

¹ Individual differences in great ape cognition across time and domains: stability, structure,
² and predictability

³ Manuel Bohn^{1,2}, Christoph Völter², Johanna Eckert², Daniel Hanus², Nico Eisbrenner², Jana
⁴ Holtmann³, & Daniel Haun²

⁵ ¹ Institute of Psychology in Education, Leuphana University Lüneburg

⁶ ² Department of Comparative Cultural Psychology, Max Planck Institute for Evolutionary
⁷ Anthropology, Leipzig, Germany

⁸ ³ Wilhelm Wundt Institute of Psychology, Leipzig University, Leipzig, Germany

¹⁰ Manuel Bohn was supported by a Jacobs Foundation Research Fellowship
¹¹ (2022-1484-00). We are grateful to thank all children and caregivers for participating in the
¹² study. We thank the Max Planck Society for the Advancement of Science.

¹³ The authors made the following contributions. Manuel Bohn: Conceptualization,
¹⁴ Formal Analysis, Writing - Original Draft Preparation, Writing - Review & Editing;
¹⁵ Christoph Völter: Conceptualization, Writing - Original Draft Preparation, Writing - Review
¹⁶ & Editing; Johanna Eckert: Conceptualization, Writing - Original Draft Preparation,
¹⁷ Writing - Review & Editing; Daniel Hanus: Conceptualization, Writing - Original Draft
¹⁸ Preparation, Writing - Review & Editing; Nico Eisbrenner: Formal Analysis, Writing -
¹⁹ Original Draft Preparation, Writing - Review & Editing; Jana Holtmann: Formal Analysis,
²⁰ Writing - Original Draft Preparation, Writing - Review & Editing; Daniel Haun:
²¹ Conceptualization, Writing - Review & Editing.

²² Correspondence concerning this article should be addressed to Manuel Bohn,
²³ Universitätsallee 1, 21335 Lüneburg, Germany. E-mail: manuel.bohn@leuphana.de

²⁴

Abstract

²⁵ 200 words

²⁶ *Keywords:* keywords

27 Individual differences in great ape cognition across time and domains: stability, structure,
28 and predictability

29 **Introduction**

30 Bohn et al. (2023) studied

31 The current study extended previous work in two important aspects. First, we study a
32 broader range of cognitive domains including social cognition, reasoning about quantities,
33 executive functions and inferential reasoning. This allows us to assess whether the results
34 obtained by Bohn et al. (2023) replicate within domains and generalize to others. Second,
35 we explored the structure of great ape cognition in more depth: we pooled the data collected
36 here with the data from Bohn et al. (2023) to study the correlations between cognitive traits
37 within and across domains.

38 **Methods**

39 **Participants**

40 A total of 48 great apes participated at least in one tasks at one time point. This
41 included 12 Bonobos (*pan paniscus*, 4 females, age 3.60 to 40.70), 24 Chimpanzees (*pan*
42 *troglodytes*, 17 females, age 3.80 to 57.80), 6 Gorillas (*gorilla gorilla*, 4 females, age 4.40 to
43 24.40), and 6 Orangutans (*pongo abelii*, 5 females, age 4.70 to 43.10). The sample size at the
44 different time points ranged from 34 to 45 for the different species (see supplementary
45 material for details). All apes participated in cognitive research on a regular basis. Apes
46 were housed at the Wolfgang Köhler Primate Research Center located in Zoo Leipzig,
47 Germany. They lived in groups, with one group per species and two chimpanzee groups
48 (group A and B). Research was noninvasive and strictly adhered to the legal requirements in
49 Germany. Animal husbandry and research complied with the European Association of Zoos
50 and Aquaria Minimum Standards for the Accommodation and Care of Animals in Zoos and
51 Aquaria as well as the World Association of Zoos and Aquariums Ethical Guidelines for the

52 Conduct of Research on Animals by Zoos and Aquariums. Participation was voluntary, all
 53 food was given in addition to the daily diet, and water was available ad libitum throughout
 54 the study. The study was approved by an internal ethics committee at the Max Planck
 55 Institute for Evolutionary Anthropology.

56 **Procedure**

57 Apes were tested in familiar sleeping or observation rooms by a single experimenter.
 58 The basic setup comprised a sliding table positioned in front of a mesh or a clear plexiglas
 59 panel. The experimenter sat on a small stool and used an occluder to cover the table (see
 60 Figure 1).

