Consistency of social interactions in sooty mangabeys and chimpanzees

Alexander Mielke[1,2,#], Anna Preis[2,3], Liran Samuni[2,4], Jan F. Gogarten[2,5], Jack Lester[6], Catherine Crockford[2,6,a], Roman M. Wittig[2,6,a,#]

1 Primate Models for Behavioural Evolution Lab, University of Oxford, UK

2 Taï Chimpanzee Project, Centre Suisse de Recherches Scientifiques en Côte d'Ivoire, Abidjan, Côte d'Ivoire

3 Wild Chimpanzee Foundation

4 Department of Human Evolutionary Biology, Harvard University, USA

5 Robert Koch Institute, P3: "Epidemiology of Highly Pathogenic Microorganisms", Berlin, Germany

6 Max-Planck-Institute for Evolutionary Anthropology, Leipzig, Germany

a C.C. and R.M.W. contributed equally to this work.

[#]correspondence: mielke.alexand@gmail.com, wittig@eva.mpg.de

ABSTRACT

Predictability of social interactions can be an important measure for the social complexity of an animal group. Predictability is partially dependent on how consistent interaction patterns are over time: does the behaviour on one day explain the behaviour on another? We developed a consistency measure that serves two functions: detecting which interaction types in a data set are so inconsistent that including them in further analyses risks introducing unexplained error; and comparatively quantifying differences in consistency within and between animal groups. We applied the consistency measure to simulated data and field data for one group of sooty mangabeys (*Cercocebus atys atys*) and to groups of Western chimpanzees (*Pan troglodytes verus*) in the Taï National Park, Côte d'Ivoire, to test its properties and compare consistency across groups. The consistency measures successfully identified interaction types whose low internal consistency would likely create analytical problems. Species-level differences in consistency were less pronounced than differences within groups: in all groups, aggression and dominance interactions were the most consistent, followed by grooming; spatial proximity at different levels was much less consistent than directed interactions. Our consistency measure can facilitate decision making of researchers wondering whether to include interaction types in their analyses or social networks and allows us to compare interaction types within and between species regarding their predictability.

INTRODUCTION

Animals living in permanent social groups must decide when and how to interact with group members, and their ability to make appropriate choices has potential fitness implications (Shettleworth, 2009). The evolution of species' cognitive apparatus is a response to selection pressures imposed by the complexity of their environment, including the social system they live in (Byrne & Whiten, 1989; Humphrey, 1976; Jolly, 1966). This hypothesis assumes that animals in more "complex" social systems must integrate more social information to out-compete others (Byrne & Whiten, 1989). However, it is unclear how to quantify social information, even though various indices have been proposed (Bergman & Beehner, 2015; Fischer et al., 2017).

45   One way to operationalize social complexity is as the amount of information necessary to
46   successfully predict future states within a system (Flack, 2012; Sambrook & Whiten, 1997).
47   Measures of interaction predictability on an individual level in group-living species would
48   facilitate examinations of factors driving evolution of complex decision-making (Aureli & Schino,
49   2019; Dunbar & Shultz, 2010).

50   Consistency of partner choice across time, i.e. repeatedly choosing to interact with the same
51   individual in the same way, enhances the predictability of future outcomes (Kalbitz et al., 2016;
52   Koski et al., 2012; Moscovice et al., 2017; Silk et al., 2006). For example, in steep linear
53   dominance hierarchies, a single interaction per dyad contains enough information to predict
54   future dyadic contests (Guillermo Paz-Y-Miño et al., 2004; Oliveira et al., 1998; Sánchez-Tójar et
55   al., 2018). Low consistency can be the result of an unpredictable distribution of social
56   interactions, frequent changes in relationships over time, or the presence of various mediating
57   factors, all challenges that might necessitate an increased need for cognitive flexibility (Barrett
58   et al., 2002).

59   Assessing predictability is complicated by the fact that we work with incomplete data, as
60   recording every interaction taking place in an animal group is not practicable. Many behavioural
61   studies depend on aggregated distributions of interactions over time: we take, for example, a
62   one-year period and calculate how many interactions were observed on an individual and
63   dyadic level. These distributions are used either as dependent or independent variables, to
64   create networks, or to create relationship indices. The fundamental assumption is that the data
65   accurately reflect what individuals were doing during the study period and that patterns are
66   consistent; the "real" distribution of interactions is unknown (Farine & Strandburg-Peshkin,
67   2015; Kasper & Voelkl, 2009; Whitehead, 2008). However, if data are sparse, estimate errors are
68   increased and robustness of the resulting distribution reduced (Lusseau et al., 2008; Shizuka &
69   Farine, 2016). Working with measures that are not accurate representations of the underlying
70   distribution can create misleading results (Davis et al., 2018). This problem is exacerbated when
71   already sparse datasets are cut into shorter time intervals (e.g. 6-month blocks), a common
72   practice in animal behaviour studies. What constitutes 'enough' data can vary depending on the

73 consistency of partner choice (Sánchez-Tójar et al., 2018; Whitehead, 2008). For many
74 researchers, it is difficult to assess whether they have collected enough data to include an
75 interaction type into their analyses. Here, we propose a shorthand.

76 In the present study, we develop a consistency measure that serves two functions: 1) allowing
77 researchers to gauge whether they have collected enough data to warrant the inclusion of an
78 interaction type in their analyses, in a social network, or when creating relationship indices. 2)
79 Compare predictability of interaction types within, between, and among species. Consistency
80 should be high if individuals choose the same partners for the same interaction type
81 independent of when they are observed, and observing an individual at one point in time allows
82 for accurate predictions of its future behaviour. Low consistency can arise if individuals show
83 weak partner preference or preference changes over time, or if insufficient data are available.

84 To explore how consistency can be used to compare social groups with different structure and
85 organisation, we first use simulations of datasets with different properties. We subsequently
86 apply the consistency measure to data from two Western chimpanzee (*Pan troglodytes verus*)
87 communities and one sooty mangabey (*Cercocebus atys atys*) community living sympatrically in
88 the Taï National Park, Côte d'Ivoire (Mielke et al., 2017, 2018). These species represent two
89 well-studied, quite different primate social systems. Sooty mangabeys have philopatric females
90 who form linear, despotic, stable matrilineal hierarchies (Mielke et al., 2017, 2018; Range, 2006;
91 Range & Noë, 2002). All mangabey directed social interactions are expected to show high
92 consistency, as they should be strongly influenced by stable parameters, especially kinship,
93 dominance rank, and sex (Range & Noë, 2002). Association patterns in this species are nearly
94 random (Mielke et al., 2020), so we predict low consistency for spatial interaction types.
95 Chimpanzees at Taï are similar in that they have been shown to have stable grooming,
96 aggression, and association patterns in both sexes. However, in contrast to the mangabeys,
97 aggression is not exclusively determined by dominance hierarchy (Wittig & Boesch, 2003), and
98 we have previously described rank changes in both sexes in the study period (Mielke et al.,
99 2019; Preis et al., 2019). Rank uncertainty and the variation in partner availability due to fission
100 fusion dynamics in chimpanzees lead us to predict that chimpanzee interactions are less

101 consistent than mangabey interactions. We developed the consistency measure with two aims:

102 a) to identify interaction types where data distributions are likely unreliable due to insufficient

103 data; and b) to draw comparisons between chimpanzees and mangabeys, and identify different

104 interaction types within species as it relates to their consistency.

105

106 **METHODS**

107 *Consistency measure*

108 To quantify consistency in an interaction type, we organised the data by collection days. Each

109 observation day is randomly assigned to one of two datasets of equal size (Sánchez-Tójar et al.,

110 2018). For each of the two resulting datasets, we calculated the dyadic interaction rates per

111 observation hour in each of the halves and calculate the non-parametric Spearman correlation

112 between the two distributions (see Fig. 1 for procedure). This allowed us to estimate how well

113 variation in one half of the dataset predicts variation in the other half. We performed 100

114 iterations, with the median correlation coefficient constituting our measure of consistency for

115 the full dataset.

116

117
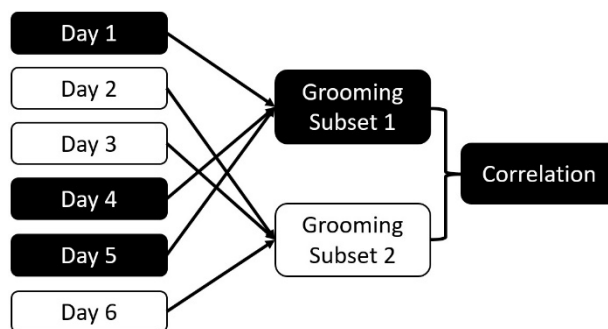
Fig. 1: Schema of the consistency measure. Data are randomly divided into two subsets based on the collection day. Dyadic interactions for all dyads are aggregated for each subset. The two subsets are correlated. This process is repeated 100 times to calculate the consistency of the overall dataset for this interaction type.

