**INTRODUCTION**

Animals living in permanent social groups must decide when and how to interact with group members and their ability to make the correct choice has potential fitness implications. It has long been hypothesised that the evolution of cognitive skills is a response to the selection pressure imposed by the complexity of the social environment (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; Dunbar, 1993). One way to operationalize social complexity, is as the amount of information necessary to successfully predict future states of a system (Flack, 2012; Sambrook & Whiten, 1997).

Measures of how predictable interactions are that are comparable between studies and species would facilitate examinations of the factors driving the evolution of complex decision-making abilities (Aureli & Schino, 2019; Dunbar & Shultz, 2010). Relationships are more predictable if partner choice is highly *consistent* over repeated interactions (Kalbitz et al., 2016; Koski et al., 2012; Moscovice et al., 2017; Silk et al., 2006) and observing a subset of interactions allows individuals to accurately predict future interactions (Whiten, 2000). An example of this are steep linear dominance hierarchies, in which a single interaction contains enough information to predict future dyadic contest (Guillermo Paz-Y-Miño et al., 2004; Oliveira et al., 1998; Sánchez-Tójar et al., 2018). Low consistency can be the result of an unpredictable distribution of social interactions or of frequent changes in relationships over time, both traits that might necessitate an increased need for cognitive flexibility (Barrett et al., 2002).

Complicating assessments of predictability, most behavioural studies depend on aggregated distributions of interaction types over time; either as dependent or independent variables,to create networks, or as relationship indices. The basic assumption is that the data accurately reflect what individuals were doing in the study period. Researchers can only observe a subset of individuals’ interactions, meaning that the “real” distribution of interactions is unknown and we have to assume that the collected data reflect it accurately (Farine & Strandburg-Peshkin, 2015; Kasper & Voelkl, 2009; Whitehead, 2008). However, if data are sparse, estimate errors are increased and robustness of the resulting distribution reduced (Lusseau et al., 2008; Shizuka & Farine, 2016). Working with distributions which are not accurate representations of the underlying distribution can lead to misleading results (Davis et al., 2018). This problem is aggravated when already sparse datasets are cut into shorter time intervals (e.g. 6-month blocks), a common practice in animal behaviour studies. What constitutes enough data can vary depending on how repeatable partner choice is (Sánchez-Tójar et al., 2018).

Here, we develop a consistency measure that has two functions: on one hand, it allows researchers to gauge whether they have collected enough data of a certain interaction type to warrant including that interaction type in their analyses, either in a social network or when creating relationship indices. On the other hand, it allows us to compare interaction types within and between species regarding their predictability. Consistency should be high if individuals regularly choose the same partners for the same interaction type and observing the individual at one point in time allows for accurate predictions of their behaviour later. Low consistency can arise if individuals show weak partner preference or preference changes over time, or if insufficient data are available.

To explore how consistency can be used to compare social groups with different structure and organisation, we apply the consistency measure to data from two Western chimpanzee (*Pan troglodytes verus*) communities and one sooty mangabey (*Cercocebus atys atys*) community living sympatrically in the Taï National Park, Côte d’Ivoire (Mielke et al., 2017, 2018). We chose these species as they represent two well-studied, quite different primate social systems. Sooty mangabeys have philopatric females who form linear, despotic, stable matrilineal hierarchies (Mielke et al., 2017, 2018; Range, 2006; Range & Noë, 2002). Interactions are predicted to show high consistency, as they should be influenced strongly by stable parameters, especially kinship, dominance rank, and sex (Range & Noë, 2002). Chimpanzees on the other hand, form male philopatric societies with stable linear hierarchies where aggression in Taï is frequently directed up the dominance hierarchy (Wittig & Boesch, 2003). While we observed chimpanzee rank changes in both sexes in the study period (Mielke et al., 2019; Preis et al., 2019),enduring social bonds have been described for both sexes (Foerster et al., 2015; Lehmann & Boesch, 2009; Mitani, 2009). For chimpanzees we predict that at least some of the relationships of each individual are predicted to fall into clear relationship types. We develop the consistency measures with two aims: a) to identify interaction types were data distributions are likely unreliable due to insufficient data; and b) to compare chimpanzees and mangabeys, and different interaction types within species, regarding their consistency, as an indicator of predictability.

