

Measuring individual differences in the understanding of gaze cues across the lifespan

Julia Prein¹, Manuel Bohn¹, Luke Maurits¹, Steven Kalinke¹, & Daniel M. Haun¹

¹ Department of Comparative Cultural Psychology, Max Planck Institute for Evolutionary
Anthropology, Leipzig, Germany

Author Note

Correspondence concerning this article should be addressed to Julia Prein, Max
Planck Institute for Evolutionary Anthropology, Deutscher Platz 6, 04103 Leipzig,
Germany. E-mail: julia_prein@eva.mpg.de

Abstract

There must be an abstract of no more than 250 words. One or two sentences providing a **basic introduction** to the field, comprehensible to a scientist in any discipline.

Two to three sentences of **more detailed background**, comprehensible to scientists in related disciplines.

One sentence clearly stating the **general problem** being addressed by this particular study.

One sentence summarizing the main result (with the words “**here we show**” or their equivalent).

Two or three sentences explaining what the **main result** reveals in direct comparison to what was thought to be the case previously, or how the main result adds to previous knowledge.

One or two sentences to put the results into a more **general context**.

Two or three sentences to provide a **broader perspective**, readily comprehensible to a scientist in any discipline.

Keywords: social cognition, individual differences, gaze cues, psychometrics

Word count: X

26 Measuring individual differences in the understanding of gaze cues across the lifespan

– POSSIBLE ROUGH STRUCTURE –

intro

individual differences in dev psy

- reliable tasks: need variation, more trials

existing tasks for social cognition

- wellman

current goal

- standardized, easy to use, continuous

methods (let's see whether we want this generic heading...)

design of our task

- training trials (touch & fam together, click on visible balloon / visible target flight, voice over trials)
- test trials
- flexible: two versions, can also use discrete
- face value
- stimulus timing

43 **implementation / development**

- 44 • JS, HTML, CSS
- 45 • parcel?!
- 46 • SVG: scalable
- 47 • webapp: portable across devices, flexible, no hard system requirements
- 48 • response collection: file format, variables saved (click responses)

49 **data processing**

- 50 • workflow for collecting data
- 51 • pipeline? downloading server data into r, converting coordinate systems

52 **does our task induce variation?**

53 **participants: kiga & prolific vali**

54 **procedure**

55 **results**

- 56 • for hedge & box
- 57 • sanity check: developmental trajectory
- 58 • remote testing: online child sample as side note

59 **do we capture variation reliably?**

60 **participants: relikiga & prolific reli**

61 **results:**

- 62 • internal consistency?
- 63 • test-retest

64 **exploratory: external validity**

65 **participants: subsample kiga & relikiga with questionnaire data**

66 **results:**

- 67
 - peer exposure (box & hedge combined?)

68 **discussion**

69 **limitations**

70 **future development / extending the task**

71 **conclusion**

Introduction

Idea for an opener :)

Developmental psychology is facing a dilemma: many research questions are questions about individual differences, yet, there is a lack of tasks to reliably measure these individual differences. For example

- individual differences in developmental psychology
- reliable tasks, variation needed, more trials
- existing tasks for social cognition: wellman
- goal of the current project

Methods

Design of our balloon finding task

- training trials (summarize touch & fam & voice over)
- test trials in detail
- two versions: continuous & discrete, flexible to use

Stimuli

TODO: cartoon style stimuli. engaging and for children and adults alike

An animal character (i.e., agent; sheep, monkey, or pig) is placed centrally in a window. A balloon (i.e., target; blue, green, yellow, or red) is located in front of them. The target then falls to the ground. The agent's gaze tracks the movement of the target, that is, the pupils and iris of the agent move in a way that their center aligns with the center of the target. Participants were then asked to locate the target's position on the screen.

TODO: phrasing training trials

There are three different trial types, depending on the visual access to the target flight and end location. In *touch training* trials, participants have *full visual access* to the target flight and the target's end location. In *familiarization trials*, participants have *partial visual access*: they witness the target flight but cannot see the target's end location. In *test trials*, participants have *no visual access* to the target flight nor the end location. When partial or no information about the target location is accessible, participants are expected to use the agent's gaze as a cue.

There are two different versions of partial and no visual access trials: participants have to indicate their estimated target location directly on the hedge (i.e., hedge version) or in one of five/eight boxes (i.e., box version). In test trials, the target flight is always covered by a hedge. In the hedge version, the hedge then shrinks to cover the target's end location. In the box version, the hedge shrinks completely. The boxes then hide the target's end location.

