaze vectors (Prein et al., 2024). Bohn 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. Nevertheless, the mode of stimulus presentation needs to be kept in mind when administering the 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.

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 the TANGO–CC indeed taps into social cognition as utilized in real life: Prein et al. (2024)

found that children's perspective-taking abilities in a personal social interaction were linked to performance in the 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 the 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 due to organizational and financial constraints. An example of assessing the TANGO's predictive validity is a study conducted in Leipzig, Germany, which used the TANGO to predict children's receptive vocabulary 6 months later (Prein et al., 2023). Future cross-cultural studies could investigate the 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 the 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-culture variability in the development of social cognition since sometimes within-culture differences exceed between-culture differences. Indeed, we found that within-group variability was greater than between-group variability. While we believe that the 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). Similarly, Elson et al. (2023) have called for standardized, psychometrically evaluated measures that can be re-used by other researchers in order 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 generalisability of psychological research findings (Li et al., 2024; Syed, 2021). Li et al. (2024) have 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 the TANGO–CC has taken. Openly sharing the TANGO–CC’s materials will allow other researchers to (hopefully) replicate the results and deepen our cumulative understanding of social-cognitive development across diverse communities.

Conclusion

The TANGO–CC captures individual differences in social-cognitive development across diverse communities. The task’s customizability, minimal language demands, and its 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 the TANGO–CC – and its pragmatic construction process – will provide a roadmap for future cross-cultural studies on cognitive development.

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

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

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. In *Advances in Child Development and Behavior* (p. S0065240722000192). Elsevier. https://doi.org/10.1016/bs.acdb.2022.04.003

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

Beaudoin, C., Leblanc, É., Gagner, C., & Beauchamp, M. H. (2020). Systematic Review and Inventory of Theory of Mind Measures for Young Children. *Frontiers in Psychology*, *10*, 2905. https://doi.org/10.3389/fpsyg.2019.02905

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., Keşşafoğlu, 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*. 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

Bürkner, P.-C. (2017). Brms: An R Package for Bayesian Multilevel Models Using Stan. *Journal of Statistical Software*, *80*(1), 1–28. https://doi.org/10.18637/jss.v080.i01

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

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

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed). L. Erlbaum Associates.

Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*(1), 155–159. https://doi.org/10.1037/0033-2909.112.1.155

Decety, J. (Ed.). (2020). *The social brain: A developmental perspective* (pp. xii, 426). The MIT Press. https://doi.org/10.7551/mitpress/11970.001.0001

Del Bianco, T., Falck-Ytter, T., Thorup, E., & Gredebäck, G. (2019). The Developmental Origins of Gaze-Following in Human Infants. *Infancy*, *24*(3), 433–454. https://doi.org/10.1111/infa.12276

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*(1), 1–4.

https://doi.org/10.1038/s44271-023-00026-9

Field, A., Miles, J., & Field, Z. (2012). *Discovering statistics using R.* Sage.

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*, 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., & Vijver, F. van de. (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), 61-83; discussion 83-135. 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), 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

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. *American Psychologist*. 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

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), 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

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

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

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

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*. PsyArXiv. https://doi.org/10.31234/osf.io/dy73a

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

Schilbach, L., Timmermans, B., Reddy, V., Costall, A., Bente, G., Schlicht, T., & Vogeley, 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*(5). https://doi.org/10.3389/fnint.2010.00005

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), 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. *Zeitschrift für Entwicklungspsychologie Und Pädagogische Psychologie*, *55*(1), 14–18. https://doi.org/10.1026/0049-8637/a000271

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

Zeigler-Hill, V., Welling, L. L., & Shackelford, T. K. (2015). How can an understanding of evolutionary psychology contribute to social psychology? In *Evolutionary perspectives on social psychology* (pp. 3–12). Springer, Cham.