

Behavioral Ecology and Sociobiology

A commentary on using extraction methods in data reduction analyses to assess social relationship structure in animals

--Manuscript Draft--

Manuscript Number:	
Full Title:	A commentary on using extraction methods in data reduction analyses to assess social relationship structure in animals
Article Type:	Methods Papers
Corresponding Author:	Blake Morton UNITED KINGDOM
Corresponding Author Secondary Information:	
Corresponding Author's Institution:	
Corresponding Author's Secondary Institution:	
First Author:	Blake Morton
First Author Secondary Information:	
Order of Authors:	Blake Morton Drew Altschul
Order of Authors Secondary Information:	
Funding Information:	
Abstract:	Data reduction analyses like factor analysis (FA) and principal components analysis (PCA) identify inter-relationships between a set of potentially correlated variables, and cluster correlated variables into fewer discrete categories called "factors" (in FA) or "components" (in PCA). Because they can provide researchers with a systematic approach to categorizing different sets of behaviors (e.g. rates of grooming or aggression between group members), data reduction analyses are increasingly being used to describe social relationship structure in animals (i.e. the number of dimensions and the characteristics of those dimensions). Before performing a data reduction analysis, one must first determine how many factors or components to extract from the analysis. This decision is critical given that it will influence how variables cluster together, thereby affecting the final solution and (hence) researchers' interpretation of those results. Using data from a previous study of brown capuchin monkeys (<i>Sapajus</i> sp.) as an example, we compare six different extraction methods (Kaiser's criterion, parallel analysis, complexity, standardized root mean square residuals, empirical bayesian information factor, very simple structure), and assess how each differentially affect the results, and thus interpretation, of capuchin social relationship structure. We use this example as a platform to make recommendations to improve the utility and reliability of data reduction techniques for comparative studies of animal social relationships.
Suggested Reviewers:	Filippo Aureli aureli.filippo@gmail.com Alexander Weiss alex.weiss@ed.ac.uk Hannah Buchanan-Smith h.m.buchanan-smith@stir.ac.uk Bonaventura Majolo bmajolo@lincoln.ac.uk Richard McFarland

	richard.mcfarland@wits.ac.za
	Thomas Bugnyar thomas.bugnyar@univie.ac.at

**A commentary on using extraction methods in data reduction analyses to assess
social relationship structure in animals**

F. Blake Morton^{a,b*}, Drew Altschul^{b,c}

^a Behaviour and Evolution Research Group, Psychology, School of Natural Sciences,
University of Stirling, Stirling, U.K.

^b Scottish Primate Research Group, U.K.

^c Department of Psychology, The University of Edinburgh, Edinburgh, U.K.

***Correspondence:**

F. B. Morton, Psychology, School of Natural Sciences, University of Stirling, Stirling FK9
4LA, U.K. E-mail address: morton.blake@gmail.com (F. B. Morton).

Abstract

Data reduction analyses like factor analysis (FA) and principal components analysis (PCA) identify inter-relationships between a set of potentially correlated variables, and cluster correlated variables into fewer discrete categories called “factors” (in FA) or “components” (in PCA). Because they can provide researchers with a systematic approach to categorizing different sets of behaviors (e.g. rates of grooming or aggression between group members), data reduction analyses are increasingly being used to describe social relationship structure in animals (i.e. the number of dimensions and the characteristics of those dimensions). Before performing a data reduction analysis, one must first determine how many factors or components to extract from the analysis. This decision is critical given that it will influence how variables cluster together, thereby affecting the final solution and (hence) researchers’ interpretation of those results. Using data from a previous study of brown capuchin monkeys (*Sapajus sp.*) as an example, we compare six different extraction methods (Kaiser’s criterion, parallel analysis, complexity, standardized root mean square residuals, empirical bayesian information factor, very simple structure), and assess how each differentially affect the results, and thus interpretation, of capuchin social relationship structure. We use this example as a platform to make recommendations to improve the utility and reliability of data reduction techniques for comparative studies of animal social relationships.

Keywords: capuchin monkey, *Sapajus apella*, principal components analysis, factor analysis, Kaiser’s criterion, scree test, parallel analysis

47

48 **Introduction**

49 Comparative studies of social relationships are critical to understanding animal
50 sociality and how it evolves (Cords and Aureli 2000; Fraser and Bugnyar 2010). One
51 challenge researchers face is how best to define and quantify social relationships such
52 that the data are systematic and comparable across studies and taxa. Data reduction
53 analyses like factor analysis (FA) and principal components analysis (PCA) identify
54 inter-relationships between a set of potentially correlated variables, and cluster
55 correlated variables into fewer discrete categories called “factors” (in FA) or
56 “components” (in PCA) (Field 2009; Gorsuch, 1983). Because they can provide
57 researchers with a systematic approach to categorizing different sets of behaviors (e.g.
58 rates of grooming or aggression between group members), data reduction analyses are
59 increasingly being used to describe social relationship structure in animals, such as
60 Japanese macaques (*Macaca fuscata*; Majolo et al. 2010), Barbary macaques (*Macaca*
61 *sylvanus*; McFarland and Majolo 2011), spider monkeys (*Ateles geoffroyi*; Rebecchini et
62 al. 2011), capuchin monkeys (*Sapajus sp.*, formerly *Cebus apella*; see Alfaro et al.
63 2012; Morton et al. 2015), chimpanzees (*Pan troglodytes*; Fraser et al. 2008; Koski et
64 al. 2012), bonobos (*Pan paniscus*; Stevens et al. 2015), and common ravens (*Corvus*
65 *corax*; Fraser and Bugnyar 2010; Loretto et al. 2012). Although data reduction analyses
66 have primarily been used to study primate social relationships, this analytic approach is
67 a potentially useful tool for studying social relationship structure across a much broader
68 range of taxonomic groups (e.g. Fraser and Bugnyar 2010).