**METHODS**

*Consistency measure*

To quantify consistency in an interaction type, we structured data by collection days. Each observation day is randomly assigned to one of two datasets of equal size (Sánchez-Tójar et al., 2018). For each of these two datasets, we then aggregated the interaction distribution of dyadic interactions per observation hour in each of the halves and calculate the non-parametric Spearman correlation between the distributions from these two datasets. This allowed us to estimate how well variation in one half of the dataset predicts variation in the other half. We repeated this procedure 100 times, with the median correlation coefficient constituting our measure of consistency for the full dataset.

As we show through data simulations (below), very inconsistent results (median correlation values between two halves of the dataset below 0.5) are usually the result of very low data density, even if the underlying ‘real’ distribution is highly consistent. Thus, if interaction types are highly inconsistent, it is unclear whether the collected data is representative for the actual distribution in the population, and whether any of the subsequent results meaningfully describe the behaviour of individuals in a community.

The overall correlation between the two halves of the dataset is likelt dependent on the data density and community size, so we explored the importance of these factors by repeating the consistency analysis using randomly selected subsets of the data. For this, we randomly selected a start date and duration for the period following that date and tested the consistency for this period for each interaction type, marking how many interactions per dyad this time period contained. For example, 10 individuals form 45 dyads; if we collect 180 aggressive events, we have a mean of 4 interactions/dyad.

This approach allows a more systematic comparison of both frequent and infrequent interaction types, or datasets of different sizes. As a standardised consistency measure, we report the median number of interactions per dyad that were sufficient to reach a cut-off correlation between halves of 0.5 for each interaction type. This measure is largely independent of data density and community size (see Simulations), and produces an interpretable result: how many interactions between two group members does an individual have to observe to reliably predict future interactions? Fewer interactions per dyads and a smaller standard deviation of values indicate higher consistency and thus higher predictability. Larger numbers of interactions per dyad and a large standard deviation indicate that interaction patterns are harder to predict. This can be either the case if partner choice less deterministic for this interaction type, or if the choice patterns change throughout the study period. All described analyses were conducted in R 3.5.1 (R Development Team 2017).

*Simulations*

We explored the impact of different group sizes, data densities, repeatability of partner choice, and changes in underlying relationships on our consistency measure using simulated datasets. We then explore whether it can be used to compare consistency across communities of different sizes. We tested whether our consistency metric is high when individuals regularly choose the same partners for the same interaction type. We also tested whether low consistency arises when individuals show weak partner preference or preference changes over time. To test how our consistency measure behaved under different conditions, and how to interpret different results, we simulated datasets with different group sizes; numbers of interactions per individual; data collection density; stereotypy of partner choice; and consistency of partner choice over time, mirroring interaction data as it could be collected in different primate communities.

Specifically, we created datasets for 10, 15, and 20 individuals in a community, for one nonspecific interaction type over a simulated time period of one year. We randomly assigned each individual between one and seven interactions per day, and each interaction took place in a “party” containing a randomly chosen number of group members. Each dyad was assigned a random likelihood to interact with each other, with three different stereotypy levels: “high certainty” (each individual has strong preference for a few group members, always chooses those when they are available), “medium certainty” (each individual prefers several group members, but can also choose non-preferred partners), and “low certainty” (the likelihood of choosing any partner is equal). Based on these dyadic values, one of the individuals in the “party” was selected as interaction partner. We explored three conditions concerning the consistency of individuals’ choice: in the first condition, dyadic interaction likelihood remained the same throughout. In the second condition, mirroring changes in relationship quality, the likelihood to choose a partner was reversed halfway through data collection, so dyads with a 0.95 likelihood of interacting in the first half had a 0.05 likelihood of interacting in the second half of data collection. In the third condition, partner choice was completely random, which should lead to an even distribution of interactions between all group members over the whole time period.

Following this procedure, we created 108 simulated datasets (three each for every combination of number of individuals, level of stereotypy, and consistency condition) that contained all interactions for all group members for each day of the data collection period. Subsequently, we simulated differences in data collection effort (Davis, Crofoot, & Farine, 2018): for each day of the sampling period, one individual was chosen as the “focal” individual whose data was retained, as would be the case in most primate datasets. We assumed a twelve-hour observation period per focal day, to calculate interaction rates. Then we simulated that data collection took place every day, 66% of days, or every third day, to test which impact low data collection density had on the consistency measure. We therefore retained 324 simulated datasets with different properties. For each of these, the proposed consistency measure – randomly selecting half of the dataset and correlating interaction rates of dyads with those of the other half, as well as repeating this procedure with subsets of the data – was carried out 10 times.

