

1 **Impact of dominance rank specification in dyadic interaction models**

2 Mielke, Alexander ^{1,2,3}

3 1 School of Psychology and Neuroscience, University of St Andrews, UK

4 2 Primate Models for Behavioural Evolution Lab, School of Anthropology and Museum Ethnography,
5 Oxford, UK

6 3 Taï Chimpanzee Project, Centre Suisse de Recherches Scientifiques en Côte d'Ivoire, Abidjan, Côte
7 d'Ivoire

8

9 **Abstract**

10 Dominance rank is a vital descriptor of social dynamics in animal societies and regularly used in
11 studies to explain observed interaction patterns. However, researchers can choose between
12 different indices and standardizations, and can specify dyadic rank relations differently when
13 studying interaction distributions. These researcher degrees of freedom potentially introduce biases
14 into studies and reduce replicability. Here, I demonstrate the impact of researcher choices by
15 comparing the performance of different combinations of rank index, standardization, and model
16 specification when explaining dyadic interaction patterns in sooty mangabeys (*Cercocebus atys atys*).
17 I show that while no combination consistently performed best across interaction types (aggression,
18 grooming, proximity, supplants), model specifications allowing for non-linear patterns performed
19 better than other models on average. Choices made in pre-processing and model building impacted
20 model performance and subsequent interpretation of results. Researchers could end up describing
21 social systems differently based on the same data. These results highlight the impact of researcher
22 choices in the processing of behavioural data and potential limitations when using indirect species
23 comparisons in animal behaviour research. To increase repeatability, researchers could make the

- 24 impact of their processing choices more transparent and report results using a variety of indices and
- 25 model specifications.

26 Introduction

27 Dominance hierarchies are one of the most conspicuous structural features of animal societies
 28 (Dugatkin & Earley, 2004). Calculating dominance hierarchies accurately, determining their linearity
 29 and temporal structure, and understanding how they influence behaviour and fitness has thus long
 30 been central to the study of animal behaviour (Drews, 1993). Dominance rank, as a measure of the
 31 way competitive advantage is distributed within a group, is used frequently as a predictor or
 32 outcome variable in models describing why animals act the way they do (Sánchez-Tójar et al., 2018).
 33 Different indices for calculating dominance hierarchies are available to researchers, with different
 34 assumptions about the nature of dominance and the underlying interaction distribution (Albers & de
 35 Vries, 2001; Bayly et al., 2006; David, 1987; Douglas et al., 2017; Goffe et al., 2016; Netto et al.,
 36 1993; Neumann et al., 2011; Vilette et al., 2019). For example, indices can assume linearity of
 37 hierarchies or not (Douglas et al., 2017), account for temporal change or not (Albers & de Vries,
 38 2001), and can weigh interaction intensity differently (Newton-Fisher, 2017). Once an index has
 39 been calculated, it can be standardised to make it more meaningful for the system under
 40 consideration, for example by creating an ordinal rank hierarchy or accounting for group size (Levy
 41 et al., 2020). Researchers want to make the correct choices when determining which values to
 42 include in their analyses; however, the large variety of solutions in circulation indicates that no
 43 optimal solution for all species and circumstances exists. The existing diversity of researcher choices
 44 poses risks to the replicability of studies (Hoffmann et al., 2021).

45 The uncertainty for researchers increases when studying the impact of dominance rank on
 46 interactions between individuals. Here, not only the individual rank is of interest; we often want to
 47 represent how the dominance ranks of both sender and receiver determine interaction patterns.
 48 Importantly, rank can influence partner choice in diverse ways: individuals can choose a partner
 49 because of the partner's rank, the simple fact that they are higher-ranking than the partner, the
 50 distance between ranks, or a non-linear interaction between their own rank and that of the partner.

51 These relationships can be captured using different specifications in statistical models: for example,
 52 we can include a fixed effect for each individual rank in a linear model, or the rank difference
 53 between them. But just as each dominance index makes assumptions about the underlying
 54 distribution of rank (Douglas et al., 2017), any decision about how individual and dyadic ranks are
 55 included in models makes assumptions about power dynamics in the group. When fitting statistical
 56 models, we decide on one way to represent power relations to describe variation in our outcome
 57 variable as accurately as possible. However, model specifications might lead researchers to different
 58 conclusions about their study system. We do not currently know the potential impact of pre-
 59 processing and analytical choices on interpretations of dyadic interaction patterns. In this study, I
 60 apply diverse ways to define dominance rank in statistical models to the same dataset of sooty
 61 mangabey (*Cercocebus atys atys*) interaction patterns to explore how differences in pre-processing
 62 and model specifications influence conclusions about the social system.

63 Researchers face two competing responsibilities to increase the credibility of their work: on one
 64 hand, they should describe the dominance structure of their study system as accurately as possible;
 65 on the other hand, their results should be replicable and comparable with the existing literature
 66 (Fraser et al., 2020). Given the variety of choices researchers can make in this context, it is easy to
 67 see that replicability between studies can suffer (Ihle et al., 2017). Currently, researcher choices are
 68 mainly hidden – readers are presented with one combination for dominance index, standardisation,
 69 and model specification. Different iterations that might have been tried in pre-processing but were
 70 not chosen might not be transparently reported (O’Dea et al., 2021), inflating the reported expected
 71 number of incorrect rejections in a frequentist framework (Lonsdorf et al., 2019; Wicherts et al.,
 72 2016). If results were conditional on researcher choices, comparing independent studies becomes
 73 strenuous. For example, when studying of infant handling in primates, ranks can be described using
 74 the rank difference between two individuals (Henzi & Barrett, 2002), ‘higher/lower-ranking than
 75 mother’ categories (Silk, 1999), or absolute rank distance (Kubenova et al., 2016). While the authors
 76 of each individual study chose their methodology with a specific hypothesis in mind, we do not know

77 whether differences between studies show species differences or reflect analytical choices. Certain
78 model specifications might also limit possible interpretations: for example, using absolute rank
79 distance assumes symmetrical effects and precludes different effects for high- and low-ranking
80 group members. A researcher designing a new study on the same subject would have to decide
81 which approach to replicate, and why.

82 While presenting the impact of different choices in this article, I argue for transparent and open
83 reporting of different analytical pathways using a ‘multiverse analysis’ approach for future studies
84 (Hoffmann et al., 2021; Steegen et al., 2016). Rather than choosing one rank
85 index/standardization/model specification combination (e.g., proportional Elo index entered as main
86 effects), researchers could consider calculating and reporting all possible combinations, to show that
87 their interpretation is not conditional on the choices they made (Lonsdorf et al., 2019). This
88 transparent approach has been highlighted as a viable strategy to handle the uncertainty arising
89 from researcher degrees of freedom in the scientific process when multiple analytical choices are
90 available (Hoffmann et al., 2021): as uncertainty cannot be removed, an honest approach can
91 counter selective reporting and increase comparability across studies. As data pre-processing and
92 analysis pipelines become more accessible with open data and scripts, including multiple results to
93 rule out conditionality of interpretations on researcher degrees of freedom becomes increasingly
94 feasible.

95 In this study, I compare the impact of different processing steps on the analysis of dyadic interaction
96 rates, by varying dominance indices, index standardisation, and model specification. Optimally,
97 these factors would have little influence, and choices along the data analysis pipeline would
98 minimally affect results and interpretations (O’Dea et al., 2021). However, the hypothesis for this
99 article is that different model specifications considerably influence model fit and interpretation
100 across different social interaction types (aggressions, spatial proximity, feeding supplants,
101 grooming). Recent studies (Balasubramaniam et al., 2013; Bayly et al., 2006; Douglas et al., 2017;

Gammell et al., 2003; Goffe et al., 2018; Newton-Fisher, 2017; Vilette et al., 2019) have compared and/or improved several dominance rank indices. Here, I focus on two commonly used dominance indices – David’s Score (David, 1987) and (optimised) Elo rating (Foerster et al., 2016) to test whether this choice influences results. I compared different standardizations (raw index values, proportional ranks, ordinal ranks) because this choice has recently been shown to influence model outcomes in primate studies (Levy et al., 2020; Schino & Lasio, 2019), but insufficient information still exists on whether and how they influence models that represent interaction rates between dyads of individuals within groups. My focus lies on how the power relations between the individuals in a dyad are represented when fitting models, given that no systematic information on the impact of this choice is currently available. My prediction is that different model specifications of rank differences vary in their prediction error of real-world primate data and lead to different interpretations of the impact of dominance rank on social interactions.

