**Guidelines for using extraction methods in data reduction analyses of social relationship structure in animals**

**F. Blake Mortona,b,c\*, Drew Altschulb,d**

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, Franklin and Marshall College, U.S.A.

d Department of Psychology, The University of Edinburgh, Edinburgh, U.K.

**\*Correspondence:**

F.B. Morton, Department of Psychology, Franklin and Marshall College, Lancaster, Pennsylvania, 17694 U.S.A. E-mail address: [morton.blake@gmail.com](mailto: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

**Introduction**

Comparative studies of social relationships are critical to understanding animal sociality and how it evolves (Cords and Aureli 2000; Fraser and Bugnyar 2010). One challenge researchers face, however, is how best to define and quantify social relationships such that the data are systematic and comparable across studies and taxa, while simultaneously highlighting the *unique* aspects of a species’ sociality.

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) (Field 2009; Gorsuch 1983). Because they provide researchers with a relatively objective approach to categorizing different sets of behavioural interactions within dyads (e.g. rates of grooming or aggression between two individuals), data reduction analyses are increasingly being used to describe social relationship structure in animals; that is, the number of dimensions and the characteristics of those dimensions (Majolo et al. 2010; McFarland and Majolo 2011; Rebeccini et al. 2011; Morton et al. 2015; Fraser et al. 2008; Koski et al. 2012; Stevens et al. 2015; Fraser and Bugnyar 2010; Loretto et al. 2012).

Social relationship structure in animals has traditionally been viewed in terms of two basic 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, however, reveal a more complex picture. In capuchin monkeys (*Sapajus sp.*), Morton et al. (2015) identified two dimensions characterized by affiliative and agonistic behaviours, respectively, supporting the traditional view of social relationship structure (Hinde,1976). By contrast, Rebecchini et al. (2011) and Stevens et al. (2015) identified two dimensions to spider monkey (*Ateles geofroyii*) and bonobo (*Pan paniscus*) social relationships, with the first component being characterized exclusively by affliative behaviors and the second component being characterized by a mixture of affiliative and agonistic behaviors (e.g. rates of embrace and aggression in spider monkeys, and rates of aggression and counter-inspection in bonobos). In all remaining studies involving data reduction analyses, researchers have reported three dimensions, some of which support a 3-component model originally proposed by Cords and Aureli (2000) – i.e. relationship “value” (i.e. immediate benefits afforded by the relationship), “compatibility” (i.e. tolerance based on partners’ shared history), and “security” (i.e. consistency and predictability in partners’ behaviour) (Fraser and Bugnyar 2010; Fraser et al. 2008; Majolo et al. 2010; McFarland and Majolo 2011).

Differences in relationship structure reported among data reduction studies may reflect differences in the natural and social world of each species. Items that “load” onto different components may indicate that those items have a different behavioural meaning and/or occur under different contexts compared to other species. However, as we will discuss, one alternative and more parsimonious interpretation is that the differences that exist between studies simply reflect instability in the structural solutions derived from the extraction methods that were used.

*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 (Ledesma and Valero-Mora 2007; Zwick and Velicer 1986). 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 more automated methods (discussed below), such as a “tie-breaker” if the plot reveals a clear and distinct drop in the eigenvalues past a certain component/factor.

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). Ruscio and Roche (2012) simulated data from abstract theoretical models with varying numbers of components, and for each simulation, tested several methods to determine how often each method selected the “correct” number of components, as defined by the theoretical models. In these simulations, Kaiser’s criterion lead to a success rate of 8.77% and failed 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, and more automatic 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 could indicate poorer fit, or several additional issues: Complexity only assesses whether a dataset fits a single-level structure; that is, high complexity suggests multi-level structure might be present. It could also indicate that FA and PCA are not appropriate analyses for the data. If no solution nears a complexity of 1, then the measure will not be informative. However, if a factor solution is very close to 1, this is strong, but not incontrovertible, evidence that the number of factors in question is the best choice.