Data reduction analyses can facilitate researchers' understanding of the overall structure of social relationships; that is, the number of dimensions and the characteristics of those dimensions. For example, while social relationships have traditionally been described in terms of two dimensions – affiliative behavior (e.g. rates of grooming, spatial tolerance) and agonistic behavior (e.g. rates of aggression) (Hinde 1976) -- studies using data reduction analyses have revealed that the social relationships of certain species may be best described in terms of three dimensions, as originally proposed by Cords and Aureli (2000), including relationship 'value' (i.e. immediate benefits afforded by the relationship, such as grooming), 'compatibility' (i.e. tolerance based on partners' shared history, such as tolerance at feeding sites), and 'security' (i.e. consistency and predictability in partners' behavior, such as rates of conflict) (Majolo et al. 2010; McFarland and Majolo 2011; Fraser et al. 2008; Fraser and Bugnyar 2010).

Data reduction analyses can also help researchers reduce the number of behavioral variables necessary to quantify each dimension of a social relationship between partners, as well as identify particular behaviors that are translatable across species. For instance, all studies to date that have used data reduction analyses to assess the structure of their subjects' social relationships have found that spatial proximity (i.e. the amount of time A spends in close proximity to B) is characteristic of a dimension reflecting relationship 'value'. Spatial proximity may therefore serve as a useful proxy for measuring relationship value within and between taxa.

Determining how many components/factors to extract from a data reduction analysis

Before performing a data reduction analysis, one must first determine how many factors or components to extract from the analysis (Field 2009). This decision is critical given that it will influence how variables cluster together, thereby affecting the final solution and (hence) researchers' interpretation of those results (Zwick and Velicer 1986; Ledesma and Valero-Mora 2007). Under-extraction can result in the loss of relevant information and distort the overall solution (Zwick and Velicer 1986). Over-extraction can result in some factors or components being unstable, making the overall solution difficult to interpret and/or replicate (Zwick and Velicer 1986).

Deciding when to stop extracting factors or components depends on when very little "random" variability remains in the final solution. Various cut-offs have been developed to help researchers make this decision, and frequently involve calculating the amount of variation that is explained by each component/factor (called "eigenvalues"; Field 2009). Two commonly used methods are Kaiser's criterion and Cattell's scree test. Kaiser's criterion retains components with eigenvalues >1.0 ; meaning, each component accounts for more variance than what is accounted for by one of the original variables (Kaiser 1960). Scree tests are a graphical technique that plots eigenvalues in a simple line plot. The number of components or factors to extract is visually estimated from the scree plot by finding the point where the line begins to level off; all components to the right of this point are considered random "noise" and should therefore be excluded (Cattell 1966).

Although scree tests and Kaiser's criterion are relatively simple to implement (perhaps contributing to their common usage by researchers), they can lead to spurious solutions. When components/factors are simple and strong, scree plots work quite well,

but they are fundamentally subjective and consequently lead to under- or over-extraction, particularly as the line of the plot begins to asymptote (Zwick and Velicer 1986). In simulations, scree tests are correct in only 41.7% of cases (Zwick and Velicer 1986). Thus, it is recommended that scree tests only be used alongside other methods.

While scree tests have maintained some utility over time, Kaiser's criterion has not, which is a fact that has been known for some time (Revelle and Rocklin 1979). First, Kaiser's criterion is only appropriate to use with components, not factors. Moreover, whereas scree tests are subjective, Kaiser's criterion is arbitrary: there is no empirical reason why a component with an eigenvalue slightly greater than 1 ought to be retained while a component with an eigenvalue just below 1 should not (Courtney 2013). Kaiser's criterion has shown tendencies, depending on the circumstances, toward over-extraction and, to a lesser degree, under-extraction (Zwick and Velicer 1986). These biases are in part due to the observation that the number of components retained by the criterion more strongly reflects the number of variables included in the analysis than any attributes of underlying latent variables (Gorsuch 1983). In simulation, Kaiser's criterion leads to a success rate of 8.77% and fails to extract the correct number of components in more than 90% of cases (Ruscio and Roche 2012). Yet, this criterion remains the default for popular statistics software, such as IBM SPSS (Field, 2009).

In light of the deficiencies associated with scree tests and Kaiser's criterion, many alternative methods have been developed. A select few include complexity (Hofmann 1978), Standardized Root Mean Square Residuals or SRMR (Hu and Bentler 1999), the Empirical Bayesian Information Factor or empirical BIC (Schwarz 1978),

Revelle and Rocklin's (1979) Very Simple Structure (VSS), and Horn's (1965) parallel analysis (PA).

Complexity represents the average number of factors needed to account for the measured variables. In an ideal solution, each factor would have a complexity of 1, meaning that every variable loaded solely on a single factor in the solution (Hofmann 1978); deviations from 1 indicate poorer fit. Complexity only assesses whether a dataset fits a single-level structure; that is, high complexity suggests multi-level structure might be present, but it could also mean that FA and PCA are not appropriate analyses for the data.

SRMR is the square root of the difference between a sample's covariance matrix and the proposed model's covariance matrix (Hooper et al. 2008). SRMR is representative of measures typically used in Confirmatory FA. SRMR is biased towards over-extraction; however, the greater the number of parameters in the model and the larger the sample size, the lower SRMR tends to be (Hu and Bentler 1999). Lower values are better; any value above 0.1 is considered unacceptable.