*Data Collection*

Behavioural data were collected in Taï National Park, Côte d’Ivoire (Boesch & Boesch-Achermann, 2000) from October 2013 to July 2015 for the chimpanzees and January 2014 to September 2015 for the mangabeys, using half- and full-day continuous focal animal sampling (Altmann, 1974) for the chimpanzees, and half-day and one-hour focal animal sampling for the mangabeys. Trained observers (AM, JFG, AP, LS) and field assistants recorded all social interactions of adult male and female chimpanzees (above 12 years of age) in the “South” (AM, AP, LS) and “East” (AP, LS) communities and adult (above 5 years) sooty mangabeys (AM, JFG). This resulted in 6441h of focal observations in South community, 5668h for East community, and 2259h for the mangabey community. We included adult individuals of both sexes in all three communities for whom sufficient focal data (at least 50 social interactions observed as focal individual) were available and who were present for at least 80% of the study period (South: 5 males, 7 females; East: 5 males, 7 females; mangabeys: 6 males, 17 females). All non-adult group members in all communities were excluded due to insufficient focal follow data in the study period, despite the impact they probably would have on all presented measures (Fedurek & Lehmann, 2017).

From the behavioural data, we extracted the duration of grooming sent and received, resting or foraging in less than 1m distance from the partner (“body contact”: used as a continuous measure with duration in the chimpanzees and an event variable in the mangabeys), resting or foraging as nearest neighbour between 1m and 3m distance (“proximity”), and both contact and noncontact aggressive interactions with one clear recipient (Preis et al., 2018). For the chimpanzee communities, we included food sharing (Samuni et al., 2018), which was not regularly observed in the mangabeys. We pant grunt vocalisations in chimpanzees and feeding supplants in mangabeys. Mutual interactions were coded as interaction given and received. We treated body contact and proximity as interaction types with the assumption that both individuals have to show sufficient tolerance to allow the other one to remain close. Body contact and proximity were only counted if no other interaction took place within 5min before or after to ensure independence of data points. We included grooming, contact aggression, noncontact aggression, pant grunts/supplants, and food sharing as directional variables, with the distribution of interactions given from each individual to every other as outcome variable. For the two spatial proximity measures, data were considered non-directional and symmetrical. Interaction distributions were standardised by focal observation time, with observation time calculated by adding the total observation times of A and B for the study period. We did not standardise interaction rates by the observation time two individuals spent in the same party, as has been suggested for fission-fusion societies (Moscovice et al., 2017), because the same standardisation would not be possible for most long-term datasets in low fission-fusion species, reducing replicability with other studies. Spatial proximity and food sharing in the chimpanzees were collected by a subset of observers and were standardised based on the focal observation time provided by those observers.

**RESULTS**

***Simulations***

We are looking for a consistency measure that can identify differences in stereotypy of partner choice and changes in interaction preference, while being independent of group size and data collection effort. We therefore present the Spearman rank correlation coefficient given the average number of interactions per dyad that were included in the dataset. The results for a dataset with medium certainty and no changes in partner preference, but with different numbers of individuals and different sampling effort can be found in Fig. S1. The consistency measure behaved the same independent of community size, even though larger communities had more dyads and therefore fewer interactions per dyad in the same period, and therefore reached a lower overall consistency. Also, datasets of different densities behaved the same way, but lower data density was characterized by lower overall consistency. This led us to two conclusions: the overall consistency of the full dataset cannot be interpreted by itself, as it is highly dependent on group size and data collection effort. However, it can be used to identify datasets in which there is simply not enough data available to make any statement about the underlying distribution of an interaction type. This is coherent with simulations showing that social network data becomes unreliable if data density per dyad sinks below a certain level (Whitehead, 2008).

