

RESEARCH ARTICLE

Personality assessment and model comparison with behavioral data: A statistical framework and empirical demonstration with bonobos (*Pan paniscus*)

Jordan S. Martin^{1,2,3}  | Scott A. Suarez^{2,4}

¹ Department of Anthropology, Emory University, Atlanta, Georgia

² Department of Anthropology, Miami University, Oxford, Ohio

³ Department of Cognitive Biology, University of Vienna, Vienna, Austria

⁴ Department of Behavioral Sciences, San Diego Mesa College, San Diego, California

Correspondence

Jordan S. Martin, Department of Anthropology, Emory University, 201 Dowman Drive, Atlanta, Georgia 30322.

Email: jsm.primatology@gmail.com

Funding information

Miami University Office for the Advancement of Research and Scholarship; The Rebecca Jeanne Andrew Memorial Award

Interest in quantifying consistent among-individual variation in primate behavior, also known as personality, has grown rapidly in recent decades. Although behavioral coding is the most frequently utilized method for assessing primate personality, limitations in current statistical practice prevent researchers' from utilizing the full potential of their coding datasets. These limitations include the use of extensive data aggregation, not modeling biologically relevant sources of individual variance during repeatability estimation, not partitioning between-individual (co)variance prior to modeling personality structure, the misuse of principal component analysis, and an over-reliance upon exploratory statistical techniques to compare personality models across populations, species, and data collection methods. In this paper, we propose a statistical framework for primate personality research designed to address these limitations. Our framework synthesizes recently developed mixed-effects modeling approaches for quantifying behavioral variation with an information-theoretic model selection paradigm for confirmatory personality research. After detailing a multi-step analytic procedure for personality assessment and model comparison, we employ this framework to evaluate seven models of personality structure in zoo-housed bonobos (*Pan paniscus*). We find that differences between sexes, ages, zoos, time of observation, and social group composition contributed to significant behavioral variance. Independently of these factors, however, personality nonetheless accounted for a moderate to high proportion of variance in average behavior across observational periods. A personality structure derived from past rating research receives the strongest support relative to our model set. This model suggests that personality variation across the measured behavioral traits is best described by two correlated but distinct dimensions reflecting individual differences in affiliation and sociability (Agreeableness) as well as activity level, social play, and neophilia toward non-threatening stimuli (Openness). These results underscore the utility of our framework for quantifying personality in primates and facilitating greater integration between the behavioral ecological and comparative psychological approaches to personality research.

KEYWORDS

bonobo, factor analysis, mixed model, personality, social behavior

1 | INTRODUCTION

Animal personality is a burgeoning interdisciplinary research program that seeks to describe and explain individual differences in cognition, affect, and behavior that are stable across time and contexts (Carere & Maestripietri, 2013). This field has been built upon the methodological

and theoretical foundations of both behavioral ecology (Mathot & Dingemanse, 2015; Réale, Reader, Sol, McDougall, & Dingemanse, 2007) and personality psychology (Uher, 2008; Vazire, Gosling, Dickey, & Schapiro, 2007), and thus it is no surprise that primate personality research has grown rapidly over the last two decades (Freeman & Gosling, 2010). Behavioral coding, referring to the

measurement of whole organism's actions and reactions (Martin & Bateson, 1993), is the most frequently utilized method for assessing primate personality (Freeman & Gosling, 2010). The expression of behavior is particularly sensitive to fluctuations in internal and external stimuli, such that animals are capable of rapid behavioral change across contexts (Duckworth, 2009). Behavioral coding therefore provides a rigorous means by which to simultaneously quantify behavioral variation across multiple levels of phenotypic, environmental, and temporal organization (Dingemanse & Araya-Ajoy, 2015).

Unfortunately, however, current approaches to the statistical analysis of primate personality prevent researchers' from utilizing the full potential of their coding datasets. For example, primatologists conducting observational studies of personality rely upon extensive data aggregation (e.g., Neumann, Agil, Widdig, & Engelhardt, 2013; Uher, Asendorpf, & Call, 2008), which enhances the temporal reliability of coding measures (Epstein, 1979; Koski, 2011a; Uher, 2011) but also diminishes statistical power and leads to the loss of biologically relevant information on within-individual variance (Janson, 2012; Pollet, Stulp, Henzi, & Barrett, 2015). The temporal consistency of behavioral measures is typically assessed without first controlling for important sources of between-individual variance such as sex, age, social group composition, and dominance rank (e.g., Carter, Marshall, Heinsohn, & Cowlshaw, 2012; Johnson et al., 2015), which can inflate the estimated contribution of individual identity to observed behavioral variance (Dingemanse & Dochtermann, 2013; Nakagawa & Schielzeth, 2010). Models of primate personality structure are currently estimated without first partitioning within- and between-individual behavioral (co)variance (e.g., Anestis, 2005; Konečná et al., 2008), leading to biased estimates of the size and even the direction of behavioral trait correlations (Brommer, 2013; Dingemanse, Dochtermann, & Nakagawa, 2012). Principal component analysis also continues to be used regularly (e.g., Seyfarth, Silk, & Cheney, 2012; Iwanicki & Lehmann, 2015), although this technique is more appropriate for data reduction than modeling personality structure (Borsboom, 2006; Budaev, 2010; Haig, 2005; Widaman, 2007). Finally, researchers currently rely upon exploratory statistical techniques to compare personality models across studies and/or data collections methods (e.g., Eckardt et al., 2015; Freeman et al., 2013). This approach increases the risk of inferential error due to multiple testing (Bell, 2007; Dochtermann, 2010) and does not draw upon the benefits of information-theoretic (I-T) model selection for statistical inference (Dingemanse, Dochtermann, & Wright, 2010).

Here, we propose a statistical framework for primate personality assessment and model comparison that attempts to address these limitations. Our approach integrates recently developed mixed-effects modeling approaches for quantifying behavioral variation (Dingemanse & Dochtermann, 2013) with I-T model selection for confirmatory personality research (Dingemanse, Dochtermann et al., 2010). We first detail this multi-step analytic procedure, describing the value of mixed-effects modeling for personality research and how I-T model selection can be used for quantitative comparisons across populations, species, and data collection methods. We then employ our framework to evaluate distinct models of social personality structure in zoo-housed

bonobos (*Pan paniscus*) using behavioral data. We end by discussing the benefits and limitations of our framework, with an emphasis upon the role of confirmatory model comparison for integrating the behavioral ecological and comparative psychological approaches to primate personality research.

2 | DESCRIPTION

The proposed statistical framework for personality assessment and model comparison consists of a multi-step process involving: (1) the estimation of behavioral trait repeatability; (2) the partitioning of among-individual (co)variance in behavior; and (3) the quantitative comparison of hypothesized structural models from prior personality research. We describe each of these steps below with the goal of providing a brief conceptual introduction to the use of mixed-effects modeling and I-T model selection procedures for personality research. References to more thorough and technical treatments of these topics are provided throughout the text. To aid in the implementation of our statistical framework, short tutorials describing our empirical demonstration with accompanying R code are also provided in the electronic supplementary material (ESM).

Due to terminological differences between psychology and biology (Koski, 2011a; Uher, 2011), it is important to explicitly define the theoretical terms employed below. We often refer to any measurable aspect of behavior as a behavioral "trait" (Carter, Feeney, Marshall, Cowlshaw, & Heinsohn, 2013), and we use the term "personality" to describe consistent individual differences in behavior (Réale et al., 2007). We therefore state that a behavioral trait "exhibits personality" or is a "personality trait" if among-individual variance is detected across temporally dispersed measurements of the relevant behavior (Mathot & Dingemanse, 2015). Finally, we use the terms "personality structure" and "structural model" to broadly encompass any model of personality trait correlations (Dingemanse, Dochtermann et al., 2010), which will typically refer to latent factor or component models (Budaev, 2010). Structural models therefore include both the behavioral syndromes derived from coding methods (Sih, Bell, & Johnson, 2004) as well as the more holistic representations of individual differences in cognition, affect, and behavior captured by rating methods (Weiss & Adams, 2013).

2.1 | Repeatability of behavioral traits

Determining the relative contribution of personality to observed behavioral variance requires partitioning among- and within-individual variance across repeated trait measurements (Nakagawa & Schielzeth, 2010). In the absence of individual variation in behavioral plasticity (Brommer, 2013; Dingemanse & Dochtermann, 2013), among-individual variance will reflect genetic and permanent environmental effects on individuals' average behavioral trait values, while within-individual variance will encapsulate both measurement error and average behavioral plasticity in response to temporary environmental changes (Falconer, 1960; Rosa, 2015). Partitioning these distinct variance components can be readily accomplished with mixed-effects

modeling (Dingemanse & Dochtermann, 2013). Mixed-effects models (MMs) provide many benefits for personality research, such as direct estimation of variance components for Gaussian and non-Gaussian data (Bolker et al., 2009; Janson, 2012; Nakagawa & Schielzeth, 2010), robust handling of unbalanced sampling (Dingemanse & Dochtermann, 2013; van de Pol & Wright, 2009), and multivariate analysis of behavioral trait covariance across multiple levels of phenotypic structure (Dingemanse and Araya-Ajoy, 2015; Dingemanse & Dochtermann, 2013; Nussey, Wilson, & Brommer, 2007). MMs also readily facilitate investigations of behavioral reaction norms, reflecting patterns of individual-environment interaction (Araya-Ajoy, Mathot, & Dingemanse, 2015; Dingemanse, Kazem, Réale, & Wright, 2010; Jaeggi, Boose, White, & Gurven, 2016; Nussey et al., 2007), and behavioral characters, which refer to the functionally coherent proximate mechanisms underlying behavioral variation (Araya-Ajoy & Dingemanse, 2014). Although MMs provide a powerful tool for the statistical analysis of behavioral data, the flexibility and potential complexity of these models can lead to easy misspecification and subsequent inferential error (Bennington & Thayne, 1994; Bolker et al., 2009; van de Pol & Wright, 2009). Researchers should therefore ensure sufficient familiarity with the assumptions and interpretation of MMs prior to data analysis (see Gelman & Hill, 2007; McElreath, 2015; West, Welch, & Gafekci, 2007; Zuur, Ieno, Walker, Saveliev, & Smith, 2009; for introductory book-length treatments).