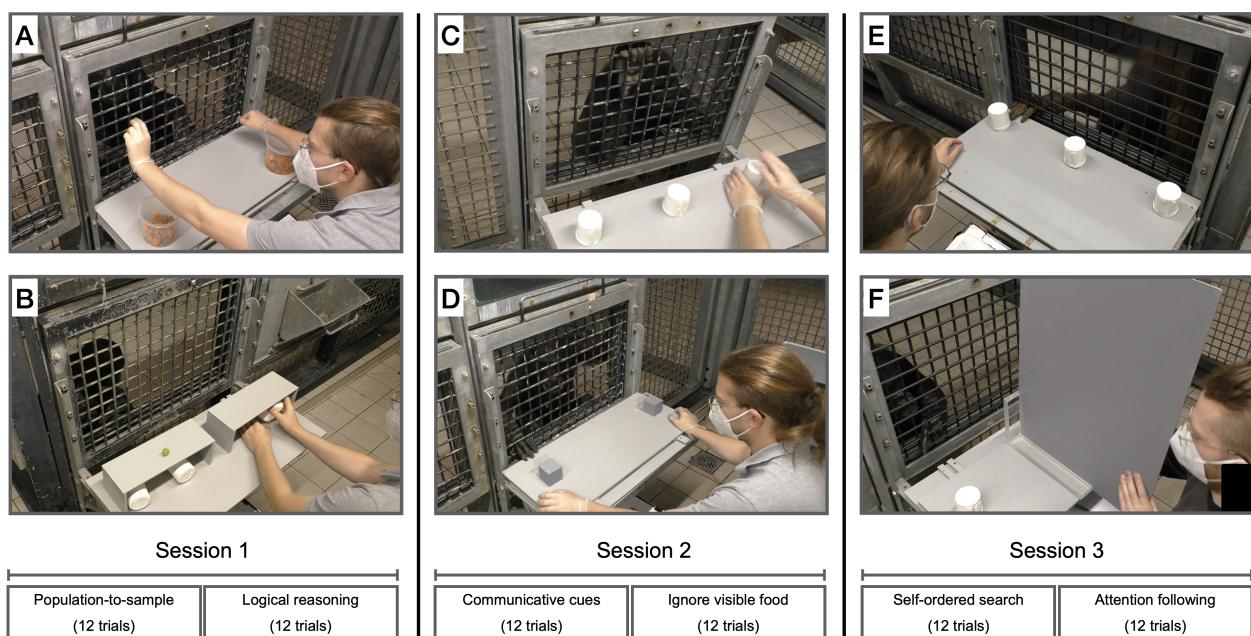


Figure 1. Setup used for the six tasks. A) population-to-sample, B) logical-reasoning, C) communicative-cues, D) ignore-visible-food, E) self-ordered-search and F) attention-following. Text at the bottom shows order of task presentation and trial numbers

61 The study involved a total of six cognitive tasks. These were based on published
 62 procedures in the field of comparative psychology. The original publications often include
 63 control conditions to rule out alternative, cognitively less demanding ways to solve the tasks.

64 We did not include such controls here and only ran the experimental conditions. For each
65 task, we refer to these papers if they want to know more about control conditions and/or a
66 detailed discussion of the nature of the underlying cognitive mechanisms. Example videos for
67 each task can be found in the associated online repository. In the following, we give a brief
68 description of each task. Additional details can be found in the supplementary material.

69 **Attention-following.** The Attention-following task was loosely modeled after
70 Kaminski, Call, and Tomasello (2004). The setup consisted of two identical cups placed on
71 the sliding table and a large opaque screen that was longer than the width of the sliding
72 table (Supplementary Figure 1F). The experimenter placed both cups on the table and
73 showed the ape that they were empty. Then, the experimenter baited both cups in view of
74 the ape and placed the opaque screen in the center between the two cups, perpendicular to
75 the mesh. Next, the experimenter moved to one side and looked at the cup in front of them.
76 Then, the experimenter pushed the sliding table forward and the ape was allowed to choose
77 one of the cups by pointing at it. If the ape chose the cup that the experimenter was looking
78 at, they received the food item. If they choose the other cup, they did not. We coded
79 whether the ape chose the side the experimenter was looking at (correct choice) or not. Apes
80 received twelve trials. The side at which the experimenter looked was counterbalanced with
81 same number of looks to each side and looks to the same side not more than two times in a
82 row. We assumed that apes follow the experimenters focus of attention to determine whether
83 or not their request could be seen and thus be successful.

84 **Communicative-cues.** This task was modeled after Schmid, Karg, Perner, and
85 Tomasello (2017). Three identical cups were placed equidistantly on a sliding table directly
86 in front of the ape (Figure 1C). In the beginning of a trial, the experimenter showed the ape
87 that all cups are empty. After placing an occluder between the subject and the cups, the
88 experimenter held up one food item and moved it behind the occluder, visiting all three cups
89 but baiting only one. Next, the occluder was lifted and E looked at the ape (ostensive cue),
90 called the name, and looked at one of the cups, while holding on to it with one hand and

91 tapping it with the other (continuous looking, 3 times tapping). Finally, the experimenter
92 pushed the sliding table forward for the ape to make a choice. If the ape chose the baited
93 cup, they received the reward – if not, not. We coded as correct choice if the ape chose the
94 indicated cup. Apes received twelve trials. The location of the indicated cup was
95 counterbalanced, with each cup being the target equally often and the same target not more
96 than two times in a row. We assumed that apes use the experimenter’s communicative cues
97 to determine where the food is hidden.

98 **Ignore-visible-food.** The task was modeled after Völter, Tinklenberg, Call, and
99 Seed (2022). The task involved two opaque cups with an additional, sealed but transparent,
100 compartment attached to the front of each cup (facing the ape). For one cup, the
101 compartment contained a preferred food item that was clearly visible, for the other cup, the
102 compartment was empty (Figure 1D). In the beginning of the trial, the two cups were placed
103 upside down on the sliding table so that the ape could see that the opaque compartments of
104 both cups were empty. Next, the experimenter baited one of the cups in full view of the
105 subject. In non-conflict trials, the baited cup was the cup with the food item in the
106 transparent compartment. In conflict trials, the baited cup was the cup with the empty
107 compartment. After baiting the experimenter pushed the sliding table forwards and the ape
108 could chose by pointing. If the baited cup was chosen, the ape received the food. Apes
109 received 14 trials, twelve conflict trials and two non-conflict trials (1st and 8th trial). Only
110 conflict trials were analyzed. The location of the cup with the baited compartment was
111 counterbalanced, with the cup not being in the same location more than twice in a row. We
112 assumed that apes need to inhibit selecting the visible food item and instead use their
113 short-term memory to remember where the food was hidden.