Methods

Data Set

I used data collected on adult female sooty mangabey social interactions in Tai Forest National Park, Cote d’Ivoire, by trained field assistants and me between January 2014 and June 2017 as part of the ongoing long-term data collection for the Tai Chimpanzee Project (Gba et al., 2019; Mielke et al., 2017, 2018, 2020). I selected three 1-year blocks of data on aggression (n = 1759 events), feeding supplants (n = 1218 events), grooming (n = 3252 events), and spatial proximity within 1m (n = 4867 events) for all adult females (above 4.5 years) present in the group for the entirety of each block (total: 25 females; range: 17-18 females per block). I only use adult females, as their hierarchies are clearly linear (Range & Noë, 2002) and they interact at relatively high rates, while the connection between male and female hierarchies is less clear and males interact at low rates in this group (Mielke et al., 2017). For each possible combination of sender and receiver, I determined the

number of interactions displayed from one to the other. We previously showed that the aggression and supplant data are internally consistent and therefore probably have low measurement error (Mielke et al., 2021). For the grooming and proximity distributions internal consistency is lower, because of low partner selectivity, the considerable number of dyads, and the relatively random distribution of proximity in this group of mangabeys (Mielke et al., 2020). For both grooming (Fruteau et al., 2011; Mielke et al., 2018; Range & Noë, 2002) and proximity (Mielke et al., 2020; Range & Noë, 2002), dominance rank has been shown to affect sooty mangabey behaviour. For the models below, the outcome variables are the number of aggression events from sender to receiver; number of supplant events from sender to receiver; minutes of grooming from sender to receiver; and occurrences of proximity within 1m between sender and receiver. In all models, offset terms were used to account for observation effort. I did not differentiate between different intensities of aggression. Proximity, in contrast to the other interaction types, was symmetrical, as no ‘sender’ and ‘receiver’ could be defined.

Dominance Indices

I calculated dominance rank using two different metrics: normalized David’s Score (David, 1987; De Vries et al., 2006) and optimized Elo ratings (Foerster et al., 2016). I chose the latter over Elo ratings without optimized k and start values as we did not observe any active rank changes between females over the course of this study (Mielke et al., 2017); thus, all changes in an individual’s rank are due to demographic processes (individuals dying or coming of age). The optimised k value for the Elo ratings was therefore confirmed and calculated as 0 for all year blocks. The Elo value at the first day of the year was chosen to represent the dominance hierarchy. Dominance hierarchies were computed using directional feeding displacements (‘supplants’), which are highly linear in this group, with only 18 out of 1218 supplants (or 0.8%) going against the established hierarchy. Normalized David’s Scores were calculated using the ‘steepness’ package in R (Leiva & De Vries, 2014). Elo

ratings were calculated using the ‘Model 3’ script provided by Foerster et al. 2016. Both indices assume linearity of hierarchy, which was given in the female mangabeys. For both normalized David’s Scores and Elo ratings, I applied different standardisations: the raw values, the ordinal hierarchy based on the raw values (highest-ranking individual has rank 1, second highest has rank 2 etc.), and the proportional rank standardised between 0 and 1, with 1 being the highest-ranking individual (Levy et al., 2020). Ordinal and proportional ranks assume equidistance of ranks.

Model Specifications

Here, I focus on six ways of representing two rank terms in statistical models of dyadic interactions, all of which have been used for animal behaviour research (Table 1). I only used numerical representations of dominance rank – categorical representations (e.g., ‘high/medium/low’) are still in use (Rosati et al., 2020) but reduce information and add even further researcher degrees of freedom (e.g., about the number of categories and cut-offs). All models include the sender rank as a main effect. Main Effects model: I included both ranks as main effects, which informs us about the variation in partner choice due to whether the sender is high or low in rank and the receiver is high or low in rank but does not represent the relation between the two. Factor Higher-ranking: I fitted a model using a factorial term defined by whether the sender was higher- or lower-ranking than the receiver (based on the raw Elo index), which captures one simple aspect of their relationship but omits the rank distance and the actual receiver rank. Rank Difference: I fitted a model with a rank difference term (subtracting the receiver rank from the sender rank), which is positive if the sender outranks the receiver but also captures information on the distance between them. Absolute Rank Difference: The rank difference term cannot be fitted in interaction with the sender rank because the two are tightly bound to each other – how much higher or lower in rank an individual can be overall is determined by the sender rank. In contrast, the absolute rank distance term is small when individuals are close in rank, and large if they are far apart, but omits information about whether the

177 receiver outranks the sender. Interaction: I fitted one model representing the interaction term
 178 between the sender and receiver ranks. The statistical interaction between the two rank variables
 179 can represent both the direct impact of sender and receiver rank but can also account for
 180 differences in partner choice between high - and low-ranking senders, as long as these are linear: for
 181 example, high-ranking senders can supplant all group members, while low-ranking senders can
 182 supplant only low-ranking group members. Squared Interaction: Lastly, I fitted a model including the
 183 interaction of the sender rank with the squared term for the partner rank. This model can capture
 184 the main effects of both ranks, and identify linear interaction effects, but should also be able to
 185 represent nonlinear partner choice, for example if both high- and low-ranking group members
 186 preferably groom individuals who are close in rank. In the following, I use all the model terms
 187 defined in Table 1 to describe the distribution of social interactions in sooty mangabeys to
 188 determine which of these captures most of the variation in observed patterns.

189 *Table 1: Different model specifications for rank in dyadic interaction models tested in this study. +*
 190 *indicates that main effects were entered independently in the model, * indicates a statistical*
 191 *interaction term.*

Model Name	Model Terms	Information about
<i>Main Effects</i>	Rank Sender + Rank Receiver	Rank Position Sender, Rank Position Receiver
<i>Factor Higher-ranking</i>	Rank Sender + Higher-ranking yes/no	Rank Position Sender, Does the Sender outrank the Receiver?
<i>Rank Difference</i>	Rank Sender + (Rank Sender – Rank Receiver)	Rank Position Sender, Rank Position Receiver, Distance in Rank
<i>Absolute Rank Difference</i>	Rank Sender * Abs(Rank Sender – Rank Receiver)	Rank Position Sender, Distance in Rank
<i>Interaction</i>	Rank Sender * Rank Receiver	Rank Position Sender, Rank Position Receiver, Linear Connection between them
<i>Squared Interaction</i>	Rank Sender * (Rank	Rank Position Sender, Rank Position Receiver,

Receiver)²

Nonlinear connection between them

192

193 *Models*

194 I tested how different scores (modified Elo ratings, David Scores), standardisation (raw, standardised
195 0-1, ordinal ranks) and model specifications (main effects, factor higher/lower ranking, rank
196 difference, absolute rank difference, interaction term, squared interaction term) influence
197 distribution of interactions in sooty mangabeys. The response variables were, respectively,
198 aggression events, supplant events, proximity events, and minutes of grooming. For each year, the
199 dataset contained each sender-receiver combination and the number of interactions observed
200 between them in that period – i.e., each sender-receiver combination is present in the dataset up to
201 three times (once per year), and each individual is included as sender and receiver for each year with
202 each possible partner. I did not control for any other variables (e.g., age). All rank variables were z-
203 standardised before being entered into the models (Schielzeth, 2010). All models were fitted using
204 Bayesian generalized linear mixed models in the ‘brms’ package (Bürkner, 2018) in R v4.1.2 (R
205 Development Core Team & R Core Team, 2020). For the aggression and supplant model, I used zero-
206 inflated Poisson error structure, as a large number of dyads had no recorded interactions for each
207 type. For proximity, I fitted the model using Poisson error structure. For grooming, the most
208 appropriate error structure was determined to be zero-inflated negative binomial. All models
209 contained the log-transformed observation effort in hours as offset term.