The degree of personality exhibited by a behavioral trait is often quantified using the common measure of repeatability, which represents the proportion of total behavioral variance accounted for by among-individual variance (Nakagawa & Schielzeth, 2010). In contrast to personality research utilizing standardized behavioral assays (e.g., Massen, Antonides, Arnold, Bionda, & Koski, 2013; Šlipogor et al., 2016), observational studies in less controlled contexts are particularly prone to low repeatability estimates (Vazire et al., 2007; Weinstein, Capitanio, & Gosling, 2008). Historically, concerns over the low repeatability of observational coding measures have motivated the use of data aggregation techniques to produce more reliable and generalizable measures (Epstein, 1979; Koski, 2011a; Uher, 2011), but the statistical analysis of mean scores leads to a loss of statistical power and biologically relevant information on within-individual variance (Janson, 2012; Pollet et al., 2015). Fortunately, Araya-Ajoy et al. (2015) have recently proposed an approach to repeatability estimation that assesses individual consistency in average behavior without the use of data aggregation. By utilizing MMs, among-individual and among-series variance can be partitioned. "Series" here refers to temporally separated periods within which multiple measurements were made on the same individual, such as observations made over a series of weeks (Uher et al., 2008), months (Neumann et al., 2013), or years (Koski, 2011b). The repeatability of individuals' average trait values can then be estimated as the proportion of behavioral variance across individuals and series accounted for by among-individual variance. This procedure can also be extended to assess the repeatability of individual differences in behavioral plasticity across series, providing an integrative approach to estimating reaction norm repeatability (Araya-Ajoy et al., 2015).

An additional benefit of MMs for personality research is the capacity to estimate and control for the effects of factors such as population membership, sex, age class, and dominance rank prior to repeatability estimation. Although these factors are expected to in part cause and/or be caused by personality (Sih et al., 2015; Weinstein et al., 2008), it is important to assess the degree of novel information personality assessment provides in addition to previously well-established findings on primate behavior and social organization (Seyfarth et al., 2012). Modeling these sources of behavioral variance facilitates the calculation of adjusted repeatabilities reflecting the contribution of among-individual variance that remains after controlling for these effects (Nakagawa & Schielzeth, 2010). MMs can also be used to control for bias in repeatability estimation resulting from other important sources of observed behavioral variance, such as unbalanced sampling (Dingemanse & Dochtermann, 2013; van de Pol & Wright, 2009), time-related change (Biro & Stamps, 2015), and measurement error (Nakagawa & Schielzeth, 2010).

2.2 | Partitioning phenotypic (co)variance

Given that the total variance of a behavioral trait reflects both among- and within-individual variance, observed behavioral covariance can also be decomposed into patterns of among- and within-individual covariance (Dingemanse et al., 2012; Searle, 1961). Previous meta-analyses of behavioral trait repeatability have found that among-individual variance on average accounts for less than 40% of observed behavioral variance (Bell, Hankison, & Laskowski, 2009; Garamszegi, Markó, & Herczeg, 2013), suggesting that within-individual covariance will tend to have a larger effect on the magnitude and direction of behavioral trait covariance patterns (Dingemanse & Dochtermann, 2013). Therefore, failing to partition these distinct components prior to the analysis of personality structure leads to biased estimates and potentially erroneous inferences regarding trait correlations across individuals (Brommer, 2013; Dingemanse et al., 2012). To avoid this issue, multivariate MMs can be utilized to directly partition and analyze among-individual covariance (Dingemanse & Dochtermann, 2013). Multivariate MMs can also be employed to assess the temporal and cross-contextual structure of behavioral plasticity (e.g., Araya-Ajoy & Dingemanse, 2014; Husby et al., 2010), although much larger sample sizes are generally required to adequately detect and precisely estimate within-individual (co)variance and individual-environment interactions (Martin, Nussey, Wilson, & Reale, 2011; van de Pol, 2012). Monte Carlo Markov Chain (MCMC) sampling is particularly useful for robust estimation of these multivariate parameters for non-Gaussian data (Hadfield, 2010; Zhao, Staudenmayer, Coull, & Wand, 2006).

2.3 | Quantitative comparison of structural models

After determining which behavioral traits exhibit personality, researchers typically utilize principal component analysis (PCA) or factor analysis (FA) to generate structural models of these behaviors. In contrast to principal component models, which seek to maximize the total variance accounted for by each component, factor models explicitly partition sources of trait covariance ("common factors") from

the remaining residual variance ("unique factors") of each behavior (Budaev, 2010; Edwards & Bagozzi, 2000; Widaman, 2007). The importance of distinguishing these factors in behavioral research is underscored by Garamszegi et al.'s (2013) finding that correlations between the most commonly studied animal personality traits tend to be of low to moderate effect size. This suggests that PCA and FA solutions are more likely to diverge in non-trivial ways when attempting to model behavioral trait covariance. Moreover, principal component model parameters are less likely to generalize across distinct measures (ethograms, behavioral assays, rating scales) of the same personality structure because of their sensitivity to unique trait variance (Widaman, 2007). Although PCA is particularly useful for data reduction, FA better facilitates investigations of the underlying causes of behavioral trait covariance (Borsboom, 2006; Budaev, 2010; Haig, 2005; Widaman, 2007). We therefore recommend that primatologists utilize FA rather than PCA for personality research.

Factor models can be statistically identified through analytic rotation techniques or a priori model specification, which are commonly referred to as modes of exploratory factor analysis (EFA) and confirmatory factor analysis (CFA), respectively (Browne, 2001; Jennrich, 2007). EFA is used to derive a posteriori hypotheses of the causal structure underlying observed trait covariance (Bollen, 2002; Haig, 2005), making EFA a highly valuable tool for early descriptive research (e.g., Koski, 2011b; Neumann et al., 2013; Staes et al., 2016). In contrast, CFA facilitates hypothesis-driven research by estimating how well prespecified structural models reproduce observed datasets (Grace, 2006; Jöreskog, 1967). This approach is particularly useful for the validation of previously hypothesized models as well as direct comparisons between competing hypotheses of personality structure (e.g., Araya-Ajoy & Dingemanse, 2014; Bókony, Kulcsár, Tóth, & Liker, 2012; Capitanio & Widaman, 2005; Dingemanse, Dochtermann et al., 2010; White, Meekan, McCormick, & Ferrari, 2013).

Evaluating the explanatory success of competing hypotheses is crucial for the process of theory building in scientific research (Elliott & Brook, 2007; Fox & Westneat, 2010; Krebs, 2000). Recently, I-T model selection techniques conducive to such hypothesis-driven research have gained prominence in evolutionary ecology (Hobbs & Hilborn, 2006; Johnson & Omland, 2004; Stephens, Buskirk, & del Rio, 2007). In contrast to the common use of implausible null hypotheses for statistical testing and inference (Fraley & Marks, 2007; Nakagawa & Cuthill, 2007), the I-T model selection paradigm encourages researchers to articulate and quantitatively compare multiple candidate models derived from prior scientific investigation (Dochtermann & Jenkins, 2011a). Information criteria such as Akaike's information criterion (AIC) are used to rank the empirical support and likelihood of each model relative to the model set (Burnham & Anderson, 2002; Hobbs & Hilborn, 2006; Symonds & Moussalli, 2011). Although, a single model may be selected because of its superior relative fit, the I-T approach also facilitates multi-model inferences when multiple hypotheses receive a similar degree of support and cannot be selected between with sufficient certainty (Burnham, Anderson, & Huyvaert, 2011; Grueber, Nakagawa, Laws, & Jamieson, 2011). As further described below, I-T model selection procedures can be utilized in combination

with CFA to facilitate rigorous model comparison and selection for personality research (Dingemanse, Dochtermann et al., 2010).

Previous research has typically relied upon a combination of EFA or PCA and pairwise correlations for model comparisons across data collection methods, populations, and/or taxa (e.g., Freeman et al., 2013; Garai, Weiss, Arnaud, & Furuichi, 2016; Morton et al., 2013). This analytic approach is limited in comparison to CFA because it increases the risk of statistical error due to multiple testing (Bell, 2007; Dochtermann, 2010) and does not facilitate structural model comparisons within the same dataset (Dingemanse, Dochtermann et al., 2010). CFA directly facilitates I-T model selection for personality structure within and across datasets, and therefore provides a valuable tool for comparative studies and model validation research. CFA can also be further embedded in larger causal hypotheses through structural equation modeling (SEM). SEM is a powerful framework capable of implementing a variety of statistical techniques for the evaluation and comparison of causal models (Grace, 2006; Grace et al., 2012; Shipley, 2016). Although, traditional rules-of-thumb suggest that sample sizes for SEM should be appreciably larger than those commonly used in primate research (Barrett, 2007), these recommendations may be overly conservative for simpler factor models (Dochtermann & Jenkins, 2011b; Wolf, Harrington, Clark, & Miller, 2013). Moreover, I-T fit measures such as AIC have been found to effectively rank multivariate behavioral models with small sample sizes (Dingemanse, Dochtermann et al., 2010; Dochtermann & Jenkins, 2011b).