114 **Logical-reasoning.** The task was modeled after Hanus and Call (2014). Three
115 identical cups were presented side-by-side on a sliding table, with the cup in the middle
116 sometimes positioned closer to the left cup and sometimes closer to the right.
117 (Supplementary Figure 1B). Two half-open boxes served as occluders to block the ape’s view

when shuffling the cups. Each trial started by showing the ape that all three cups (one on one side of the table, two on the other) were empty. After placing the occluders over both sides of the table, the experimenter put one piece of food on top of each occluder. Next, the experimenter hid each piece of food under the cup(s) behind the occluders. In case of the occluder with the two cups, the food was randomly placed under one of the two cups while both cups were visited and even shuffled. Finally, both occluders were lifted and the table pushed forwards, allowing the ape to choose one of the three cups, from which they then received the content. We coded whether the ape chose the certain cup (i.e. the cup from the side of the table with only one cup). Apes received 12 trials. The side with one cup was counterbalanced, with the same constellation appearing not more than two times in a row on the same side. We assumed that apes would infer that the cup from the tray with only one cup certainly contains food while the other cups contain food only in 50% of cases.

Population-to-sample. The task was modeled after Eckert, Call, Hermes, Herrmann, and Rakoczy (2018). During the test, apes saw two transparent buckets filled with pellets and carrot pieces (the carrot pieces had roughly the same size and shape as the pellets). Each bucket contained 80 food items. The distribution of pellets to carrot pieces was 4:1 in bucket A, and 1:4 in bucket B. Pellets are preferred food items in comparison to carrots. The experimenter placed both buckets on a table, one left, one right (Figure 1A). In the beginning of a trial, the experimenter picked up the bucket on the right side, tilted it forward so the ape could see inside, placed it back on the table and turned it around 360°. The same procedure was repeated with the other bucket. Next, the experimenter looked at the ceiling, inserted each hand in the bucket in front of it and drew one item from the bucket without the ape seeing which type (E picked always of the majority type). The food items remained hidden in the experimenter's fists. Next, the experimenter extended the arms (in parallel) towards the ape who was then allowed to make a choice by pointing to one of the fists. The ape received the chosen sample. In half of the trials, the experimenter crossed arms when moving the fists towards the ape to ensure that the apes made a choice between

samples and not just chose the side where the favorable population was still visible. In between trials, the buckets were refilled to restore the original distributions. Apes received twelve trials. We coded whether the ape chose the sample from the population with the higher number of high quality food items. The location of the buckets (left and right) was counterbalanced, with the buckets in the same location no more than two times in a row. The crossing of the hands was also counterbalanced with no more than two crossings in a row. We assumed that apes reasoned about the probability of the sample being a high quality item based on observing the ratio in the population.

Self-ordered-search. The task was modeled after Völter, Mundry, Call, and Seed (2019; Diamond, Prevor, Callender, and Druin, 1997; see also Petrides, 1995). Three identical cups were placed equidistantly on a sliding table directly in front of the ape (Supplementary Figure 1E). The experimenter baited all three cups in full view of the ape. Next, the experimenter pushed the sliding table forwards for the ape to choose one of the cups by pointing. After the choice, the table was pulled back and the ape received the food. After a 3s pause, the table was pushed forward again for a second choice. This procedure was repeated for a third choice. If the ape chose a baited cup, they received the food, if not, not. We coded the number of times the ape chose an empty cup (i.e. chose a cup they already chose before). Please note that this outcome variable differed from the other tasks in two ways: first, possible values were 0, 1, and 2 (instead of just 0 and 1) and second, a lower score indicated better performance. Apes received twelve trials. No counterbalancing was needed. We assumed that apes use their working memory abilities to remember where they had already searched and which cups still contained food.

Predictor variables. In addition to the data from the cognitive tasks, we collected data for a range of predictor variables to predict individual differences in performance in the cognitive tasks. Predictors could either vary with the individual (stable individual characteristics: group, age sex, rearing history, and time spent in research), vary with individual and time point (variable individual characteristics: rank, sickness, and sociality),

172 vary with group membership (group life: time spent outdoors, disturbances, and life events),
173 or vary with the testing arrangements and thus with individual, time point and session
174 (testing arrangements: presence of an observer, participation in other studies on the same
175 day or since the last time point). Predictors were collected from the zoo handbook with
176 demographic information about the apes, via a diary that the animal caretakers filled out on
177 a daily basis, or via proximity scans of the whole group. We provide a detailed description of
178 these variables in the supplementary material.

179 Data collection

180 Data collection started on April 28th, 2022, lasted until October 7th, 2023 and included
181 10 time points. One time point meant running all tasks with all participants. Within each
182 time point, the tasks were organized in three sessions (see Fig. 1). Session 1 included the
183 population-to-sample and logical-reasoning tasks, session 2 the communicative-cues and
184 ignore-visible-food tasks and session 3 the self-ordered-search and attention-following tasks.