To keep participants engaged and interested, the presentation of events was accompanied by cartoon-like sounds. Each trial started with an eye-blinking sound, while pupils and iris were enlarged and changed in opacity for 0.3 sec. The landing of the target was accompanied by a tapping sound. After the response was registered, a short plop sound played and a small orange circle confirmed the participants' location choice. Once the target landed, the instructor's voice asked "Where is the balloon?" If no response was registered, an audio prompt reminded the participant to respond and click on the balloon/the hedge/a box.

Study instructions were pre-recorded and constant across participants. No interaction with the experimenter was necessary and children needed minimal assistance from their caregivers. Participants received 19 trials with one touch training, two familiarization trials, and 16 test trials. The first trial of each type comprised a voice-over description of

the presented events. Touch and familiarization trials, as well as voice-over trials, were not included in the analysis. Therefore, we conducted our analyses with 15 test trials.

Randomization. All agents and target colors appeared equally often and could not be repeated in more than two consecutive trials. The randomization of the target end location depended on the study version. In the hedge version, the full width of the screen was divided into ten bins. Exact coordinates within each bin were randomly generated. In the box version, the target randomly landed in one of the boxes. Each bin/box occurred equally often. The same bin/box could only occur twice in a row.

Dependent variable. The dependent variable depends on the study version: for the hedge version, our dependent variable is continuous. Here, the dependent variable is imprecision, which is defined as the absolute difference between the true x coordinate of the target and the x coordinate of the participant's click on the screen. For the box version, we use our categorical outcome (i.e., which box was clicked) to calculate the proportion of correct responses.

Implementation

- Parcel?
- benefits of our choices: easy to adapt, portable across devices, webapp: no installation needed
- response collection: file formats, information we save

Material is presented as an interactive web-app that is accessible for computers and tablets and runs on any web browser. The code is open-source (<https://github.com/ccp-eva/gafo-demo>) and a live demo version can be found under: <https://ccp-odc.eva.mpg.de/gafo-demo/>.

The web-app generates two files: (1) a text file (.json) containing meta-data, trial specifications and participants' click responses, and (2) a video file (.webm) of the

participant’s webcam recording. Data gets automatically collected and safely stored on local servers located in Leipzig, Germany.

The web-app was programmed in JavaScript (ECMAScript 2015, i.e., ES6), HTML5 and CSS. For stimulus presentation, a scalable vector graphic (SVG) composition was parsed. This way, the composition scales according to the user’s viewport without loss of quality, while keeping the aspect ratio and relative object positions constant. Furthermore, SVGs allow us to precisely control and calculate the size and position of all composite parts of the scene (e.g., pupil of the agent).

The GreenSock Animation Platform (GSAP; TODO insert citation) library was used to animate the movement of single SVG elements.

Data processing

- workflow server to R, converting coordinate systems We used R [Version 4.1.2; R Core Team (2021)] and the R-package *papaja* [Version 0.1.0.9997; Aust and Barth (2020)] for all our analyses.

Does the balloon finding task induce variation?

Participants: kiga & prolific vali

Procedure

Results

- for hedge & box
- sanity check: developmental trajectory
- remote testing: online child sample as side note

Can we capture variation reliably?

Participants: religa kiga & prolific reli

Procedure: 14 ± 3 days

- what is different compared to variation? nr of trials, fixed order ## Results
- internal consistency
- test-retest

Exploring the external validity of our task

Participants: subset kiga & relikiga that filled out questionnaires

Results

- peer exposure (hedge & box combined?)

Discussion

Limitations

Future development / extending the task

Conclusion

Declarations

Open practices statement

The web application (<https://ccp-odc.eva.mpg.de/gafo-demo/>) described here is open source (<https://github.com/ccp-eva/gafo-demo>). The datasets generated during and/or analysed during the current study are available in the [gazecues-methods] repository,

(<https://github.com/jprein/gazecues-methods>). All experiments were preregistered
(<https://osf.io/zjhsc/>).

Funding

This study was funded by the Max Planck Society for the Advancement of Science, a
noncommercial, publicly financed scientific organization (no grant number). We thank all
the children and parents who participated in the study.

Conflicts of interest

The authors declare that they have no conflict of interest.

Ethics approval

Consent to participate

Informed consent was obtained from all individual participants included in the study
or their legal guardians.