118

119 The overall correlation between halves of the dataset is likely dependent on the data collection

120 effort and community size, making it difficult to compare communities and interaction types. To

121 mitigate this challenge, we developed a standardised version of the consistency measure (Fig. 2)

5

122   by repeatedly selecting subsets of the data that differ in length and the amount of data

123   included, followed by randomly selecting a start date and duration for the period following that

124   date. We tested the consistency for this period for each interaction type, marking how many

125   interactions per dyad the subset contained. For example, 10 individuals form 45 dyads; if we

126   collect 180 aggressive events, we have a mean of 4 interactions/dyad. We then collate the

127   consistency of all datasets with the same number of interactions per dyad – *e.g.*, we could have

128   100 consistency values based on datasets that contain 3 interactions per dyad, 120 based on 4

129   interactions per dyad, and so on. For each interaction per dyad value, we plot and report the
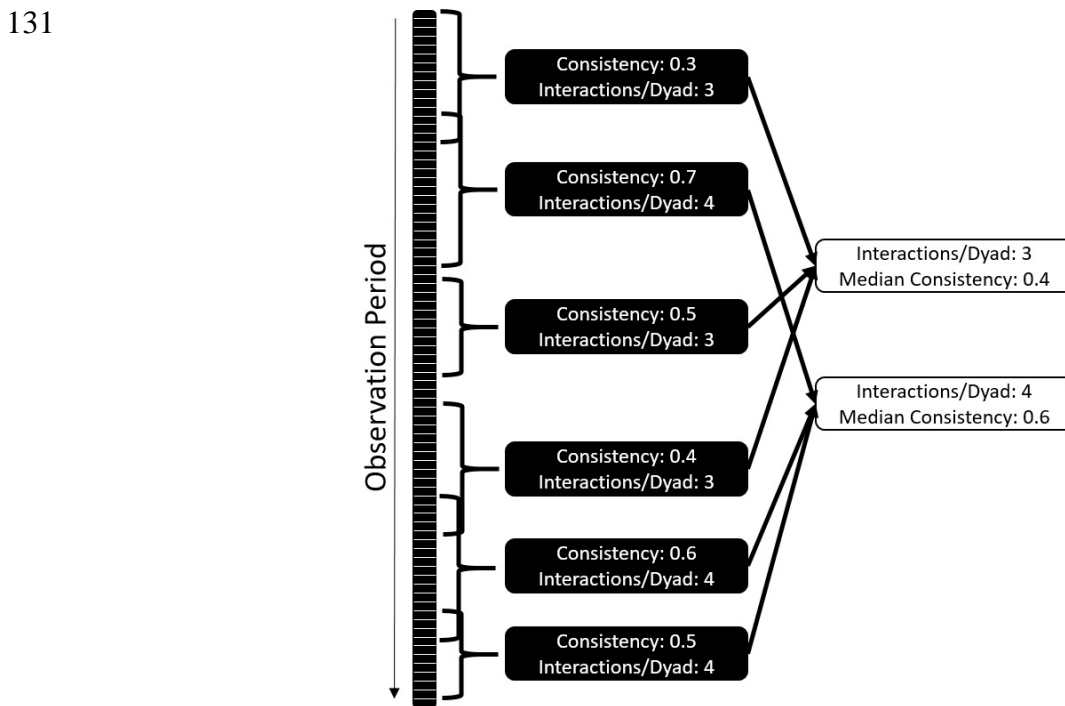
130   median of the consistency values.

131

Fig. 2: Standardisation of consistency as a comparative measure: each dataset is randomly cut into shorter time windows of changing size and starting point. Consistency and amount of interactions per dyad are established. The median of consistencies across different interactions per dyad values are established. The comparative value is the number of interactions per dyad where the consistencies cross 0.5.

132

133   This approach allows a systematic comparison of both frequent and infrequent interaction

134   types, *i.e.*, datasets of different sizes. Analyses comprising differing group sizes are possible

135   because we compare the behaviour of datasets that contain the same number of interactions

136   per dyad. As a standardised consistency measure, we report the number of interactions per

137   dyad needed to get a median consistency value of 0.5; although this value has no strong

138   biological justification, in simulations it was reliable in distinguishing interactions types that

139   were consistent from those that had insufficient data or were inconsistent. This measure is

140   largely independent of data density and community size, and produces an interpretable result:

141   how many interactions between two group members does an individual need to observe to

142   reliably predict future interactions? Fewer interactions per dyad and a smaller standard

143   deviation of values indicate higher consistency in partner choice and thus higher predictability.

144   Larger numbers of interactions per dyad and a large standard deviation indicate that interaction

145   patterns are harder to predict. This can be the case if either the partner choice is less

146   deterministic for the interaction type, or the choice patterns change throughout the study

147   period.

148

149   *Simulations*

150   All described analyses were conducted in R 4.0.0 (R Development Core Team & R Core Team,

151   2020). Scripts can be found in the associated GitHub repository. We explored the impact of

152   different group sizes, data densities, repeatability of partner choice, and changes in underlying

153   relationships on our consistency measure using simulated datasets. We then explore whether it

154   can be used to compare consistency across communities of different sizes. We tested whether

155   our consistency metric is high when individuals regularly choose the same partners for the same

156   interaction type. We also tested whether low consistency arises when individuals show weak

157   partner preference or when preference changes over time. To test how our consistency

158   measure performed under different conditions, and how to interpret different results, we

159   simulated datasets with different group sizes; numbers of interactions per individual; data

160   collection density; stereotypy of partner choice; and consistency of partner choice over time,

161   mirroring interaction data as it could be collected in different social animal species.

162   Specifically, we created datasets for 10, 15, and 20 individuals in a community, for one

163   nonspecific interaction type over a simulated period of one year. We randomly assigned each

7

164   individual between 1 and 10 interactions per day, and for each interaction the partner was
165   chosen from a random chosen subset of group members (to simulate animal groups, in which
166   not all group members are always physically available). To simulate different underlying
167   probability distributions of who interacts with whom, each dyad was assigned a random
168   likelihood to interact with each other, with three different stereotypy levels: "high certainty"
169   (each individual has strong preference for a few group members, always chosen those when
170   they are available), "medium certainty" (each individual prefers several group members, but can
171   also choose non-preferred partners), and "low certainty" (the likelihood of choosing any partner
172   is relatively equal). Based on these dyadic values, one of the individuals in the "party" was
173   selected as interaction partner. We explored three conditions concerning the consistency of
174   individuals' choice: in the first condition, dyadic preferences remained the same throughout. In
175   the second condition, to simulate changes in interaction patterns, all likelihoods of partner
176   choice were reversed halfway through data collection, so dyads with a 0.95 likelihood of
177   interacting in the first half had a 0.05 likelihood of interacting in the second half of data
178   collection. In the third condition, partner choice was completely random, which should lead to
179   an even distribution of interactions between all group members over the whole time.

180   Following this procedure, we created 108 simulated datasets (three each for every combination
181   of number of individuals, level of stereotypy, and consistency condition) that contained all
182   interactions for all group members for each day of the data collection period. Subsequently, we
183   simulated differences in data collection effort (Davis, Crofoot, & Farine, 2018): for each day of
184   the sampling period, one individual was chosen as the "focal" individual whose data were
185   retained, as would be the case in most animal datasets. We assumed a twelve-hour observation
186   period per focal day, to calculate interaction rates. Then we simulated that data collection took
187   place every day, 66% of days, or every third day (33%), to test the impact of low data collection
188   density on the consistency measure. We therefore retained 324 simulated datasets with
189   different properties. For each of these, the proposed consistency measure – randomly selecting
190   half of the dataset and correlating interaction rates of dyads with those of the other half, as well
191   as repeating this procedure with subsets of the data – was carried out 100 times.

