Title:

Integrative personality assessment in wild Assamese macaques (*Macaca assamensis*)

Authors:

Anja Ebenau[1,2,3], Christoph von Borell[3,4], Lars Penke[3,4], Julia Ostner[1,2,3] and Oliver Schülke[1,2,3]

Authors affiliation:

[1] Department of Behavioral Ecology, Johann-Friedrich-Blumenbach Institute for Zoology and Anthropology, University of Goettingen, Kellnerweg 6, 37077 Goettingen, Germany

[2] Research Group Social Evolution in Primates, German Primate Center, Leibniz Institute for Primate Research, Kellnerweg 4, 37077 Goettingen, Germany

[3] Leibniz ScienceCampus Primate Cognition, Kellnerweg 4, 37077 Goettingen, Germany

[4] Biological Personality Psychology, Georg Elias Mueller Institute of Psychology, University of Goettingen, Goßlerstraße 14, 37073 Goettingen, Germany

Correspondence concerning this article should be addressed to Oliver Schülke, Department of Behavioral Ecology, Johann-Friedrich-Blumenbach Institute for Zoology and Anthropology, University of Goettingen, Kellnerweg 6, 37077 Goettingen, Germany, Phone: +49 551 39 33926, Fax: +49 551 39 9637, E-Mail: oschuel@gwdg.de

ACKNOWLEDGEMENTS

We thank the National Research Council of Thailand (NRCT) and the Department of National Parks, Wildlife and Plant Conservation (DNP) for permission to conduct this study and for all the support granted (permit 0002/2424). We are thankful to three anonymous reviewers for their constructive and helpful comments on the manuscript. We are grateful to I. Chatchawarn, J. Prabnasuk, K. Nitaya T. Wongsnak, M. Pongjantarasatien, K. Kreetiyutanont, M. Kumsuk and W. Saenphala (PKWS) for their cooperation over the years and permission to carry out this study. We thank A. Koenig and C. Borries, who developed the field site. Our special thanks goes to S. Jumrudwong, W. Nueorngshiyos, N. Juntuch, J. Wanart, R. Intalo, T. Kilawit, N. Pongangan, B. Klaewklar, N. Bualeng, D. Gutleb, P. Saisawatdikul, K. Srithorn, M. Swagemakers and T. Wisate for their excellent help in the field. This research was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Project number 254142454 / GRK 2070.

Abstract

In nonhuman animals, individuals of the same sex and age differ in their behavior patterns consistently across time, comparable with human personality differences. To draw conclusions about the adaptive value of behavior traits, it is essential to study them also in the wild where animals are subject to the ecological pressures that promoted the evolution of behavior strategies. This study was conducted in the Phu Khieo Wildlife Sanctuary, Thailand, on four groups of habituated wild Assamese macaques by observers who had familiarized themselves with the subjects over the course of an ongoing long-term study. We used a multi-method approach enabling the most comprehensive understanding of variation in stable inter-individual differences in a species-typical ecological setting. We combined trait ratings (TR), assessed with observer-report questionnaires (54 item HPQ) of 107 individuals of diverse age-sex classes, with behavior codings (BC) of 24 adult males. We found male and female personality constructs to be congruent and examined reliability and construct validity. Combining methods, we found two solutions with five factors to best describe the personality structure of the males: one structure comprised the dimensions $\text{Gregariousness}_{BC}$, $\text{Aggressiveness}_{BC}$, Sociability_{BC} and Vigilance_{BC}, complemented by a Confidence_{TR} domain and the other structure Opportunism_{TR}, Confidence_{TR}, Friendliness_{TR}, Activity_{TR} complemented with Vigiance_{BC}. We discuss our findings with regard to the importance of construct validity and reproducibility in the context of method development and standardization in nonhuman animal personality research.