Empirical BIC is an information theoretical assessment of fit that evaluates the parsimony of any model (Schwarz 1978). A solution with more components/factors will very often have a better absolute fit, but the BIC applies a penalty based on the number of parameters, and because solutions with more components/factors have more parameters, BIC measures are an effective statistic for comparing many models. BIC is widely used in model building across many fields, and has been shown to be a superior statistic among information theory measures (Posada and Buckley 2004). In

simulations, BIC identifies the correct number of factors more than 60% of the time (Ruscio and Roche 2012).

VSS examines how well the individual components/factors fit within many solutions, where each progressive solution has one more component/factor than the last (Revelle and Rocklin 1979). VSS can be used in an entirely objective fashion, by finding maxima, but it can be viewed subjectively as well, like a scree plot. However, VSS is best at identifying simple structures (i.e. those with a single-level of factors); it is probably not appropriate if the complexities of some items are greater than two (Revelle 2015), and to the best of our knowledge it has not been compared to alternative modern methods in simulation studies (Courtney 2013).

PA is unique in that it has survived and been improved upon since it was first developed (Horn 1965), and therefore remains unambiguously one of the best tests available for component/factor extraction. PA is a procedure based on generating random eigenvalues that “parallel” the observed data in terms of sample size and the number of variables (Zwick and Velicer 1986). A component/factor is retained if its eigenvalue is greater than the 95th percentile of the distribution of eigenvalues generated from the random data (Horn 1965). This technique improves upon most other methods, both subjective (e.g. scree test) and objective (e.g. empirical BIC, Complexity), by taking into account sampling error, which is left unpartitioned from total variance in other methods (Horn 1965). PA is not arbitrary: the “parallel” data it generates can be resampled from the empirical data themselves, and the technique is robust. Both resampled and simulated parallel data do not yield substantively different results (Revelle 2015). Moreover, PA is flexible, having been modified and improved

upon since its conception, and is capable of assessing factor and component structures, as well as both ratio and ordinal data (Garrido et al. 2012). Finally, PA is noteworthy when contrasted with other, modern factor number tests because unlike even the best alternatives, e.g. Comparison Data (Ruscio and Roche 2012), it is completely unbiased (cf. Courtney 2013, Table 1).

This study

No single extraction test should be used as the sole method to determine how many components/factors to extract from a data reduction analysis. This is predominately because few datasets yield an immediate and clear solution. Ideally, multiple tests should be implemented and compared; if multiple tests agree on the same number of components/factors to extract, then researchers can be confident with their decisions about extraction.

To date, most studies using data reduction analyses to describe the social relationship structure of animals have relied solely on Kaiser's criterion (Majolo et al. 2010; McFarland and Majolo 2011; Rebecchini et al. 2011; Fraser et al. 2008; Fraser and Bugnyar 2010), with a handful of studies supplementing this technique with a scree test (Koski et al. 2012). By contrast, all other extraction methods are rarely used (Morton et al. 2015; Stevens et al. 2015). Therefore, using data from a previous study of brown capuchin monkeys (*Sapajus sp.*) as an example, we compare how each of the six extraction methods (described above) can differentially affect the results, and thus interpretation, of capuchin social relationship structure. We use this example as a platform to make recommendations about component/factor extraction, and improve the

utility and reliability of data reduction techniques for comparative studies of social relationships.

Methods and materials

Study site and subjects

Eighteen brown capuchin monkeys were studied at the Living Links to Human Evolution Research Centre (LL), located within the Royal Zoological Society of Scotland (RZSS), U.K. (Macdonald and Whiten, 2011). Subjects were from two breeding groups. The 'East' group contained four adult males, three adult females, one juvenile male and five infants (following age–sex categories in Frigaszy et al. 2004). The 'West' group contained four adult males, three adult females, two juvenile males, one juvenile female and five infants. Infants dependent on their mothers (i.e. those less than a year old) were not included as study subjects. Subjects' ages ranged from 2 to 40 years for males (mean \pm SD = 10.79 \pm 8.55 years, N = 11) and 3 to 14 years for females (mean \pm SD = 8.86 \pm 3.63 years, N = 7). All group members were captive born except an adult male from East group, who was hand-reared, and the original wild-caught alpha male of West group; both individuals came to LL as established members of their groups. Both breeding groups were housed separately in identically designed 189 m³ indoor enclosures with natural light and near-permanent access to a 900 m² outdoor enclosure containing trees and other vegetation, providing ample opportunity to engage in natural behaviors. All monkeys received commercial TrioMunch pellets supplemented with fresh fruits and vegetables three times daily and were given cooked chicken and

hardboiled eggs once a week. Water was available *ad libitum* at all times. Further details of housing and husbandry are provided in Leonardi et al. (2010).

Ethical note

This study was entirely observational except for one aspect of data collection involving puzzle feeders, which were placed within the monkeys' outdoor enclosures (see 'Behavioral sampling'). Subjects could interact freely with the puzzle feeders, which were made entirely of non-hazardous material. The feeders provided a source of food snacks (raisins) and enrichment to subjects. This study was approved by Edinburgh Zoo and the ethics committee of the Psychology Department at the University of Stirling, and complied with the ASAB (2012) Guidelines.

Behavioral sampling

Behavioral data come from a previous study by Morton et al. (2015). Fifty-four hours of focal observations were recorded between May and August of 2011, totalling 3h per individual. Behaviors (Table 1) were recorded daily per focal monkey for 10 min. Monkeys were sampled evenly between 0900 and 1730 hours. Incidences of aggression, coalitions, scrounging and food sharing were recorded continuously; all other behaviors were recorded at 1 min intervals using point sampling (Martin and Bateson 2007). In each point sample, group members within two body lengths from the focal were recorded. The total number of sampling points was the same for all subjects.

Between 15 May 2011 and 8 June 2011 five puzzle feeders were introduced to the outdoor enclosures of each group. Monkeys could freely interact with the feeders.