3 | EMPIRICAL EXAMPLE

3.1 | General overview

We now provide an empirical demonstration of our statistical framework using behavioral data collected on zoo-housed bonobos. In contrast to the large literature on chimpanzee (*Pan troglodytes*) personality (e.g., Freeman et al., 2013; Koski, 2011b; Massen et al., 2013; Weiss et al., 2009), only a small number of exploratory studies have been conducted on bonobo personality. Investigations of chimpanzee personality have helped to establish the often multi-dimensional structure of primate social personality traits, a topic which has thus far received less attention in the non-primate literature (Koski, 2014). Bonobos also exhibit a highly diverse repertoire of social behavior, however, and although these sister species share many basic behavioral similarities (e.g., Bullinger, Burkart, Melis, & Tomasello, 2013; Cronin, De Groot, & Stevens, 2015; Jaeggi, Stevens, & Van Schaik, 2010; Stanford, 1998), important differences in their cognition and behavior (e.g., Heilbronner, Rosati, Stevens, Hare, & Hauser, 2008; Herrmann, Hare, Call, & Tomasello, 2010; Palagi, 2006; Wobber, Wrangham, & Hare, 2010) suggest that bonobos may exhibit a unique social personality structure. Addressing both the conserved and derived features of these species' personality is crucial for furthering our currently limited understanding of the phylogeny of human personality traits (Weiss, Adams, & Widdig, 2011) as well as the

evolutionary and ecological mechanisms that produce cognitive, affective, and behavioral trait correlations in primate populations (Sih et al., 2015; Wolf and Weissing, 2010).

Early research on bonobo personality by Uher et al. (2008) and Murray (2011) demonstrated stable individual differences in captive bonobos' behavior, but due to small sample sizes did not address broader personality dimensions. Recent exploratory studies conducted with larger samples using both rating and coding methods (Garai et al., 2016; Staes et al., 2016; Staes, Eens, Weiss, & Stevens, in press; Weiss et al., 2015) have produced multiple distinct hypotheses of bonobo personality structure. Weiss et al. (2015) found that ratings of the traits playful, curious, and active reflect an Openness personality factor, while ratings of sociable and affectionate reflect a separate Agreeableness factor. Ratings of solitary also loaded onto a distinct Extraversion factor (Weiss et al., 2015). In contrast, Garai et al.'s (2016) regularized analysis of wild bonobo ratings produced a factor encompassing traits related to sociability, activity, play, and affection, which bears appreciable similarity to the Extraversion dimension uncovered in chimpanzee rating research (King, Weiss, & Farmer, 2005). Staes et al.'s (2016) behavioral data suggest that activity, play, and allogrooming are each aspects of distinct personality dimensions. Similar personality factors were uncovered in captive chimpanzee research using behavioral coding (Koski, 2011b). Staes et al. (in press) also found that allogrooming and the number of neighbors during focal observation predicted Conscientiousness ratings, reflecting aspects of friendliness toward conspecifics (Weiss et al., 2015), while activity, approach, and play behavior strongly correlated with ratings of both Openness and Conscientiousness. Given current uncertainty about the structure of bonobo personality, we applied our statistical framework to independently assess multiple structural models of bonobo social personality variation.

3.2 | Methods

This research was approved by the Miami University Institutional Animal Care Committee and adhered to the guidelines set forth in the American Society of Primatologists Principles for the Ethical Treatment of Non Human Primates. Our study was reviewed and approved by the Cincinnati Zoo and Botanical Garden Jungle Trails and Columbus Zoo and Aquarium Congo Expedition departments. These zoos are accredited by the Association of Zoos and Aquariums and are members of the Bonobo Species Survival Plan.

3.2.1 | Subjects and housing

We conducted behavioral observations on two colonies of bonobos housed at the Cincinnati Zoo and Botanical Garden (CZBG) and Columbus Zoo and Aquarium (CZA). We used Kano's (1992) age definitions (adult: ≥ 15 ; adolescent: 7–14; juvenile: 2–6; infant: 0–1) for the following classifications. At the time of observation, CZBG housed 12 bonobos (three adult males, three adult females, one adolescent male, one juvenile male, two juvenile females, two infant females) and CZA housed 16 bonobos (four adult males, four adult females, two adolescent males, one adolescent female, two male juveniles, two juvenile females, one infant female). One bonobo from CZBG and four

bonobos from CZA were wild-born. Data were not collected on infants or dependent juveniles in these colonies (age < 3 years), resulting in a total sample size of 24 individuals across CZBG ($N = 10$) and CZA ($N = 14$). All analyses were conducted on data collected in the CZBG and CZA on-exhibit, indoor enclosures to facilitate observation during the winter season as well as to better standardize environmental conditions across zoos.

3.2.2 | Data collection

A total of 191 hr of data were collected at CZBG during fall 2014 (Period 1: 9/19/14–12/17/14; 106 hr) and spring 2015 (Period 2: 2/28/15–6/1/15; 85 hr), and 195 hr of data were collected at the Columbus Zoo during early (Period 1: 6/5/15–7/9/15; 100 hr) and late (Period 2: 7/28/15–8/20/15; 95 hr) summer 2015. We conducted 1 hr focal animal observations using a stratified random sampling procedure balancing total hours observed with the hour of the day during observation (CZBG: 10:00–17:00; CZA: 09:00–17:00). This resulted in 386 hr of observational data, with an average of 16 hr of observation per individual (range 10–23 hr). Previous simulation studies (Dingemanse & Dochtermann, 2013; Martin et al., 2011; van de Pol, 2012) suggest that this sampling design provides appropriate power and accuracy for detecting moderate among-individual (co) variance.

The ethogram for this study was composed of both state and event social behaviors, including various forms of social play (Palagi & Paoli, 2007) and affiliation (Palagi & Norscia, 2013) exhibited by captive bonobos, as well as allogrooming, interactions with enrichment objects, and indicators of activity level (see Table 1). We analyzed specific play, activity, and affiliative behaviors in larger functional coding categories to better address the broad behavioral domains of interest. A combination of instantaneous and one-zero sampling with 1 min intervals was used to code behavior. Given historic concerns about the utility of one-zero sampling in comparison to other sampling regimes (Altmann, 1974), we simultaneously coded continuous and one-zero interval frequencies for events and bouts during the 195 hr of behavioral data collected at CZA. Using Fisher's r -to- Z transformation to average across Pearson correlations, continuous and interval frequencies were found to strongly correlate with one another (mean $r = 0.97$), showing that our one-zero data effectively quantified the amount of behavior exhibited (Martin & Bateson, 1993). A total of 3 hr of observation were independently coded by two trained research assistants to determine the inter-observer reliability of our measurements. Interobserver reliability was estimated by two-way mixed, absolute agreement, single-measure intraclass correlation coefficients. Using Fisher's r -to- Z transformation, mean inter-rater reliability across observations (mean ICC = 0.91) indicated highly reliable measurement.

3.2.3 | Data analysis

Potential sources of variance across focal observations were assessed for each behavioral variable using univariate generalized linear mixed models (GLMMs) with Markov Chain Monte Carlo sampling in the "MCMCglmm" package (Hadfield, 2010) for R statistical environment.

TABLE 1 Bonobo ethogram

Behavior (code)	Description	Method
Activity (ACT)	Individual stands, walks, runs, forages, or climbs	INST
Allogrooming (GRM)	Individual manually or orally cleans the hair of another	INST
Resting in proximity (RSP)	Individual sits or lays within ~2 m of others	INST
Resting alone (RSA)	Individual sits or lays more than ~2 m away from others	INST
Affinitive behavior (AFF)	Individual touches, embraces, sits in contact, kisses, or engages in socio-sexual behavior with another	1-0
Social play (PLY)	Individual grabs, pushes, gently bites, slaps, tickles, chases, pirouettes onto, stamps, holds down, brusque rushes, or dangles, climbs, and jumps from environmental supports with another	1-0
Enrichment object interaction (OBJ)	Individual hits, holds, throws, or manually inspects an enrichment object	1-0

INST, instantaneous sampling; 1-0, one-zero sampling.

All models included sex, age, zoo population (CZBG or CZA), time of day, and observational period nested within zoo as fixed effects. Although, there was little among-subject variance in the average time of day during focal observation (0.08 hr), we centered time of day within subjects to provide a less biased estimate of possible within-individual effects (van de Pol & Wright, 2009). We also grand mean-centered age. The continuous fixed effects of age ($SD = 11.27$ years) and time of day ($SD = 1.95$ hr) were then divided by two standard deviations to facilitate more direct comparison with the binary effects of sex and zoo population (Gelman, 2008). Random effects included individual identity, social group composition during focal observations, and observational period as our time series unit. Unfortunately, we could not control for dominance rank because counts of aggression and fleeing upon aggression, a common indicator of linear rank in bonobos (Vervaecke, de Vries, & van Elsacker, 2000), were rare or absent during observational periods.