A close up of a map

Description automatically generated

*Figure 1: Results of the data simulation with varying group size and data collection density, while having medium stereotypy of partner preference and no preference changes throughout the dataset.*

To test how the stereotypy of partner choice influenced the consistency measure, we present the results for the three different conditions (high, medium, low certainty) for datasets containing 15 group members, 100% data density, and no changes in preference throughout the sampling period (Fig. S2). Our results show that the consistency measure differentiates between the conditions, but again not with the overall consistency, but the slope at which this consistency is reached. If partner choice is highly stereotypical, already a small number of interactions was sufficient to predict partner choice in half of the data with that of the other half; with increasing uncertainty, more interactions per dyad are necessary.

A close up of a map

Description automatically generated

*Figure 2: Results of the data simulation with varying stereotypy of dyadic preference, while having consistent group size, data density, and no preference changes throughout the dataset.*

Last, we investigated how changes in partner preference over the study period would influence the consistency measure in a dataset with 15 individuals, 100% data density, and high stereotypy of partner choice. Here, we compare three conditions: one where no changes took place, one where the partner preference was reversed halfway through the study, and one where partner choice was randomised. Again, we found differences in the slope with which the consistency increased with increasing data density. Additionally, the different conditions could be differentiated by the spread of consistency values: when partner choice was consistent, selecting subsets of the same size at different points of the sampling period resulted in very similar consistency values. If partner choice changed throughout the sampling period, the predictive value of different subsets differed. As seen before, random partner choice could be identified because the consistency of the full dataset never increased above a certain threshold.

A close up of a map

Description automatically generated

*Figure S3: Results of the data simulation with preference changes 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 of dyadic preference, while random choice indicates that all partners were chosen with the same likelihood.*

Based on these observations, the consistency measure can be used to compare the predictability of interactions between groups of different sizes and different interaction types within the same community. We suggest using it the following way: the overall consistency, using the entire dataset, was heavily influence by the amount of interactions available per dyad, and thus does not make a good measure. However, it offers a good shorthand for researchers to decide whether to include an interaction type in their analyses at all, for example for social network analysis. In our simulations, even if the underlying distribution of interactions was highly stereotyped and consistent, the consistency measure remained low if data density was low, indicating that one half of the dataset was not a good predictor of the other half. Thus, if the Spearman rank correlation between halves of the same dataset does not reach 0.5, it is likely that not enough data has been collected to make statements about the underlying distribution of an interaction type in a population, unless that distribution is random. We therefore propose to remove interaction types with an overall consistency below r=0.5, as their interpretation is unclear.

For all other interactions, we propose the following standardised consistency measure: To make the interpretation of consistency comparable across communities of different sizes, we suggest the average number of interactions per dyad necessary to reach a median consistency of r=0.5 as a good measure. This accounts for the fact that the slope at which consistency rises when cutting the dataset in ever smaller subsets, rather than the overall consistency, can detect both differences in stereotypy of partner choice and changes in dyadic preference. The resulting value is directly interpretable – for example, in the “high certainty” condition, individuals need to observe around one interaction per dyad on average to be able to somewhat predict future interactions, whereas individuals in the “low certainty” condition need ten or more interactions per dyad to discern a pattern. Valuable information also arises from the spread of values of the repeated comparisons between halves of the dataset: if dyadic preference remained stable throughout, it does not matter much what subset of the data is selected, the consistency is relatively stable for subsets of the same size. However, if dyadic preference was not stable and changed throughout, the correlation between halves varies even for datasets of the same size.

***Field Data***

Our measure of consistency, repeatedly selecting half of focal observation days and correlating the interaction rates with those of the other half, proved useful in identifying interaction types with insufficient information available. In all three communities, contact aggressions and food sharing stood out as having the lowest number of interactions (around or below 1 average interaction per dyad). Both also stood out as for all communities, consistency values were below r=0.5 (Tab. 1). In our simulations, such low values only occurred when insufficient data was available to successfully approximate the underlying distributions of interactions, even in cases where the underlying distribution was highly consistent. For all other interaction types, consistency values were above r=0.5 in all communities.