8

192

193   *Data Collection*

194   Behavioural data were collected in Taï National Park, Côte d'Ivoire (Wittig & Boesch, 2019) from
195   October 2013 to July 2015 for the chimpanzees and January 2014 to September 2015 for the
196   mangabeys, using half- and full-day continuous focal animal sampling (Altmann, 1974) for the
197   chimpanzees, and half-day and one-hour focal animal sampling for the mangabeys. Scripts and
198   data can be found in the associated GitHub repository. Trained observers and field assistants
199   recorded all social interactions of adult male and female chimpanzees (above 12 years of age) in
200   the "South" and "East" communities and adult (above 4.5 years) sooty mangabeys. This resulted
201   in 6441h of focal observations in South community, 5668h for East community, and 2259h for
202   the mangabey community. We included adult individuals of both sexes in all three communities
203   for whom sufficient focal data (at least 50 social interactions observed as focal individual) were
204   available and who were present for at least 80% of the study period (South: 5 males, 7 females;
205   East: 5 males, 7 females; mangabeys: 6 males, 17 females).

206   From the behavioural data, for each dyad, we extracted the duration of grooming sent, resting
207   or foraging in less than 1m distance from the partner ("body contact": used as a continuous
208   measure with duration in the chimpanzees and an event variable in the mangabeys), resting or
209   foraging as nearest neighbour between 1m and 3m distance ("proximity"), and both contact
210   and noncontact aggressive interactions with one clear recipient (Preis et al., 2018). For the
211   chimpanzee communities, we included food sharing (Samuni et al., 2018), which was not
212   regularly observed in the mangabeys. We used pant grunt vocalisations in chimpanzees and
213   feeding supplants in mangabeys as additional interaction types. Mutual interactions were coded
214   as interactions given in both directions. We treated body contact and proximity as 'interaction
215   types' with the assumption that both individuals have to show sufficient tolerance to allow the
216   other one to remain close. Body contact and proximity were only counted if no other interaction
217   took place within 5min before or after to ensure independence of data points. We included
218   grooming, contact aggression, noncontact aggression, pant grunts/supplants, and food sharing
219   as directional variables, with the distribution of interactions given from each individual to every

9

220   other. For the two spatial proximity measures, data were considered non-directional and

221   symmetrical. Interaction distributions were standardised by focal observation time, with

222   observation time calculated by adding the total observation times of A and B. Spatial proximity

223   and food sharing in the chimpanzees were collected by a subset of observers and were

224   standardised based on the focal observation time provided by those observers. Scripts to create

225   the consistency measure and plots can be found in the Supplementary.

226

227   **RESULTS**

228   *Simulations*

229   We sought a consistency measure that can identify differences in stereotypy of partner choice

230   and changes in interaction preference, while being independent of group size and data

231   collection effort. Our standardised consistency measure performed the same, independent of

232   community size (Fig. 3.1). Meanwhile, datasets of different data collection density followed the

233   same trajectory, but lower data density was characterized by lower overall consistency. The

234   overall consistency cannot be interpreted alone, as it is highly dependent on group size and data

235   collection effort. This is consistent with simulations showing that social network data becomes

236   unreliable if data density per dyad sinks below a certain level (Whitehead, 2008).

237
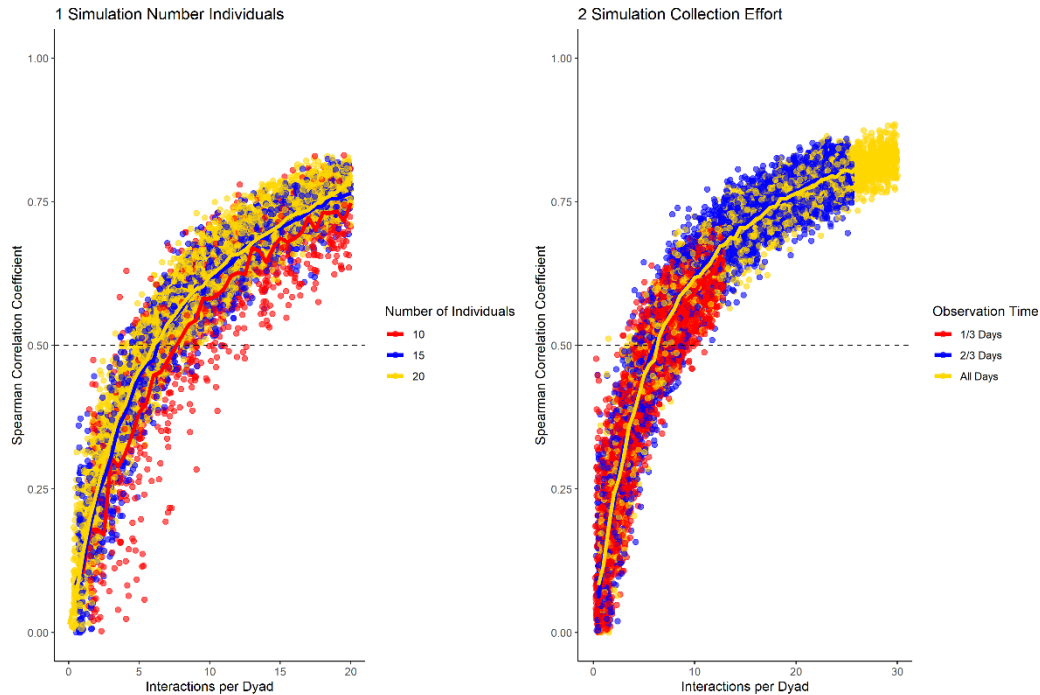
238 *Figure 3: Results of the data simulation with varying group size (1) and data collection density (2), while having medium*
239 *stereotypy of partner preference and no preference changes throughout the dataset. Horizontal line marks a correlation of*
240 *halves of 0.5. The number of interactions per dyad allows to compare datasets of different density and number of individuals.*

241

242

243 To test how the stereotypy of partner choice influenced the consistency measure, we present

244 the results for the three different conditions (high, medium, low certainty) in datasets

245 containing 20 group members, with 100% data density, and no changes in preference

246 throughout the sampling period (Fig. 4). Our results showed that the consistency measure

247 differentiated between the conditions, using the slope at which the chosen cut-off value is

248 reached. If partner choice is highly stereotypical, a small number of interactions was sufficient

249 to predict partner choice in half of the data with that of the other half; with increasing

250 uncertainty, more interactions per dyad were necessary. For low certainty of partner choice,

251 more than the number of simulated interactions would have been necessary to reach the cut-

252 off of 0.5.

253

254
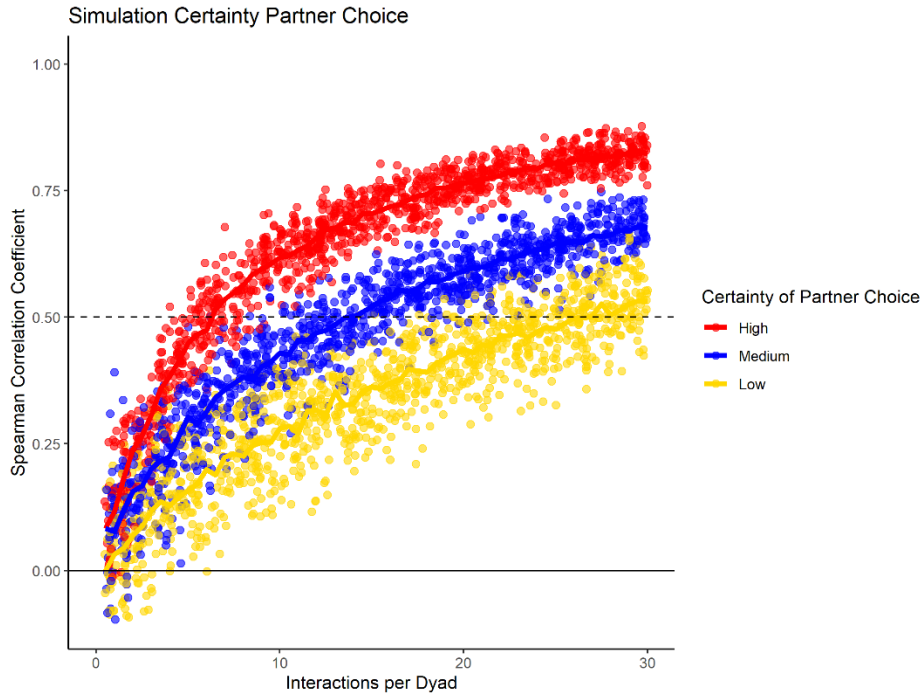
255 *Figure 2: Data simulation of varying stereotypy of partner choice, while having consistent group size, data density, and no*
256 *preference changes throughout the dataset.*

257

258 Last, we investigated how changes in partner preference over the study period would influence

259 the consistency measure in a dataset with 20 individuals, with 100% data density, and high

260 stereotypy of partner choice. Here, we compare three conditions: one where no changes took

261 place, one where the partner preference was reversed halfway through the study, and one

262 where partner choice was randomized. Again, we found differences in the slope whereby the

263 consistency increased with increasing data density (Fig. 5). Additionally, the conditions could be

264 differentiated by the spread of consistency values: when partner choice was consistent,

265 selecting subsets of the same size at different points of the sampling period resulted in very

266 similar consistency values. If partner choice changed throughout the sampling period, variation

267 was much higher. Also, the consistency of the full dataset was smaller than that of some shorter

268 subsets, with the highest levels for subsets that were roughly half the total size – mirroring our

269 built-in change of interaction likelihood after half the 'collection period'. As seen before, random

270 partner choice could be identified because the consistency of the full dataset never increased

271 above a certain threshold.