All behaviors were modeled with Poisson errors and inverse-Wishart priors. A total of 1,000 MCMC samples exhibiting acceptable autocorrelation and model convergence were extracted to estimate parameters for each univariate model (see Text S1 in the ESM for further details). An effect was considered statistically "significant" if the estimated 95% Bayesian credibility interval excluded zero. Average phenotype repeatability was calculated for individual consistency in behavior across observational periods (Araya-Ajoy et al., 2015). We then fit a multivariate GLMM for all behavioral traits with individual identity as a random effect. A parameter expanded prior was utilized to enhance MCMC mixing properties and weakly inform (co)variance parameter estimates (Gelman, 2006). Each of the 1,000 among-individual (co)variance parameter MCMC samples were then extracted from this model (see ESM Text S2). This facilitated the estimation of Bayesian credibility intervals for all confirmatory factor model parameters as uncertainty in the multivariate GLMM was brought forward into subsequent analyses (Araya-Ajoy and Dingemanse, 2014; Hadfield, Wilson, Garant, Sheldon, & Kruuk, 2010).

We developed seven distinct personality models based upon prior bonobo rating and coding research (see Figure 1) and fit them as structural equation models in the "lavaan" package in R (Rosseel, 2012). Both orthogonal and oblique (correlated) factor structures were

examined for each model. The models were quantitatively compared to one another as well as a null model of no trait covariance using differences in the bias-corrected AIC ($\Delta AICc$), which converges upon the uncorrected AIC in large samples and should generally be preferred to avoid overfitting models (Burnham & Anderson, 2004). $\Delta AICc \leq 2$ indicate substantial support relative to the best fitting model, and $\Delta AICc \geq 10$ suggest little to no support (Burnham & Anderson, 2002). Further details on model development and comparison are provided in the ESM Text S3 and S4, respectively.

4 | RESULTS

Significant differences in behavior due to sex, age, zoo, and time of day were detected across traits (see Table 2). Females engaged in significantly higher frequencies of allogrooming and affinitive behavior, while males engaged in significantly higher frequencies of social play and resting alone. Younger individuals exhibited higher rates of activity affinitive behavior, social play, enrichment and object interaction. Bonobos at CZBG were on average more active than the CZA bonobos. Higher rates of enrichment object interaction and lower rates of resting in proximity occurred during the spring observational period at CZA. Resting in proximity and affinitive behavior tended to decrease across the day, whereas rates of allogrooming significantly increased. All behaviors were found to exhibit moderate to high average phenotype repeatability (mean $R = 0.54$), providing considerable evidence for personality across the measured traits (see Table 2). Social group composition accounted for less variance than individual identity for all behaviors except play and resting in proximity.

The oblique Model 7 personality structure was found to exhibit the greatest strength of evidence relative to the model set (see Figure 1). All other models received little to no support relative to the best fitting model ($\Delta AICc > 10$). Statistically significant factor loadings were found for all behavioral traits in Model 7, and all traits correlated in the direction predicted by the original rating-derived model on which this hypothesis was based (Weiss et al., 2015). Although the Model 7 oblique factor structure was more strongly supported than the orthogonal structure (mean $\Delta AICc = 263.69$), the degree of correlation

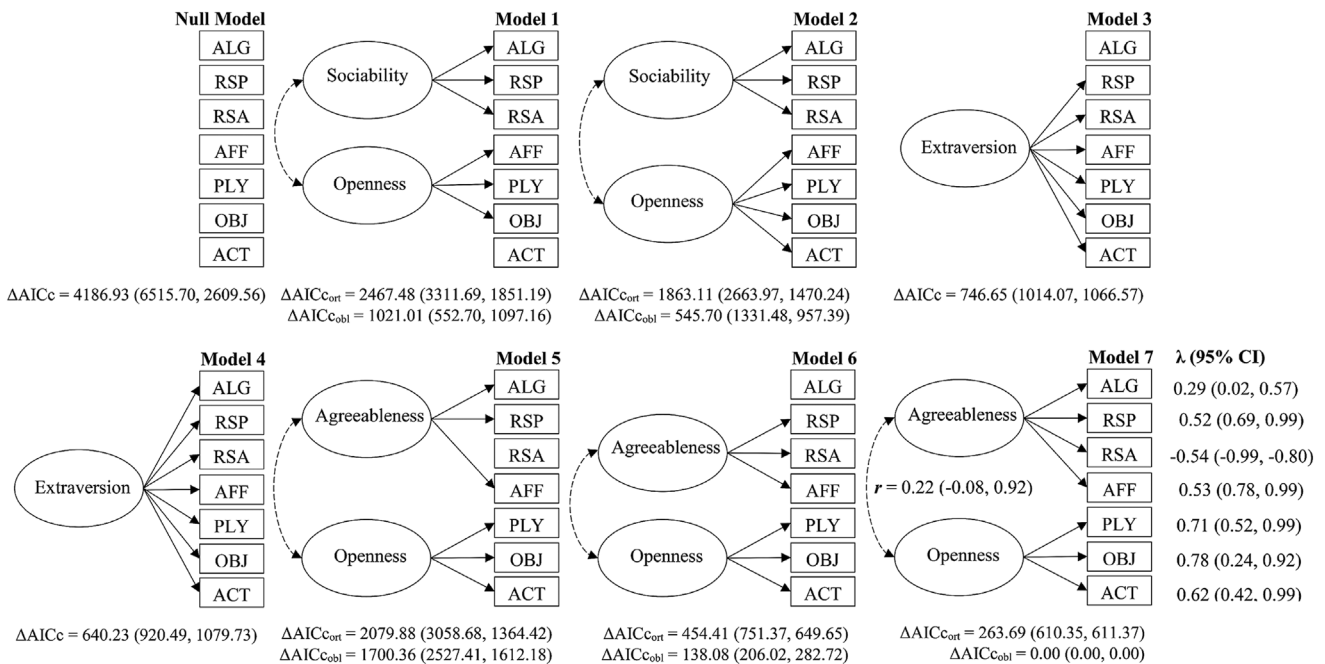


FIGURE 1 Hypothesized models of bonobo personality. Within each structural equation model, circles represent latent variables and squares represent observed behavioral variables. Dashed double-headed arrows indicate that both orthogonal and oblique factor structures were examined. For the best supported model, standardized factor loadings (λ) and the inter-factor correlation (r) are presented adjacent to the trait labels and dashed double-headed arrow respectively. Highest posterior density credibility intervals are presented in parentheses. AICc, bias-corrected Akaike's information criterion; ΔAIC_c , the difference between the model AICc and the minimum model AICc for both orthogonal ($\Delta AIC_{c_{ort}}$) and oblique ($\Delta AIC_{c_{obl}}$) solutions; G, allogrooming; RSP, resting in proximity; RSA, resting alone; AFF, affiliative behavior; PLY, social play; OBJ, enrichment object interaction; ACT, activity

between the Agreeableness and Openness factors could not be precisely determined due to uncertainty in the posterior distribution (mean $r = 0.22$; 95%CI [-0.08, 0.92]).

5 | DISCUSSION

In this paper, we proposed an integrative statistical framework for primate personality assessment and model comparison with behavioral coding data. Building upon recent developments in the mixed-effects modeling of behavioral variation (Dingemanse & Dochtermann, 2013) and I-T model selection for confirmatory personality research (Dingemanse, Dochtermann et al., 2010), this approach not only facilitates greater integration of primate personality models across taxa and data collection methods, but also identifies and provides solutions for current limitations in the statistical analysis of primate personality data. These limitations include the common practice of aggregating data rather than using mixed-effects modeling techniques (Janson, 2012; Pollet et al., 2015), not modeling biologically relevant sources of individual variance during repeatability estimation (Dingemanse & Dochtermann, 2013), not partitioning between-individual (co)variance prior to modeling personality structure (Brommer, 2013; Dingemanse et al., 2012), the continued misuse of PCA (Borsboom, 2006; Budaev, 2010; Widaman, 2007), and an over-reliance upon exploratory statistical techniques for model comparisons (Dingemanse, Dochtermann et al., 2010). Applying

our multi-step framework to observational data on zoo-housed bonobos revealed that differences between sexes, ages, zoos, time of observation, and social group composition contributed to significant behavioral variance across most traits. Independently of these effects, however, personality nonetheless accounted for a moderate to high proportion of variance in average behavior across observational periods (see Table 2).

Through the application of I-T model selection for CFA, we found that a personality model derived from prior rating data (Weiss et al., 2015) received the strongest support relative to our model set. This model suggests that personality variation in the measured behavioral traits (see Table 1) is best described by two correlated but distinct dimensions reflecting individual differences in affiliation and sociability (Agreeableness) as well as activity level, social play, and neophilia toward non-threatening stimuli (Openness). The greater support found for this model in comparison to Model 2 and Model 3 suggests that dispositional allogrooming and affiliative behavior may be more strongly associated in bonobos than chimpanzees (Koski, 2011b). The superior fit of such a two-factor model relative to a general factor model (Model 4) also underscores the often multidimensional structure of personality variation in primate social behavior (Koski, 2014). Moreover, as noted by Weiss et al. (2015), the loading of activity level and social play onto a common factor distinct from aspects of sociability and affiliation suggests that bonobo Openness more closely resembles Openness in mountain gorillas (*Gorilla beringei beringei*; Eckardt et al., 2015) and tufted capuchins (*Sapajus apella*;

TABLE 2 Univariate behavioral trait models.

