1   Supplements for the manuscript

2   "TANGO: A reliable, open-source, browser-based task to assess individual differences in

3   gaze understanding in 3 to 5-year-old children and adults"

4   Julia Prein[1], Manuel Bohn[1], Steven Kalinke[1], & Daniel B. M. Haun[1]

5   [1] Department of Comparative Cultural Psychology, Max Planck Institute for Evolutionary

6   Anthropology, Leipzig, Germany

7   Author Note

8   Correspondence concerning this article should be addressed to Julia Prein, Max

9   Planck Institute for Evolutionary Anthropology, Deutscher Platz 6, 04103 Leipzig,

10   Germany. E-mail: julia_prein@eva.mpg.de

11   Supplements for the manuscript

12   "TANGO: A reliable, open-source, browser-based task to assess individual differences in

13   gaze understanding in 3 to 5-year-old children and adults"

14   **Child sample**
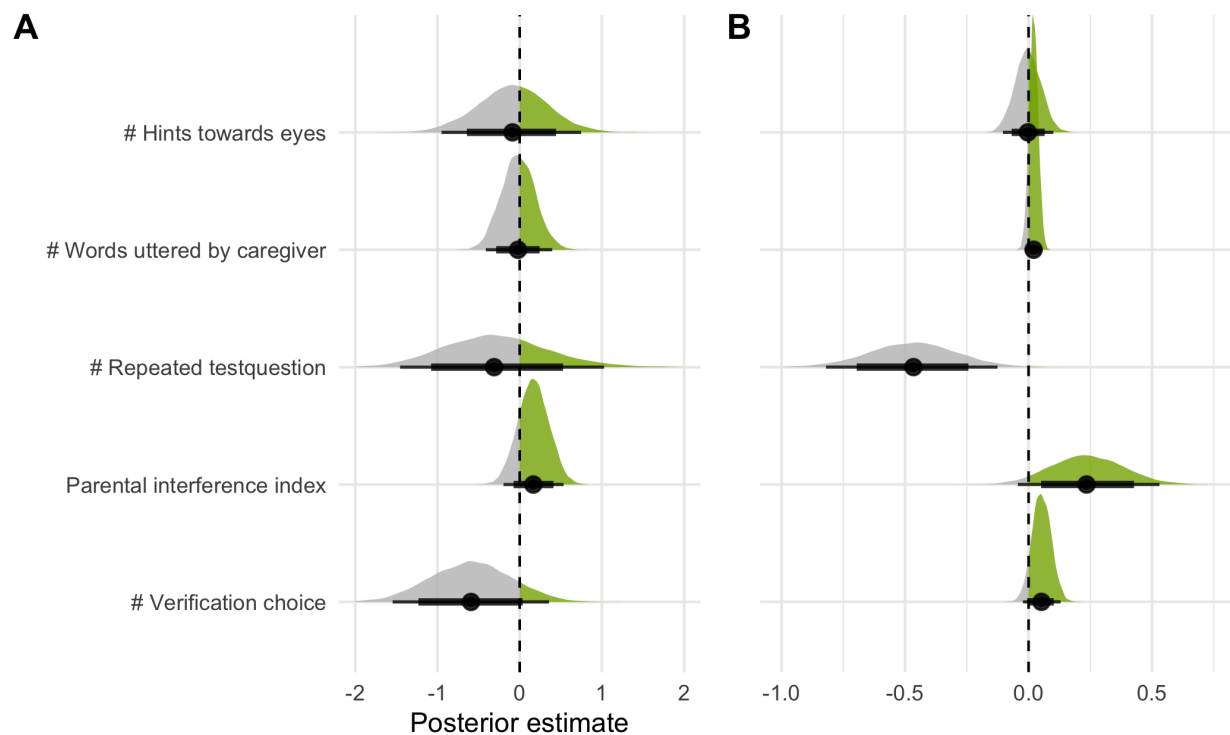
15 **Webcam coding**

Figure 1. **Model comparison for exploratory webcam coding of parental interference**. Factors of parental interference and their influence on the probability of responding correctly. The graph shows the estimated density curves of a model's predictor coefficient. Models are ordered according to their WAIC scores in the trial-by-trial analysis, with the uppermost winning the model comparison. (A) Analysis on a trial-by-trial level. (B) Analysis on a subject level (i.e., average across trials per subject).