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).

All methods of course have their drawbacks (Ruscio and Roche 2012); there is no “one size fits” all approach. Even if some methods are demonstrably more accurate than others, e.g. PA vs. Kaiser’s criterion, few datasets will produce an immediate and clear solution. Therefore, it is paramount that no single automated extraction test should be used as the sole method to determine how many components/factors to extract from a data reduction analysis. Ideally, multiple automated 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 (Gorsuch 1983).

Of the methods discussed, we recommend PA and BIC because of their strong performance under simulation (Ruscio and Roche 2012), but novel methods are being developed with surprising frequency, and we encourage readers to explore the literature for newly verified methods.

*This study*

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; Rebeccini 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 (but see Morton et al. 2015; Stevens et al. 2015). This is likely facilitated by the fact that popular statistical packages (e.g. SPSS) use Kaiser’s criterion and scree tests as the default method for determining component extraction. Stevens et al. (2015), however, compared Kaiser’s criterion to a parallel analysis to determine component extraction number, and found that the latter approach lead to a more conservative structure (2 vs. 3 components). Again, Kaiser’s criterion increases the risk of over-extraction compared to more automated tests, which in turn can lead researchers to make entirely different conclusions about their results.

To encourage the future use of more robust extraction techniques, we use data from a previous study of brown capuchin monkeys (*Sapajus sp.*) by Morton et al. (2015) to compare how each of the six extraction methods (described above) can differentially affect the results, and thus interpretation, of social relationship structure. We provide instructions on how to perform each of the automated extraction tests, and provide recommendations about their use. As initially described by Horn (1965), parallel analysis 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, the results can differ considerably. So while five of the extraction methods used in our study need not distinguish between factors and components, PA must be adjusted to support FA (Revelle 2015). We therefore chose to use PCA for data reduction to allow us to compare all six analyses within the same data reduction analysis.

**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 Fragaszy 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 ± SD = 10.79 ± 8.55 years, *N* = 11) and 3 to 14 years for females (mean ± SD = 8.86 ± 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. Koski et al. 2012; Rebeccini et al. 2011), a set of behavioral measures (Table 1) were calculated as events per monkey dyad and subjected to a principal components analysis (PCA). Our solutions were rotated with the orthogonal Varimax rotation, as this has been recommended for small samples (Budaev 2010), and maximizes comparability with previous studies, which have predominately used orthogonal rotations (e.g. Stevens et al. 2015, MacFarland & Majolo 2011). 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 using a separate function – “fa.parallel” – which automatically generates random PA eigenvalues and plots them on a scree plot.

Following previous studies, we used PCA to identify the underlying structure of capuchins’ social relationships. In keeping with broad recommendations for data reduction studies on sociality (Stevens, 2012), component loadings greater than |0.4| were considered salient (e.g. Koski et al. 2012). However, components were not considered robust unless they possessed high loadings (i.e. |0.7|) and/or four or more loadings greater than |0.4| (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).

Internal consistency of the component solutions was assessed with “omega-heirachical” – ωh. ωh is considered to be superior to older, and perhaps better known, measures of internal consistency, such as Cronbach’s *α* and Revelle’s *β* (Zinbarg et al. 2005). ωh represents the proportion of observed variance in the measures that is due to a general factor common to all of the items, so a small ωh suggests that the proposed structure is robust to the addition of a general factor, and as ωh approaches 1, the proposed model becomes increasingly unsuited to the data. The psych package includes the function “omega” to calculate ωh, along with related consistency statistics.

*Instructions on how to perform each automated extraction method*

These instructions are specific to the R programming language because of its wide use and robust, well-maintained feature set. All commands are available from base R, or the “psych” package (Revelle 2015).