185 The interval between two time points was planned to be eight weeks. However, it was
186 not always possible to follow this schedule so that some intervals were longer or shorter (see
187 supplementary material for details). The order of tasks was the same for all subjects. So was
188 the counterbalancing within each task. This exact procedure was repeated at each time point
189 so that the results would be comparable across participants and time points.

190 Analysis, results and discussion

191 To get an overview of the results, we first visualized the data (Fig. 2). Performance
192 was consistently above chance in the communicative-cues, ignore-visible-food and
193 population-to-sample tasks. For attention-following, this was the case only from time point 7
194 onward and for logical-reasoning, performance was, if anything, below chance. For the
195 self-ordered-search task, performance was below chance but here lower values reflect better
196 performance (i.e. systematic avoidance of the visible food item). For attention-following,

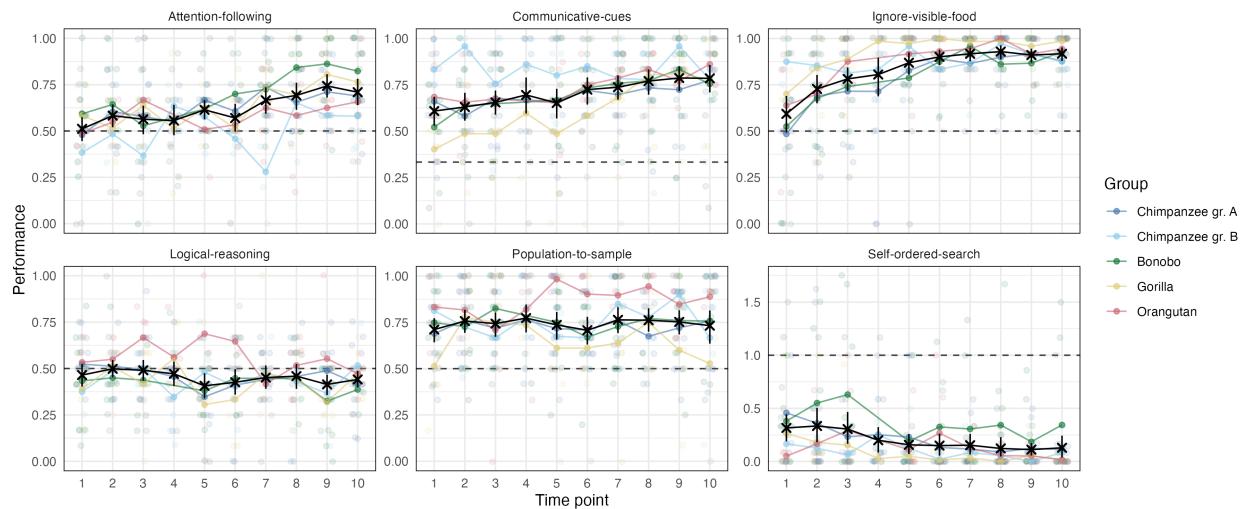


Figure 2. Results from the six cognitive tasks across time points. Black crosses show mean performance at each time point across species (with 95% CI). The sample size varied between time points and can be found in Supplementary Figure 1. Colored dots show mean performance by species. Dashed line shows chance level performance.

197 ignore-visible-food, communicative-cues and self-ordered-search there was a steady
198 improvement in performance over time.

199 In the following, we link performance in the tasks across time points to latent variables
200 representing cognitive abilities. We first ask how stable these abilities are over time and how
201 reliably they are measured. Next, we study the correlations between different abilities to
202 explore the internal structure of great ape cognition. Finally, we link performance in the
203 tasks to external predictors to shed light on the sources of individual differences in abilities.
204 Each section uses different statistical techniques which we describe in the respective section.

205 Stability and reliability

206 ... describe models.

207 The six tasks differed substantially in what they revealed about group- and
208 individual-level variation. What stands out is the widespread change in performance over