Each feeder was made out of a cylindrical piece of white piping (length: 76.2 cm; diameter: 5.08 cm), with approximately 8–10 holes drilled into it (see Appendix Fig. A1 in Morton et al., 2015). Feeders were attached vertically to trees, 2–10 m apart. For each feeder, the bottom of the pipe was left open while the top of the pipe was closed. Ten paper packets, each containing five raisins, were placed in the top portion of each feeder, and wooden sticks were inserted into the holes of the pipes to prevent the packets from falling out from the bottom. The packets dropped freely from the pipe once all the wooden sticks had been removed by the monkeys. Feeders were introduced 4 days a week for approximately 30 min each day or until all of the puzzle feeders had been solved. During sessions, all instances in which a monkey approached another monkey at a feeding site were recorded, as well as the behavioral response of the receiving monkey (i.e. by avoiding or staying). East group underwent 8 sessions and West group underwent 10 sessions.

Following previous studies of study social relationship structure (e.g. Rebecchini et al. 2011; Koski et al. 2012), a set of behavioral measures (Table 1) were calculated as events per monkey dyad and subjected to a principal components analysis (PCA) with varimax rotation. Overall mean numbers of social dyadic interactions are provided in Appendix Table A1 in Morton et al. (2015).

Statistical analyses

All analyses were conducted in the R programming language, using the psych package (Revelle 2015). The “nfactors” function of the psych package produces a variety of preliminary statistics and an accompanying descriptive chart (Figure 1), which

includes VSS, Complexity, SRMR, and empirical BIC. These statistics are popular, well-documented, and useful for comparisons to other tests, e.g. Kaiser's criterion and PA; these two tests were visualized on a scree plot using a separate function – "fa.parallel".

Following previous studies, we used PCA to identify the underlying structure of capuchins' social relationships. Component loadings greater than |0.4| were considered salient (e.g. Koski et al. 2012). Components with high loadings (i.e. |0.7|) and/or those with four or more loadings greater than |0.4| were considered robust (Guadagnoli and Velicer 1988). Seventy-three dyads and 10 behavioral measures were entered into each PCA, which meets previous recommendations for having a fixed ratio of at least 5 between the sample size and number of variables (Gorsuch 1983). Components with high loadings (i.e. |0.7|) and/or those with four or more loadings greater than |0.4| were considered robust (Guadagnoli and Velicer 1988). The 2-component PCA solution comes from Morton et al. (2015).

As initially described by Horn (1965), PA is not appropriate for FA, only PCA. FA and PCA often produce very similar solutions in practice, but the underlying matrix algebra differs such that when each procedure is repeated, as in PA simulations, and the results can differ considerably. So while many tests need not distinguish between factors and components, PA must be adjusted to support FA (Revelle 2015). Since we chose to use PCA for data reduction, this was not an issue in the current analysis, but we wish to highlight the nuance of all such analyses.

Results

Determining the number of factors or components

Kaiser's criterion suggested that 3 components ought to be extracted. A scree plot suggested only 2 components should be retained in the solution (Figure 2). The results of our call to the "nfactors" function are shown in Figure 1. SRMR indicated that these data were comprised of at least 2 components. For VSS, the sharp rise from 1 to 2 components, and the flattening out of the curve from 2 components onwards, suggested that 2 components were the most appropriate number to extract. Complexity rose when a third component was added, suggesting that 2 components were a better fit than 3. The Empirical BIC suggested that 2 components were the best fit for these data, since empirical BIC reached a minimum with the 2-component solution. Finally, PA recommended that 2 components should be retained in the solution (Figure 2).

Collectively, whereas Kaiser's criterion suggested a 3-component model, comparisons between all of the other extraction tests largely recommended a 2-component solution. We therefore extracted both the 2 and 3 component solutions, and compared their fit and interpretability.

Extracted Solutions

The 2-component solution collectively explained 55.0% of the total variance. Component 1 (Table 2) had an eigenvalue of 3.72, explained 37.22% of the variance, and was characterized by moderate to high loadings ($\geq|0.45|$) on behaviors related to social affiliation (e.g. proximity, social foraging, food sharing, and grooming). Component 2 (Table 2) had an eigenvalue of 1.78, explained 17.8% of the variance, and was characterized by high loadings ($\geq|0.869|$) from agonistic behaviors, i.e. conflict

and conflict symmetry. Correlations between these two components were only weakly correlated ($r=-0.072$).

The 3-component solution had eigenvalues over 1.0, and explained 67.32% of the total variance. The first component was moderately correlated with the second component ($r=0.493$), and weakly correlated with the third component ($r=-0.106$). The second component was weakly correlated with the third component ($r=0.01$).

Component 1 (Table 2) had an eigenvalue of 2.56, explained 25.6% of the variance, and was characterized by moderate to high loadings (>0.4) by behaviors reflecting the importance of the relationship in terms of social affiliation (i.e. proximity, social foraging) and direct benefits gained from this affiliation (i.e. grooming, grooming symmetry, coalitions). Component 2 (Table 2) had an eigenvalue of 2.45, explained 24.48% of the variance, and was characterized by moderate to high loadings (>0.4) from behaviors related to tolerance to approaches (avoid-stay symmetry), tolerance at feeding sites (social foraging, food sharing, food sharing symmetry), and coalitionary support.

Component 3 (Table 2) had an eigenvector of 1.72, explained 17.24% of the variance, and was characterized by high loadings (>0.89) from behaviors indicating a lack of stability or predictability in the relationship (i.e. conflict and conflict symmetry).