Keywords: personality, wild, macaques, behavior coding, trait rating, integrative assessment

Introduction

"Personality in the broadest sense is the internal organisation of behaviour that is stable over considerable time periods in the individual yet varies among the individuals of a population on latent dimensions" (Uher 2008 p. 476). Personality has a moderately heritable component and is systematically associated with differences in fitness parameters, like survival, reproductive success, and health in both animals and humans (Deary, Weiss, & Batty, 2010; Penke & Jokela, 2016; Wolf & Weissing, 2012). Animal personality became an expanding field of behavior ecology as well as comparative psychology in the last two decades (Carere & Maestripieri, 2013; Dingemanse, Kazem, Réale, & Wright, 2010; Mehta & Gosling, 2008; Uher, 2008). Yet, there are still no consistent definitions for personality dimensions, or standardized assessment methods, and differences in species-specific behavior repertoires make cross-species comparisons challenging (Gosling, 2001; Lilley, Kuczaj, & Yeater, 2017).

The two most common assessment methods in animal personality research are trait ratings and behavior coding from naturalistic observations or experimental data (Réale, Reader, Sol, McDonald, & Dingemanse 2007; Vazire, Gosling, Dickey, & Schapiro, 2007). In human psychology, personality is mainly assessed via self-reported statement ratings on personality inventories (Koski, 2011a; Uher, 2011) and expressed in broad bipolar or monopolar dimensions, often referred to as the so-called "Big Five" Extraversion, Agreeableness, Conscientiousness, Neuroticism, and Openness to Experience (Digman, 1990; John & Srivastava, 1999). In a top-down approach personality questionnaires have been modified for animals to allow trait ratings by human observers familiar with the individuals (e.g., Hominoid Personality Questionnaire; Freeman & Gosling, 2010; Weiss et al., 2009). To

avoid issues with validity possible anthropomorphic projections in trait ratings (Freeman et al., 2013; Uher, 2018) or naturalistic observational approaches use previously established ethograms defining behaviors and situations at the species level to quantify single behaviors or behaviors in a behavioral situation (Uher, 2015, Massen et al. 2013). Experimental observational approaches create situations to tap into one (Dammhahn, 2012) or several personality dimensions at a time (Carter et al., 2012b; Koski & Burkart, 2015; Massen et al., 2013; Réale et al., 2007; Vazire et al., 2007).

The choice of method typically depends on time and feasibility. Trait ratings by familiar observers are less time-consuming and can cover a broader range of individual traits compared to behavior codings in experimental or natural settings. However, trait ratings require that several raters are familiar with the individuals, because different observers know an individual from different situations and may have unique observer biases (Freeman, Gosling, & Schapiro, 2011). Experimentation is better able to assess non-social personality dimensions, such as boldness and exploration because relevant situations occur rarely and unpredictable in naturalistic observations (Massen, Antonides, Arnold, Bionda, & Koski, 2013), but is not always feasible, especially in wild populations (but see: Neumann, Agil, Widdig, Engelhardt, & Kalueff, 2013; Tkaczynski et al., 2018). When several methods have been applied to assess personality of the same subjects, convergent validity across methods was often low (Freeman et al., 2011; John & Soto, 2007; Uher & Visalberghi, 2016).

Such disparity may result from the disadvantages of behavior codings, which may struggle with reliably detecting individual differences in behaviors that occur at low frequencies (Freeman et al., 2013) and behaviors that change from day to day, across seasons, or with changing social and physical environment (Brommer & Class, 2017; Uher, 2011).

Aggregation may overcome these shortcomings in behavior measurements, since a higher aggregation level inevitably leads to higher reliability scores (for discussion about the *Principle of Aggregation* see Uher, 2011). In trait ratings, however, variability in behavior due to random variance is partly cancelled out, because observers implicitly aggregate an animal's behavior over time when they form an image of a subject's personality and subject it to their own memory. This aggregation comes at a cost though, because such images may be biased towards specific, more memorable events, and the assessment of subjects further may be influenced by discussions among raters, and may be modified through recalled memories (which are reshaped every time they are recalled). The advantage of behavior coding studies is that behaviors are directly perceivable and measurable. Hence they are not susceptible to most of the biases that can influence trait ratings (Freeman et al., 2011; van Aken & Asendorpf, 2018). The above mentioned problems can be overcome by repeated experimentation and long-term observations, when even rare behaviors occur often enough to reliably assess individual variation and all individuals are assessed across a variety of social and ecological contexts.