272
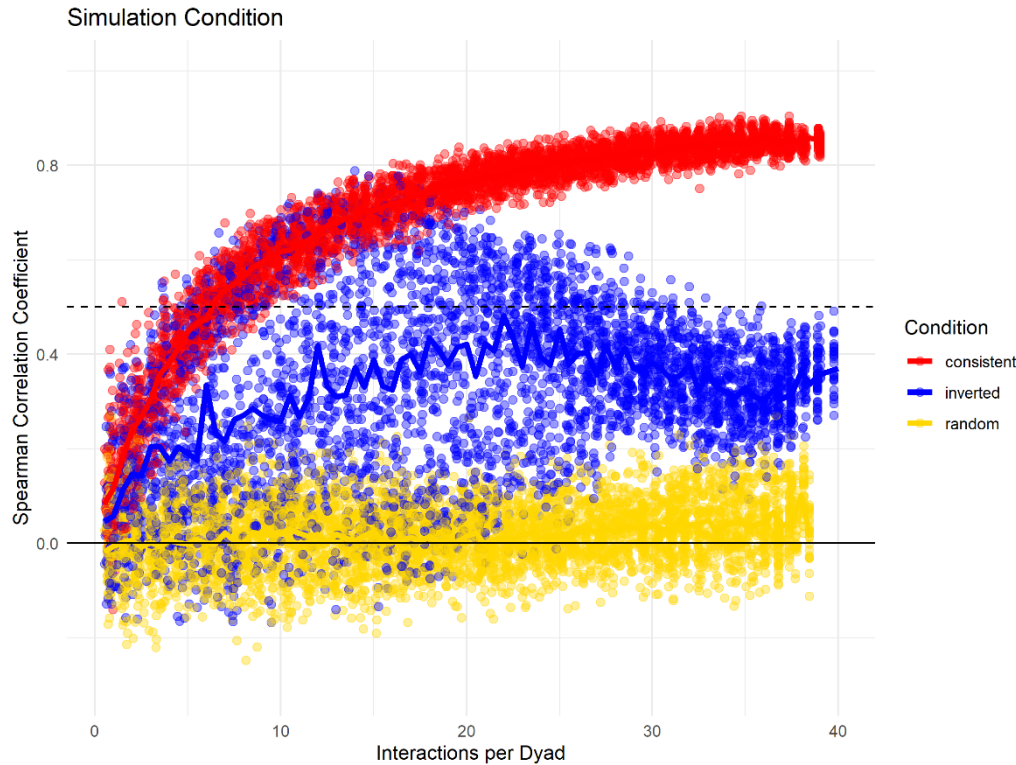
Fig. 5: Data simulation changes potential changes of interaction distributions throughout the dataset while having consistent group size, data density, and stereotypy of dyadic preference. "Consistent Choice" indicates no changes in preferences throughout, "Inverted Choice" indicates one reversal for all dyadic preferences, while random choice indicates that all partners were chosen with the same likelihood.

In sum, based on these simulations, the consistency measure can be used to compare the predictability of interactions. Using the entire dataset, the overall consistency was heavily influenced by the amount of interactions available per dyad, and thus does not make a good comparative measure. In our simulations, even if the underlying distribution of interactions was highly stereotyped and consistent, the consistency measure remained low if few data points were available per dyad, indicating that one half of the dataset was not a good predictor of the other half due to random sampling error. Thus, if the Spearman rank correlation between halves of the same dataset does not reach r_s = 0.5, it is likely that not enough data have been collected to make statements about the underlying distribution of an interaction type in a population (unless that distribution is random). We therefore suggest using interaction types with an overall consistency below r_s = 0.5 with care or remove them from analyses where possible, as

289   their interpretation is unclear. For all other interactions, we propose the described standardised

290   consistency measure, the average number of interactions per dyad necessary to reach a median

291   consistency of r_s = 0.5 as a good measure. Valuable information also arose from the spread of

292   values of the repeated comparisons between halves of the dataset: if dyadic preference

293   remained stable throughout, the consistency is relatively stable for subsets of the same size.

294   However, if dyadic preference was not stable, the correlation between halves varies even for

295   datasets of the same size.

296

297   **_Field Data_**

298   _Mangabeys_

299   For the mangabeys, noncontact aggression rates (3 interactions/dyad) and supplants (3

300   interactions/dyad) were highly consistent, as was grooming (4.5 interactions/dyad), indicating

301   that individuals observing a subset of interactions in the community would be able to predict

302   future interactions (Fig. 6, 7, 8; Tab. 1). Body contact (17 interactions/dyad) was much less

303   consistent, and proximity (being within 3m of each other) did not reach the threshold of 0.5,

304   despite having among the highest number of data points available for any interaction type in

305   this study. Given the trend of the graph, proximity would probably have reached the threshold if

306   more data had been available, but this still suggests a highly inconsistent distribution across the

307   data collection period. For contact aggression, only a small number of cases were available, and

308   the graph did not reach the consistency threshold. In our simulations, such low values occurred

309   when insufficient data were available to successfully approximate the underlying distributions

310   of interactions, even in cases where the underlying distribution was highly consistent; or when

311   distribution of interaction was random or close to random.

312

313   _Table 1: Overview of consistency scores in chimpanzee and mangabey social interactions: datasets for each interaction type and_

314   _group, and the results of the consistency measures. "Overall consistency" is the median of the repeated correlation between_

315   _randomly selected halves for the full dataset available for an interaction type. "Standardised Consistency" and the standard_

316   _deviation are the result of the repeated random selection of halves of subsets of different lengths, with number of interactions_

317   _per dyad where the median consistency exceeds r_s= 0.5 as measure of how much information is needed to predict future_

318   _interactions in a community. Interaction types were the r_s= 0.5 was not exceeded are marked with '-'._

14

| Group	Interaction Type	IDs	Interactions	Overall Consistency	Standardised Consistency (Interactions/dyad)	SD
Mangabey	Grooming	23	1162	0.52	4.5	0.03
Mangabey	Body Contact	23	2218	0.53	17.0	0.03
Mangabey	Proximity	23	4373	0.48	-	-
Mangabey	Aggression Non-contact	23	971	0.58	3.0	0.03
Mangabey	Aggression Contact	23	210	0.24	-	-
Mangabey	Supplant	23	1219	0.64	3.0	0.03
East	Grooming	12	3099	0.77	8.5	0.07
East	Body Contact	12	1935	0.71	9.0	0.11
East	Proximity	12	2796	0.79	12.0	0.12
East	Aggression Non-contact	12	693	0.65	4.0	0.07
East	Aggression Contact	12	126	0.37	-	-
East	Food Sharing	12	151	0.41	-	-
East	Pant Grunt	12	2429	0.90	1.5	0.10
South	Grooming	12	4693	0.81	6.0	0.09
South	Body Contact	12	1669	0.55	27.0	0.08
South	Proximity	12	2579	0.71	19.0	0.08
South	Aggression Non-contact	12	768	0.80	2.5	0.08
South	Aggression Contact	12	173	0.43	-	-
South	Food Sharing	12	153	0.37	-	-
South	Pant Grunt	12	3350	0.92	2.5	0.10