To assess the internal consistency of the solutions, we applied ω_h to both the 2 and 3 component solutions (Dunn et al. 2014). For the 2 component solution, $\omega_h = 0.08$, which suggests that there were no large effects left unaccounted for by missing latent variables (e.g. a third component, cf. Cords and Aureli 2000). On the other hand, for the 3 component model, $\omega_h = 0.64$. An ω_h of this magnitude suggests that a single-level 3 component structure alone does not adequately model the data, and points towards

there being an additional, general latent variable (e.g. a combination of two of the extracted constructs).

Discussion

The PCA solution derived using Kaiser's criterion resembled the 3-component model proposed by Cords and Aureli (2000). Component 1 resembled relationship 'value' (e.g. grooming and food sharing), component 2 resembled relationship 'compatibility' (e.g. tolerance at feeding sites), and component 3 resembled relationship 'security' (e.g. rates of aggression). By contrast, the PCA solution derived using the other extraction tests recommended two components, i.e. basic affiliative and agonistic components, which, as previously discussed, reflects a more traditional approach to describing social relationship structure (e.g. Hinde 1976).

Similar findings have recently been reported in bonobos by Stevens et al. (2015), who identified three components resembling the 3-component model proposed by Cords and Aureli (2000) when using Kaiser's criterion and a scree test, but ended up retaining only two components (labelled 'value' and 'compatibility', respectively) based on a parallel analysis. Importantly, their findings illustrate how a scree test, which in our current study recommended the same number of components as a PA, can be more liberal than the unbiased parallel analysis.

As previously noted, most studies have relied solely on Kaiser's criterion to interpret and compare the social relationship structures of their animals. These studies, including the current study, typically find 3 components to subjects' social relationships – which, again, globally resemble Cords and Aureli's (2000) 3-component model (but

see Rebecchini et al. 2011). However, if one closely examines the individual loadings of the behavioral measures that are entered into each of these analyses (most of which are the same measures used across all of these studies), there are nevertheless striking differences in terms of how certain items load onto each of the three components. For example, in capuchins and Japanese macaques, aggression loads positively onto a component resembling relationship 'security' (this study; Majolo et al. 2010), whereas in chimpanzees, Barbary macaques, and corvids, the same measure loads positively onto a component resembling relationship 'compatibility' (Fraser et al. 2008; Fraser and Bugnyar 2010; Koski et al. 2012). Also, in capuchins and Japanese macaques, grooming symmetry loads positively onto a component resembling relationship 'value' (this study; Majolo et al. 2010), whereas in chimpanzees and Barbary macaques, the same measure loads positively onto a component resembling relationship 'security' (Fraser et al. 2008; McFarland and Majolo 2011).

Two possible explanations may underlie these differences. First, most species studied to date (with the exception of spider monkeys; Rebecchini et al. 2011) may indeed have three dimensions to their social relationship structure that broadly resemble the 3-component model proposed by Cords and Aureli (2000); however, for species like capuchins and Barbary macaques, certain items like aggression and grooming symmetry may have different "social meanings" compared to other species, and therefore cluster onto different components compared to solutions derived in other species. Alternatively, differences between studies may reflect instability in the PCA solutions derived using Kaiser's criterion. This is not to say that using Kaiser's criterion is "wrong" *per se*, or that species do not exhibit three dimensions to their social

relationships. Rather, as noted previously, one major disadvantage to using Kaiser's criterion is that it often leads to unstable solutions as a result of over-extraction (Fabrigar et al. 1999). In other words, Kaiser's criterion is easily susceptible to sampling variability (Ruscio and Roche 2012) and spreads item loadings "too thin" across components/factors. Therefore, within the context of social relationship studies that have used Kaiser's criterion, structural differences in item loadings across studies could be biologically meaningful (i.e. they could reflect differences in sociality), but could just as likely reflect structural instability brought about by sampling variability. So far, studies that have used more robust extraction methods like PA support a more tradition view of social relationship structure – i.e. affiliative and agonistic dimensions (Hinde 1976).

Although scree tests are subjective, they can be useful when automated methods disagree on the number of components/factors to extract. In such cases, a scree test may be used as a "tie-breaker" if the plot reveals a clear and distinct drop in the eigenvalues past a certain component/factor. Such instances, however, are becoming increasingly rare as automated methods are improved upon. The bottom line is that scree tests should only be used alongside less subjective and automated tests, not Kaiser's criterion. If different tests recommend different extraction numbers, more sophisticated analyses like Everett's tests may be needed to determine which model to use for subsequent analyses after extracting multiple solutions with differing numbers of components/factors (Everett 1988).

PA has a long, published record of success in multiple simulation studies (Zwick and Velicer 1986; Fabrigar 1999; Ruscio and Roche 2012), but again, no single method should be relied upon when deciding how many components/factors to extract. Using

Kaiser's criterion as the sole method for component/factor extraction is not appropriate. PA is a recommended starting point for researchers exploring their data in preparation for a data reduction analysis because of its robustness, but also due to the ease with which it can be directly compared to the scree tests (Figure 2), and other quantitative methods (e.g. empirical BIC).

Conclusion

The current example should serve as a cautionary note to researchers wishing to use data reduction analyses to study social relationship structure in animals. In particular, careful decisions must be made when determining how many components or factors to retain in one's analysis. In light of the well-known deficiencies associated with Kaiser's criterion, we recommend that researchers refrain from using this technique in future work on social relationship structure. Although we recommend a wider use of PA, all methods of course have their drawbacks (Ruscio and Roche 2012). Therefore, as discussed, PA should not be the sole method used to determine how many components/factors to extract. Rather, this approach should be supplemented with other extraction techniques, such as scree tests and more robust and automated tests (e.g. empirical BIC, VSS, Comparison Data). If these tests recommend the same number of components/factors, then researchers can be confident about their decisions to extract. Avoiding Kaiser's criterion and supplementing scree tests with more robust and automated tests will greatly improve the utility and reliability of data reduction techniques for comparative studies of animal social relationships.