210 Models were fitted using 3000 iterations on three chains. For all fixed effects in all models, I used
211 weakly informative, normally distributed priors (Lemoine, 2019) with a mean of 0 and a standard
212 deviation of 1. All models contained the year, sender identity, and receiver identity as random
213 effects. Model performance was evaluated and compared using the leave-one out cross validation
214 information criterion as available in the ‘loo’ package (Vehtari et al., 2017), as a measure of
215 prediction error of the model posterior distribution. My assumption here is that, given that the

outcome remains constant, the model with the smallest LOO-IC predicts the outcome most accurately. All models within a social interaction type were compared against each other, and I report the difference between the LOO-IC of each model with the best model in the set (deltaLOO); the best model therefore has a deltaLOO of 0. I report Bayesian R² values for the best model to give a sense of explained variance (Gelman et al., 2019). I also report the interpretation researchers would have arrived at using the fixed effect results of each of the models, to demonstrate that these choices matter for the publication process. Effects were interpreted as meaningful if the 95% credible interval of the posterior distribution did not include 0, and plots were used to establish the direction of effects.

Results

Aggression

Comparing all models for the distribution of aggression interactions in the group, the simple factor describing whether the sender was higher- or lower-ranking than the receiver had the highest prediction accuracy (Fig. 1). This is explained by the distribution of zeros – lower-ranking individuals almost never attack higher-ranking individuals in this group. The explained variance of the best model for aggression was $R^2 = 0.53$. Of the other models, everything else being equal, the Elo-based models on average outperformed the David's Score based models (Fig. 1A). The proportional and ordinal standardizations did not show consistent differences, but both outperformed the raw values (Fig. 1B). All these differences were minor compared to the impact of the model specification (Fig. 1C). Models that can represent non-linear rank distances, especially the squared interaction and interaction models, but also the model including absolute rank distance, performed considerably better than those using only main effects or the rank difference. This would considerably affect interpretations of results (Table 2): Most well-performing models found that high-ranking senders showed more aggression, low-ranking receivers were victims more often, and that this created a

241 situation where aggression was directed down the hierarchy but preferentially against closely
 242 ranked group members. The model using only main effects, modelling only whether either individual
 243 was high- or low-ranking, failed to capture these dynamics. The higher/lower rank factor and the
 244 rank difference models would have captured the down-the-hierarchy aspect, implying that all
 245 individuals indiscriminately attack lower-ranking group members. The absolute rank distance would
 246 have captured the increased aggression towards closely ranked group members but would indicate
 247 that this effect was distributed evenly for high- and low-ranking senders. The interaction and
 248 squared interaction models allowed for more complex interpretations and different patterns at
 249 different points in the hierarchy, but interpretations were also more ambiguous based on plots
 250 alone. Here, it seemed that low-ranking individuals receive more aggression because they are
 251 victims both of other low-ranking and high-ranking individuals, while low-ranking never attack high-
 252 ranking individuals.

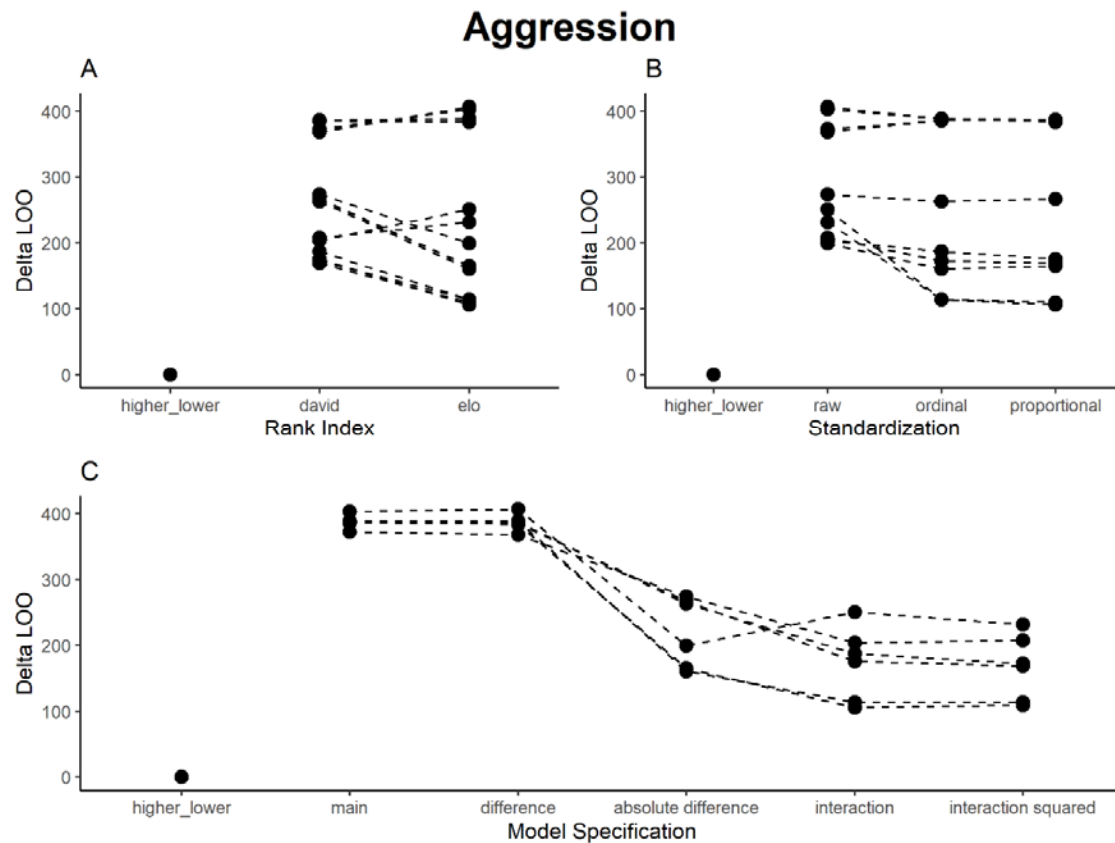


Figure 1: Results of model comparison for Aggression interactions. The y-axis portrays delta LOO-ICs (distance to best model) - the best model is therefore set to 0, and higher scores indicate poorer performance. Points indicate models (split by the respective category), while lines connect otherwise comparable models (e.g., the main effects ordinal scale models for both Elo ratings and David's Scores).

Table 2: Interpretation of Elo results for Aggression. Model interpretations are displayed for all models using Elo ratings only to facilitate interpretation. Downwards arrows indicate that lower-ranking individuals show higher rates, upward arrows indicate that higher-ranking individuals show higher rates. Interactions can show Down The Hierarchy (DTH; Targeting lower-ranking individuals), Closely-Ranked Receiver (CRR; Targeting individuals with similar rank); Up The Hierarchy (UTH; Targeting higher-ranking individuals), or a mix of those.

Model	Standardization	DeltaLOO	Sender Rank	Receiver Rank	Interaction
Factor Higher-ranking	-	0	↓	-	DTH
Main Effects	Raw	403.3	↑	-	-
	Ordinal	388.6	-	↑	-
	Proportional	387.5	-	↑	-
Rank Difference	Raw	407.0	↑	-	-
	Ordinal	388.7	-	-	DTH
	Proportional	383.9	-	-	DTH