319
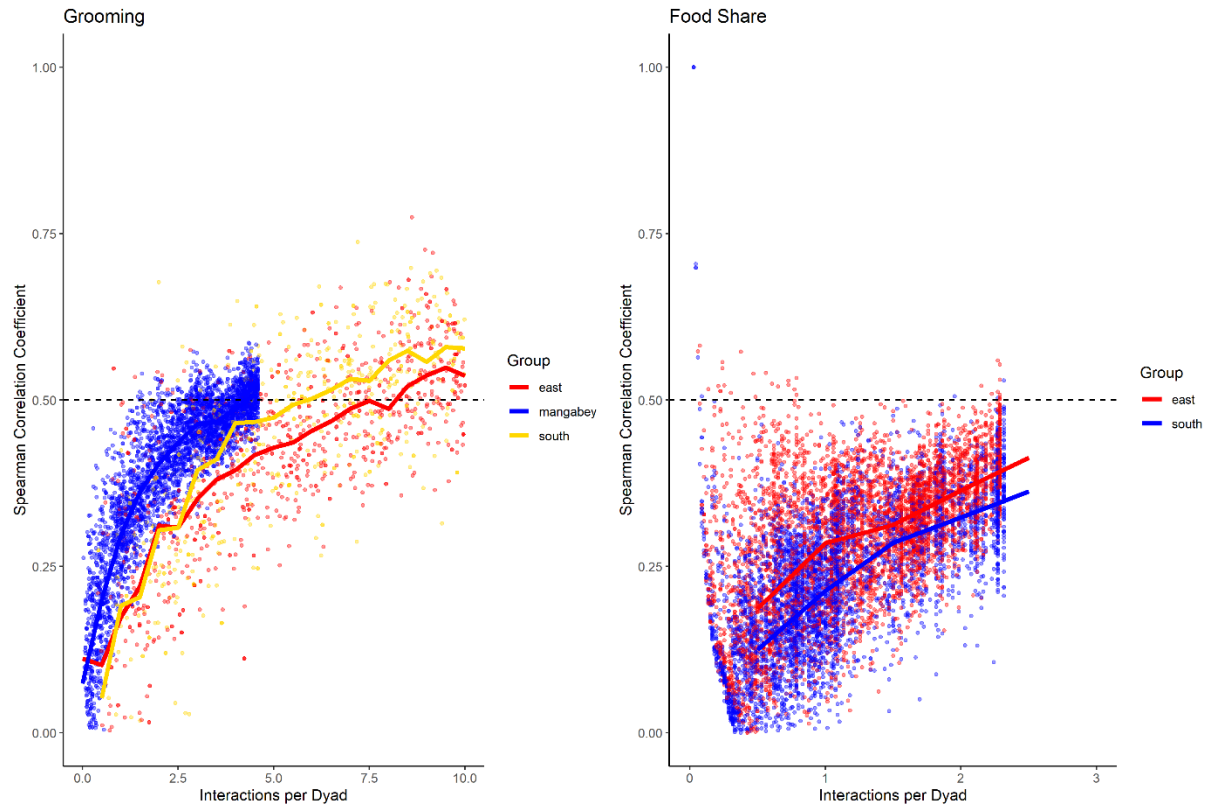
Figure 6: Spearman correlation between two halves of randomly selected subsets of the datasets for mangabeys (blue), East chimpanzee community (red) and South chimpanzee community (golden) for grooming and food sharing (chimpanzees only). The standardised consistency is marked by the number of interactions per dyad where the median of correlation coefficients exceeds r=0.5. If that value is reached with fewer interactions per dyad, the distribution of interaction rates is more consistent. Distributions of correlation coefficients with a large spread indicate changes in interaction preference over time.

Chimpanzees

As in the mangabeys, noncontact aggression rates were highly consistent in both chimpanzee communities (Table 1), more so in South (2.5 interactions/dyad) than in East (4.0 interactions/dyad). As in the mangabeys, contact aggression occurred so infrequently that now consistent representation of the distribution existed. The larger standard deviation in the chimpanzees and wider spread of the graph compared to the mangabeys might indicate changes of aggression patterns over time. Pant grunt interactions in both communities showed the most predictable patterns (East: 1.5 interactions/dyad; South: 2.5 interactions/dyad). Grooming was less consistent than in the mangabeys (East: 8.5 interactions/dyad; South: 6.0

335   interactions/dyad). Body contact showed considerable variation between groups, with East (9

336   interactions/dyad) being the most consistent, while South (27.0 interactions/dyad) being the

337   least consistent of the three groups. Proximity (East: 12.0 interactions/dyad; South: 19.0

338   interactions/dyad) was more predictable than in the mangabeys. Body contact and proximity

339   were considerably less predictable than the directed interaction types. This indicates that in all

340   three communities, most dyads will feed and rest in proximity with a wide variety of partners,

341   while they direct interactions at a smaller and more stable subset of group members.

342

343
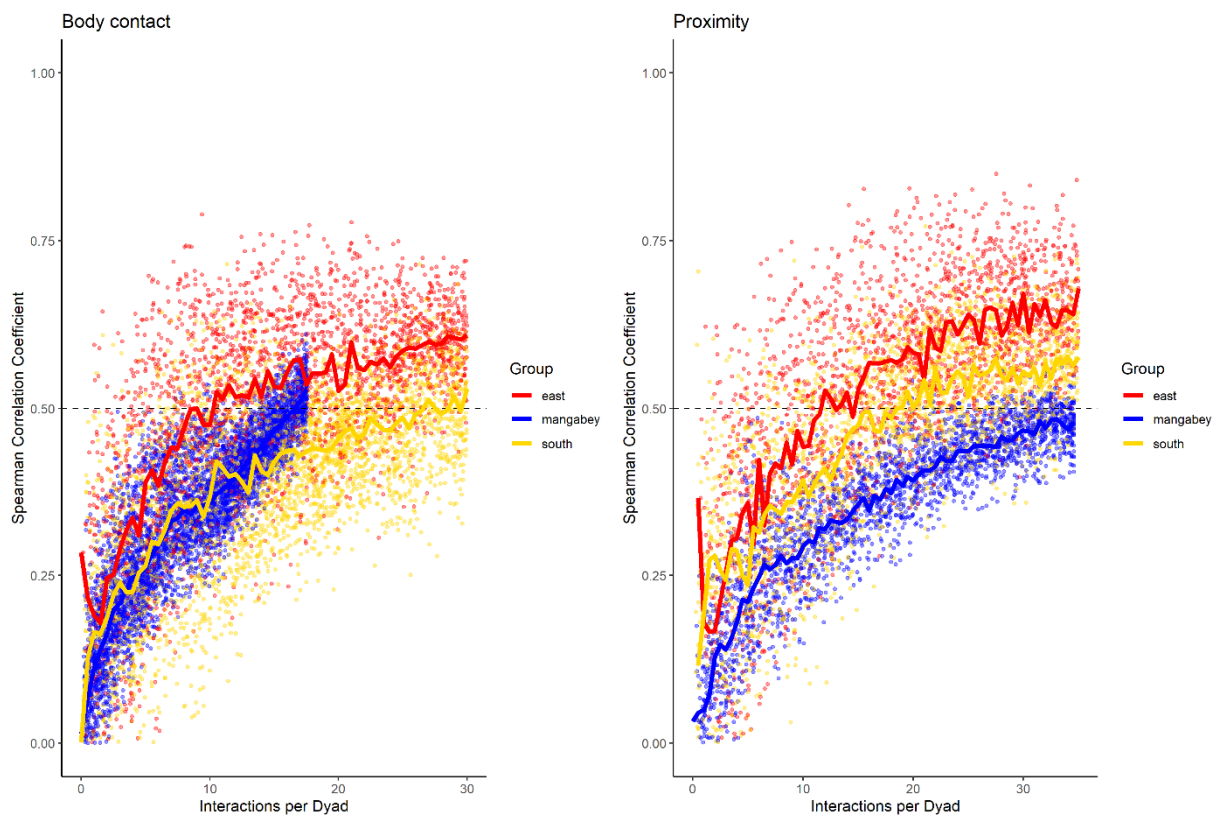
344   *Figure 7: Spearman correlation between two halves of randomly selected subsets of the datasets for mangabeys (blue), East*

345   *chimpanzee community (red) and South chimpanzee community (golden) for body contact and proximity. The standardised*

346   *consistency is marked by the number of interactions per dyad where the median of correlation coefficients exceeds r_s= 0.5. If*

347   *that value is reached with fewer interactions per dyad, the distribution of interaction rates is more consistent. Distributions of*

348   *correlation coefficients with a large spread indicate changes in interaction preference over time.*