Acknowledgements

We thank Prof. Andrew Whiten, Director of Living Links, for permission to conduct research at LL, and the LL students and staff for support and assistance during data collection. Special thanks go to Phyllis Lee, Hannah Buchanan-Smith, Alexander Weiss, Sarah Vick, and Nicola Koyama for their useful comments on earlier drafts of the manuscript. We also thank Alexander Weiss for help with the parallel analysis. FBM thanks the University of Stirling and the Primate Society of Great Britain for funding. We declare no conflict of interest.

References

- Alfaro JWL, Silva JDSE, Rylands AB (2012) How different are robust and gracile capuchin monkeys? An argument for the use of *Sapajus* and *Cebus*. *American Journal of Primatology*, doi: 10/1002/ajp.22007.
- ASAB (2012) Guidelines for the treatment of animals in behavioural research and teaching. *Animal Behaviour*, 83:301-309.
- Cattell RB (1966) The scree test for the number of factors. *Multivariate Behavioural Research*, 1:245-276.
- Cords M, Aureli F (2000) Reconciliation and relationship qualities. In: *Natural conflict resolution* (Aureli F. & de Waal F.B.M., eds.). University of California Press, Berkley, CA, p. 177-198.
- Courtney, M. G. R. (2013). Determining the number of factors to retain in EFA: Using the SPSS R-Menu v2. 0 to make more judicious estimations. *Practical Assessment, Research & Evaluation*, 18(8), 1-14.

457 Dunn, T. J., Baguley, T., & Brunsden, V. (2014). From alpha to omega: A practical
 458 solution to the pervasive problem of internal consistency estimation. *British*
 459 *Journal of Psychology*, 105(3), 399-412.

460 Everett, J. E. (1983). Factor comparability as a means of determining the number of
 461 factors and their rotation. *Multivariate Behavioral Research*, 18(2), 197-218.

462 Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating
 463 the use of exploratory factor analysis in psychological research. *Psychological*
 464 *methods*, 4(3), 272.

465 Field A (2009) *Discovering statistics using SPSS*, third ed. London, SAGE.

466 Fragaszy DM, Visalberghi E, Fedigan LM (2004) *The complete capuchin*. Cambridge:
 467 Cambridge University Press.

468 Fraser ON, Bugnyar T (2010) The quality of social relationships in ravens. *Animal*
 469 *Behaviour*, 79:927-933.

470 Fraser ON, Schino G, Aureli F (2008) Components of relationship quality in
 471 chimpanzees. *Ethology*, 114:834-843.

472 Garrido, L. E., Abad, F. J., & Ponsoda, V. (2013). A new look at Horn's parallel analysis with
 473 ordinal variables. *Psychological methods*, 18(4), 454.

474 Gorsuch RL (1983) *Factor analysis*, 2nd Edition. Hillsdale, NJ:LEA.

475 Gouadagnoli E, Velicer WF (1988) Relation of sample size to the stability of component
 476 patterns. *Psychological Bulletin*, 103:265-275.

477 Hinde RA (1976) Interactions, relationships, and social structure. *Man*, 11:1-17.

478 Hofmann, R. J. (1978). Complexity and simplicity as objective indices descriptive of factor
 479 solutions. *Multivariate Behavioral Research*, 13(2), 247-250.

480 Hooper, D., Coughlan, J., & Mullen, M.R. (2008). Structural equation modelling: Guidelines for
 481 determining model fit. *Journal of Business Research Methods*, 6, 53–60.

482 Horn JL (1965) A rationale and test for the number of factors in factor analysis. *Psychometrika*,
 483 30:179-185.

484 Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis:
 485 Conventional criteria versus new alternatives. *Structural equation modeling: a*
 486 *multidisciplinary journal*, 6(1), 1-55.

487 Kaiser HF (1960) The application of electronic computers to factor analysis. *Ed. Psychol. Meas.*,
 488 20:141-151.

489 Koski SE, de Vries H, van de Kraats A, Sterck EHM (2012) Stability and change of social
 490 relationship quality in captive chimpanzees. *International Journal of Primatology*, 33:905-
 491 921.

492 Ledesma RD, Valeromora P (2007) Determining the number of factors to retain in EFA: an easy-
 493 to-use computer program for carrying out parallel analysis. *Practical Assessment*,
 494 *Research & Evaluation*, 12:1-11.

495 Leonardi, R., Buchanan-Smith, H. M., Dufour, V., MacDonald, C., & Whiten, A. (2010).
 496 Living together: Behavior and welfare in single and mixed species groups of
 497 capuchin (*Cebus apella*) and squirrel monkeys (*Saimiri sciureus*). *American Journal*
 498 *of Primatology*, 72, 33–47.

499 Loretto, M.-C., Fraser, O. N., & Bugnyar, T. (2012). Ontogeny of social relations and
 500 coalition formation in common ravens (*Corvus corax*). *International Journal of*
 501 *Comparative Psychology*, 25, 180–194.

502 Macdonald, C., & Whiten, A. (2011). The 'Living Links to Human Evolution' research
 503 centre in Edinburgh zoo: A new endeavour in collaboration. *International Zoo*
 504 *Yearbook*, 45, 7–17.

505 Majolo, B., Ventura, R., & Schino, G. (2010). Asymmetry and dimensions of relationship
 506 quality in the japanese macaque (*Macaca fuscata yakui*). *International Journal of*
 507 *Primatology*, 31, 736–750.