Consent for publication

Open access

Authors' contributions

optional: please review the submission guidelines from the journal whether
statements are mandatory

References

- Aust, F., & Barth, M. (2020). *papaja: Create APA manuscripts with R Markdown*. Retrieved from <https://github.com/crsh/papaja>
- R Core Team. (2021). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from <https://www.R-project.org/>

Supplements

Adult sample

Recruitment. We recruited participants using the online participant recruitment service *Prolific* from the University of Oxford. *Prolific*'s subject pool consists of a mostly European and US-american sample although subjects from all over the world are included. The recruitment platform realises ethical payment of participants, which requires researchers to pay participants a fixed minimum wage of £5.00 (around US\$6.50 or €6.00) per hour. We decided to pay all participants the same fixed fee which was in relation to the estimated average time taken to complete the task. *Prolific* distributed our study link to potential participants, while the hosting of the online study was done by local servers in the Max Planck Institute for Evolutionary Anthropology, Leipzig. Therefore, study data was saved only on our internal servers, while *Prolific* provided demographic information of the participants. Participants' *Prolific* ID was forwarded to our study website using URL parameters. This way, we could match participant demographic data to our study data. The same technique was used to confirm study completion: we redirected participants from our study website back to the *Prolific* website using URL parameters. We used *Prolific*'s inbuilt prescreening filter to include only participants who were fluent in English and could therefore properly understand our written and oral study instructions.

Study 1 - Validation hedge version. The aim of Study 1 was to validate the hedge version of our balloon finding task. The pre-registration can be found here: <https://osf.io/r3bhn>. We recruited participants online by advertising the study on *Prolific*.

50 adults participated in the study. One additional subject returned their submission, i.e., decided to leave the study early or withdrew their submission after study completion. Data collection took place in May 2021.

Participants were compensated with £1.25 for completing the study. We estimated

an average completion time of 6 minutes, resulting in an estimated hourly rate of £10.00.

In average, participants took 05:56min to complete the study.

Participants were required to complete the study on a tablet or desktop. Participation on mobile devices was disabled since the display would be too small and would harm click precision. It was indicated that the study required audio sound.

We stored *Prolific*'s internal demographic information, while not asking for additional personal information.

After clicking the study title, participants were directed to our online study website.

...

Study 2 - Validation box version. As in study 1, we recruited participants on *Prolific*, and employed the same methodology. However, this time we focussed on validating the box version of the task in an adult sample. Participants were presented with eight boxes in which the target could land.

50 adults participated in the study. One additional subject returned their submission, i.e., decided to leave the study early or withdrew their submission after study completion. Data collection took place in June 2021.

Participants were compensated with £1.00 for completing the study. We estimated an average completion time of 6 minutes, resulting in an estimated hourly rate of £10.00. In average, participants took 04:43min to complete the study.

Study 3 - Reliability hedge version. In study 3 and 4, we assessed the test-retest reliability of our balloon-finding task in an adult sample. The pre-registration can be found here: <https://osf.io/nu62m>. We tested the same participants twice with a delay of two weeks. The testing conditions were as specified in Study 1 and 2. However, the target locations as well as the succession of animals and target colors was randomized once. Each participant then received the same fixed randomized order of target location, animal, and target color. Participants received 30 test trials without voice-over description,

so that each of the ten bins occurred exactly three times.

In addition to the beforementioned prescreening settings, we used a whitelist. *Prolific* has a so-called *custom allowlist prescreening filter* where one can enter the *Prolific* IDs of participants who completed a previous study. Only these subjects are then invited to participate in a study. This way, repeated measurements can be implemented, collecting data from the same subjects at different points in time.

In a first round, 60 participants took part on the first testday. Additional two subjects returned their submission, i.e., decided to leave the study early or withdrew their submission after study completion. One additional participant timed out, i.e., did not finish the survey within the allowed maximum time. The maximum time is calculated by *Prolific*, based on the estimated average completion time. For this study, the maximum time amounted to 41 minutes. For the first testday, participants were compensated with £1.25. We estimated an average completion time of 9 minutes, resulting in an estimated hourly rate of £8.33. In average, participants took 07:11min to complete the first part.

Of the 60 participants that completed testday 1, 41 subjects finished testday 2. One additional participant timed out, i.e., did not finish the survey within the allowed maximum time. Participants were compensated with £1.50 for completing the second part of the study. We estimated an average completion time of 9 minutes, resulting in an estimated hourly rate of £10. In average, participants took 06:36min to complete the second part of the study.

Since we aimed for a minimum sample size of 60 subjects participating on both testdays, we reran the first testday with additional 50 participants. Additional seven subjects returned their submission, i.e., decided to leave the study early or withdrew their submission after study completion. Two additional participants timed out, i.e., did not finish the survey within the allowed maximum time. Again, participants were compensated with £1.25 for completing the first part of the study (estimated average completion time 9

minutes, estimated hourly rate of £8.33). In average, participants took 06:51min to complete the first part.

Of the additional 50 participants that completed testday 1, 29 subjects finished testday 2. Again, participants were compensated with £1.50 for completing the second part of the study (estimated average completion time 9 minutes, estimated hourly rate of £10). In average, participants took 06:26min to complete the second part of the study.