16   Comparing the performances of children across our two data collection modes, we

17 found that children participating remotely were slightly more precise. This difference was

18 especially prominent in younger participants in the box version of the task. It is

19 conceivable that caregivers were especially prone to influence the behavior of younger

20 children. In the box version, caregivers might have had more opportunities to interfere

21 since they carried out the clicking for their children. In an exploratory analysis, we coded

22 parental behavior and environmental factors during remote unsupervised testing. Due to

23 the time consuming nature of hand coding videos frame by frame, we focused on the

24 subsample with the greatest performance difference between data collection modes: the

25 three-year-olds in the box version of the task (n = 16). We reasoned that if parental

26 interference cannot explain the greatest performance difference in our sample, the effects

27 would be negligible in the remaining sample. A trial was defined as the time between two

28 eye blinking sounds. We transcribed all utterances by parents and children and counted

29 the words uttered by each. We then classified the utterances into several categories:

30 question asked by child, repeated test questions by caregiver, hints towards agents (how

31 many times the caregivers guided the child's attention to the agent), hints towards eyes

32 (how many times the caregivers guided the child's attention to the agent's eyes),

33 verification of choice (how many times the caregiver questioned or double checked the

34 child's response), mentioning of screen (how many times the caregiver verbally guided the

35 child's attention to the screen), pointing to screen (how many times the caregiver pointed

36 towards the screen), positive & negative feedback, motivational statements, and

37 incomprehensible utterances. In addition, we coded how many adults and children were

38 present, whether a response click was obviously conducted by the caregiver themselves, and

39 whether children took a break during the trial. We conducted a model comparison to

40 estimate the effects of parental interference. Our null model explained the response

41 behavior by age, while including random effects for subject and target position (model

42 notation in R: `correct ~ age + (1 | subjID) + (1 | targetPosition)`.[1]

43   We compared this null model to models including the number of words uttered by the

---

[1] Attentive readers might notice that we simplified the structure of random effects. Compared to our models in the *Individual differences* section, this model does not include the random slope for symmetric target position within participants. We decided to do so since we had limited amount of data from few participants.

caregiver, number of repeated testquestions, verification of choice, or hints towards eyes as fixed effects. Furthermore, we calculated an parental interference index by summing up number of repeated testquestions, verification of choice, and hints towards eyes, with the sign matching the variable's direction of effect. Remaining variables that we coded for were not included since there was not enough variation and/or occurrences in our sample. We compared models using WAIC (widely applicable information criterion) scores and weights. As an indicator of out-of-sample predictive accuracy, lower WAIC scores stand for a better model fit. WAIC weights represent the probability that the model in question provides the best out-of-sample prediction compared to the other models. On the trial level, the model including the verification of choice as a main effect performed best: here, the less the caregivers asked for children's responses again, the more likely children clicked on the correct box. Interestingly, the effect reversed on a subject level - possibly due to greater learning effects for the children that were most likely to click incorrectly in the beginning and then receiving most parental comments. On the subject level, the model including number of repeated test questions performed best: the more caregivers asked again where the target landed, the more likely children were to respond to the incorrect box. In all cases, however, ELPD difference scores were smaller than their standard errors. Similarly, 95% CI of the model estimates included zero and were rather wide (Table 1). Therefore, we conclude that the effect of parental interference was negligable and could, most likely, be explained as described above.

Table 1

Model comparison for exploratory webcam coding of parental interference

Predictor	WAIC	Weight	ELPD	SE
By trial - # Verification choice	262.98	0.57	0.00	0.00
By trial - Null model	263.25	0.43	-0.13	1.19
By trial - Parental interference index	264.32	0.00	-0.67	0.97
By trial - # Repeated testquestion	265.34	0.00	-1.18	1.26
By trial - # Words uttered by caregiver	266.27	0.00	-1.65	1.21
By trial - # Hints eyes	266.33	0.00	-1.68	1.27
By subject - # Repeated testquestion	83.83	0.70	0.00	0.00
By subject - Parental interference index	88.50	0.00	-2.34	3.56
By subject - Null model	89.62	0.00	-2.90	3.60
By subject - # Verification choice	89.75	0.30	-2.96	4.46
By subject - # Words uttered by caregiver	90.46	0.00	-3.32	4.21
By subject - # Hints eyes	92.90	0.00	-4.54	3.73

Note. All models included random intercepts for participant and target position.

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

77 parameters. This way, we could match participant demographic data to our study data.

78 The same technique was used to confirm study completion: we redirected participants from

79 our study website back to the *Prolific* website using URL parameters. We used *Prolific*'s

80 inbuilt prescreening filter to include only participants who were fluent in English and could

81 therefore properly understand our written and oral study instructions.

Study 1 - Validation hedge version

83 The aim of Study 1 was to validate the hedge version of our gaze understanding task.

84 The pre-registration can be found here: https://osf.io/r3bhn. We recruited participants

85 online by advertising the study on *Prolific*.

86 50 adults participated in the study. One additional subject returned their submission,

87 i.e., decided to leave the study early or withdrew their submission after study completion.

88 Data collection took place in May 2021. Participants were compensated with £1.25 for

89 completing the study. We estimated an average completion time of 6 minutes, resulting in

90 an estimated hourly rate of £10.00. On average, participants took 05:56min to complete

91 the study. Participants were required to complete the study on a tablet or desktop.