Fixed effects	ACT		ALG		RSP		RSA		AFF		PLY		OBJ	
	β (95%CI)		β (95%CI)		β (95%CI)		β (95%CI)		β (95%CI)		β (95%CI)		β (95%CI)	
Intercept	2.18 (1.72, 2.70)*		-2.61 (-4.20, -1.19) ^a		2.63 (2.10, 3.13) ^a		2.82 (2.33, 3.21) ^a		0.59 (-0.19, 1.32)		-0.39 (-1.93, 1.06)		0.16 (-0.79, 1.03)	
Sex	0.01 (-0.41, 0.46)		1.67 (0.47, 2.96) ^a		0.42 (-0.04, 0.85)		-0.37 (-0.76, -0.02) ^a		1.10 (0.34, 1.80) ^a		-1.79 (-2.98, -0.39) ^a		-0.33 (-1.08, 0.41)	
Age	-0.66 (-1.08, -0.24) ^a		0.45 (-1.11, 1.63)		-0.15 (-0.57, 0.27)		0.33 (-0.03, 0.71)		-0.81 (-1.51, -0.10) ^a		-2.24 (-3.38, -0.94) ^a		-1.11 (-1.89, -0.36) ^a	
Zoo	-0.72 (-1.25, -0.12) ^a		0.94 (-0.83, 2.56)		-0.30 (-0.91, 0.31)		0.31 (-0.13, 0.72)		-0.48 (-1.35, 0.51)		-0.91 (-2.69, 0.94)		-0.17 (-1.24, 0.80)	
CZBG:period	0.09 (-0.28, 0.52)		0.41 (-1.26, 1.93)		-0.36 (-0.81, 0.08)		0.19 (-0.22, 0.64)		0.27 (-0.47, 0.97)		-0.26 (-1.46, 0.90)		-0.07 (-0.86, 0.78)	
CZA:period	0.08 (-0.30, 0.47)		0.24 (-1.00, 1.75)		-0.49 (-0.89, -0.02) ^a		0.18 (-0.21, 0.54)		-0.55 (-1.22, 0.11)		0.19 (-1.19, 1.79)		1.16 (0.48, 1.94) ^a	
Time of day	0.07 (-0.09, 0.24)		0.67 (0.03, 1.40) ^a		-0.39 (-0.53, -0.22) ^a		0.06 (-0.07, 0.24)		-0.44 (-0.71, -0.19) ^a		-0.06 (-0.62, 0.45)		-0.24 (-0.54, 0.09)	
Random effects	σ^2 (95%CI)		σ^2 (95%CI)		σ^2 (95%CI)		σ^2 (95%CI)		σ^2 (95%CI)		σ^2 (95%CI)		σ^2 (95%CI)	
Individual	0.19 (0.07, 0.39)		0.83 (0.04, 2.14)		0.18 (0.05, 0.34)		0.11 (0.03, 0.22)		0.49 (0.06, 1.03)		1.17 (0.09, 2.65)		0.44 (0.05, 0.96)	
Social group	0.08 (0.03, 0.15)		0.44 (0.05, 1.15)		0.20 (0.07, 0.37)		0.08 (0.03, 0.15)		0.26 (0.06, 0.55)		1.99 (0.37, 3.84)		0.33 (0.07, 0.68)	
Series	0.10 (0.03, 0.19)		1.47 (0.10, 3.14)		0.10 (0.04, 0.19)		0.11 (0.04, 0.20)		0.40 (0.08, 0.78)		0.67 (0.06, 1.81)		0.41 (0.07, 0.83)	
Residual	0.50 (0.39, 0.62)		7.40 (5.18, 9.62)		0.46 (0.37, 0.56)				1.10 (0.82, 1.37)		3.89 (2.66, 5.25)		1.67 (1.26, 2.19)	
Repeatability	R (95%CI)		R (95%CI)		R (95%CI)		R (95%CI)		R (95%CI)		R (95%CI)		R (95%CI)	
Average phenotype	0.63 (0.35, 0.87)		0.37 (0.03, 0.78)		0.61 (0.31, 0.85)		0.49 (0.22, 0.73)		0.54 (0.20, 0.89)		0.63 (0.17, 0.97)		0.51 (0.18, 0.86)	

Age and time of day were mean-centered and scaled to two standard deviations prior to analysis. Reference categories for categorical fixed effects were set to male (sex), CZBG (zoo), and the first observational period (Period). Colons indicate a nested fixed effect. Posterior mean values are reported for all parameter estimates. ACT, activity; ALG, allogrooming; RSP, resting in proximity; RSA, resting alone; AFF, affiliative behavior; PLY, social play; OBJ, enrichment object interaction.

^aStatistically significant fixed effects (95% Bayesian credibility interval excludes 0).

Morton et al., 2013) than in chimpanzees (Weiss et al., 2009). Although the rating-derived Openness factor (Weiss et al., 2015) is positively correlated with behavioral responses to both threatening and non-threatening novel stimuli (Staes et al., in press), Staes et al. (2016) also found that these measures reflect distinct coding-derived personality factors of Boldness and Openness, respectively. Similar results were found in behavioral research on chimpanzee personality structure (Massen et al., 2013), suggesting that Openness and Boldness are distinct dimensions in both species. Given that we only measured behavioral interactions with enrichment objects, the Openness factor supported by our dataset does not adjudicate between these hypotheses. Nonetheless, such findings highlight the potential evolution of distinct, contextualized exploratory and risk-taking tendencies (Carter et al., 2013; Réale et al., 2007), suggesting that future studies should further investigate the domain-specificity of Openness in bonobos.

Synthesizing the results of past research for model comparison and theoretical integration can be an appreciably complex task (Dochtermann & Jenkins, 2011a; Uher, 2008). Nonetheless, rigorous validation of a population or species' personality structure is critical for understanding not only the phylogeny (Gosling & Graybeal, 2007; Weiss et al., 2011) and function (Wolf & Weissing, 2010) of personality traits, but also for enhancing the welfare of captive primates (e.g., Coleman, 2012; Izzo, Bashaw, & Campbell, 2011). Unfortunately, although research on primate personality has rapidly increased over the past two decades (Freeman & Gosling, 2010), the behavioral ecological and comparative psychological approaches to personality remain disconnected (Koski, 2011a; Weiss & Adams, 2013). This has hampered theory-building and resulted in method-specific models that are often not compared across studies.

Some researchers have suggested that either behavioral coding or rating methods provide intrinsically more reliable and valid measures (Uher, 2013; Vazire et al., 2007), or that both methods capture distinct aspects of personality (Carter, Marshall, Heinsohn, & Cowlshaw, 2011). While these claims may be true in particular contexts, understanding the proximate and ultimate bases of personality will require cohesive and integrative models that draw upon the insights provided by both approaches (Carter et al., 2013; Nettle & Penke, 2010). Although exploratory tools are essential for hypothesis generation and early descriptive research using both methods, we see the utilization of a confirmatory statistical framework as the most rigorous means by which this goal can be achieved. Nonetheless, researchers should remain aware that appropriate model comparisons may be hampered due to the unique limitations of rating and coding methods, such as the influence of differential item interpretation and sociocultural biases on ratings (Uher, Werner, & Gosselt, 2013; Uher & Visalberghi, 2016). By employing rating and coding measures in tandem, the current drawbacks of each method can be more clearly addressed (Freeman & Gosling, 2010; Koski, 2011a; Uher, 2008).

There are important limitations to our framework and empirical demonstration. Although CFA facilitates a rigorous and hypothesis-driven approach to personality measurement, the statistical assumptions of this model have been criticized for being overly restrictive (Asparouhov & Muthén, 2009). Recently proposed analytic techniques

such as exploratory structural equation modeling (Marsh, Morin, Parker, & Kaur, 2014) and generalized latent variable modeling (Epskamp, Rhemtulla, & Borsboom, in press) can overcome these restrictions and should be considered when traits are expected to exhibit high unique factor covariance and/or cross-loadings onto multiple common factors. The ethogram utilized in this study was restricted to a subset of social behaviors exhibited by zoo-housed bonobos. We did not utilize adjective ratings or behavioral assays to measure other pertinent traits, counts of agonistic behavior were too low to address the effects of dominance rank, and we also opted for a lumping strategy rather than analyzing behaviors separately. Although this simplified the complexity of our models, the results of these analyses were subsequently narrower in scope and should be treated as a subset of the broader bonobo personality structure. Finally, given the size of our dataset, we only fit an individual identity random effect to our multivariate GLMM and did not assess residual covariance or potential among-individual variance in behavioral plasticity. Future studies utilizing comprehensive ethograms and/or rating inventories with larger samples will therefore be crucial for addressing other sources of (co)variance in bonobo social behavior.

Despite these limitations, our results underscore the applicability of the proposed framework for personality assessment and model comparison with behavioral data. Using a combination of MMs and CFA to implement I-T model selection allows researchers to more rigorously analyze the validity of distinct personality models and to carry this knowledge forward into future studies. Indeed, one of the principal benefits of an I-T approach to model selection is its emphasis upon the continuous assessment and reevaluation of multiple working hypotheses (Burnham & Anderson, 2002; Elliott & Brook, 2007). For this reason, future bonobo personality research should continue to compare the fit of each of the models addressed in the present study, as well as others that may be devised through exploratory methods using novel rating or behavioral indicators. More generally, standardization of methods should be secondary to the primary goal of developing accurate personality models that transcend the particular means of measurement. As demonstrated in our empirical example, models derived across taxa from distinct coding or rating measures can be quantitatively compared. Further comparative research employing this statistical framework can therefore be used to build a more integrative science of animal personality (Weiss & Adams, 2013), as well as more generalizable theories of personality structure and evolution.