349

350

17

351
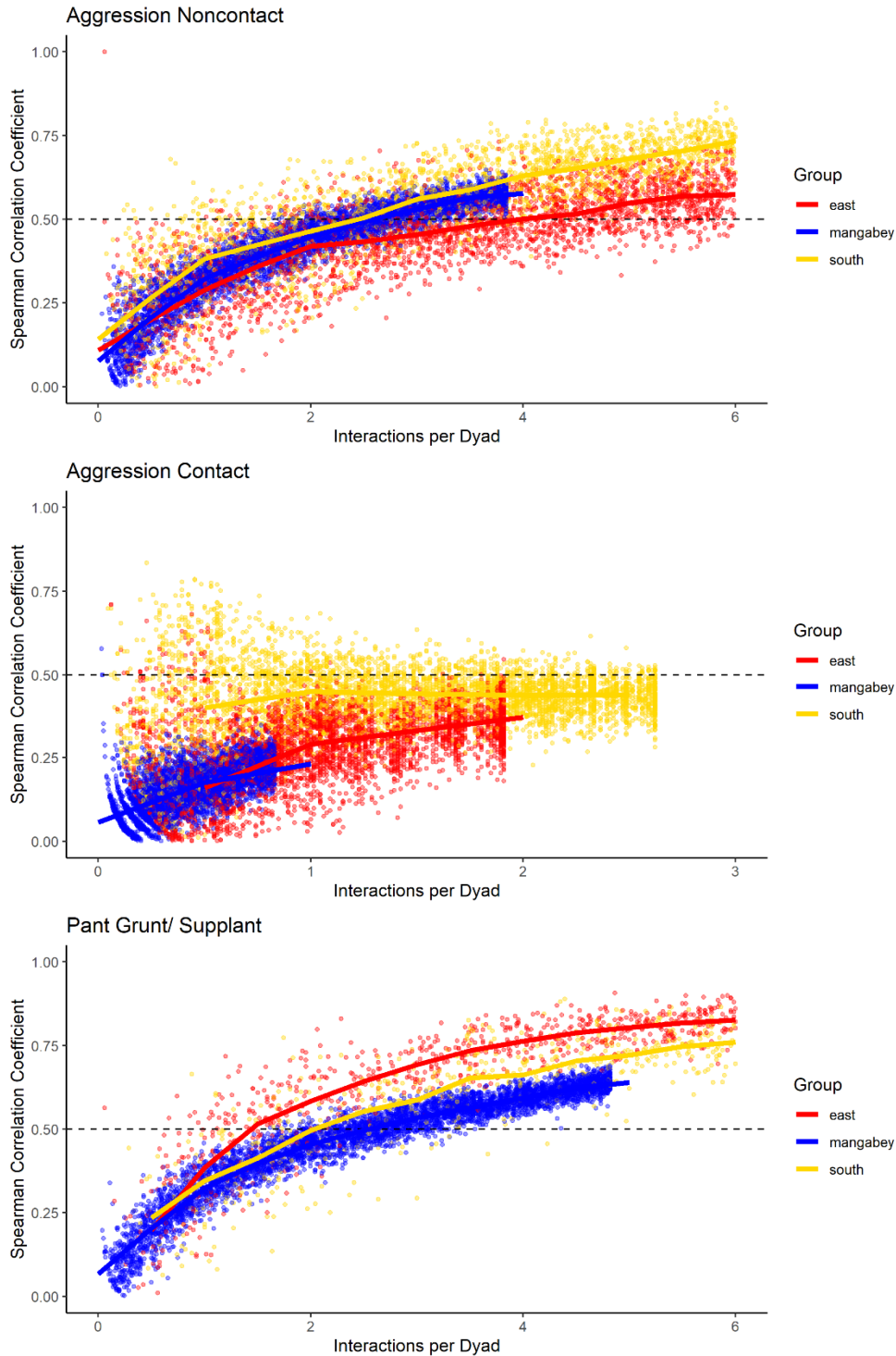
352 Figure 8: Spearman correlation between two halves of randomly selected subsets of the datasets for mangabeys (blue), East
353 chimpanzee community (red) and South chimpanzee community (golden) for noncontact aggression, contact aggression, and
354 pant grunts/supplants. The standardised consistency is marked by the number of interactions per dyad where the median of
355 correlation coefficients exceeds $r_s = 0.5$. If that value is reached with fewer interactions per dyad, the distribution of interaction
356 rates is more consistent. Distributions of correlation coefficients with a large spread indicate changes in interaction preference

357   *over time.*

358

359

Summary Standardised Consistency
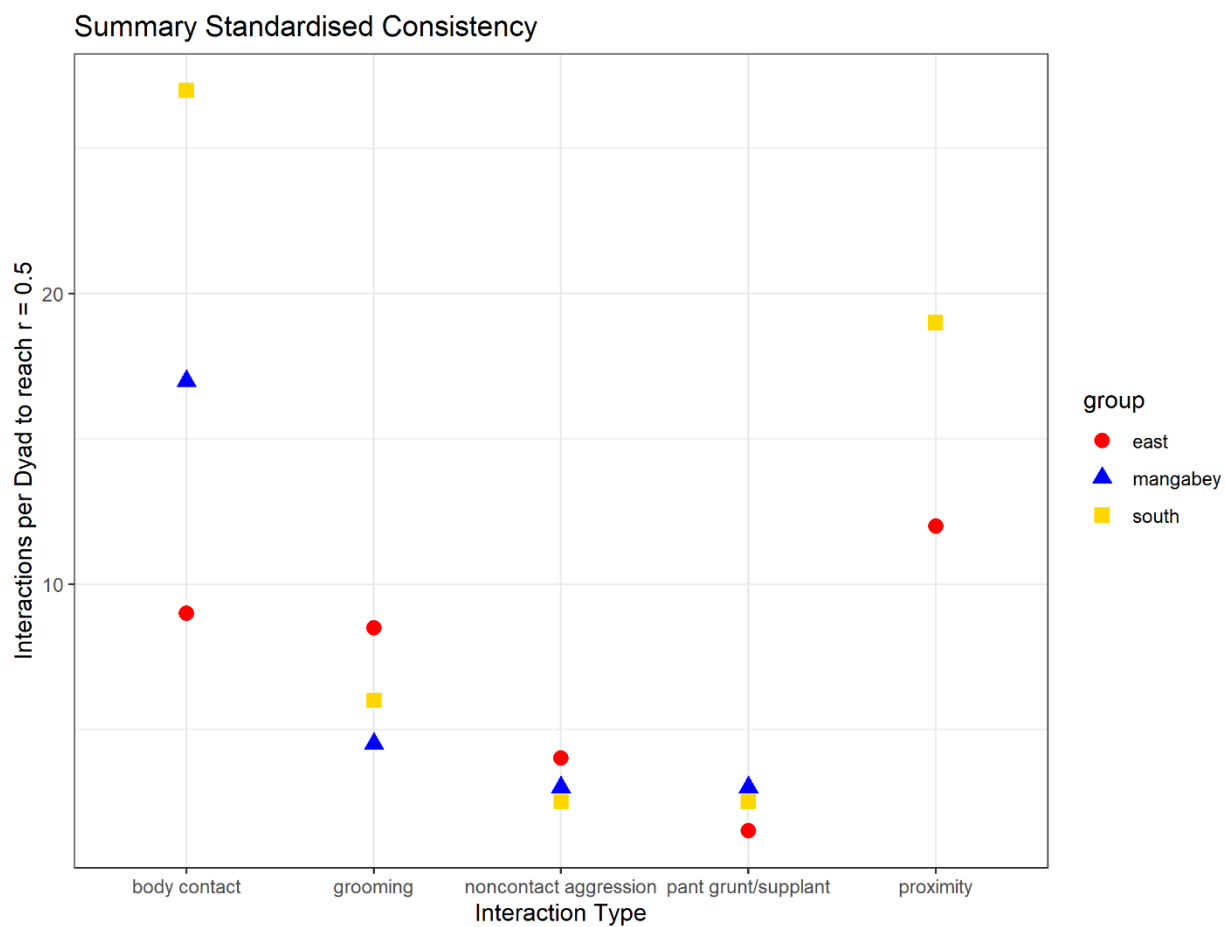
Figure 9: Summary of the mean number of interactions needed per dyad to reach correlations between halves of $r_s = 0.5$ (mangabeys: blue triangles, East: red points, South: golden square).

360

361 **DISCUSSION**

362 Establishing measures of predictability of social interactions between individuals is necessary to

363 understand the complexity of a social group from the perspective of the individual (Dunbar &

364 Shultz, 2010; Lukas & Clutton-Brock, 2018). Here, our premise was that interactions are more

365 predictable for participants and bystanders if interaction distributions are consistent over time.

366 Our results showed that across communities and species, interaction types vary in predictability,

367 indicating yet again that animal lives cannot be captured using one simplistic measure of

368 complexity: challenges differ within and between species, and we need multi-dimensional

369 measures to quantify where 'complexity' really arises.