92 Participation on mobile devices was disabled since the display would be too small and

93 would harm click precision. It was indicated that the study required audio sound.

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

95 personal information.

Study 2 - Validation box version

97 As in study 1, we recruited participants on *Prolific*, and employed the same

98 methodology. However, this time we focussed on validating the box version of the task in

99 an adult sample. Participants were presented with eight boxes in which the target could

100 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. On average, participants took 04:43min to complete the study.

Study 3 - Reliability hedge version

In stuyy 3 and 4, we assessed the test-retest reliability of our gaze understanding 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

126 £1.25. We estimated an average completion time of 9 minutes, resulting in an estimated

127 hourly rate of £8.33. On average, participants took 07:11min to complete the first part.

128   Of the 60 participants that completed testday 1, 41 subjects finished testday 2. One

129 additional participant timed out, i.e., did not finish the survey within the allowed

130 maximum time. Participants were compensated with £1.50 for completing the second part

131 of the study. We estimated an average completion time of 9 minutes, resulting in an

132 estimated hourly rate of £10. On average, participants took 06:36min to complete the

133 second part of the study.

134   Since we aimed for a minimum sample size of 60 subjects participating on both

135 testdays, we reran the first testday with additional 50 participants. Additional seven

136 subjects returned their submission, i.e., decided to leave the study early or withdrew their

137 submission after study completion. Two additional participants timed out, i.e., did not

138 finish the survey within the allowed maximum time. Again, participants were compensated

139 with £1.25 for completing the first part of the study (estimated average completion time 9

140 minutes, estimated hourly rate of £8.33). On average, participants took 06:51min to

141 complete the first part.

142   Of the additional 50 participants that completed testday 1, 29 subjects finished

143 testday 2. Again, participants were compensated with £1.50 for completing the second

144 part of the study (estimated average completion time 9 minutes, estimated hourly rate of

145 £10). On average, participants took 06:26min to complete the second part of the study.

146 **Study 4 - Reliability box version**

147   As in study 3, we recruited participants on *Prolific*, and employed the same

148 methodology. However, this time participants were presented with the box version of the

149 task. Participants received 32 test trials without voice-over description, so that each of the

150 eight boxes occurred exactly four times. As in study 2, we employed eight boxes in which

151 the target could land.

152   In a first round, 60 participants took part on the first testday. Additional five

153 subjects returned their submission, i.e., decided to leave the study early or withdrew their

154 submission after study completion. For the first testday, participants were compensated

155 with £1.25. We estimated an average completion time of 9 minutes, resulting in an

156 estimated hourly rate of £8.33. On average, participants took 07:33min to complete the

157 first part.

158   Of the 60 participants that completed testday 1, 41 subjects finished testday 2.

159 Participants were compensated with £1.50 for completing the second part of the study. We

160 estimated an average completion time of 9 minutes, resulting in an estimated hourly rate of

161 £10. On average, participants took 07:50min to complete the second part of the study.

162   Since we aimed for a minimum sample size of 60 subjects participating on both

163 testdays, we reran the first testday with additional 50 participants. Additional eight

164 subjects returned their submission, i.e., decided to leave the study early or withdrew their

165 submission after study completion. One additional participant timed out, i.e., did not

166 finish the survey within the allowed maximum time. Again, participants were compensated

167 with £1.25 for completing the first part of the study (estimated average completion time 9

168 minutes, estimated hourly rate of £8.33). On average, participants took 07:37min to

169 complete the first part.

170   Of the additional 50 participants that completed testday 1, 28 subjects finished

171 testday 2. Additional three subjects returned their submission, i.e., decided to leave the

172 study early or withdrew their submission after study completion. One additional

173 participant timed out, i.e., did not finish the survey within the allowed maximum time.

174 Again, participants were compensated with £1.50 for completing the second part of the

175 study (estimated average completion time 9 minutes, estimated hourly rate of £10). On

176 average, participants took 06:30min to complete the second part of the study.