*Table 1: Overview of the datasets for each interaction type and group, and the results of the consistency measures. “Overall consistency” is the median of the repeated correlation between randomly selected halves for the full dataset available for an interaction type. “Standardised Consistency” and the standard deviation are the result of the repeated random selection of halves of subsets of different lengths, with the average number of interactions per dyad for which the median correlation coefficient exceeds r=0.5 as measure of how much information is needed to predict future interactions in a community.*

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Group** | **Interaction Type** | **IDs** | **Interactions** | **Overall Consistency** | **Standardised Consistency (Interactions/dyad)** | **SD** |
| Mangabey | Grooming | 23 | 1162 | 0.52 | 1.8 | 0.4 |
| Mangabey | Body Contact | 23 | 2218 | 0.51 | 7.4 | 1.4 |
| Mangabey | Proximity | 23 | 4373 | 0.51 | 14.5 | 2.5 |
| Mangabey | Aggression Non-contact | 23 | 971 | 0.58 | 1.3 | 0.3 |
| Mangabey | Aggression Contact | 23 | 210 | 0.24 | - | - |
| Mangabey | Supplant | 23 | 1219 | 0.69 | 2.1 | 0.5 |
| East | Grooming | 12 | 3099 | 0.75 | 4.7 | 2.0 |
| East | Body Contact | 12 | 1935 | 0.64 | 11.7 | 5.6 |
| East | Proximity | 12 | 2796 | 0.69 | 13.1 | 6.2 |
| East | Aggression Non-contact | 12 | 693 | 0.65 | 2.2 | 0.8 |
| East | Aggression Contact | 12 | 126 | 0.33 | - | - |
| East | Food Sharing | 12 | 151 | 0.40 | - | - |
| East | Pant Grunt | 12 | 2429 | 0.92 | 0.9 | 0.1 |
| South | Grooming | 12 | 4693 | 0.81 | 4.2 | 2.5 |
| South | Body Contact | 12 | 1669 | 0.62 | 12.0 | 5.1 |
| South | Proximity | 12 | 2579 | 0.69 | 14.0 | 5.6 |
| South | Aggression Non-contact | 12 | 768 | 0.80 | 1.3 | 0.6 |
| South | Aggression Contact | 12 | 173 | 0.42 | - | - |
| South | Food Sharing | 12 | 153 | 0.34 | - | - |
| South | Pant Grunt | 12 | 3350 | 0.93 | 1.4 | 0.00 |

To be able to compare groups of different sizes, and interaction types with different occurrence rates, we tested how many interactions per dyad were necessary to reach a consistency value of r=0.5. Interaction types that reached this cut-off with fewer average interactions per dyad were considered more predictable than those that require more interactions per dyad to reach the cut-off (Fig. 4, 5, 6).

A close up of a map

Description automatically generated

*Figure 4: Spearman correlation between two halves of randomly selected subsets of the datasets for mangabeys (green), East chimpanzee community (yellow) and South chimpanzee community (blue) for grooming, body contact, proximity, and pant grunts/supplants. 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.*

A close up of a map

Description automatically generated

*Figure 5: Spearman correlation between two halves of randomly selected subsets of the datasets for mangabeys (green), East chimpanzee community (yellow) and South chimpanzee community (blue) for nonphysical aggression, physical aggression, and food sharing. 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.*

*Mangabeys*

For the mangabeys, overall aggression rates (1.3 interactions/dyad, SD=0.2), grooming (1.8 interactions/dyad, SD=0.4), and supplants (2.1 interactions/dyad, SD=0.5), were among the most consistent behaviours we observed for all groups, indicating that individuals observing a subset of interactions in the community would be able to predict future interactions (Fig. 4,5; Tab. 1). Body contact (7.4 interactions/dyad, SD=1.4) and proximity (14.5 interactions/dyad, SD=2.5) were much less consistent.

*Chimpanzees*

As in the mangabeys, overall aggression rates were highly consistent in both chimpanzee communities (Table 1), more so in South (1.3 interactions/dyad, SD=0.6) than in East (2.6 interactions/dyad, SD=0.9). The larger standard deviation in the chimpanzees compared to the mangabeys might indicate changes of aggression patterns over time. Pant grunt interactions in both communities showed the most predictable patterns (East: 0.9 interactions/dyad, SD=0.1; South: 1.4 interactions/dyad, SD=0.00). Grooming was less consistent than in the mangabeys (East: 4.7 interactions/dyad, SD=2.0; South: 4.2 interactions/dyad, SD=2.5), with the large standard deviation indicating changes in interaction patterns over the course of the study period. Body contact (East: 11.7 interactions/dyad, SD=5.6; South: 12.0 interactions/dyad, SD=5.1) and proximity (East: 13.1 interactions/dyad, SD=6.2; South: 14.0 interactions/dyad, SD=5.6) were considerably less predictable than the directed interaction types. This indicates that in all three communities, most dyads will feed and rest in close proximity with a wide variety of partners, while they direct interactions at a smaller and more stable subset of group members.