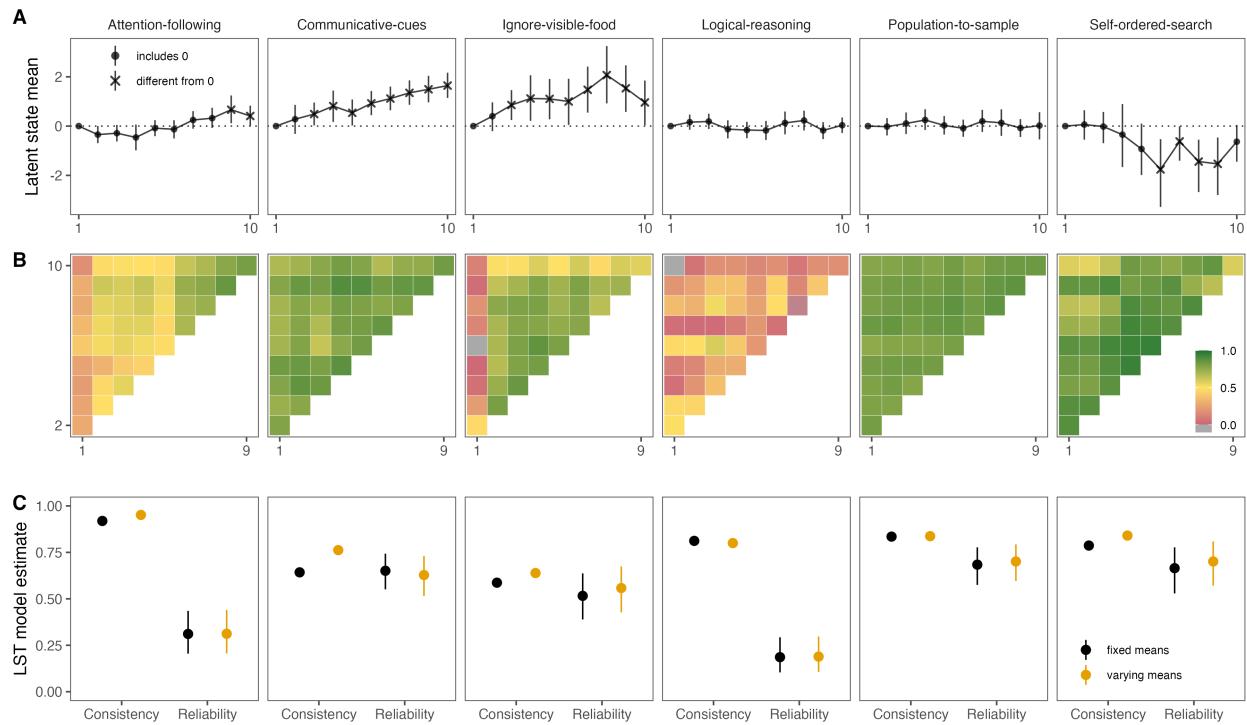


Figure 3. A) Latent mean estimates for each time point by task based on latent state model. Means at time point 1 are set to zero. Shape denotes whether the 95% CrI included zero (dashed line). The sample size varied between time points and can be found in Supplementary Fig. 1. B) Correlations between subject-level latent state estimats for the different time points by task. C) Mean estimates from latent state-trait models with fixed and varyin means (color codeed) with 95% CrI. Consistency refers to the proportion of (measurement-error-free) variance in performance explained by stable trait differences. Reliability refers to the proportion of true score variance to variance in raw scores.

time. For all tasks except population-to-sample and logical-reasoning we observed an improvement in performance over time. This group-level change, however, has different individual-level interpretations for the different tasks. For communicative-cues, ignore-visible-food and self-ordered-search, individual differences remained relatively stable despite the group-level change suggesting stable individual differences combined with a systematic learning effect across individuals. In contrast, for attention-following, there was little stability in individual differences at earlier time points and only towards the end emerged a more stable ordering of individuals. In combination with the low reliability at earlier time points, this suggests that at least some individuals changed their response strategy in the course of the study. The combination of low reliability, chance-level performance and low correlation of latent states for logical-reasoning suggests that this task is not suited to probe individual differences in logical reasoning abilities in great apes. It is also noteworthy that the reliability estimates are on average lower compared to a previous study testing the same individuals on different tasks (Bohn et al., 2023). One explanation might be the increase in performance over time. At the beginning of the study, more individuals might have chosen randomly instead of using the available information provided in the task setup and the demonstrations. By definition, random variation is not reliable. With time, more and more individuals started using the available information so that inter-individual differences in how good they are in using it could be detected.

228 Structure

229 To explore the structure of great ape cognition we correlated latent trait estimates for
230 each task. In contrast to raw performance scores, these estimates take into account the
231 reliability of measurement and are considered to be measurement-error free. Bohn et al.
232 (2023) tested the same individuals and we therefore also include the data from tasks reported
233 there (data from phase 2). Even though the data in the two studies was collected at different
234 time points, we think it is justifiable to analyse them jointly because the trait estimates

represent stable, time-invariant individual differences in cognitive abilities. The estimates were computed ...

Figure4 shows the correlations between trait estimates for the different tasks. Overall, most correlations were not significantly different from zero (i.e. the 95% CI did include zero). Because of this low average level of correlations, we decided not to explore models with higher-order factors and will only interpret the qualitative patterns.