177   **Instructions and voice-over descriptions**

178   This is the content of our audio recordings that were played as instructions and
179   during voice-over trials.

Timeline	German	English	Filename
welcome	Hallo! Schön, dass du da bist. Wir spielen jetzt das Ballon-Spiel! Siehst du die Tiere auf dem Bild da? Wir möchten gleich zusammen mit den Tieren mit einem Ballon spielen. Was genau passiert, erklären wir dir jetzt ganz in Ruhe.	Hello! Great that you're here. We'll now play a balloon game. Can you see the animals in the picture over there? We want to play together with the animals using the balloon. We'll now talk you through exactly what will happen.	welcome.mp3

touch	Schau mal, da steht ein Tier im Fenster. Und siehst du den Ballon da? Der Ballon fällt immer runter und landet auf dem Boden. Und du musst ihn dann finden. Das Tier hilft Dir und schaut immer den Ballon an.	Look, an animal is standing in the window. And can you see the balloon over there? The balloon always falls down and lands on the ground. And you have to find it! The animal helps you and always looks at the balloon.	touch-1.mp3
	Wo ist der Ballon? Drück auf den Ballon!	Where is the balloon? Click on the balloon!	prompt-touch-long.mp3

fam - HEDGE	Klasse, das war super! Jetzt spielen wir weiter. Siehst du wieder das Tier und den Ballon da? Der Ballon fällt wieder runter. Diesmal fällt er hinter eine Hecke. Du musst ihn wieder finden. Das Tier hilft dir und schaut immer den Ballon an.	Perfect, that was great! Now, we'll continue playing. Can you see the animal and the balloon again? The balloon will fall down again. This time, it will fall behind a hedge. And you have to find it! The animal helps you and looks at the balloon.	fam-hedge-1.mp3
	Wo ist der Ballon? Drücke auf die Hecke - wo der Ballon ist.	Where is the balloon? On the hedge, click where the balloon is.	prompt-hedge-long.mp3

fam - BOX	Klasse, das war super! Jetzt spielen wir weiter. Siehst du wieder das Tier und den Ballon da? Der Ballon fällt wieder runter. Diesmal fällt er in eine Kiste. Du musst ihn wieder finden. Das Tier hilft dir und schaut immer den Ballon an.	Perfect, that was great! Now, we'll continue playing. Can you see the animal and the balloon again? The balloon falls down again. This time, it falls into a box. And you have to find it! The animal helps you and looks at the balloon.	fam-box-1.mp3
	Wo ist der Ballon? Drücke auf die Kiste mit dem Ballon.	Where is the balloon? Click on the box with the balloon.	prompt-box-long.mp3
test - HEDGE	Klasse , das hast du toll gemacht! Nun spielen wir weiter. Da sind wieder der Ballon, das Tier und die Hecke. Die Hecke wächst jetzt hoch.	Nice, good job! Now, we'll continue playing. There is the balloon, the animal and the hedge. The hedge is growing a bit now.	test-hedge-1.mp3

	Der Ballon ist nun hinter der Hecke. Du kannst das nicht sehen - das Tier aber! Jetzt fällt der Ballon auf den Boden und du musst ihn wieder finden. Denk dran - das Tier schaut immer den Ballon an.	The balloon is behind the hedge now. You can't see it - but the animal can! The balloon falls to the ground and you have to find it. Remember - the animal always looks at the balloon!	test-hedge-2.mp3
	Dann schrumpft die Hecke. Drücke auf die Hecke - wo der Ballon ist.	Now, the hedge is shrinking. On the hedge, click where the balloon is.	test-hedge-3.mp3
test - BOX	Klasse , das hast du toll gemacht! Nun spielen wir weiter. Da sind wieder der Ballon, das Tier und die Kisten. Jetzt wächst eine Hecke hoch.	Nice, good job! Now, we'll continue playing. There is the balloon and the animal. Now, a hedge is growing.	test-box-1.mp3

	Der Ballon ist nun hinter der Hecke. Du kannst das nicht sehen - das Tier aber! Jetzt fällt der Ballon in eine Kiste und du musst ihn wieder finden. Denk dran - das Tier schaut immer den Ballon an.	The balloon is behind the hedge now. You can't see it - but the animal can! The balloon falls into a box and you have to find it. Remember - the animal always looks at the balloon!	test-box-2.mp3
	Dann schrumpft die Hecke. Drücke auf die Kiste mit dem Ballon.	Now, the hedge is shrinking. Click on the box with the balloon.	test-box-3.mp3
goodbye	Geschafft! Die Tiere sind schon ganz glücklich vom Spielen! Vielen Dank für deine Hilfe! Bis zum nächsten Mal und liebe Grüße vom Schwein, Affen und Schaf	The animals are super happy after playing. Thanks a lot for your help! See you soon and goodbye from the pig, monkey and sheep	goodbye.mp3
general prompt	Wo ist der Ballon?	Where is the balloon?	prompt-general.mp3

touch - no response	Drück auf den Ballon!	Click on the balloon!	prompt-touch.mp3
hedge - no response	Drücke auf die Hecke - wo der Ballon ist!	On the hedge, click where the balloon is!	prompt-hedge.mp3
box - no response	Drücke auf die Kiste mit dem Ballon!	Click on the box with the balloon!	prompt-box.mp3
landing sound of balloon	-	-	balloon-lands.mp3
sound of blinking eyes	-	-	blink.mp3
sound for target click	-	-	positive-feedback.mp3