A picture containing sky, text, map

Description automatically generated

Figure 6: Summary of the mean number of interactions needed per dyad to reach correlations between halves of r = 0.5 (mangabeys: green triangles, East: red point, South: blue square).

**DISCUSSION**

Establishing measures of predictability of social interactions and the resulting relationships between individuals is necessary to understand the complexity of a social group from the perspective of the individual (Dunbar & Shultz, 2010; Lukas & Clutton-Brock, 2018). Here, our premise was that interactions are more predictable for participants and bystanders if interaction distributions are consistent over time. Our results showed that across communities and species, interaction types vary in predictability, indicating yet again that animal lives cannot be capture using one simplistic measure of complexity: challenges differ within and between species, and we need multi-dimensional measures to quantify where ‘complexity’ really arises.

This study introduces a consistency measure, repeatedly dividing the dataset into halves and comparing how well these predict each other, which serves two functions. Researchers can use it to find out whether they have collected sufficient data for their dataset to be internally consistent, given a community of a certain size and an interaction type with a specific diversity of partner choice (Sánchez-Tójar et al., 2018). In our sample, despite pooling 18 months of data, food sharing and contact aggressions were observed at such low rates in all three communities that observing the group at a certain time point would make it impossible to predict their behaviour at another time point. We generally assume that randomly selected focal follows allow us to also make statements about interaction rates on those days on which we do not observe an individual (Davis et al., 2018), but this might not be the case for rare interaction types. One way to understand patterns driving interaction patterns for rare interaction types would be to test the factors influencing decisions by comparing each partner choice an individual makes against the alternatives they did not choose (Mielke et al., 2018; Samuni et al., 2018). In order to assess changes in relationships over time, there has been a trend to cut datasets into smaller subsets and then compare network overlap between these, assuming that the data in each is sufficient to depict the underlying distribution in the community. With our consistency measure, seasonality and change could be established if smaller subsets would show higher consistency than larger subsets, as random subsets retained consistent time intervals. This was not the case for any interaction type, even though some interaction types showed high standard variation, and indication that consistency is high during some times but not others.

The standardized consistency measure allowed us to segregate interaction types into those that needed either large or small amounts of information to predict future interactions. We use the number of interactions per dyad at which the majority of subset correlations exceeds the value 0.5; while the value 0.5 itself is arbitrary, using it across species and interaction types allows researchers to make comparative statements. Feeding supplants and pant grunts, which are used to create hierarchies in the respective species, were highly consistent in the two species, indicating generally stable hierarchies (Sánchez-Tójar et al., 2018). Consistency of aggression distribution did not vary between species. Despite being the larger community, mangabey interactions were generally more predictable than chimpanzee interaction patterns, with fewer grooming and body contact interactions per dyad needed to reach the consistency cut-off in mangabeys. Directed interactions (grooming, aggression, pant grunts/supplants) were consistent despite the inclusion of 18 months of data per community, indicating that most dyads interacted at relatively constant rates throughout the study period. The higher consistency of mangabey interactions could be the result of a highly stable dominance hierarchy (Mielke et al., 2017), few demographic changes (Borgeaud et al., 2017), and a relatively low diversity of partner choice for most individuals (Duboscq et al., 2017; Silk et al., 2013), while both chimpanzee communities displayed rank changes throughout and all group members interacted with each other. Difference in fission-fusion dynamics could be responsible for less consistent interaction patterns in the chimpanzees, as individuals have to show behavioural flexibility when they are constrained in their access to preferred partners, and interact with individuals based on availability (Amici et al., 2008). Spatial proximity was much less predictable than directed interactions, most likely because a certain level of tolerance in foraging and resting extends to most group members, adding random noise that is not present in directed interactions, a pattern mirroring previous results on spatial association in the mangabey group (Mielke et al., 2020). Alternatively, factors influencing proximity might be more variable than can be captured in aggregated measures. An individual that has to accurately predict future spatial proximity between individuals would therefore have to observe more events than individuals predicting aggressive interactions. Just like rare interaction types, common but highly inconsistent interaction types could add noise to social relationship indices or when comparing network overlap.