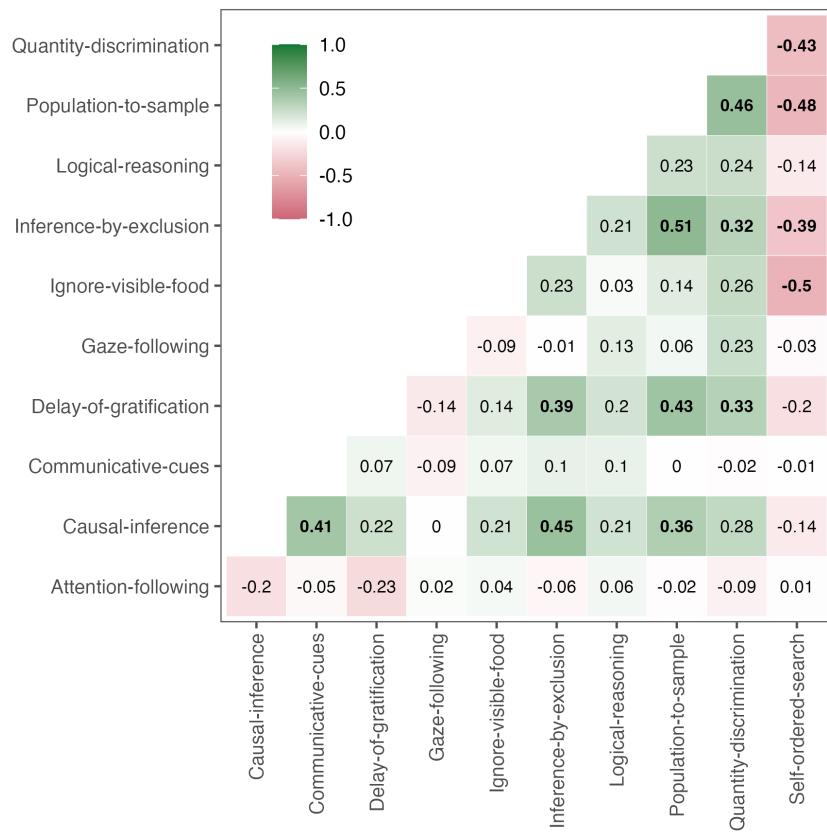


Figure 4. Correlations between ... trait estimates. Bold correlations have 95% CrI non overlapping with zero.

Conceptually, the tasks can be clustered in the following broader domains: *social cognition* (attention-following, gaze-following, communicative-cues), *reasoning about quantities* (quantity-discrimination, population-to-sample), *executive functions* (delay-of-gratification, self-ordered-search, ignore-visible-food) and *inferential reasoning*

245 (logical-reasoning, causal-inference, inference-by-exclusion). As a first step, we will evaluate
246 whether we find evidence for such a clustering in the data.

247 There was no significant correlation between any of the social cognition tasks.

248 Furthermore, attention-following and gaze-following did not correlate significantly with any
249 of the other tasks and communicative-cues correlated only with causal-inference – a result we
250 will discuss below. Thus, and in line with previous work (Herrmann, Hernández-Lloreda,
251 Call, Hare, & Tomasello, 2010), we found no evidence for shared cognitive processes in tasks
252 measuring different aspects of social cognition.

253 The two tasks measuring reasoning about quantities did correlate significantly. Both
254 tasks require discriminating between different quantities, directly in the case of
255 quantity-discrimination and as part of the decision making process in the case of
256 population-to-sample. Deciding between the samples from the two populations requires
257 discriminating between the relative quantities within each bucket from which the samples
258 were drawn.

259 Within the executive functions measures, self-ordered-search and inhibit-visible-food
260 were significantly correlated but none of the two correlated with delay-of-gratification. The
261 significant correlation can be explained by the need to inhibit a premature response
262 (selecting visible food or a cup that was previously rewarded) in both tasks. It has been
263 argued that delay-of-gratification requires self control (tolerating a longer waiting time to
264 gain a more valuable reward) over and above behavioral inhibition (Beran, 2015). From this
265 point of view, individual differences in the delay-of-gratification task might be due to
266 differences in self control and less due to differences in inhibition.

267 Finally, for the three inferential reasoning measures we found a correlation between
268 inference-by-exclusion and causal-inference. Logical-reasoning did not correlate with either
269 (neither did it with any other task). This is not surprising given the results reported above:

270 the observed variation in the logical-reasoning task was largely noise and did not reflect
271 systematic individual differences. The correlation between causal-inference and
272 inference-by-exclusion is most likely due to the fact that both tasks involve making
273 inferences about the location of food based on reasoning about its physical properties.

274 Next we turn to the correlations across domains. Perhaps the most surprising finding is
275 the correlation between causal-inference and communicative-cues. On a closer look, the
276 origin might be task impurity in that there are two ways to solve the causal-inference task:
277 first, as hypothesized, by using the rattling sound to infer the location of the food. Second,
278 by interpreting the experimenter's shaking of the cup as a communicative cue, which is very
279 similar to the communicative-cues task. Thus, we suspect that at least some individuals
280 solved the task via the second route.

281 Finally, there was a cluster of significant correlations between delay-of-gratification,
282 self-ordered-search, inference-by-exclusion, causal-inference, population-to-sample and
283 quantity discrimination. Of the 15 possible correlations, only four were non-significant. One
284 commonality between these tasks that might – in part – explain this pattern is that they all
285 benefit from sustained attention to the task. Sustained attention facilitates the processing of
286 the experimenter's demonstrations (population-to-sample, inference-by-exclusion,
287 causal-inference, delay-of-gratification), ones one actions on the setup (self-ordered-search) or
288 visually complex stimuli (quantity discrimination). Tentative support for this idea comes
289 from the analysis of relevant predictors (see Bohn et al., 2023 and below) in which **time**
290 **spent in research** was selected as a relevant predictor of performance for all of these tasks
291 except causal-inference. This predictor reflects individual's experience with experimental
292 studies, which often involve sustained attention to distributions of food items, actions of
293 conspecifics and/or demonstrations by experimenters. Next, we turn to the sources of the
294 individual differences analysed here.

295 Predictability

296 In this section, we analysed which external variables accounted for inter- and
297 intra-individual differences in task performance. That is, we asked which of the predictor
298 variables described above predicted performance in the different tasks. Given the large
299 number of predictor variables (14), this question translates to a variable selection problem:
300 selecting a subset of variables from a larger pool. We used the projection predictive inference
301 (Piironen, Paasiniemi, & Vehtari, 2020) approach because it is a state-of-the-art procedure
302 that provides an excellent trade-off between model complexity and accuracy (Pavone,
303 Piironen, Bürkner, & Vehtari, 2020; Piironen & Vehtari, 2017). The projection prediction
304 approach is a two-step process: The first step consists of building the best predictive model
305 possible, called the reference model. In our case, the reference model is a Bayesian multilevel
306 regression model – fit via `brms` (Bürkner, 2017) – including all available predictors (Catalina,
307 Bürkner, & Vehtari, 2020). In the second step, the goal is to replace the posterior
308 distribution of the reference model with a simpler distribution containing fewer predictors
309 compared to the reference model. The importance of a predictor is assessed by inspecting
310 the mean log-predictive density (`elpd`) and root-mean-squared error (`rmse`) of models
311 containing the predictor vs. not.