Study 4 - Reliability box version. As in study 3, we recruited participants on *Prolific*, and employed the same methodology. However, this time participants were presented with the box version of the task. Participants received 32 test trials without voice-over description, so that each of the eight boxes occurred exactly four times. As in study 2, we employed eight boxes in which the target could land.

In a first round, 60 participants took part on the first testday. Additional five subjects returned their submission, i.e., decided to leave the study early or withdrew their submission after study completion. For the first testday, participants were compensated with £1.25. We estimated an average completion time of 9 minutes, resulting in an estimated hourly rate of £8.33. In average, participants took 07:33min to complete the first part.

Of the 60 participants that completed testday 1, 41 subjects finished testday 2. Participants were compensated with £1.50 for completing the second part of the study. We estimated an average completion time of 9 minutes, resulting in an estimated hourly rate of £10. In average, participants took 07:50min to complete the second part of the study.

Since we aimed for a minimum sample size of 60 subjects participating on both testdays, we reran the first testday with additional 50 participants. Additional eight subjects returned their submission, i.e., decided to leave the study early or withdrew their submission after study completion. One additional participant timed out, i.e., did not finish the survey within the allowed maximum time. Again, participants were compensated

with £1.25 for completing the first part of the study (estimated average completion time 9 minutes, estimated hourly rate of £8.33). In average, participants took 07:37min to complete the first part.

Of the additional 50 participants that completed testday 1, 28 subjects finished testday 2. Additional three subjects returned their submission, i.e., decided to leave the study early or withdrew their submission after study completion. One additional participant timed out, i.e., did not finish the survey within the allowed maximum time. Again, participants were compensated with £1.50 for completing the second part of the study (estimated average completion time 9 minutes, estimated hourly rate of £10). In average, participants took 06:30min to complete the second part of the study.

Child sample

Study 1 - Validation Remote and in-person. The validation of our task in a in-person and remote child sample can be found here: <https://osf.io/snju6>. We chose to have at least 20 data points per cell (i.e. unique combination of data collection mode, study version, and age-group). Across the two data collection modes, a total of 267 children participated. Participants received a small gift as thank you for their participation in the study.

For our in-person supervised testing sample, we went to kindergartens in Leipzig and surroundings that cooperate with the Max Planck Institute for Evolutionary Anthropology. For our remote unsupervised testing sample, families were recruited on a voluntary basis via email from the database of the Max Planck Institute for Evolutionary Anthropology. Children in both sub samples live in Leipzig, Germany or surrounding areas and grow up in an industrialized, urban Central-European context. Information on socioeconomic status was not formally recorded, although the majority of families come from mixed, mainly mid to high socioeconomic backgrounds with high levels of parental education. Written

informed consent was obtained from at least one caregiver prior to testing. TODO: how to phrase for kiga testing?

Procedure Remote Testing

In the beginning of the online study, families were invited to enter “our virtual institute” and were welcomed by an introductory video of the study leader, shortly describing the research background and further procedure. Then, caregivers were informed about data security and were asked for their informed consent. They were asked to enable the sound and seat their child centrally in front of their device. Subsequently, a brief demographic questionnaire was displayed, asking for (1) the total number of household members, (2) the number of children, (3) age of the other children, (4) whether the child was in day care, and if yes, (5) since when and (6) for how long on an average day. Before the study started, families were instructed how to setup their webcam and enable the recording permissions. Study participation was video recorded whenever possible in order to ensure that the answers were generated by the children themselves. Then, families were guided through the online study with pre-recorded audio instructions. After completion, families received a little crafting / coloring sheet as a small thank-you gift.

Depending on the participant’s device, the website automatically presents the hedge or box version of the study. For families that use a tablet with touchscreen, the hedge version is shown. Here, children can directly click on the screen themselves to indicate where the target is. For families that use a computer without touchscreen, the website presents the box version of the task. We assumed that younger children in our sample would not be acquainted with the usage of a computer mouse. Therefore, we asked children to point to the screen. Caregivers were then asked to act as the “digital finger” of their children and click on the indicated box. In order to facilitate the translation of children’s pointing and caregivers’ clicking, we decided to implement this categorical version of the task.

Our in-person supervised testing sample involved 120 children, including 40 3-year-olds (mean = 41.45 months, SD = 3.85, range = 36 - 47, 22 girls), 40 4-year-olds (mean = 54.6 months, SD = 3.1, range = 48 - 59, 19 girls), and 40 5-year-olds (mean = 66.95 months, SD = 3.39, range = 60 - 71, 22 girls).

Study 2 - Reliability in-person. <https://osf.io/xqm73>