While many animal species are studied at great detail, and vast amounts of long-term data are available, it is surprisingly difficult to convey the structure of social interactions across sites and species. Our consistency measure may help by providing a standardised way to convey the flexibility in interaction patterns over time and identify interaction types that likely differ in complexity between species. Further, many researchers use multilevel social network analysis and create relationship indices including different interaction types, unsure whether all of them will be equally reliable. This consistency measure, like similar efforts for hierarchies (Sánchez-Tójar et al., 2018), can be a useful tool to make these decisions while conveying important information about the study species.

**Data Availability.** Data and R scripts for the consistency analysis are available: xxxxxx

**Competing Interests.** We have no competing interests

**Funding.** AM, AP, LS, CC, RMW were supported by the Max Planck Society; AM was supported by the Wenner Gren Foundation (Grant Number 9095) and the British Academy Newton International Fellowship; AP was supported by the Leakey Foundation; LS was supported by the Minerva Foundation; JFG was supported by an NSF Graduate Research Fellowship (DGE-1142336), the Canadian Institutes of Health Research’s Strategic Training Initiative in Health Research’s Systems Biology Training Program, an NSERC Vanier Canada Graduate Scholarship (CGS), and a long-term Research Grant from the German Academic Exchange Service (DAAD-91525837-57048249). C. C. was supported by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement no. 679787). RMW was supported by DFG Researcher Unit (FOR 2136) ‘Sociality and Health in Primates’ (WI 2637/3-1). Research at the Taï Chimpanzee Project has been funded by the Max Planck Society since 1997.

**Research Ethics.** This study was purely observational with no manipulation of animals. Methods were approved by the Ethikrat der Max-Planck-Gesellschaft (4.08.2014).

**Permission to carry out fieldwork.** Permissions to conduct the research were granted by the Ministries of Research and Environment of Ivory Coast (379/MESRS/GGRSIT/tm) and Office Ivorien des Parcs et Reserves.

**Acknowledgements.** We thank the Ivorian Ministry of Environment and Forests and Ministry of Higher Education and Scientific Research and the Office Ivoirien des Parcs et Reserves of Côte d’Ivoire. We thank Simon Kannieu, Daniel Bouin, Gnimion Florent, Fabrice Blé, Florent Goulei and the team of the TCP for field work support and data collection.

**References**

Altmann, J. (1974). Observational study of behavior: Sampling methods. *Behaviour*, *49*(3), 227–266. https://doi.org/10.1080/14794802.2011.585831

Amici, F., Aureli, F., & Call, J. (2008). Fission-fusion dynamics, behavioral flexibility, and inhibitory control in primates. *Current Biology*, *18*(18), 1415–1419. https://doi.org/10.1016/j.cub.2008.08.020

Aureli, F., & Schino, G. (2019). Social complexity from within: how individuals experience the structure and organization of their groups. In *Behavioral Ecology and Sociobiology* (Vol. 73, Issue 1, pp. 1–13). Springer Verlag. https://doi.org/10.1007/s00265-018-2604-5

Barrett, L., Gaynor, D., & Henzi, S. P. P. (2002). A dynamic interaction between aggression and grooming reciprocity among female chacma baboons. *Animal Behaviour*, *63*(6), 1047–1053. https://doi.org/10.1006/anbe.2002.3008

Boesch, C., & Boesch-Achermann, H. (2000). *The chimpanzees of the Taï Forest: Behavioural ecology and evolution*. Oxford University Press.

Borgeaud, C., Sosa, S., Sueur, C., & Bshary, R. (2017). The influence of demographic variation on social network stability in wild vervet monkeys. *Animal Behaviour*, *134*, 155–165. https://doi.org/10.1016/j.anbehav.2017.09.028

Byrne, R. W., & Whiten, A. (1989). *Machiavellian intelligence: social expertise and the evolution of intellect in monkeys, apes, and humans*. Oxford University Press. http://www.amazon.com/exec/obidos/redirect?tag=citeulike07-20&path=ASIN/0198521758

Castles, M., Heinsohn, R., Marshall, H. H., Lee, A. E. G., Cowlishaw, G., & Carter, A. J. (2014). Social networks created with different techniques are not comparable. In *Animal Behaviour* (Vol. 96, pp. 59–67). Academic Press. https://doi.org/10.1016/j.anbehav.2014.07.023