Data should be organized in a “data.frame” format which is native to R – we will call our example data.frame: “df”. The first column of the data.frame should contain the names of individual and/or dyads. Many functions require only numeric input, and the first column can be subsetted out of the data.frame with the command “df[,-1]”. So for an example, to examine the correlation matrix of the data for suitability, the entire command “cor(df[,-1])” will display the numeric correlation matrix. We also suggest using “corPlot” in exactly the same way, to view the correlation matrix graphically.

Executing the command “nfactors(df[,-1])” will display graphical representations of VSS, Complexity, eBIC, and SRMR (e.g. Figure 1), plus, it will produce a myriad of other fit statistics which may be of utility to the advanced user. Executing fa.parallel(df[,-1])” will display a plot, like in Figure 2, as well as give a specific recommendation of how many factors or components to retain for extraction.

**Results**

*Suitability of the data*

Before extracting any solutions, we inspected the data for suitable factorability. The first step in doing this is manually generating the correlation matrix of the measured variables, and inspecting it for any abnormalities. The highest inter-item correlation was strong (*r =* 0.75), and every item was moderately correlated (*r >* 0.48) with at least one other item in the matrix.Additional diagnostics indicated acceptable sampling adequacy (Kaiser–Meyer–Olkin measure (KMO) = 0.7, sphericity test: *χ2 =* 302.6, df = 45, *p* < 0.0001). Collectively, we found no evidence that the data were unsuitable for PCA.

*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. 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. SRMR indicated that these data were comprised of at least 2 components; using solutions with SRMR > 0.1 is not recommended. Finally, PA recommended that 2 components should be retained in the solution (Figure 2): when interpreting a PA scree plot, only components which are above the PC resampling line (and outside the error bars) should be retained for extraction.

*Extracted Solutions*

W

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 (>|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 (>|0.869|) from agonistic behaviors, i.e. conflict and conflict symmetry. Loadings between these two components were weakly correlated (r = -0.072), indicating discrete components with very little overlap.

The 3-component solution had eigenvalues over 1.0, and explained 67.32% of the total variance. 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). indicating that the item loadings from both components overlapped to a considerable degree. The first component was , and t, indicating discrete components with very little overlap in item loadings.

To assess the internal consistency of the solutions,i.e. the degree to which items within components are “redundant” to the overall solution (MacDonald 1999), we applied ωh to both the 2 and 3 component solutions (Dunn et al. 2014). For the 2 component solution, ωh = 0.08, which suggests that there were no large effects left unaccounted for by missing latent variables (e.g. a third component). On the other hand, for the 3 component model, ωh = 0.64. An ωh 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**

*PCA solutions derived from each extraction method*

Our findings are comparable to previous studies in various ways depending on which extraction method was used. First, based on the 3-component solution derived using Kaiser’s criterion, one might conclude that capuchins possess three components to their social relationships as found in corvids (refs), chimpanzees (refs), and various macaque species (refs). Upon examining the individual loadings of the 3-component solution to other 3-component solutions in previous studies, one might also conclude that there are striking similarities and differences in terms of how certain items load onto components across species (refs). For instance, XXXXX. Our analysis of structure using more robust and automated tests, however, recommended a more conservative 2-component structure. Thus, while structural comparisons between our 3-component solution and those reported in other studies *may* be socially meaningful, they are more likely to reflect the structural instability of this solution brought on by over-extraction and spreading items “too thin” across components. This would explain, for instance, why the first and second components contained overlapping items with moderately high loadings (Table 2); the internal consistencies of both components suggested that a more general latent variable (e.g. a combination of two of the extracted constructs) could underlie the overall solution.

By contrast, the PCA solution derived using the other extraction tests recommended two components. Very little overlap was found between components, indicating that each component contained items with very distinct behavioural characteristics. Overall, this structure reflects a more traditional view of animal social relationships, i.e. basic affiliative and agonistic components (e.g. Hinde 1976), and may be more robust (and thus meaningful for comparative studies) than the 3-component solution.