ACKNOWLEDGMENTS

We thank the Cincinnati Zoo and Botanical Garden and Columbus Zoo and Aquarium for allowing us to observe their bonobos. In particular, we extend our deepest thanks to zookeepers Stephanie Schuler and Jessica Linek for all of their assistance. We also thank Katlyn Mann and Tanner Csonka for their aid during data collection, as well as Dr. Yimen Aray-Ajoy for sharing R code with us. We are sincerely grateful to Jeff and Melanie Andrew, the parents of Rebecca Jeanne Andrew, for their kindness and generosity. This research would not have been possible without their support. This work was also supported by the Miami University Office for the Advancement of Research and

Scholarship, and adhered to the American Society of Primatologists principles for the ethical treatment of primates.

REFERENCES

- Altmann, J. (1974). Observational study of behavior: Sampling methods. *Behaviour*, 49, 227–266.
- Anestis, S. F. (2005). Behavioral style, dominance rank, and urinary cortisol in young chimpanzees (*Pan troglodytes*). *Behaviour*, 142, 1245–1268.
- Araya-Ajoy, Y. G., & Dingemanse, N. J. (2014). Characterizing behavioural 'characters': An evolutionary framework. *Proceedings of the Royal Society B*, 281, 20132645.
- Araya-Ajoy, Y. G., Mathot, K. J., & Dingemanse, N. J. (2015). An approach to estimate short-term, long-term and reaction norm repeatability. *Methods in Ecology and Evolution*, 6, 1462–1473.
- Asparouhov, T., & Muthén, B. (2009). Exploratory structural equation modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, 16, 397–438.
- Barrett, P. (2007). Structural equation modeling: Adjusting model fit. *Personality and Individual Differences*, 42, 815–824.
- Bell, A. M. (2007). Future directions in behavioural syndromes research. *Proceedings of the Royal Society of London B: Biological Sciences*, 274, 755–761.
- Bell, A. M., Hankison, S. J., & Laskowski, K. L. (2009). The repeatability of behaviour: A meta-analysis. *Animal Behaviour*, 77, 771–783.
- Bennington, C. C., & Thayne, W. V. (1994). Use and misuse of mixed model analysis of variance in ecological studies. *Ecology*, 75, 717–722.
- Biro, P. A., & Stamps, J. A. (2015). Using repeatability to study physiological and behavioural traits: Ignore time-related change at your peril. *Animal Behaviour*, 105, 223–230.
- Bókony, V., Kulcsár, A., Tóth, Z., & Líker, A. (2012). Personality traits and behavioral syndromes in differently urbanized populations of house sparrows (*Passer domesticus*). *PLOS ONE*, 7, e36639.
- Bolker, B. M., Brooks, M. E., Clark, C. J., Geange, S. W., Poulsen, J. R., Stevens, M. H. H., & White, J. S. S. (2009). Generalized linear mixed models: A practical guide for ecology and evolution. *Trends in Ecology & Evolution*, 24, 127–135.
- Bollen, K. A. (2002). Latent variables in psychology and the social sciences. *Annual Review of Psychology*, 53, 605–634.
- Borsboom, D. (2006). The attack of the psychometricians. *Psychometrika*, 71, 425–440.
- Brommer, J. E. (2013). On between-individual and residual (co) variances in the study of animal personality: Are you willing to take the "individual gambit"? *Behavioral Ecology and Sociobiology*, 67, 1027–1032.
- Browne, M. W. (2001). An overview of analytic rotation in exploratory factor analysis. *Multivariate Behavioral Research*, 36, 111–150.
- Budaev, S. V. (2010). Using principal components analysis and factor analysis in animal behavior research: Caveats and guidelines. *Ethology*, 116, 472–480.
- Bullinger, A. F., Burkart, J. M., Melis, A. P., & Tomasello, M. (2013). Bonobos, *Pan paniscus*, chimpanzees, *Pan troglodytes*, and marmosets, *Callithrix jacchus*, prefer to feed alone. *Animal Behaviour*, 85, 51–60.
- Burnham K. P., & Anderson D. R. (2002). *Model selection and multimodel inference: A practical information-theoretic approach* (p. 488). New-York: Springer-Verlag.
- Burnham, K. P., & Anderson, D. R. (2004). Multimodel inference understanding AIC and BIC in model selection. *Sociological Methods & Research*, 33, 261–304.
- Burnham, K. P., Anderson, D. R., & Huyvaert, K. P. (2011). AIC model selection and multimodel inference in behavioral ecology: Some background, observations, and comparisons. *Behavioral Ecology and Sociobiology*, 65, 23–35.
- Capitanio, J. P., & Widaman, K. F. (2005). Confirmatory factor analysis of personality structure in adult male rhesus monkeys (*Macaca mulatta*). *American Journal of Primatology*, 65, 289–294.
- Carere C., & Maestripieri D. (2013). *Animal personalities: Behavior, physiology, and evolution*. Chicago: University of Chicago Press.
- Carter, A. J., Feeney, W. E., Marshall, H. H., Cowlshaw, G., & Heinsohn, R. (2013). Animal personality: What are behavioural ecologists measuring? *Biological Reviews of the Cambridge Philosophical Society*, 88, 465–475.
- Carter, A. J., Marshall, H. H., Heinsohn, R., & Cowlshaw, G. (2011). Evaluating animal personalities: Do observer assessments and experimental tests measure the same thing? *Behavioral Ecology and Sociobiology*, 66, 153–160.
- Carter, A. J., Marshall, H. H., Heinsohn, R., & Cowlshaw, G. (2012). How not to measure boldness: Novel object and antipredator responses are not the same in wild baboons. *Animal Behaviour*, 84, 603–609.
- Coleman, K. (2012). Individual differences in temperament and behavioral management practices for nonhuman primates. *Applied Animal Behaviour Science*, 137, 106–113.
- Cronin, K. A., De Groot, E., & Stevens, J. M. (2015). Bonobos show limited social tolerance in a group setting: A comparison with chimpanzees and a test of the Relational Model. *Folia Primatologica*, 86, 164–177.
- Dingemanse, N. J., & Araya-Ajoy, Y. G. (2015). Interacting personalities: Behavioural ecology meets quantitative genetics. *Trends in Ecology & Evolution*, 30, 88–97.
- Dingemanse, N. J., & Dochtermann, N. A. (2013). Quantifying individual variation in behaviour: Mixed-effect modelling approaches. *Journal of Animal Ecology*, 82, 39–54.
- Dingemanse, N. J., Dochtermann, N., & Nakagawa, S. (2012). Defining behavioural syndromes and the role of 'syndrome deviation' in understanding their evolution. *Behavioral Ecology and Sociobiology*, 66, 1543–1548.
- Dingemanse, N. J., Dochtermann, N., & Wright, J. (2010). A method for exploring the structure of behavioural syndromes to allow formal comparison within and between data sets. *Animal Behaviour*, 79, 439–450.
- Dingemanse, N. J., Kazem, A. J., Réale, D., & Wright, J. (2010). Behavioural reaction norms: Animal personality meets individual plasticity. *Trends in Ecology & Evolution*, 25, 81–89.
- Dochtermann, N. A. (2010). Behavioral syndromes: Carryover effects, false discovery rates, and a priori hypotheses. *Behavioral Ecology*, 21, 437–439.
- Dochtermann, N. A., & Jenkins, S. H. (2011a). Developing multiple hypotheses in behavioral ecology. *Behavioral Ecology and Sociobiology*, 65, 37–45.
- Dochtermann, N. A., & Jenkins, S. H. (2011b). Multivariate methods and small sample sizes. *Ethology*, 117, 95–101.
- Duckworth, R. A. (2009). The role of behavior in evolution: A search for mechanism. *Evolutionary Ecology*, 23, 513–531.
- Eckardt, W., Steklis, H. D., Steklis, N. G., Fletcher, A. W., Stoinski, T. S., & Weiss, A. (2015). Personality dimensions and their behavioral correlates in wild Virunga mountain gorillas (*Gorilla beringei beringei*). *Journal of Comparative Psychology*, 129, 26–41.
- Edwards, J. R., & Bagozzi, R. P. (2000). On the nature and direction of relationships between constructs and measures. *Psychological Methods*, 5, 155–174.
- Elliott, L. P., & Brook, B. W. (2007). Revisiting Chamberlin: Multiple working hypotheses for the 21st century. *Bioscience*, 57, 608–614.