508 Martin, P., Bateson, P. (2007). Measuring behaviour: an introductory guide. Cambridge,
 509 U.K.: Cambridge University Press.

510 McFarland, R., & Majolo, B. (2011). Exploring the components, asymmetry and
 511 distribution of relationship quality in wild barbary macaques (*Macaca sylvanus*).
 512 *PLoS One*, 6, e28826.

513 Morton FB, Weiss A, Buchanan-Smith HM, Lee PC (2015) Capuchin monkeys with
 514 similar personalities have higher-quality relationships independent of age, sex,
 515 kinship, and rank. *Animal Behaviour*. Doi: 10.1016/j.anbehav.2015.04.013

516 Posada, D., & Buckley, T. R. (2004). Model selection and model averaging in
 517 phylogenetics: advantages of Akaike information criterion and Bayesian
 518 approaches over likelihood ratio tests. *Systematic biology*, 53(5), 793-808.

519 Rebecchini, L., Schaffner, C. M., & Aureli, F. (2011). Risk is a component of social
 520 relationships in spider monkeys. *Ethology*, 117, 691–699.

521 Revelle, W., & Rocklin, T. (1979). Very simple structure: An alternative procedure for
522 estimating the optimal number of interpretable factors. *Multivariate Behavioral*
523 *Research*, 14(4), 403-414.

524 Revelle, W. (2015) psych: Procedures for Personality and Psychological Research,
525 Northwestern University, Evanston, Illinois, USA, [http://CRAN.R-](http://CRAN.R-project.org/package=psych)
526 [project.org/package=psych](http://CRAN.R-project.org/package=psych) Version = 1.5.4.

527 Ruscio, J., & Roche, B. (2012). Determining the number of factors to retain in an
528 exploratory factor analysis using comparison data of known factorial structure.
529 *Psychological assessment*, 24(2), 282.

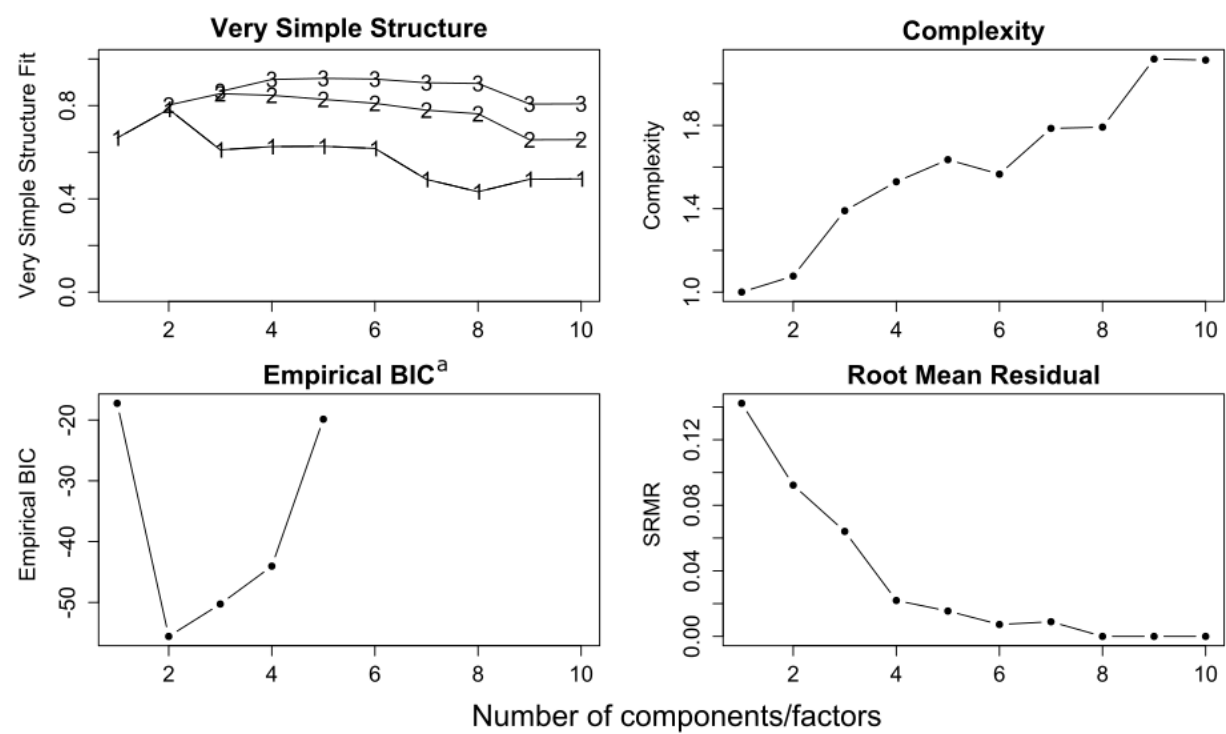
530 Schwarz, G. (1978). Estimating the dimension of a model. *The annals of statistics*, 6(2),
531 461-464.

532 Stevens JMG, de Groot E, Staes N (2015) Relationship quality in captive bonobo
533 groups. *Behaviour*, 152:259-283.

534 Zwick WR, Velicer WF (1986) Comparison of five rules for determining the number of
535 components to retain. *Psychological Bulletin*, 99: 432-442.