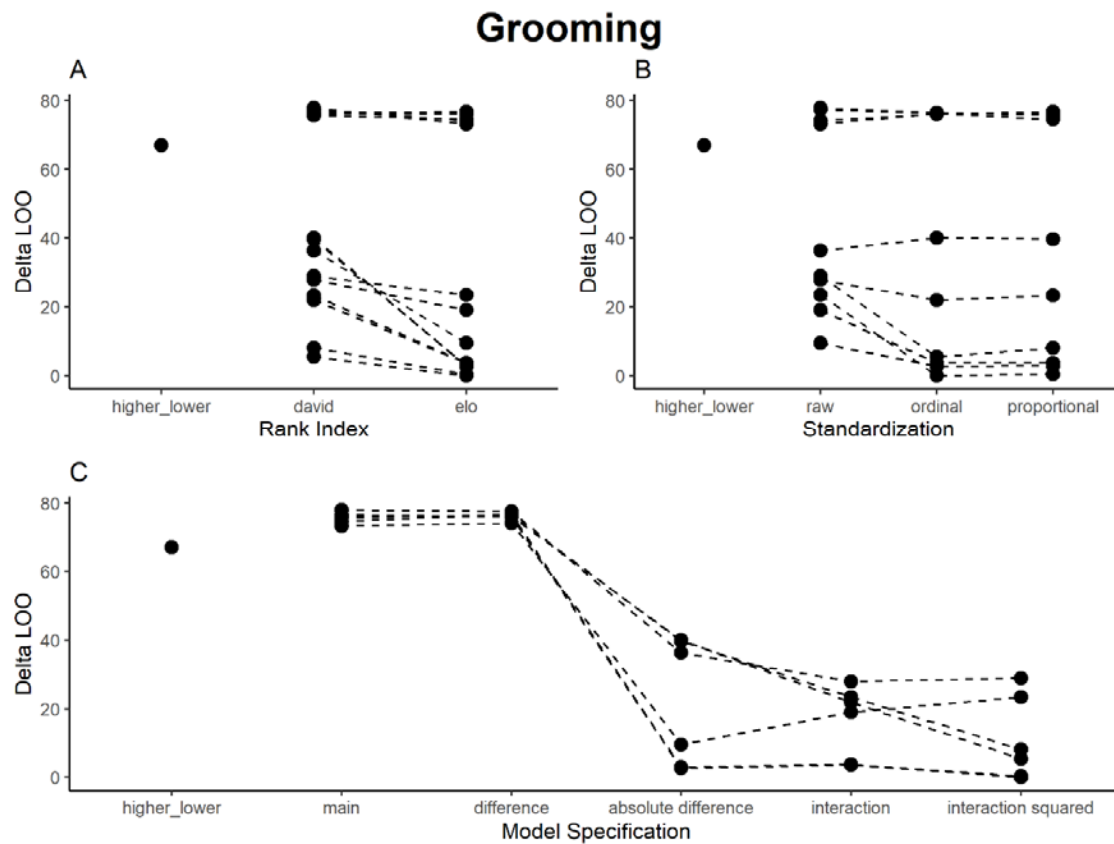
<i>Absolute Rank Difference</i>	Raw	199.8	↑	-	CRR (especially high-ranking sender)
	Ordinal	160.4	↑	-	CRR (especially high-ranking sender)
	Proportional	165.4	↑	-	CRR (especially high-ranking sender)
<i>Interaction</i>	Raw	250.8	↑	↓	High-ranking: CRR and DTH; Low-ranking: CRR
	Ordinal	101.8	↑	↓	High-ranking: CRR and DTH; Low-ranking: CRR
	Proportional	106.4	↑	↓	High-ranking: CRR and DTH; Low-ranking: CRR
<i>Squared Interaction</i>	Raw	232.0	-	↓	High-ranking: CRR and DTH; Low-ranking: CRR
	Ordinal	106.4	-	↓	High-ranking: CRR and DTH; Low-ranking: CRR
	Proportional	109.7	-	↓	High-ranking: CRR and DTH; Low-ranking: CRR

265

266 *Grooming*

267 For grooming, models using Elo ratings generally performed better than those using David's Scores
268 (Fig. 2A), and again raw values performed worse than both ordinal and proportional values (Fig. 2B).
269 The best models (with small differences between them) were those that could quantify closeness in
270 rank (Fig. 2C): the absolute rank difference (both proportional and ordinal) and the (squared)
271 interaction terms. The explained variance of the best model for grooming was $R^2 = 0.19$. Main
272 effects models, the higher/lower factor, and the rank distance consistently performed worse than
273 other models. Again, using different models would have led to different interpretations. The
274 consensus was that there was a tendency to groom individuals who are close in rank and that lower-
275 ranking individuals additionally groomed up the hierarchy (Table 3). Main effects models, rank
276 distance, and the higher/lower factor model showed that high-ranking individuals received more
277 grooming, indicating that grooming went up the hierarchy. The other models (absolute rank
278 difference, interaction and squared interaction models) indicated grooming of those close in rank,
279 with the squared interaction model also indicating that low-ranking individuals sent more grooming
280 and preferred higher-ranking partners.

281



282

283 *Figure 2: Results of model comparison for Grooming interactions. The y-axis portrays delta LOO-ICs (distance to best model)*
 284 *- the best model is therefore set to 0, and higher scores indicate poorer performance. Points indicate models (split by the*
 285 *respective category), while lines connect otherwise comparable models (e.g., the main effects ordinal scale models for both*
 286 *Elo ratings and David's Scores).*

287

288 *Table 3: Interpretation of Elo results for Grooming. Model interpretations are displayed for all models*
 289 *using Elo ratings (similar results for David's Scores). Downward arrows indicates that lower-ranking*
 290 *individuals show higher rates, upward arrows indicates that higher-ranking individuals show higher*
 291 *rates. Interactions can show Down The Hierarchy (DTH; Targeting lower-ranking individuals), Closely-*
 292 *Ranked Receiver (CRR; Targeting individuals with similar rank); Up The Hierarchy (UTH; Targeting*
 293 *higher-ranking individuals), or a mix of those.*

Model	Standardization	DeltaLOO	Sender Rank	Receiver Rank	Interaction
Factor Higher-ranking	-	67.4	-	-	UTH
Main Effects	Raw	72.8	-	↑	-
	Ordinal	75.8	-	↑	-
	Proportional	74.2	-	↑	-
Rank Difference	Raw	73.7	-	-	UTH

	Ordinal	76.0	-	-	UTH
	Proportional	76.3	-	-	UTH
<i>Absolute Rank Difference</i>	Raw	9.2	-	-	CRR
	Ordinal	2.6	-	-	CRR
	Proportional	2.5	-	-	CRR
<i>Interaction</i>	Raw	18.6	-	-	CRR
	Ordinal	8.9	-	-	CRR
	Proportional	3.4	-	-	CRR
<i>Squared Interaction</i>	Raw	23.1	↓	↑	High-ranking: CRR; Low-ranking: CRR and UTH
	Ordinal	7.1	↓	↑	High-ranking: CRR; Low-ranking: CRR and UTH
	Proportional	0	↓	↑	High-ranking: CRR; Low-ranking: CRR and UTH

294

295 Proximity

296 The proximity data were symmetrical ($A \rightarrow B \Rightarrow B \rightarrow A$), allowing us to test how models cope with
297 this additional oddity in the data. For proximity, models using David's Scores and Elo ratings
298 performed roughly similar, with a slight advantage for the latter (Fig. 3A), and again raw values
299 performed worse than proportional values, which in turned performed worse than ordinal values
300 (Fig. 3B). In fact, across models, raw values detected different patterns than the two
301 standardizations (Table 4), and anyone basing their interpretation on those results would describe a
302 different social system. In raw values, the sender and receiver main effects both had a strong
303 positive effect (high-ranking senders and receivers show higher levels of proximity), while for both
304 standardizations, the two main effects had strong negative effects. The raw value models indicated a
305 preference for closely ranked partners for proximity across sender ranks, while the other models
306 indicated a preference for closely-ranked partners in low- and medium-ranking individuals, while
307 high-ranking individuals showed no clear preference. This might be because I removed males from
308 the analysis. The different result for raw values is concerning, given the raw values were highly
309 correlated (>0.9) with both standardizations across indices. Given the inferior performance of the
310 raw values in the model comparison, these models likely fail to predict the actual group patterns,
311 but this would not be apparent to a researcher focusing exclusively on these values. The best model

to detect proximity was the squared interaction model for the ordinal Elo ratings (explained variance: $R^2 = 0.60$), followed by the same model for David's Scores (Fig. 3C). Main effects, rank distance, and the higher/lower factor again performed poorly in general – not surprisingly, given that the proximity data were symmetrical. Despite the symmetry, the higher/lower factor and rank difference would have indicated directional effects. The absolute difference also performed worse than the interaction models.