- Epskamp, S., Rhemtulla, M., & Borsboom, D. (in press). Generalized network psychometrics: Combining network and latent variable models. *Psychometrika*. Available online at: <https://arxiv.org/abs/1605.09288>
- Epstein, S. (1979). The stability of behavior: 1. On predicting most of the people much of the time. *Journal of Personality & Social Psychology*, 37, 1097–1126.
- Falconer D. S. (1960). *Introduction to quantitative genetics*. New York: The Ronald Press Company.
- Fox C. W., & Westneat D. F., (2010). Adaptation. In C. W. Fox, & D. F. Westneat, (Eds.), *Evolutionary behavioral ecology* (pp. 16–31). New York: Oxford University Press.
- Frabley R. C., & Marks M. J., (2007). The null hypothesis significance testing debate and its implications for personality research. In R. W. Robins, R. C. Frabley, & R. F. Krueger, (Eds.), *Handbook of research methods in personality psychology* (pp. 149–169). New York: Guilford.
- Freeman, H. D., Brosnan, S. F., Hopper, L. M., Lambeth, S. P., Schapiro, S. J., & Gosling, S. D. (2013). Developing a comprehensive and comparative questionnaire for measuring personality in chimpanzees using a simultaneous top-down/bottom-up design. *American Journal of Primatology*, 75, 1042–1053.
- Freeman, H. D., & Gosling, S. D. (2010). Personality in nonhuman primates: A review and evaluation of past research. *American Journal of Primatology*, 72, 653–671.
- Garai, C., Weiss, A., Arnaud, C., & Furuichi, T. (2016). Personality in wild bonobos (*Pan paniscus*). *American Journal of Primatology*, 78, 1178–1189.
- Garamszegi, L. Z., Markó, G., & Herczeg, G. (2013). A meta-analysis of correlated behaviors with implications for behavioral syndromes: Relationships between particular behavioral traits. *Behavioral Ecology*, 24, 1068–1080.
- Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis*, 1, 515–533.
- Gelman, A. (2008). Scaling regression inputs by dividing by two standard deviations. *Statistics in Medicine*, 27, 2865–2873.
- Gelman A., & Hill J. (2007). *Data analysis using regression and multilevel/hierarchical models*. Cambridge: Cambridge University Press.
- Gosling, S. D., & Graybeal, A. (2007). Tree thinking: A new paradigm for integrating comparative data in psychology. *Journal of General Psychology*, 134, 259–277.
- Grace J. B. (2006). *Structural equation modeling and natural systems*. Cambridge: Cambridge University Press.
- Grace, J. B., Schoolmaster, D. R., Guntenspergen, G. R., Little, A. M., Mitchell, B. R., Miller, K. M., & Schweiger, E. W. (2012). Guidelines for a graph-theoretic implementation of structural equation modeling. *Ecosphere*, 3, 1–44.
- Grueber, C. E., Nakagawa, S., Laws, R. J., & Jamieson, I. G. (2011). Multimodel inference in ecology and evolution: Challenges and solutions. *Journal of Evolutionary Biology*, 24, 699–711.
- Hadfield, J. D. (2010). MCMC methods for multi-response generalized linear mixed models: The MCMCglmm R package. *Journal of Statistical Software*, 33, 1–22.
- Hadfield, J. D., Wilson, A. J., Garant, D., Sheldon, B. C., & Kruuk, L. E. (2010). The misuse of BLUP in ecology and evolution. *The American Naturalist*, 175, 116–125.
- Haig, B. D. (2005). Exploratory factor analysis, theory generation, and scientific method. *Multivariate Behavioral Research*, 40, 303–329.
- Heilbronner, S. R., Rosati, A. G., Stevens, J. R., Hare, B., & Hauser, M. D. (2008). A fruit in the hand or two in the bush? Divergent risk preferences in chimpanzees and bonobos. *Biology Letters*, 4, 246–249.
- Herrmann, E., Hare, B., Call, J., & Tomasello, M. (2010). Differences in the cognitive skills of bonobos and chimpanzees. *PLoS ONE*, 5, e12438.
- Hobbs, N. T., & Hilborn, R. (2006). Alternatives to statistical hypothesis testing in ecology: A guide to self teaching. *Ecological Applications*, 16, 5–19.
- Husby, A., Nussey, D. H., Visser, M. E., Wilson, A. J., Sheldon, B. C., & Kruuk, L. E. (2010). Contrasting patterns of phenotypic plasticity in reproductive traits in two great tit (*Parus major*) populations. *Evolution*, 64, 2221–2237.
- Iwanicki, S., & Lehmann, J. (2015). Behavioral and trait rating assessments of personality in common marmosets (*Callithrix jacchus*). *Journal of Comparative Psychology*, 129, 205–217.
- Izzo, G. N., Bashaw, M. J., & Campbell, J. B. (2011). Enrichment and individual differences affect welfare indicators in squirrel monkeys (*Saimiri sciureus*). *Journal of Comparative Psychology*, 125, 347–352.
- Jaeggi, A. V., Boose, K. J., White, F. J., & Gurven, M. (2016). Obstacles and catalysts of cooperation in humans, bonobos, and chimpanzees: Behavioural reaction norms can help explain variation in sex roles, inequality, war and peace. *Behaviour*, 153, 1015–1051.
- Jaeggi, A. V., Stevens, J. M., & Van Schaik, C. P. (2010). Tolerant food sharing and reciprocity is precluded by despotism among bonobos but not chimpanzees. *American Journal of Physical Anthropology*, 143, 41–51.
- Janson, C. H. (2012). Reconciling rigor and range: Observations, experiments, and quasi-experiments in field primatology. *International Journal of Primatology*, 33, 520–541.
- Jennrich R. I., (2007). Rotation methods, algorithms, and standard errors. In R. Cudeck, & R. C. MacCallum, (Eds.), *Factor analysis at 100: Historical developments and future directions* (pp. 315–335). Mahwah, New Jersey: Lawrence Erlbaum Associates.
- Johnson, J. B., & Omland, K. S. (2004). Model selection in ecology and evolution. *Trends in Ecology & Evolution*, 19, 101–108.
- Johnson, Z., Brent, L., Alvarenga, J. C., Comuzzie, A. G., Shelledy, W., Ramirez, S., . . . Kaplan, J. R. (2015). Genetic influences on response to novel objects and dimensions of personality in Papio baboons. *Behavior Genetics*, 45, 215–227.
- Jöreskog, K. G. (1967). A general approach to confirmatory maximum likelihood factor analysis. *Psychometrika*, 34, 183–202.
- Kano T. (1992). *The last ape: Pygmy chimpanzee behavior and ecology*. Stanford: Stanford University Press.
- King, J., Weiss, A., & Farmer, K. (2005). A chimpanzee (*Pan troglodytes*) analogue of cross-national generalization of personality structure: Zoological parks and an African sanctuary. *Journal of Personality*, 73, 389–410.
- Konečná, M., Lhota, S., Weiss, A., Urbánek, T., Adamová, T., & Pluháček, J. (2008). Personality in free-ranging Hanuman langur (*Semnopithecus entellus*) males: Subjective ratings and recorded behavior. *Journal of Comparative Psychology*, 122, 379–389.
- Koski S. E., (2011a). How to measure animal personality and why does it matter? Integrating psychological and biological approaches to animal personality. In M. Inoue-Murayama, S. Kawamura, & A. Weiss, (Eds.), *From genes to animal behavior: Social structures, personalities, communication by color* (pp. 115–136). Tokyo: Springer.
- Koski, S. E. (2011b). Social personality traits in chimpanzees: Temporal stability and structure of behaviourally assessed personality traits in three captive populations. *Behavioral Ecology and Sociobiology*, 65, 2161–2174.
- Koski, S. E. (2014). Broader horizons for animal personality research. *Frontiers in Ecology and Evolution*, 2, 70.