370 This study introduces a consistency measure, repeatedly dividing the dataset into halves and

371 comparing how well these predict each other, which serves two functions. Researchers can use

372 it to find out whether they have collected sufficient data for their dataset to be internally

373 consistent, given a community of a certain size and an interaction type with a specific diversity

374 of partner choice (Sánchez-Tójar et al., 2018). In our sample, despite pooling 18 months of data,

375 food sharing and contact aggression were observed at such low rates in all three communities

376 that observing the group at a certain time point would make it impossible to predict their

377 behaviour at another time point. We do not know whether the error bars around the observed

378 values are biological or statistical, but they can introduce unexplained uncertainty in our

379 subsequent analyses. We generally assume that randomly selected focal follows allow us to also

380 make statements about interaction rates on those days on which we do not observe an

381 individual (Altmann, 1974), but this might not be the case for rare interaction types or for

382 interaction types that are naturally almost randomly distributed (Davis et al., 2018). If the

383 distribution of interactions in the group is not even consistent within an interaction type in a

384 period, correlating it with other interaction types or across periods would probably produce

385 spurious results (Whitehead, 2008).

386 The standardized consistency measure allowed us to identify interaction types that needed

387 either large or small amounts of information to predict interactions on other collection days. We

388 used the number of interactions per dyad at which the majority of subset correlations exceeds

389   the value r_s = 0.5; while the value r_s = 0.5 itself is arbitrary, using it across species and
390   interaction types allows researchers to make comparative statements, and it is high enough to
391   not fall into random variation. We did not find generalizable species differences using our
392   consistency measure: differences within species were much more pronounced and followed the
393   same trends between species. Chimpanzee distributions had generally larger standard
394   deviations, potentially indicating changes in partner choice over time. Feeding supplants and
395   pant grunts, which are used to create hierarchies in the respective species, were highly
396   consistent, indicating generally stable hierarchies (Sánchez-Tójar et al., 2018). Consistency of
397   aggression distributions did not vary strongly between species. Despite being the larger
398   community, mangabey grooming interactions were generally more predictable than chimpanzee
399   interaction patterns. Directed interactions (grooming, noncontact aggression, pant
400   grunts/supplants) were consistent despite the inclusion of 18 months of data per community,
401   indicating that most dyads interacted at relatively constant rates throughout the study period.
402   Body contact and proximity showed lower consistency than directed interactions, most likely
403   because a certain level of tolerance in foraging and resting extends to most group members,
404   adding random noise that is not present in directed interactions. For body contact, no clear
405   species-specific pattern emerged, but proximity (3m distance) was much less consistent in
406   mangabeys than in chimpanzees, a result in line with recent findings regarding high levels of
407   randomness in sooty mangabey spatial association patterns (Mielke et al., 2020). Just like rare
408   interaction types, common but highly inconsistent interaction types could add noise to social
409   relationship indices or when comparing network overlap.

410   While many animal species are studied in great detail, and vast amounts of long-term data are
411   available, it is surprisingly difficult to convey the structure of social interactions across sites and
412   species. Our consistency measure may help by providing a standardised way to convey the
413   flexibility in interaction patterns over time and identify interaction types that likely differ in
414   complexity between species. Further, many researchers use multilevel social network analysis
415   and create relationship indices including different interaction types, unsure whether all of them
416   will be equally reliable. This consistency measure, like similar efforts for hierarchies (Sánchez-

21

417   Tójar et al., 2018), can be a useful tool to make these decisions while conveying important
418   information about the study species. Importantly, these results further cement that researchers
419   need to report sample sizes not only of their outcome variable, but also for interaction types
420   that might have gone into creating relationship indices or network measures, because this gives
421   readers the ability to judge the error associated with this predictor variable or network. To
422   assess changes in relationships over time, there has been a trend to cut datasets into smaller
423   subsets and then compare network overlap between these, assuming that the data in each are
424   sufficient to depict the underlying distribution in the community. With our consistency
425   measure, seasonality and change could be established if smaller subsets would show higher
426   consistency than larger subsets, as random subsets retained consistent time intervals. This was
427   not the case for any interaction type, even though some interaction types showed large
428   variation, an indication that consistency is high during some times but not others.

429   Predictability is an important aspect of social complexity: an individual living in a system where
430   all future social interactions are largely pre-defined by a few re-occurring factors needs little
431   information to make decisions about its own behaviour (Flack, 2012; Sambrook & Whiten,
432   1997). Our consistency measure captures one aspect of predictability: if individuals distribute
433   their social interactions the same across time, it is likely easy for group members to predict
434   future social choices. This measure can easily be combined with other standardised approaches
435   to social complexity and should mirror patterns (Fischer et al., 2017; Thierry et al., 2008). We
436   did not find one consistent pattern of consistency difference between sooty mangabeys and
437   chimpanzees; rather, variation within species was larger than between species, and each species
438   showed higher consistency in some of the interaction types. One-dimensional measures of
439   social complexity, such as group size, are thus probably insufficient to capture species
440   differences in social complexity, as 'complexity' probably does not affects all aspects of life in a
441   species uniformly: different species face different challenges, creating uncertainty in different
442   areas of their social lives. Our consistency measure can detect which areas these are. Dyadic
443   distributions of aggression and dominance interactions were highly predictable across groups.
444   Spatial proximity was the least predictable aspect for all three groups, but as we have reported

22

445   before, mangabey association beyond body contact contains large uncertainty (Mielke et al.,

446   2020). Grooming interactions were less predictable in chimpanzees, indicating more varied

447   grooming partner choice or changes over time. Many challenges are shared between primate

448   species, especially regarding dyadic interaction patterns: It is therefore worth in a next step to

449   consider the challenges arising from structuring interactions as sequences in time and the

450   uncertainty arising when third parties influence decision making (Wittig et al., 2014). Our

451   consistency measure offers a valuable piece in the puzzle of social complexity across animal

452   species.

468   **Research Ethics.** This study was purely observational with no manipulation of animals. Methods

469   were approved by the Ethikrat der Max-Planck-Gesellschaft (4.08.2014).

470   **Permission to carry out fieldwork.** Permissions to conduct the research were granted by the

471   Ministries of Research and Environment of Ivory Coast (379/MESRS/GGRSIT/tm) and Office

472   Ivorien des Parcs et Reserves.

478

479   **References**

480   Altmann, J. (1974). Observational study of behavior: Sampling methods. *Behaviour*, *49*(3), 227–266.

481   https://doi.org/10.1080/14794802.2011.585831

482   Aureli, F., & Schino, G. (2019). Social complexity from within: how individuals experience the structure

483   and organization of their groups. In *Behavioral Ecology and Sociobiology* (Vol. 73, Issue 1, pp. 1–13).

484   Springer Verlag. https://doi.org/10.1007/s00265-018-2604-5

485   Barrett, L., Gaynor, D., & Henzi, S. P. P. (2002). A dynamic interaction between aggression and grooming

486   reciprocity among female chacma baboons. *Animal Behaviour*, *63*(6), 1047–1053.

487   https://doi.org/10.1006/anbe.2002.3008

488   Bergman, T. J., & Beehner, J. C. (2015). Measuring social complexity. *Animal Behaviour*, *103*, 203–209.

489   https://doi.org/10.1016/j.anbehav.2015.02.018

490   Byrne, R. W., & Whiten, A. (1989). *Machiavellian intelligence: social expertise and the evolution of*

491   *intellect in monkeys, apes, and humans*. Oxford University Press.

492   Davis, G. H., Crofoot, M. C., & Farine, D. R. (2018). Estimating the robustness and uncertainty of animal

493   social networks using different observational methods. *Animal Behaviour*, *141*, 29–44.

494   https://doi.org/10.1016/j.anbehav.2018.04.012

495   Dunbar, R. I. M., & Shultz, S. (2010). Bondedness and sociality. In *Behaviour* (Vol. 147, Issue 7, pp. 775–

496   803). https://doi.org/10.1163/000579510X501151

497   Farine, D. R., & Strandburg-Peshkin, A. (2015). Estimating uncertainty and reliability of social network

498   data using Bayesian inference. *Royal Society Open Science*, *2*(9), 150367.

499   https://doi.org/10.1098/rsos.150367

500   Fischer, J., Farnworth, M. S., Sennhenn-Reulen, H., & Hammerschmidt, K. (2017). Quantifying social

501   complexity. *Animal Behaviour*, *130*, 57–66. https://doi.org/10.1016/j.anbehav.2017.06.003

502   Flack, J. C. (2012). Multiple time-scales and the developmental dynamics of social systems. *Philosophical*

503   *Transactions of the Royal Society B: Biological Sciences*, *367*(1597), 1802–1810.