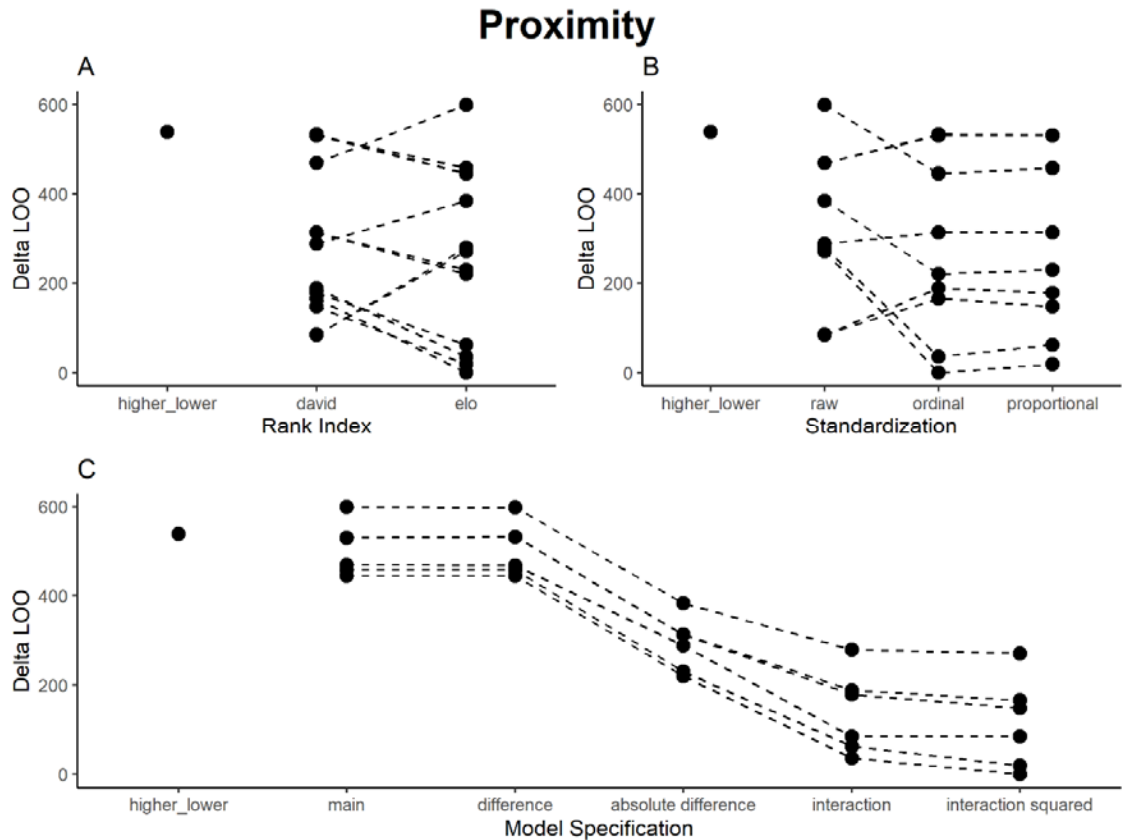


Figure 3: Results of model comparison for Proximity interactions. The y-axis portrays delta LOO-ICs (distance to best model) - the best model is therefore set to 0, and higher scores indicate poorer performance. Points indicate models (split by the respective category), while lines connect otherwise comparable models (e.g., the main effects ordinal scale models for both Elo ratings and David's Scores).

Table 4: Interpretation of Elo results for Proximity. Model interpretations are displayed for all models using Elo ratings (similar results for David's Scores). Downward arrows indicate that lower-ranking individuals show higher rates, upward arrows indicate that higher-ranking individuals show higher

328 rates. Interactions can show Down The Hierarchy (DTH; Targeting lower-ranking individuals), Closely-
329 Ranked Receiver (CRR; Targeting individuals with similar rank); Up The Hierarchy (UTH; Targeting
330 higher-ranking individuals), or a mix of those.

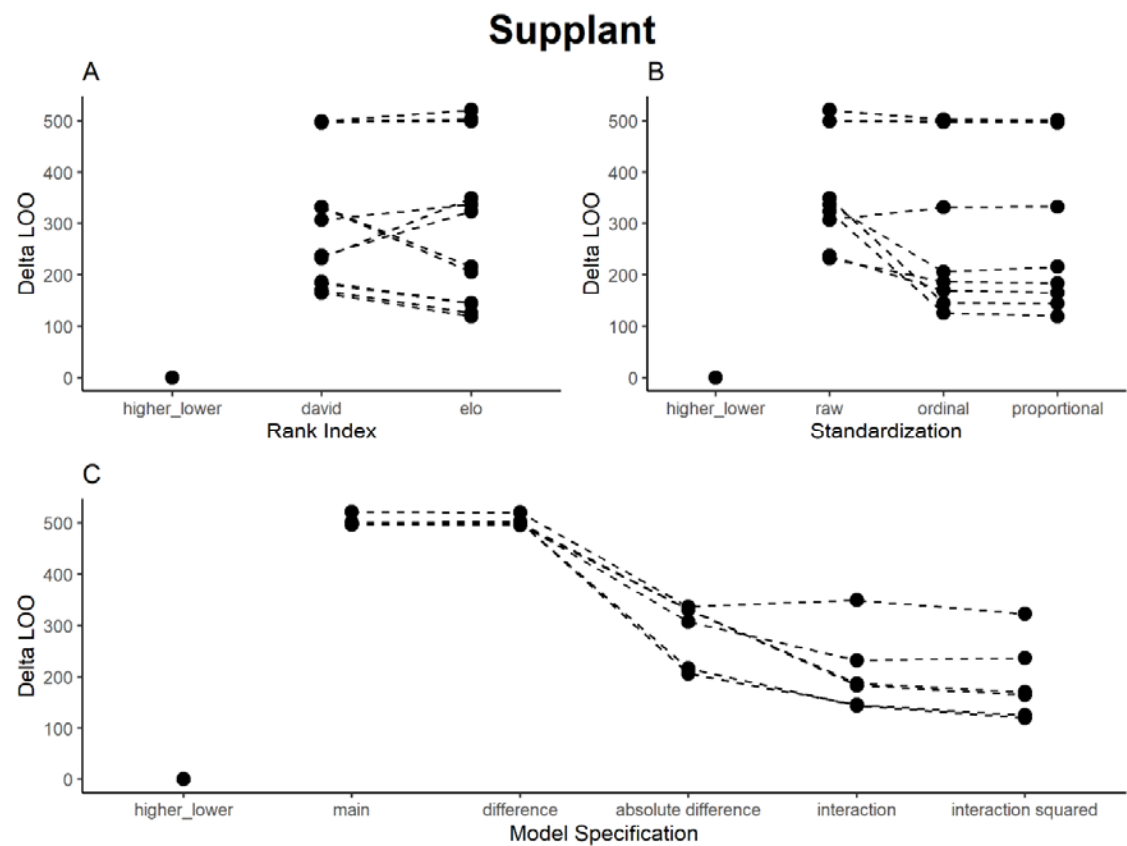
Model	Standardization	DeltaLOOs	Sender Rank	Receiver Rank	Interaction
<i>Factor Higher-ranking</i>	-	517.5	↓	-	DTH
<i>Main Effects</i>	Raw	580.1	↑	↑	-
	Ordinal	421.2	↓	↓	-
	Proportional	439.0	↓	↓	-
<i>Rank Difference</i>	Raw	579.3	↑	-	DTH
	Ordinal	424.7	↓	-	UTH
	Proportional	438.5	↓	-	UTH
<i>Absolute Rank Difference</i>	Raw	365.1	↑	-	CRR
	Ordinal	203.3	↓	-	CRR (stronger for low-ranking)
	Proportional	211.3	↓	-	CRR (stronger for low-ranking)
<i>Interaction</i>	Raw	260.5	↑	↑	CRR
	Ordinal	40.3	↓	↓	CRR for low-ranking
	Proportional	43.4	↓	↓	CRR for low-ranking
<i>Squared Interaction</i>	Raw	251.8	↑	↑	CRR
	Ordinal	25.7	↓	↓	CRR for low-ranking
	Proportional	0	↓	↓	CRR for low-ranking

331

332 *Supplants*

333 For supplants, models using Elo ratings generally performed better than David's Scores (Fig. 4A), and
334 raw values performed worse than both ordinal and proportional values (Fig. 4B). However, as in
335 aggression, the model with a simple factor denoting that the sender was higher-ranking performed
336 best (explained variance: $R^2 = 0.57$) – not surprisingly, given that supplants are used to make the
337 dominance hierarchy. Among the other models, those that encode that supplants generally go down
338 the hierarchy and differences in patterns across the dominance hierarchy – interactions and squared
339 interactions – fared better than the absolute difference, rank difference, and main effects models
340 (Fig.4C). The best models (apart from the higher/lower factor model) were the squared interaction
341 models of proportional and ordinal David's Scores and Elo ratings. Results closely resembled the

aggression models: Main effects, the higher/lower factor, and the rank difference would have indicated supplants going down the hierarchy, the absolute rank difference would have indicated supplants to those close in rank, and the interaction and squared interaction models captured a mix of the two effects (Table 5).