- Krebs C. J., (2000). Hypothesis testing in ecology. In L. Boitani, & T. K. Fuller, (Eds.), *Research techniques in animal ecology* (pp. 1–14). New York: Columbia University Press.
- Martin, J. G., Nussey, D. H., Wilson, A. J., & Reale, D. (2011). Measuring individual differences in reaction norms in field and experimental studies: A power analysis of random regression models. *Methods in Ecology and Evolution*, 2, 362–374.
- Martin P., & Bateson P. (1993). *Measuring behavior: an introductory guide* (2nd ed.). Cambridge: University of Cambridge Press.
- Marsh, H. W., Morin, A. J., Parker, P. D., & Kaur, G. (2014). Exploratory structural equation modeling: An integration of the best features of exploratory and confirmatory factor analysis. *Annual Review of Clinical Psychology*, 10, 85–110.
- Massen, J. J. M., Antonides, A., Arnold, A. K., Bionda, T., & Koski, S. E. (2013). A behavioral view on chimpanzee personality: Exploration tendency, persistence, boldness, and tool-orientation measured with group experiments. *American Journal of Primatology*, 75, 947–958.
- Mathot K. J., & Dingemanse N. J., (2015). Plasticity and personality. In L. B. Martin, C. K. Ghalambor, & H. A. Woods, (Eds.), *Integrative organismal biology* (pp. 55–69). New Jersey: Wiley.
- McElreath R. (2015). *Statistical rethinking: A Bayesian course with examples in R and Stan*. Boca Raton: CRC Press.
- Morton, F. B., Lee, P. C., Buchanan-Smith, H. M., Brosnan, S. F., Thierry, B., Paukner, A., ... Weiss, A. (2013). Personality structure in brown capuchin monkeys (*Sapajus apella*): comparisons with chimpanzees (*Pan troglodytes*), orangutans (*Pongo spp.*), and rhesus macaques (*Macaca mulatta*). *Journal of Comparative Psychology*, 127, 282–298.
- Murray L., (2011). Predicting primate behavior from personality ratings. In A. Weiss, J. E. King, & L. Murray, (Eds.), *Personality and temperament in nonhuman primates* (pp. 129–168). Berlin: Springer.
- Nakagawa, S., & Cuthill, I. C. (2007). Effect size, confidence interval and statistical significance: A practical guide for biologists. *Biological Reviews*, 82, 591–605.
- Nakagawa, S., & Schielzeth, H. (2010). Repeatability for gaussian and non-gaussian data: A practical guide for biologists. *Biological Reviews*, 85, 935–956.
- Nettle, D., & Penke, L. (2010). Personality: Bridging the literatures from human psychology and behavioural ecology. *Philosophical Transactions of the Royal Society of London B*, 365, 4043–4050.
- Neumann, C., Agil, M., Widdig, A., & Engelhardt, A. (2013). Personality of wild male crested macaques (*Macaca nigra*). *PLoS ONE*, 8, e 69383.
- Nussey, D. H., Wilson, A. J., & Brommer, J. E. (2007). The evolutionary ecology of individual phenotypic plasticity in wild populations. *Journal of Evolutionary Biology*, 20, 831–844.
- Palagi, E. (2006). Social play in bonobos (*Pan paniscus*) and chimpanzees (*Pan troglodytes*): Implications for natural social systems and interindividual relationships. *American Journal of Physical Anthropology*, 129, 418–426.
- Palagi, E., & Paoli, T. (2007). Play in adult bonobos (*Pan paniscus*): Modality and potential meaning. *American Journal of Physical Anthropology*, 134, 219–225.
- Palagi, E., & Norscia, I. (2013). Bonobos protect and console kin. *PLOS ONE*, 8, e79290.
- Pollet, T. V., Stulp, G., Henzi, S. P., & Barrett, L. (2015). Taking the aggravation out of data aggregation: A conceptual guide to dealing with statistical issues related to the pooling of individual-level observational data. *American Journal of Primatology*, 77, 727–740.
- Réale, D., Reader, S. M., Sol, D., McDougall, P. T., & Dingemanse, N. J. (2007). Integrating animal temperament within ecology and evolution. *Biological Reviews*, 82, 291–318.
- Rosa G. J. M., (2015). Heritability and repeatability. In H. Khatib, (Ed.), *Molecular and quantitative animal genetics* (pp. 39–42). Hoboken: Wiley Blackwell.
- Rossee, Y. (2012). Lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, 48, 1–36.
- Searle, S. R. (1961). Phenotypic, genetic, and environmental correlations. *Biometrics*, 17, 474–480.
- Seyfarth, R. M., Silk, J. B., & Cheney, D. L. (2012). Variation in personality and fitness in wild female baboons. *Proceedings of the National Academy of Sciences*, 109, 16980–16985.
- Shipley B. (2016). *Cause and correlation in biology* (2nd ed.). Cambridge: Cambridge University Press.
- Sih, A., Bell, A., & Johnson, J. (2004). Behavioral syndromes: An ecological and evolutionary overview. *Trends in Ecology & Evolution*, 19, 372–378.
- Sih, A., Mathot, K. J., Moirón, M., Montiglio, P. O., Wolf, M., & Dingemanse, N. J. (2015). Animal personality and state-behaviour feedbacks: A review and guide for empiricists. *Trends in Ecology & Evolution*, 30, 50–60.
- Šlipogor, V., Gunhold-de Oliveira, T., Tadić, Z., Massen, J. J., & Bugnyar, T. (2016). Consistent inter-individual differences in common marmosets (*Callithrix jacchus*) in Boldness-Shyness, Stress-Activity, and Exploration-Avoidance. *American Journal of Primatology*, 78, 961–973.
- Staes, N., Weiss, A., Helsen, P., Korody, M., Eens, M., & Stevens, J. M. (2016). Bonobo personality traits are heritable and associated with vasopressin receptor gene 1a variation. *Scientific Reports*, 6, 38193.
- Staes N., Eens M., Weiss A., & Stevens J. M. G., (in press). Bonobo personality: Age and sex effects and links with behavior and dominance. In B. Hare, & S. Yamamoto, (Eds.), *Bonobos: Unique in mind, brain and behaviour*. Oxford University Press, Available online at: [http://www.research.ed.ac.uk/portal/en/publications/bonobo-personality\(c9805261-a12f-4f06-a66e-72fce7bcb774\).html](http://www.research.ed.ac.uk/portal/en/publications/bonobo-personality(c9805261-a12f-4f06-a66e-72fce7bcb774).html)
- Stanford, C. (1998). The social behavior of chimpanzees and bonobos. *Current Anthropology*, 39, 399–407.
- Stephens, P. A., Buskirk, S. W., & del Rio, C. M. (2007). Inference in ecology and evolution. *Trends in Ecology & Evolution*, 22, 192–197.
- Symonds, M. R., & Moussalli, A. (2011). A brief guide to model selection, multimodel inference and model averaging in behavioural ecology using Akaike's information criterion. *Behavioral Ecology and Sociobiology*, 65, 13–21.
- Uher, J. (2008). Comparative personality research: Methodological approaches. *European Journal of Personality*, 22, 427–455.
- Uher, J. (2011). Individual behavioral phenotypes: An integrative meta-theoretical framework. Why “behavioral syndromes” are not analogs of “personality”. *Developmental Psychobiology*, 53, 521–548.
- Uher, J., Asendorpf, J., & Call, J. (2008). Personality in the behaviour of great apes: Temporal stability, cross-situational consistency and coherence in response. *Animal Behaviour*, 75, 7599–75.
- Uher, J. (2013). Personality psychology: Lexical approaches, assessment methods, and trait concepts reveal only half of the story—why it is time for a paradigm shift. *Integrative Psychological and Behavioral Science*, 47, 1–55.
- Uher, J., Werner, C. S., & Gosselt, K. (2013). From observations of individual behaviour to social representations of personality: Developmental pathways, attribution biases, and limitations of questionnaire methods. *Journal of Research in Personality*, 47, 647–667.
- Uher, J., & Visalberghi, E. (2016). Observations versus assessments of personality: A five-method multi-species study reveals numerous

- biases in ratings and methodological limitations of standardised assessments. *Journal of Research in Personality*, 61, 61–79.
- van de Pol, M. (2012). Quantifying individual variation in reaction norms: How study design affects the accuracy, precision and power of random regression models. *Methods in Ecology and Evolution*, 3, 268–280.
- van de Pol, M., & Wright, J. (2009). A simple method for distinguishing within-versus between-subject effects using mixed models. *Animal Behaviour*, 77, 753–758.
- Vazire S., Gosling S. D., Dickey A. S., & Schapiro S. J., (2007). Measuring personality in nonhuman animals. In R. W. Robins, R. C. Fraley, & R. F. Krueger, (Eds.), *Handbook of research methods in personality psychology* (pp. 190–206). New York: Guilford Press.
- Vervaecke, H., de Vries, H., & van Elsacker, L. (2000). Dominance and its behavioral measures in captive groups of bonobos (*Pan paniscus*). *International Journal of Primatology*, 21, 47–68.
- Weinstein T. A., Capitanio J. P., & Gosling S. D., (2008). Personality in animals. In O. P. John, R. W. Robins, & L. A. Pervin, (Eds.), *Handbook of personality: Theory and research* (pp. 328–348). New York: Guilford Press.
- Weiss A., & Adams M. J., (2013). Differential behavioral ecology: The structure, life history, and evolution of primate personality. In C. Carere, & D. Maestripieri, (Eds.), *Animal personalities: Behavior, physiology, and evolution* (pp. 96–117). Chicago: University of Chicago Press.
- Weiss, A., Adams, M. J., & Widdig, A. (2011). Rhesus macaques (*Macaca mulatta*) as living fossils of hominoid personality and subjective well-being. *Journal of Comparative Psychology*, 125, 72–83.
- Weiss, A., Inoue-Murayama, M., Hong, K., Inoue, E., Udono, T., Ochiai, T., . . . King, J. E. (2009). Assessing chimpanzee personality and subjective well-being in Japan. *American Journal of Primatology*, 71, 283–292.
- Weiss, A., Staes, N., Pereboom, J. J., Inoue-Murayama, M., Stevens, J. M., & Eens, M. (2015). Personality in bonobos. *Personality Science*, 25, 1430–1439.
- West B. T., Welch K. B., & Gatecki A. T. (2007). *Linear mixed models: A practical guide using statistical software*. Boca Raton: Taylor & Francis.
- White, J. R., Meekan, M. G., McCormick, M. I., & Ferrari, M. C. (2013). A comparison of measures of boldness and their relationships to survival in young fish. *PLoS ONE*, 8, e68900.
- Widaman K. F., (2007). Common factors versus components: Principals and principles, errors and misconceptions. In R. Cudeck, & R. C. MacCallum, (Eds.), *Factor analysis at 100: Historical developments and future directions* (pp. 177–203). Mahwah, New Jersey: Lawrence Erlbaum Associates.
- Wobber, V., Wrangham, R., & Hare, B. (2010). Bonobos exhibit delayed development of social behavior and cognition relative to chimpanzees. *Current Biology*, 20, 226–230.
- Wolf, E. J., Harrington, K. M., Clark, S. L., & Miller, M. W. (2013). Sample size requirements for structural equation models an evaluation of power, bias, and solution propriety. *Educational and Psychological Measurement*, 73, 913–934.
- Wolf, M., & Weissing, F. J. (2010). An explanatory framework for adaptive personality differences. *Philosophical Transactions of the Royal Society B*, 365, 3959–3968.
- Zhao, Y., Staudenmayer, J., Coull, B. A., & Wand, M. P. (2006). General design Bayesian generalized linear mixed models. *Statistical Science*, 21, 35–51.
- Zuur A. F., Ieno E. N., Walker N. J., Saveliev A. A., & Smith G. M. (2009). *Mixed Effect Models and Extensions in Ecology with R*. New York: Springer.

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

How to cite this article: Martin JS, Suarez SA. Personality assessment and model comparison with behavioral data: A statistical framework and empirical demonstration with bonobos (*Pan paniscus*). *Am J Primatol*. 2017;e22670. <https://doi.org/10.1002/ajp.22670>