504   https://doi.org/10.1098/rstb.2011.0214

505   Guillermo Paz-Y-Miño, C., Bond, A. B., Kamil, A. C., & Balda, R. P. (2004). Pinyon jays use transitive

506   inference to predict social dominance. *Nature*, *430*(7001), 778–781.

507 https://doi.org/10.1038/nature02723

508 Humphrey, N. K. (1976). *The Social Function of Intellect*. https://doi.org/10.2307/375925

509 Jolly, A. (1966). Lemur social behavior and primate intelligence. *Science*, *153*(3735), 501–506.
510 https://doi.org/10.1126/science.153.3735.501

511 Kalbitz, J., Ostner, J., & Schülke, O. (2016). Strong, equitable and long-term social bonds in the dispersing
512 sex in Assamese macaques. *Animal Behaviour*, *113*, 13–22.
513 https://doi.org/10.1016/j.anbehav.2015.11.005

514 Kasper, C., & Voelkl, B. (2009). A social network analysis of primate groups. *Primates*, *50*(4), 343–356.
515 https://doi.org/10.1007/s10329-009-0153-2

516 Koski, S. E., de Vries, H., van de Kraats, A., & Sterck, E. H. M. (2012). Stability and Change of Social
517 Relationship Quality in Captive Chimpanzees (Pan troglodytes). *International Journal of Primatology*,
518 *33*(4), 905–921. https://doi.org/10.1007/s10764-012-9623-2

519 Lukas, D., & Clutton-Brock, T. (2018). Social complexity and kinship in animal societies. *Ecology Letters*,
520 *21*(8), 1129–1134. https://doi.org/10.1111/ele.13079

521 Lusseau, D., Whitehead, H., & Gero, S. (2008). Incorporating uncertainty into the study of animal social
522 networks. In *Animal Behaviour* (Vol. 75, Issue 5, pp. 1809–1815). Academic Press.
523 https://doi.org/10.1016/j.anbehav.2007.10.029

524 Mielke, A., Crockford, C., & Wittig, R. (2019). Rank changes in female chimpanzees in Taï National Park. In
525 C. Boesch & R. M. Wittig (Eds.), *The Chimpanzees of the Taï Forest* (pp. 290–300). Cambridge University
526 Press. https://doi.org/10.1017/9781108674218.019

527 Mielke, A., Crockford, C., & Wittig, R. M. (2020). Predictability and variability of association patterns in
528 sooty mangabeys. *Behavioral Ecology and Sociobiology*, *74*(4), 1–13. https://doi.org/10.1007/s00265-
529 020-2829-y

530 Mielke, A., Preis, A., Samuni, L., Gogarten, J. F., Wittig, R. M., & Crockford, C. (2018). Flexible decision-
531 making in grooming partner choice in sooty mangabeys and chimpanzees. *Royal Society Open Science*,
532 *5*(7), 172143. https://doi.org/10.1098/rsos.172143

533 Mielke, A., Samuni, L., Preis, A., Gogarten, J. F., Crockford, C., & Wittig, R. M. (2017). Bystanders
534 intervene to impede grooming in western chimpanzees and sooty mangabeys. *Royal Society Open*

26

535 *Science, 4*(11), 171296. https://doi.org/10.1098/rsos.171296

536 Moscovice, L. R., Douglas, P. H., Martinez-Iñigo, L., Surbeck, M., Vigilant, L., & Hohmann, G. (2017).
537 Stable and fluctuating social preferences and implications for cooperation among female bonobos at
538 LuiKotale, Salonga National Park, DRC. *American Journal of Physical Anthropology, 163*(1), 158–172.
539 https://doi.org/10.1002/ajpa.23197

540 Oliveira, R. F., McGregor, P. K., & Latruffe, C. (1998). Know thine enemy: Fighting fish gather information
541 from observing conspecific interactions. *Proceedings of the Royal Society B: Biological Sciences,*
542 *265*(1401), 1045–1049. https://doi.org/10.1098/rspb.1998.0397

543 Preis, A., Samuni, L., Deschner, T., Crockford, C., & Wittig, R. M. (2019). Urinary cortisol, aggression,
544 dominance and competition in wild, West African male chimpanzees. *Frontiers in Ecology and Evolution,*
545 *7*(APR). https://doi.org/10.3389/fevo.2019.00107

546 Preis, A., Samuni, L., Mielke, A., Deschner, T., Crockford, C., & Wittig, R. M. (2018). Urinary oxytocin levels
547 in relation to post-conflict affiliations in wild male chimpanzees (Pan troglodytes verus). *Hormones and*
548 *Behavior.* https://doi.org/10.1016/j.yhbeh.2018.07.009

549 R Development Core Team, & R Core Team. (2020). R: A language and environment for statistical
550 computing. *R Foundation for Statistical Computing Vienna Austria, 0,* {ISBN} 3-900051-07-0.
551 https://doi.org/10.1038/sj.hdy.6800737

552 Range, F. (2006). Social behavior of free-ranging juvenile sooty mangabeys (Cercocebus torquatus atys).
553 *Behavioral Ecology and Sociobiology, 59*(4), 511–520. https://doi.org/10.1007/s00265-005-0076-x

554 Range, F., & Noë, R. (2002). Familiarity and dominance relations among female sooty mangabeys in the
555 Taï National Park. *American Journal of Primatology, 56*(3), 137–153. https://doi.org/10.1002/ajp.1070

556 Sambrook, T., & Whiten, A. (1997). On the Nature of Complexity in Cognitive and Behavioural Science.
557 *Theory & Psychology, 7*(2), 191–213. https://doi.org/10.1177/0959354397072004

558 Samuni, L., Preis, A., Mielke, A., Deschner, T., Wittig, R. M., & Crockford, C. (2018). Social bonds facilitate
559 cooperative resource sharing in wild chimpanzees. *Proceedings of the Royal Society B: Biological*
560 *Sciences, 285*(1888), 20181643. https://doi.org/10.1098/rspb.2018.1643

561 Sánchez-Tójar, A., Schroeder, J., & Farine, D. R. (2018). A practical guide for inferring reliable dominance
562 hierarchies and estimating their uncertainty. *Journal of Animal Ecology, 87*(3), 594–608.

563   https://doi.org/10.1111/1365-2656.12776

564   Shettleworth, S. J. (2009). *Cognition, evolution, and behavior*. Oxford University Press.

565   Shizuka, D., & Farine, D. R. (2016). Measuring the robustness of network community structure using

566   assortativity. *Animal Behaviour*, *112*, 237–246. https://doi.org/10.1016/j.anbehav.2015.12.007

567   Silk, J. B., Alberts, S. C., & Altmann, J. (2006). Social relationships among adult female baboons (Papio

568   cynocephalus) II. Variation in the quality and stability of social bonds. *Behavioral Ecology and*

569   *Sociobiology*, *61*(2), 197–204. https://doi.org/10.1007/s00265-006-0250-9

570   Thierry, B., Aureli, F., Nunn, C. L., Petit, O., Abegg, C., & de Waal, F. B. M. (2008). A comparative study of

571   conflict resolution in macaques: insights into the nature of trait covariation. *Animal Behaviour*, *75*(3),

572   847–860. https://doi.org/10.1016/j.anbehav.2007.07.006

573   Whitehead, H. (2008). Precision and power in the analysis of social structure using associations. *Animal*

574   *Behaviour*, *75*(3), 1093–1099. https://doi.org/10.1016/j.anbehav.2007.08.022

575   Wittig, R. M., & Boesch, C. (2003). "Decision-making" in conflicts of wild chimpanzees (Pan troglodytes):

576   An extension of the Relational Model. *Behavioral Ecology and Sociobiology*, *54*(5), 491–504.

577   https://doi.org/10.1007/s00265-003-0654-8

578   Wittig, R. M., & Boesch, C. (2019). Observation protocol and long-term data collection in Taï. In C. Boesch

579   & R. M. Wittig (Eds.), *The Chimpanzees of the Taï Forest* (pp. 44–57). Cambridge University Press.

580   https://doi.org/10.1017/9781108674218.004

581   Wittig, R. M., Crockford, C., Langergraber, K. E., & Zuberbühler, K. (2014). Triadic social interactions

582   operate across time: a field experiment with wild chimpanzees. *Proceedings of the Royal Society B:*

583   *Biological Sciences*, *281*(1779), 20133155. https://doi.org/10.1098/rspb.2013.3155

584