346

Figure 4: Results of model comparison for Supplant interactions. The y-axis portrays delta LOO-ICs (distance to best model) - the best model is therefore set to 0, and higher scores indicate poorer performance. Points indicate models (split by the respective category), while lines connect otherwise comparable models (e.g., the main effects ordinal scale models for both Elo ratings and David's Scores).

351

Table 5: Interpretation of Elo results for Supplants. Model interpretations are displayed for all models using Elo ratings (similar results for David's Scores). Downward arrows indicate that lower-ranking individuals show higher rates, upward arrows indicate that higher-ranking individuals show higher rates. Interactions can show Down The Hierarchy (DTH; Targeting lower-ranking individuals), Closely-Ranked Receiver (CRR; Targeting individuals with similar rank); Up The Hierarchy (UTH; Targeting higher-ranking individuals), or a mix of those.

Model	Standardization	DeltaLOO	Sender Rank	Receiver Rank	Interaction
-------	-----------------	----------	-------------	---------------	-------------

<i>Factor Higher-ranking</i>	-	0	↑	-	DTH
<i>Main Effects</i>	Raw	521.4	-	-	-
	Ordinal	501.6	-	↓	-
	Proportional	498.8	↑	↓	-
<i>Rank Difference</i>	Raw	520.9	-	-	-
	Ordinal	502.2	-	-	DTH
	Proportional	501.5	-	-	DTH
<i>Absolute Rank Difference</i>	Raw	336.8	↑	-	CRR (especially high-ranking sender)
	Ordinal	205.9	↑	-	CRR (especially high-ranking sender)
	Proportional	216.1	-	-	CRR (especially high-ranking sender)
<i>Interaction</i>	Raw	349.5	-	↓	High-ranking: CRR and DTH; Low-ranking: CRR
	Ordinal	145.4	↑	↓	High-ranking: CRR and DTH; Low-ranking: CRR
	Proportional	143.9	↑	↓	High-ranking: CRR and DTH; Low-ranking: CRR
<i>Squared Interaction</i>	Raw	323.1	↑	↓	High-ranking: CRR and DTH; Low-ranking: CRR
	Ordinal	125.5	↑	↓	High-ranking: CRR and DTH; Low-ranking: CRR
	Proportional	119.5	↑	↓	High-ranking: CRR and DTH; Low-ranking: CRR

358

359

360 Discussion

361 My results showed that the choice of dominance rank index, standardization, and model
362 specification matter when interpreting how ranks affect dyadic interaction patterns in primate
363 societies. This is true for all models using dominance rank in some form – it becomes even more
364 central for dyadic models, which have to capture the interactional nature of animal social life. For all
365 interaction types, the results showed that different researchers, using the same data, would have
366 reached different interpretations of the social life of sooty mangabeys. Only a handful of studies
367 exist for most animal species, comparisons between species are usually indirect, and results across

studies are aggregated to make higher-level statements about evolutionary forces underlying sociality (e.g., socio-ecological models, Koenig et al., 2013). Without sufficient control, there might be a risk of creating non-replicable results that come to represent what we know about a species and social evolution more broadly (O'Dea et al., 2021). While I make some of the choices in this study explicit and show their impact, other researcher choices remained fixed here but certainly influenced results (such as the exclusion of certain age/sex classes; Fedurek & Lehmann, 2017).

Studies of animal sociality and evolution are, at their core, comparative: we study one species with the goal of understanding underlying evolutionary processes across species. For this, it is vital that results across studies are comparable (Fraser et al., 2020; Ihle et al., 2017). A positive note of this study is that, while Elo ratings and David's scores differed in predictive power, they would have led to a similar interpretation of results. The same was true for the use of proportional and ordinal dominance ranks (while using raw values would have led to different results). Thus, at least for systems with clearly linear hierarchy and sufficient data, these choices might only weakly influence interpretation. They might however still be relevant in cases where one of them leads to significant results in a frequentist framework, while the others do not, and researchers select reported models based on this hidden multiple comparison (Wicherts et al., 2016). More worrisome, I found that differences in model specification caused considerable differences in results and interpretation. Most model specifications can only represent one type of relationship. For example, a researcher using the absolute rank difference would find that all interaction types in sooty mangabeys are directed at closely ranked group members. A researcher using the simple rank difference would find that aggression and supplants go down the hierarchy, while grooming goes up the hierarchy, and would not be able to make meaningful statements about proximity at all. Most likely, these effects are both present in most systems, but we would class the same social system differently based on these results.

In my comparisons, no combination of index/standardization/model specification outperformed the others throughout – however, some choices consistently performed worse than others did. Raw Elo ratings and David’s Scores performed poorly, indicating that the distance between individual values generated by those scores was not meaningful, with equidistance the most parsimonious assumption. Main effects and simple rank distance models also performed poorly throughout – often, these choices had poor predictive accuracy and failed to detect the most likely pattern (based on models with higher predictive power). The absolute rank distance and higher/lower factor models performed well for a subset of the models but are limited in the information they could encode – they will invariably lead to the same, simple interpretation, no matter the underlying data. The higher/lower factor even ‘found’ a directional pattern in symmetrical proximity data. The interaction and squared interaction terms were able to detect relatively complex patterns and performed reasonably well for all interaction types, so if researchers would want to apply one model specification without previous knowledge of which patterns to expect, interaction terms would be the best choice. This is largely due to non-linear patterns in interaction distributions, probably arising out of the interplay of competition and kinship patterns (Seyfarth, 1977). However, it is hard to interpret (squared) interaction terms unambiguously.

Neither Elo ratings nor David’s Scores performed better across the board, and the same was true for ordinal or proportional values. Often, within the same comparison, some ordinal models performed better than the proportional models while some performed worse, with no clear pattern explaining these differences (especially given that the two standardizations correlated more than -0.99). This highlights the danger of using model comparisons to identify the ‘best predictor’ for any given distribution (Levy et al., 2020): once we complexify the picture, it soon becomes hard to find meaningful explanations for the observed patterns, and any additional choice could upend the interpretation. Recently, studies have interpreted the difference in model performance using different rank variables or standardizations as a sign that the power structures or individual decisions within the group make the same assumptions as that index (Levy et al., 2020; Schino &

Lasio, 2019). However, these conclusions have to be drawn with care: comparisons between models indicate which model has lower error in-sample, but they do not provide evidence that either model is particularly good at representing the social structure or decision-making of a group (Appleby, 1993). Higher predictive accuracy does not allow us to assume causation (McElreath, 2018). For example, most indices we currently use assume linear hierarchies (Douglas et al., 2017), which is often not assessed (Schino & Lasio, 2019). We do not, currently, know how much sampling and measurement errors bias dominance indices and their performance. Most indices and standardizations are correlated very highly (Vilette et al., 2019), especially when sufficient data are available and the linearity assumption is fulfilled, so any difference in their performance could be the result of random variation of few data points. In my models, the higher/lower factor often performed best, but would entirely fail to predict aggression or supplant patterns within the subset of dyads where the sender was higher-ranking.

For the mangabeys, these results paint a complex picture of the impact of rank on social interactions. The results are broadly in line with earlier studies for this species (Fruteau et al., 2011; Mielke et al., 2018, 2020, 2021; Range & Noë, 2002, 2005). My previous use of absolute rank differences to describe preferred association patterns (Mielke et al., 2020) might have omitted important information about differences in social behaviour between individuals of different ranks. In this study, the effect we previously found for female-female association patterns (preferred spatial association with closely-ranked group members) only held for lower-ranking group members – possibly because high-ranking females associate more closely with males, which were omitted here (Mielke et al., 2020). Social interactions (both socio-positive and negative) occurred largely with closely ranked group members. Aggression and supplants were directed down the hierarchy by high-ranking group members, while low-ranking individuals groomed up the hierarchy. The grooming results (strong preference for close rank, more pronounced in high-ranking individuals) are a replication of earlier results using a different analytical approach (Mielke et al., 2018) and are in line with some predictions for cercopithecine monkeys with similar social systems (Seyfarth, 1977).