*Sample limitations*

As is common in samples of animal behavior, our sample size was not overly large, and because observations were dyadic, they were not fully independent. Independence of observations is an important consideration for all researchers; though dyadic observations have been shown to have validity, even if they do not capture information about the individual (Moore et al. 2013). Nevertheless, incorporation of both individual and dyadic observations within future studies will be an important advance.

LIMITATIONS: Items

*What if multiple automated methods disagree?*

We found that all the automated tests generally agreed on the number of components to extract. However, what if multiple automated methods disagree on the number of components to extract? This is not uncommon.

As previously noted, in such cases a scree test may be used as a quick and easy “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. For example, additional,

Researchers should always keep in mind the theory they wish to test, and where theory is well-established, it can be used to guide choices in how many components/factors to extract. If the analysis is wholly exploratory, or theories are at odds, there is nothing wrong with extracting multiple factor structures and comparing them, as we have done in this paper. Factor interpretability can be assessed post-extraction, and depending on what variables are of interest, investigating additional associations (Altschul et al. 2015) may indicate which structure is the most useful.

Finally, basic PCA or FA may not be the best method for all situations. More complex, potentially hierarchical data may require a more advance modelling approach. FA is itself a specific implementation of more general Structural Equation Modelling (SEM), which allows users to specify latent variables (such as our components) and all paths between latent and measured variables. If one suspects that a simple, one-level factor or component model is not sufficient to explain the data, e.g. there are unambiguous sources of non-independence, then SEM ought to be considered. However, while SEM gives users a much greater degree of control, it also makes SEM better suited for confirmatory, not exploratory, models.

**Conclusion**

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. We also recommend a wider use of more robust and automated extraction techniques (e.g. PA, 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. If they disagree, then there are multiple avenues to take to aid decision-making on extraction and modelling framework. 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.

**Compliance with Ethical Standards**

**Funding:** The study was funded by the University of Stirling and the Primate Society of Great Britain (no grant numbers were issued).

**Conflict of Interest:** Both authors declare no conflict of interest.

**Ethical approval:** All applicable international, national, and institutional guidelines for the care and use of animals were followed. This study was approved by Edinburgh Zoo and the ethics committee of the Psychology Department at the University of Stirling, and complied with the ethical standards of the ASAB (2012) Guidelines. This article does not contain any studies with human participants performed by any of the authors.

**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*. Am J

Primatol 74:273-286

ASAB (2012) Guidelines for the treatment of animals in behavioural research and

teaching. Anim Behav 83:301-309

Cattell RB (1966) The scree test for the number of factors. Multivariate Behav

Res 1:245-276

Cords M, Aureli F (2000) Reconciliation and relationship qualities. In: Aureli F and de

Waal FBM (eds) Natural conflict resolution. University of California Press,

Berkley, pp 177-198

Courtney MGR (2013) Determining the number of factors to retain in EFA: Using the SPSS R-Menu v2. 0 to make more judicious estimations. PARE 18:1-14

Dunn TJ, Baguley T, Brunsden V (2014) From alpha to omega: A practical solution to the pervasive problem of internal consistency estimation. Br J Psychol 105:399-412

Everett JE (1983) Factor comparability as a means of determining the number of factors and their rotation. Multivariate Behav Res 18:197-218

Fabrigar LR, Wegener DT, MacCallum RC, Strahan EJ (1999) Evaluating the use of exploratory factor analysis in psychological research. Psychol Methods 4:272

Field A (2009) Discovering statistics using SPSS, 3rd edn. SAGE, London

Fragaszy DM, Visalberghi E, Fedigan LM (2004) The complete capuchin. Cambridge

University Press, Cambridge

Fraser ON, Bugnyar T (2010) The quality of social relationships in ravens. Anim Behav

79:927-933

Fraser ON, Schino G, Aureli F (2008) Components of relationship quality in

chimpanzees. Ethol 114:834-843

Garrido LE, Abad FJ, Ponsoda V (2013) A new look at Horn’s parallel analysis with ordinal variables. Psychol Methods 18:454