Figure 1

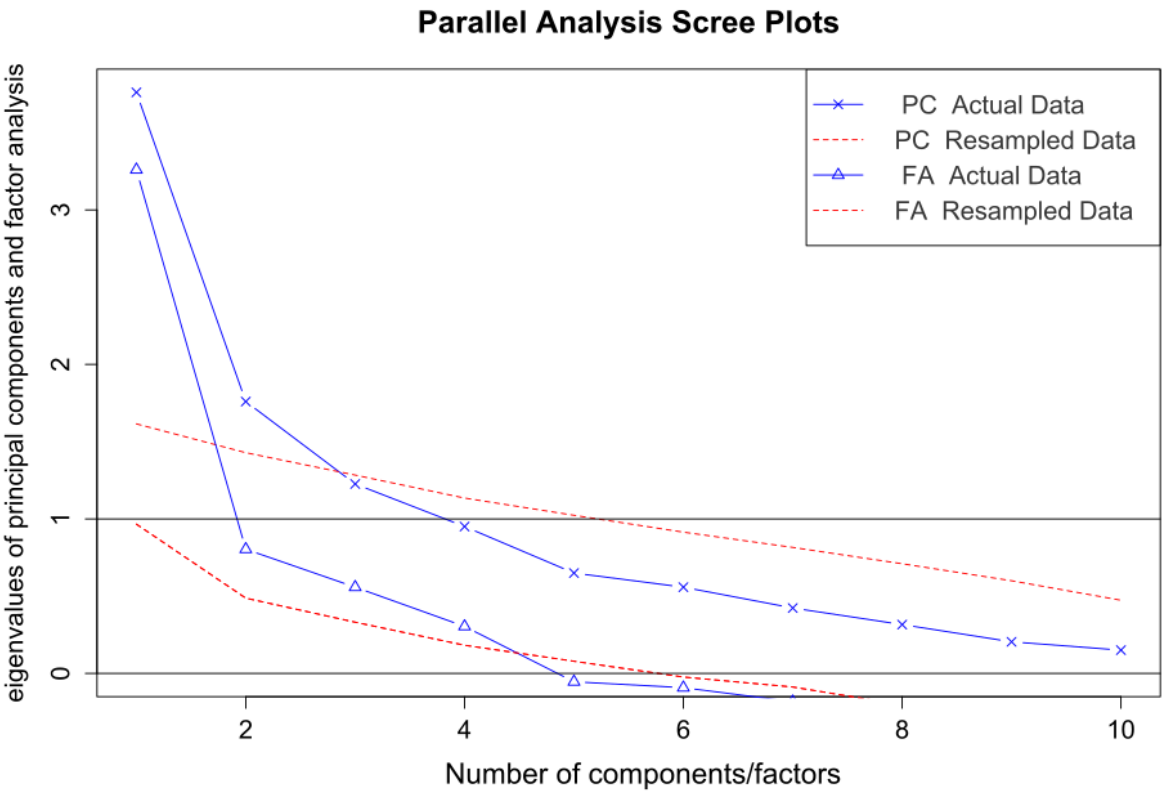
Figure 1. Plotted result of the R psych package “nfactors” function.



^a The number of variables (10) limits the calculation of empirical BIC to solutions of at most 5 components/factors.

Figure 2

Figure 2. Results of parallel analysis, on a scree plot.



Note. Triangles indicate eigenvalues for components; X's indicate adjusted eigenvalues for factors. Dashed lines represent random simulated eigenvalues for the corresponding factor or component procedures. The horizontal black line at 1 is Kaiser's criterion.

Table 1. Behavioral measures calculated for each monkey dyad (reproduced with permission from Morton et al. 2015). *

Behavioral measure	Calculation
Avoid/Stay Symmetry	$(\text{no. times A approaches B}) / [(\text{no. times A approaches B}) + (\text{no. times B approaches A})]$
Coalitions	$[(\text{no. times A supports B}) + (\text{no. times B supports A})]$
Conflict	$[(\text{no. times A aggresses B}) + (\text{no. times B aggresses A})]$
Conflict Symmetry	$(\text{no. times A aggresses B}) / [(\text{no. times A aggresses B}) + (\text{no. times B aggresses A})]$
Food Sharing	$[(\text{no. times A gives to B}) + (\text{no. times B gives to A})]$
Food Sharing Symmetry	$(\text{no. times A gives to B}) / [(\text{no. times A gives to B}) + (\text{no. times B gives to A})]$
Grooming	$[(\text{no. minutes A grooms B}) + (\text{no. minutes B grooms A})]$
Grooming Symmetry	$(\text{no. minutes A grooms B}) / [(\text{no. minutes A grooms B}) + (\text{no. minutes B grooms A})]$
Social Foraging	$[(\% \text{ of time A within proximity of B}) + (\% \text{ of time B within proximity of A})]$
Spatial Proximity	$[(\% \text{ of time A within proximity of B}) + (\% \text{ of time B within proximity of A})]**$

*Modified from Rebecchini et al. (2011) and Koski et al. (2012). **These calculations do not include time spent grooming or time spent social foraging.

Table 2. Varimax-rotated PCA structure of 10 behavioral measures based on Kaiser’s criterion, a scree test, and parallel analysis^a

Behavioral measure	Kaiser’s criterion			Parallel analysis ^{b,c}	
	PC1	PC2	PC3	PC1	PC2
Spatial Proximity	.863	.274	.028	.803	-.112
Grooming	.806	.285	.050	.772	-.077
Grooming Symmetry	-.743	.099	.168	-.450	.343
Avoid/Stay Symmetry	.048	-.763	-.036	-.507	-.188
Food Sharing	.247	.680	-.202	.651	-.120
Food Sharing Symmetry	.088	.658	.175	.532	.272
Coalitions	.434	.658	-.064	.771	-.033
Social Foraging	.590	.607	-.011	.846	-.026
Conflict Symmetry	.006	.033	.899	.049	.865
Conflict	-.086	-.019	.898	-.053	.875

^aSalient loadings (>|0.4|) per behavior are in bold; PC=principal component;

^bA scree test, complexity, empirical BIC, and VSS also recommended that two components be extracted; ^c reproduced with permission from Morton et al. (2015).