However, models including other factors would be necessary to determine whether the preference for closely ranked individuals is the result of genuine preference for closely-ranked partners, attraction to kin, reciprocity, spatial proximity, or priority of access.

Given that many of the steps taken here are rarely described in detail in ecological studies, and data and scripts are still mostly unavailable in this field (Culina et al., 2020), we face further challenges to replicability. One way to counter these problems to replicability would be to report all possible dominance rank index/standardization/model specification combinations in some form of ‘multiverse’ analysis (Hoffmann et al., 2021; Steegen et al., 2016). Given that analyses these days are done using some statistical software, repeating the analyses with all possible combinations does not in itself pose a computational problem (even though it might dramatically increase computation times). This approach has the advantage of increasing transparency and making results comparable across studies because researchers can compare the same model specifications against each other, rather than different specifications. Interpretation can become more difficult and ambiguous, as I have shown in this study – at the same time, researchers can demonstrate that the interpretation they are presenting is not conditional on the choices they made in the data pipeline (O’Dea et al., 2021). Further developing this framework will be an ongoing process to improve research in ecology and evolutionary biology (Hoffmann et al., 2021; O’Dea et al., 2021).

462 **Funding**

463 This research was supported by the Max Planck Society, the Wenner-Gren Foundation, British
464 Academy, and Leverhulme Trust. Research at the Tai Chimpanzee Project has been funded by the
465 Max Planck Society since 1997.

466 **Acknowledgements**

467 The authors thank the Ivorian Ministry of Environment and Forests and Ministry of Higher Education
468 and Scientific Research and the Office Ivoirien des Parcs et Reserves of Côte d'Ivoire. I thank Simon
469 Kannieu, Daniel Bouin, Gnimion Florent, and the team of the TCP for field work support and data
470 collection. Many thanks go to Roman M Wittig and Catherine Crockford for making this research
471 possible and comments on the manuscript, and to Delphine de Moor for helpful comments.

472

473 **Data Availability**

474 Scripts and Data are available here: <https://github.com/AlexMielke1988/Mielke-Mangabey-Ranks>

475 **References**

- 476 Albers, P. C. H., & de Vries, H. (2001). Elo-rating as a tool in the sequential estimation of dominance
477 strengths. *Animal Behaviour*, 61(2), 489–495. <https://doi.org/10.1006/anbe.2000.1571>
- 478 Appleby, M. C. (1993). How animals perceive a hierarchy: Reactions to Freeman et al. *Animal*
479 *Behaviour*, 46(6), 1232–1233. <https://doi.org/10.1006/anbe.1993.1316>
- 480 Balasubramaniam, K. N., Berman, C. M., de Marco, A., Dittmar, K., Majolo, B., Ogawa, H., Thierry, B.,
481 & de Vries, H. (2013). Consistency of dominance rank order: A comparison of David's scores
482 with I&SI and Bayesian methods in macaques. *American Journal of Primatology*, 75(9), 959–
483 971. <https://doi.org/10.1002/ajp.22160>
- 484 Bayly, K. L., Evans, C. S., & Taylor, A. (2006). Measuring social structure: A comparison of eight
485 dominance indices. *Behavioural Processes*, 73(1), 1–12.
486 <https://doi.org/10.1016/j.beproc.2006.01.011>
- 487 Bürkner, P. C. (2018). Advanced Bayesian multilevel modeling with the R package brms. *R Journal*,
488 10(1), 395–411. <https://doi.org/10.32614/rj-2018-017>
- 489 Culina, A., van den Berg, I., Evans, S., & Sánchez-Tójar, A. (2020). Low availability of code in ecology:
490 A call for urgent action. *PLOS Biology*, 18(7), e3000763.
491 <https://doi.org/10.1371/journal.pbio.3000763>
- 492 David, H. A. (1987). Ranking from unbalanced paired-comparison data. *Biometrika*, 74(2), 432–436.
493 <https://doi.org/10.1093/biomet/74.2.432>
- 494 De Vries, H., Stevens, J. M. G., & Vervaecke, H. (2006). Measuring and testing the steepness of
495 dominance hierarchies. *Animal Behaviour*, 71(3), 585–592.
496 <https://doi.org/10.1016/j.anbehav.2005.05.015>
- 497 Douglas, P. H., Ngonga Ngomo, A. C., & Hohmann, G. (2017). A novel approach for dominance
498 assessment in gregarious species: ADAGIO. In *Animal Behaviour* (Vol. 123, pp. 21–32).
499 Academic Press. <https://doi.org/10.1016/j.anbehav.2016.10.014>
- 500 Drews, C. (1993). The concept and definition of dominance in animal behaviour. *Behaviour*, 125(3–
501 4), 283–313. <https://doi.org/10.1163/156853993X00290>
- 502 Dugatkin, L. A., & Earley, R. L. (2004). Individual recognition, dominance hierarchies and winner and
503 loser effects. *Proceedings of the Royal Society B: Biological Sciences*, 271(1547), 1537–1540.
504 <https://doi.org/10.1098/rspb.2004.2777>
- 505 Fedurek, P., & Lehmann, J. (2017). The effect of excluding juveniles on apparent adult olive baboons
506 (*Papio anubis*) social networks. *PLoS ONE*, 12(3), e0173146.
507 <https://doi.org/10.1371/journal.pone.0173146>
- 508 Foerster, S., Franz, M., Murray, C. M., Gilby, I. C., Feldblum, J. T., Walker, K. K., & Pusey, A. E. (2016).
509 Chimpanzee females queue but males compete for social status. *Scientific Reports*, 6(1), 1–11.
510 <https://doi.org/10.1038/srep35404>
- 511 Fraser, H., Barnett, A., Parker, T. H., & Fidler, F. (2020). The role of replication studies in ecology.
512 *Ecology and Evolution*, 10(12), 5197–5207. <https://doi.org/10.1002/ece3.6330>

- 513 Fruteau, C., Lemoine, S., Hellard, E., van Damme, E., & Noë, R. (2011). When females trade grooming
514 for grooming: Testing partner control and partner choice models of cooperation in two primate
515 species. *Animal Behaviour*, 81(6), 1223–1230. <https://doi.org/10.1016/j.anbehav.2011.03.008>
- 516 Gammell, M. P., De Vries, H., Jennings, D. J., Carlin, C. M., & Hayden, T. J. (2003). David's score: A
517 more appropriate dominance ranking method than Clutton-Brock et al.'s index. In *Animal*
518 *Behaviour* (Vol. 66, Issue 3, pp. 601–605). Academic Press.
519 <https://doi.org/10.1006/anbe.2003.2226>
- 520 Gba, B. C., Bene, J.-C. K., Gone Bi, Z. B., Mielke, A., & Kone, I. (2019). Within-group spatial position
521 and activity budget of wild sooty mangabeys (*Cercocebus atys*) in Taï National Park, Côte
522 d'Ivoire. *International Journal of Biological and Chemical Sciences*, 13(7), 2991–3008.
523 <https://doi.org/10.4314/ijbcs.v13i7.2>
- 524 Gelman, A., Goodrich, B., Gabry, J., & Vehtari, A. (2019). R-squared for Bayesian Regression Models.
525 *American Statistician*, 73(3), 307–309. <https://doi.org/10.1080/00031305.2018.1549100>
- 526 Goffe, A. S., Fischer, J., & Sennhenn-Reulen, H. (2018). Bayesian inference and simulation
527 approaches improve the assessment of Elo-ratings in the analysis of social behaviour. *Methods*
528 *in Ecology and Evolution*, 9(10), 2131–2144. <https://doi.org/10.1111/2041-210X.13072>
- 529 Goffe, A. S., Zinner, D., & Fischer, J. (2016). Sex and friendship in a multilevel society: behavioural
530 patterns and associations between female and male Guinea baboons. *Behavioral Ecology and*
531 *Sociobiology*, 70(3), 323–336. <https://doi.org/10.1007/s00265-015-2050-6>
- 532 Henzi, S. P., & Barrett, L. (2002). Infants as a commodity in a baboon market. *Animal Behaviour*,
533 63(5), 915–921. <https://doi.org/10.1006/anbe.2001.1986>
- 534 Hoffmann, S., Schönbrodt, F., Elsas, R., Wilson, R., Strasser, U., & Boulesteix, A.-L. (2021). The
535 multiplicity of analysis strategies jeopardizes replicability: lessons learned across disciplines.
536 *Royal Society Open Science*, 8(4), rsos.201925. <https://doi.org/10.1098/rsos.201925>
- 537 Ihle, M., Winney, I. S., Krystalli, A., & Croucher, M. (2017). Striving for transparent and credible
538 research: Practical guidelines for behavioral ecologists. *Behavioral Ecology*, 28(2), 348–354.
539 <https://doi.org/10.1093/beheco/axx003>
- 540 Koenig, A., Scarry, C. J., Wheeler, B. C., & Borries, C. (2013). Variation in grouping patterns, mating
541 systems and social structure: What socio-ecological models attempt to explain. *Philosophical*
542 *Transactions of the Royal Society B: Biological Sciences*, 368(1618).
543 <https://doi.org/10.1098/rstb.2012.0348>
- 544 Kubenova, B., Konecna, M., Majolo, B., Smilauer, P., Ostner, J., & Schülke, O. (2016). Triadic
545 awareness predicts partner choice in male-infant-male interactions in Barbary macaques.
546 *Animal Cognition*, 20(2), 1–12. <https://doi.org/10.1007/s10071-016-1041-y>
- 547 Leiva, D., & De Vries, H. (2014). *steepness: Testing Steepness of Dominance Hierarchies*. R package
548 version 0.2-2.
- 549 Levy, E. J., Zippel, M. N., Mclean, E., Campos, F. A., Dasari, M., Fogel, A. S., Franz, M., Gesquiere, L. R.,
550 Gordon, J. B., Grieneisen, L., Habig, B., Jansen, D. J., Learn, N. H., Weibel, C. J., Altmann, J.,
551 Alberts, S. C., & Archie, E. A. (2020). A comparison of dominance rank metrics reveals multiple
552 competitive landscapes in an animal society: Dominance rank & competitive landscapes.