Gorsuch RL (1983) Factor analysis, 2nd edn. LEA, Hillsdale

Gouadagnoli E, Velicer WF (1988) Relation of sample size to the stability of component

patterns. Psychological Bull 103:265-275

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

Hofmann RJ (1978) Complexity and simplicity as objective indices descriptive of factor solutions. Multivariate Behav Res 13:247-250

Hooper D, Coughlan J, Mullen, MR (2008) Structural equation modelling: Guidelines for determining model fit. EJBRM 6:53-60

Horn JL (1965) A rationale and test for the number of factors in factor analysis. Psychometrika

30:179-185

Hu LT, Bentler PM (1999) Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. Struct Equ Modeling 6:1-55

Kaiser HF (1960) The application of electronic computers to factor analysis. Ed Psychol Meas,

20:141-151

Koski SE, de Vries H, van de Kraats A, Sterck EHM (2012) Stability and change of social

relationship quality in captive chimpanzees. Int J Primatol 33:905-921

Ledesma RD, Valeromora P (2007) Determining the number of factors to retain in EFA: an easy-

to-use computer program for carrying out parallel analysis. PARE 12:1-11

Leonardi R, Buchanan-Smith HM, Dufour V, MacDonald C, Whiten A (2010) Living together: Behavior and welfare in single and mixed species groups of capuchin (*Cebus apella*) and squirrel monkeys (*Saimiri sciureus*). Am J Primatol *72*:33-47

Loretto MC, Fraser ON, Bugnyar T (2012) Ontogeny of social relations and coalition formation in common ravens (*Corvus corax*). Int J Comp Psychol *25*:180-194

Macdonald C, Whiten A (2011) The ‘Living Links to Human Evolution’ research centre in Edinburgh zoo: A new endeavour in collaboration. Int Zoo Yearb 45:7-17

Majolo B, Ventura R, Schino G (2010) Asymmetry and dimensions of relationship quality in the japanese macaque (*Macaca fuscata yakui*). Int J Primatol 31:736-750

Martin P, Bateson P (2007) Measuring behaviour: an introductory guide. Cambridge

University Press, Cambridge

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

Morton FB, Weiss A, Buchanan-Smith HM, Lee PC (2015) Capuchin monkeys with similar personalities have higher-quality relationships independent of age, sex, kinship, and rank. Anim Behav 105:163-171

Posada D, Buckley TR (2004) Model selection and model averaging in phylogenetics: advantages of Akaike information criterion and Bayesian approaches over likelihood ratio tests. Syst Biol 53:793-808

Rebecchini L, Schaffner CM, Aureli F(2011) Risk is a component of social relationships in spider monkeys. Ethol 117:691-699

Revelle W, Rocklin T (1979) Very simple structure: An alternative procedure for estimating the optimal number of interpretable factors. Multivariate Behav Res 14:403-414

Revelle W (2015) psych: Procedures for Personality and Psychological Research. Northwestern University, Evanston, Illinois, USA. http://CRAN.R-project.org/package=psych Version = 1.5.4. Accessed 26 May 2015

Ruscio J, Roche B (2012) Determining the number of factors to retain in an exploratory factor analysis using comparison data of known factorial structure. Psychol Assess 24:282

Schwarz G (1978) Estimating the dimension of a model. Ann Stat 6:461-464

Stevens JMG, de Groot E, Staes N (2015) Relationship quality in captive bonobo

groups. Behav 152:259-283

Zwick WR, Velicer WF (1986) Comparison of five rules for determining the number of

components to retain. Psychol Bull 99: 432-442

**Figure Captions**

**Fig. 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

**Fig. 2**

Results of parallel analysis, on a scree plot. Triangles indicate eigenvalues for factors; X’s indicate adjusted eigenvalues for components. The dashed and dotted lines represent random resampled eigenvalues for the corresponding factor or component procedures, respectively. The horizontal black line at 1 is Kaiser’s criterion