312 The output of the procedure is a ranking of the different predictors. That is, for each
313 task, we get a ranking of how important a predictor is for constructing the simpler
314 replacement distribution. In addition, we can make a qualitative assessment of whether or
315 not a predictor is relevant or not. In addition to the global assessment, we also inspected the
316 projected posterior distribution of the predictors classified as relevant to see how they
317 influenced performance. In the supplementary material we provide a detailed description of
318 the procedure including how the different variables were handled and how the importance of
319 each predictor was assessed.

320 In addition to the external predictors, the models also included a random intercept

321 term for subject ((1 | subject) in brms notation). This predictor was handled in a special
 322 way in that it was always considered last because it would otherwise have soaked up most of
 323 the variance before the other predictors would have had a chance to explain any of it.

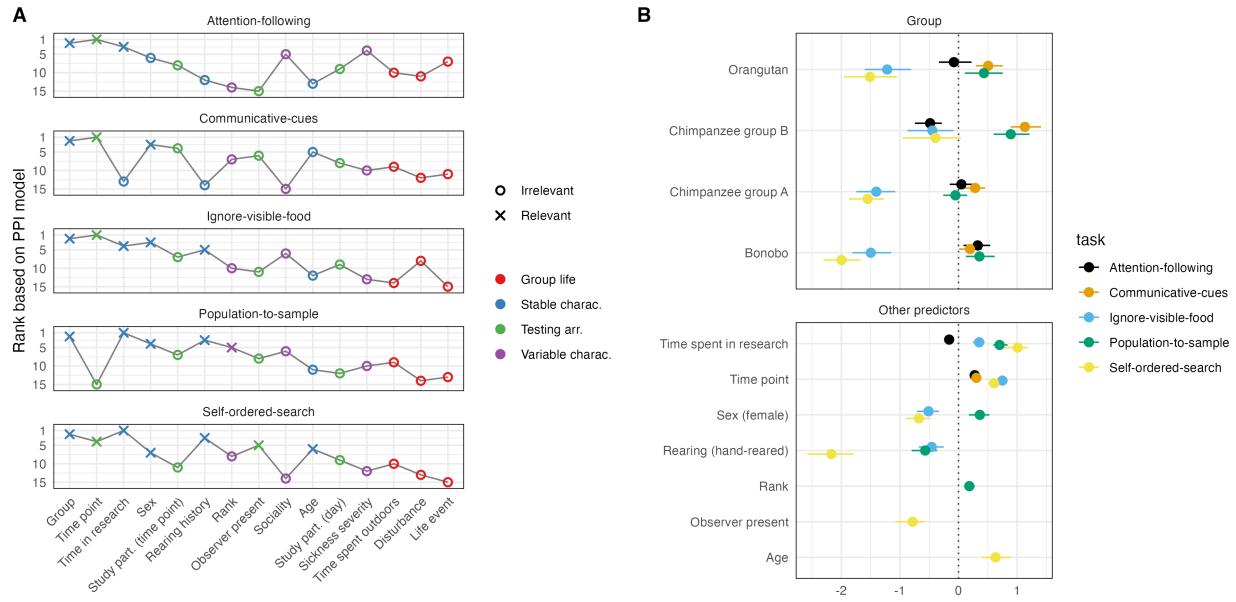


Figure 5. A) Predictor ranking and selection based on PPI models. Crosses mark predictors that were selected to be relevant based on the PPI models. Color shows the broader category each predictor belongs to. The x-axis is sorted by the average rank across tasks. B) Posterior model estimates for the selected predictors for each task based on data. Points show means with 95% Credible Interval. Color denotes task. For categorical predictors, the estimate gives the difference compared to the reference level (Gorilla for group).

324 Fig. 5A summarizes the selected predictors across tasks. For all tasks, the random
 325 intercept term improved model fit the most (not shown in Fig. 5A). In line with results
 326 reported by Bohn et al. (2023), this suggests that idiosyncratic developmental processes or
 327 genetic pre-dispositions, which operate on time-scales longer than what we captured in our
 328 study, accounted for a substantial portion of the variance in cognitive abilities between
 329 individuals.

330 However, for two tasks, other predictors had an comparable explanatory power –

331 something that was not observe in Bohn et al. (2023). For population-to-sample, **time**
332 **spent in research** improved the model fit even more than adding the random intercept at
333 the end did. This could be interpreted that performance in this task strongly depends on
334 having learned to pay attention to stimuli and the human experimenter. For
335 ignore-visible-food, **time point** had an influence exceeding that of the random intercept
336 term. We think this result reflects the strong within-task learning effect across subjects.
337 Because performance increased substantially with time, most of the variation captured by
338 **time point** exceeded the variation between individuals.