553 *Proceedings of the Royal Society B: Biological Sciences*, 287(1934).
554 <https://doi.org/10.1098/rspb.2020.1013>

555 Lonsdorf, T. B., Klingelhöfer-Jens, M., Andreatta, M., Beckers, T., Chalkia, A., Gerlicher, A., Jentsch, V.
556 L., Drexler, S. M., Mertens, G., Richter, J., Sjouwerman, R., Wendt, J., & Merz, C. J. (2019).
557 Navigating the garden of forking paths for data exclusions in fear conditioning research. *ELife*,
558 8. <https://doi.org/10.7554/eLife.52465>

559 McElreath, R. (2018). Statistical Rethinking: A Bayesian Course with Examples in R and Stan. In
560 *Statistical Rethinking: A Bayesian Course with Examples in R and Stan*. Chapman & Hall/CRC.
561 <https://doi.org/10.1201/9781315372495>

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

565 Mielke, A., Preis, A., Samuni, L., Gogarten, J. F., Lester, J. D., Crockford, C., & Wittig, R. M. (2021).
566 Consistency of Social Interactions in Sooty Mangabeys and Chimpanzees. *Frontiers in Ecology*
567 *and Evolution*, 8. <https://doi.org/10.3389/fevo.2020.603677>

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

571 Mielke, A., Samuni, L., Preis, A., Gogarten, J. F., Crockford, C., & Wittig, R. M. (2017). Bystanders
572 intervene to impede grooming in western chimpanzees and sooty mangabeys. *Royal Society*
573 *Open Science*, 4(11), 171296. <https://doi.org/10.1098/rsos.171296>

574 Netto, W. J., Hanegraaf, P. L. H., & De Vries, H. (1993). Matman: a Program for the Analysis of
575 Sociometric Matrices and Behavioural Transition Matrices. *Behaviour*, 125(3–4), 157–175.
576 <https://doi.org/10.1163/156853993x00218>

577 Neumann, C., Duboscq, J., Dubuc, C., Ginting, A., Irwan, A. M., Agil, M., Widdig, A., & Engelhardt, A.
578 (2011). Assessing dominance hierarchies: Validation and advantages of progressive evaluation
579 with Elo-rating. *Animal Behaviour*, 82(4), 911–921.
580 <https://doi.org/10.1016/j.anbehav.2011.07.016>

581 Newton-Fisher, N. E. (2017). Modeling Social Dominance: Elo-Ratings, Prior History, and the Intensity
582 of Aggression. *International Journal of Primatology*, 38(3), 427–447.
583 <https://doi.org/10.1007/s10764-017-9952-2>

584 O’Dea, R. E., Parker, T. H., Chee, Y. E., Culina, A., Drobnjak, S. M., Duncan, D. H., Fidler, F., Gould, E.,
585 Ihle, M., Kelly, C. D., Lagisz, M., Roche, D. G., Sánchez-Tójar, A., Wilkinson, D. P., Wintle, B. C., &
586 Nakagawa, S. (2021). Towards open, reliable, and transparent ecology and evolutionary
587 biology. *BMC Biology*, 19(1), 68. <https://doi.org/10.1186/s12915-021-01006-3>

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

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

594 Range, F., & Noë, R. (2005). Can simple rules account for the pattern of triadic interactions in
595 juvenile and adult female sooty mangabeys? *Animal Behaviour*, 69(2), 445–452.
596 <https://doi.org/10.1016/j.anbehav.2004.02.025>

597 Rosati, A. G., Hagberg, L., Enigk, D. K., Otali, E., Thompson, M. E., Martin, N. M., Wrangham, R. W., &
598 Machanda, Z. P. (2020). Social selectivity in aging wild chimpanzees. *Science*, 370(6515), 473–
599 476. <https://doi.org/10.1126/science.aaz9129>

600 Sánchez-Tójar, A., Schroeder, J., & Farine, D. R. (2018). A practical guide for inferring reliable
601 dominance hierarchies and estimating their uncertainty. *Journal of Animal Ecology*, 87(3), 594–
602 608. <https://doi.org/10.1111/1365-2656.12776>

603 Schielzeth, H. (2010). Simple means to improve the interpretability of regression coefficients.
604 *Methods in Ecology and Evolution*, 1(2), 103–113. <https://doi.org/10.1111/j.2041->
605 210x.2010.00012.x

606 Schino, G., & Lasio, F. (2019). Mandrills represent their own dominance hierarchy on a cardinal, not
607 ordinal, scale. *Animal Cognition*, 22(6), 1159–1169. <https://doi.org/10.1007/s10071-019->
608 01308-8

609 Seyfarth, R. M. (1977). A model of social grooming among adult female monkeys. *Journal of*
610 *Theoretical Biology*, 65(4), 671–698. [https://doi.org/10.1016/0022-5193\(77\)90015-7](https://doi.org/10.1016/0022-5193(77)90015-7)

611 Silk, J. B. (1999). Why are infants so attractive to others? The form and function of infant handling in
612 bonnet macaques. *Animal Behaviour*, 57(5), 1021–1032.
613 <https://doi.org/10.1006/anbe.1998.1065>

614 Steegen, S., Tuerlinckx, F., Gelman, A., & Vanpaemel, W. (2016). Increasing Transparency Through a
615 Multiverse Analysis. *Perspectives on Psychological Science*, 11(5), 702–712.
616 <https://doi.org/10.1177/1745691616658637>

617 Vehtari, A., Gelman, A., & Gabry, J. (2017). Practical Bayesian model evaluation using leave-one-out
618 cross-validation and WAIC. *Statistics and Computing*, 27(5), 1413–1432.
619 <https://doi.org/10.1007/S11222-016-9696-4/FIGURES/12>

620 Vilette, C., Bonnell, T. R., Henzi, S. P., & Barrett, L. (2019). Determining Temporal Stability in
621 Dominance Hierarchies. In *bioRxiv*. <https://doi.org/10.1101/692384>

622 Wicherts, J. M., Veldkamp, C. L. S., Augusteyn, H. E. M., Bakker, M., van Aert, R. C. M., & van Assen,
623 M. A. L. M. (2016). Degrees of Freedom in Planning, Running, Analyzing, and Reporting
624 Psychological Studies: A Checklist to Avoid p-Hacking. *Frontiers in Psychology*, 7, 1832.
625 <https://doi.org/10.3389/fpsyg.2016.01832>

626