Davis, G. H., Crofoot, M. C., & Farine, D. R. (2018). Estimating the robustness and uncertainty of animal social networks using different observational methods. *Animal Behaviour*, *141*, 29–44. https://doi.org/10.1016/j.anbehav.2018.04.012

Duboscq, J., Neumann, C., Agil, M., Perwitasari-Farajallah, D., Thierry, B., & Engelhardt, A. (2017). Degrees of freedom in social bonds of crested macaque females. *Animal Behaviour*, *123*, 411–426. https://doi.org/10.1016/j.anbehav.2016.11.010

Dunbar, R. I. M. (1993). Coevolution of neocortical size, group size and language in humans. *Behavioral and Brain Sciences*, *16*(4), 681–694. https://doi.org/10.1017/S0140525X00032325

Dunbar, R. I. M., & Shultz, S. (2010). Bondedness and sociality. In *Behaviour* (Vol. 147, Issue 7, pp. 775–803). https://doi.org/10.1163/000579510X501151

Farine, D. R., & Strandburg-Peshkin, A. (2015). Estimating uncertainty and reliability of social network data using Bayesian inference. *Royal Society Open Science*, *2*(9), 150367. https://doi.org/10.1098/rsos.150367

Fedurek, P., & Lehmann, J. (2017). The effect of excluding juveniles on apparent adult olive baboons (Papio anubis) social networks. *PLoS ONE*, *12*(3), e0173146. https://doi.org/10.1371/journal.pone.0173146

Flack, J. C. (2012). Multiple time-scales and the developmental dynamics of social systems. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*(1597), 1802–1810. https://doi.org/10.1098/rstb.2011.0214

Foerster, S., McLellan, K., Schroepfer-Walker, K., Murray, C. M., Krupenye, C., Gilby, I. C., & Pusey, A. E. (2015). Social bonds in the dispersing sex: Partner preferences among adult female chimpanzees. *Animal Behaviour*, *105*, 139–152. https://doi.org/10.1016/j.anbehav.2015.04.012

Guillermo Paz-Y-Miño, C., Bond, A. B., Kamil, A. C., & Balda, R. P. (2004). Pinyon jays use transitive inference to predict social dominance. *Nature*, *430*(7001), 778–781. https://doi.org/10.1038/nature02723

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

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

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

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

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

Lehmann, J., & Boesch, C. (2009). Sociality of the dispersing sex: the nature of social bonds in West African female chimpanzees, Pan troglodytes. *Animal Behaviour*, *77*(2), 377–387. https://doi.org/10.1016/j.anbehav.2008.09.038

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

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

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

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

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

Mitani, J. C. (2009). Male chimpanzees form enduring and equitable social bonds. *Animal Behaviour*, *77*(3), 633–640. https://doi.org/10.1016/j.anbehav.2008.11.021

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

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

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

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

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

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

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

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

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

Shizuka, D., & Farine, D. R. (2016). Measuring the robustness of network community structure using assortativity. *Animal Behaviour*, *112*, 237–246. https://doi.org/10.1016/j.anbehav.2015.12.007

Silk, J. B., Alberts, S. C., & Altmann, J. (2006). Social relationships among adult female baboons (Papio cynocephalus) II. Variation in the quality and stability of social bonds. *Behavioral Ecology and Sociobiology*, *61*(2), 197–204. https://doi.org/10.1007/s00265-006-0250-9

Silk, J. B., Cheney, D., & Seyfarth, R. (2013). A practical guide to the study of social relationships. *Evolutionary Anthropology*, *22*(5), 213–225. https://doi.org/10.1002/evan.21367

Whitehead, H. (2008). Precision and power in the analysis of social structure using associations. *Animal Behaviour*, *75*(3), 1093–1099. https://doi.org/10.1016/j.anbehav.2007.08.022

Whiten, A. (2000). Social complexity and social intelligence. In *The nature of intelligence* (Vol. 233, pp. 185–196; discussion 196-201). Wiley. http://www.ncbi.nlm.nih.gov/pubmed/11276903

Wittig, R. M., & Boesch, C. (2003). “Decision-making” in conflicts of wild chimpanzees (Pan troglodytes): An extension of the Relational Model. *Behavioral Ecology and Sociobiology*, *54*(5), 491–504. https://doi.org/10.1007/s00265-003-0654-8