339 For the remaining predictors, the most highly-ranked and frequently selected ones
340 came from the group of stable individual characteristics. The big exception being **time**
341 **point**, which was ranked second across tasks. This pattern aligns with the SEM results, in
342 which we saw that most of the variance in performance could be traced back to stable trait
343 differences between individuals. The remaining occasion specific variation was largely due to
344 improvement over time, most likely reflecting task-specific learning processes. The remaining
345 time-varying predictors did not account for much variation.

346 The predictor selected most often was **group**. It was the only predictor that was
347 selected as relevant for all tasks. However, differences between groups were variable in that
348 the ranking of the groups changed from task to task (Fig. 5B). For example, Gorillas
349 performed best in ignore-visible-food and self-ordered-search, the Chimpanzee group B
350 performed best in communicative-cues and population-to-sample and the Bonobos performed
351 best in attention-following. This speaks against clear species or group differences in general
352 cognitive performance. Again, the most likely explanation for group differences is an
353 interaction between species specific dispositions and individual- / task-level developmental
354 processes.

355 The predictors that were selected more than once influenced performance in variable
356 ways (Fig. 5B). As mentioned above, **time point** always had a positive effect because

357 performance increased with time. Whenever **rearing** was selected to be relevant,
358 mother-reared individuals outperformed others. **Time spent in research** had a positive
359 effect, suggesting that more experience with research leads to better performance, except for
360 attention-following. The effect of **sex** was variable in that females outperformed males in
361 population-to-sample but males outperformed females in self-ordered-search and
362 ignore-visible-food.

363 **General Discussion**

364 **Conclusion**

365

References

- 366 Beran, M. J. (2015). The comparative science of “self-control”: What are we talking about?
- 367 *Frontiers in Psychology*, 6, 51.
- 368 Bohn, M., Eckert, J., Hanus, D., Lugauer, B., Holtmann, J., & Haun, D. B. (2023). Great
- 369 ape cognition is structured by stable cognitive abilities and predicted by developmental
- 370 conditions. *Nature Ecology & Evolution*, 7(6), 927–938.
- 371 Bürkner, P.-C. (2017). brms: An R package for Bayesian multilevel models using Stan.
- 372 *Journal of Statistical Software*, 80(1), 1–28.
- 373 Catalina, A., Bürkner, P.-C., & Vehtari, A. (2020). Projection predictive inference for
- 374 generalized linear and additive multilevel models. *arXiv Preprint arXiv:2010.06994*.
- 375 Diamond, A., Prevor, M. B., Callender, G., & Druin, D. P. (1997). Prefrontal cortex
- 376 cognitive deficits in children treated early and continuously for PKU. *Monographs of the*
- 377 *Society for Research in Child Development*, i–206.
- 378 Eckert, J., Call, J., Hermes, J., Herrmann, E., & Rakoczy, H. (2018). Intuitive statistical
- 379 inferences in chimpanzees and humans follow weber’s law. *Cognition*, 180, 99–107.
- 380 Hanus, D., & Call, J. (2014). When maths trumps logic: Probabilistic judgements in
- 381 chimpanzees. *Biology Letters*, 10(12), 20140892.
- 382 Herrmann, E., Hernández-Lloreda, M. V., Call, J., Hare, B., & Tomasello, M. (2010). The
- 383 structure of individual differences in the cognitive abilities of children and chimpanzees.
- 384 *Psychological Science*, 21(1), 102–110.
- 385 Kaminski, J., Call, J., & Tomasello, M. (2004). Body orientation and face orientation: Two
- 386 factors controlling apes’ begging behavior from humans. *Animal Cognition*, 7, 216–223.
- 387 Pavone, F., Piironen, J., Bürkner, P.-C., & Vehtari, A. (2020). *Using reference models in*
- 388 *variable selection*. Retrieved from <https://arxiv.org/abs/2004.13118>
- 389 Petrides, M. (1995). Impairments on nonspatial self-ordered and externally ordered working
- 390 memory tasks after lesions of the mid-dorsal part of the lateral frontal cortex in the
- 391 monkey. *Journal of Neuroscience*, 15(1), 359–375.

- 392 Piironen, J., Paasiniemi, M., & Vehtari, A. (2020). Projective inference in high-dimensional
393 problems: Prediction and feature selection. *Electronic Journal of Statistics*, 14(1),
394 2155–2197. <https://doi.org/10.1214/20-EJS1711>
- 395 Piironen, J., & Vehtari, A. (2017). Comparison of bayesian predictive methods for model
396 selection. *Statistics and Computing*, 27, 711–735.
397 <https://doi.org/10.1007/s11222-016-9649-y>
- 398 Schmid, B., Karg, K., Perner, J., & Tomasello, M. (2017). Great apes are sensitive to prior
399 reliability of an informant in a gaze following task. *PLoS One*, 12(11), e0187451.
- 400 Völter, C. J., Mundry, R., Call, J., & Seed, A. M. (2019). Chimpanzees flexibly update
401 working memory contents and show susceptibility to distraction in the self-ordered search
402 task. *Proceedings of the Royal Society B*, 286(1907), 20190715.
- 403 Völter, C. J., Tinklenberg, B., Call, J., & Seed, A. M. (2022). Inhibitory control and cue
404 relevance modulate chimpanzees'(*pan troglodytes*) performance in a spatial foraging task.
405 *Journal of Comparative Psychology*, 136(2), 105.