The strengths and weaknesses of trait rating and behavior coding have been repeatedly discussed (Freeman et al. 2011; Koski 2014; Uher et al. 2013; Uher & Visalberghi 2016; Uher 2018; Weiss 2017). It has been argued that the most comprehensive understanding of variation in stable inter-individual differences may come from studies combining approaches to compensate for the weaknesses that each technique individually has (Koski 2014). Multi-method approaches in field and lab work use item selection based on reliability measures (e.g., test-retest) and analyses of construct validity (Eckardt et al., 2015; Garai, Weiss, Arnaud, & Furuichi, 2016), and may generate integrative/complemented personality structures to provide a more complete picture (e.g., behavior codings and experiments:

Massen & Koski, 2014; Neumann et al., 2013; Uher, Addessi, & Visalberghi, 2013; behavior codings and trait ratings: Iwanicki & Lehmann, 2015; Uher, Werner, & Gosselt, 2013; all three methods: Tkaczynski et al., 2018). As can be expected from the differences in strength and weaknesses of the methods, personality structures derived from trait ratings and behavior codings do not always clearly correspond to each other. Aggressiveness related trait rating dimensions are mainly well supported with aggressive behavior (Pritchard et al., 2014; Uher, Werner, et al., 2013; Vazire et al., 2007), yet aggressive behavior may also correlate with a Dominance dimension (Freeman et al., 2013; Uher, Werner, et al., 2013; Uher & Visalberghi, 2016) or Excitability (Tkaczynski et al., 2018). Sociability mostly displays sociable or affiliative facets of behavior, yet grooming also correlates with Confident (Capitanio et al., 1999) or Irritability (Garai et al., 2016). The behaviors active and playful are often correlated (Uher & Visalberghi, 2016; Vazire et al., 2007), but playful also correlates with anxious and scratch (Iwanicki & Lehmann, 2015). Further, being physically active (i.e. not resting) does not always correlate with social activity (Koski, 2011). Thus, multi-method approaches may reveal more subtle and complete personality structures because each method captures aspects of personality that are not fully grasped by the other (Garai et al., 2016; Uher & Visalberghi; 2016; Tkaczynski et al., 2018) .

Different species of the macaque genus vary in their social style, i.e. in aspects of affiliation, aggression, dominance, nepotism, maternal behavior and socialization (Thierry, Singh, & Kaumanns, 2004). Adams and colleagues (2015) showed that similarities in personality dimensions capturing aggression and social competence are related to similarities in social styles (ranging from despotic via intermediate to egalitarian), and that the personality structure of the intermediately tolerant Assamese macaques (*Macaca assamensis*) studied here clustered with that of egalitarian crested macaques (*Macaca nigra*). In contrast to many

other male primates, male Assamese macaques form differentiated non-kin social relationships with other males (Schülke, Bhagavatula, Vigilant, & Ostner, 2010) with females (Haunhorst, Schülke, & Ostner, 2016), and with immatures (Minge, Berghänel, Schülke, & Ostner, 2016) and engage in frequent coalition formation and different affiliative behaviors, like grooming, gentle touch, and male–infant–male interactions (Kalbitz, Schülke, & Ostner, 2017; Ostner & Schülke, 2014).

The aim of this study was to explore the personality structure of male Assamese macaques in their natural habitat to provide the baseline for future studies on the socioecology of inter-individual differences. We employed behavior coding to emphasize prosocial personality traits that received less attention in previous studies (Koski, 2014), but may influence social partner choice (Massen & Koski, 2014) and individual fitness (Seyfarth & Cheney, 2013). The behavior coding was combined with trait rating to develop a more comprehensive personality model that integrates the strengths of both methods. Our discussion will contribute to the ongoing debate about personality assessment in nonhuman animals (e.g., Iwanicki & Lehmann, 2015; Tkaczynski et al., 2018; Uher & Visalberghi, 2016).

Methods

Study Site and Subjects

Fieldwork took place in the Phu Khieo Wildlife Sanctuary (PKWS: 16°5' – 35'N, 101°20' – 55'E) which is part of the ca. 6,500 km^2 interconnected and well-protected Western Isaan Forest Complex in north-eastern Thailand (Borries, Larney, Kreetiyutanont, & Koenig, 2002). The study area is covered by hill evergreen forest and harbors a diverse community of large mammals and predators (Borries et al., 2002). Data for this study were collected on four

multimale-multifemale groups from April 2014 (ASM and AOM groups) or October 2014 (ASS and AOS groups) to March 2016. Group sizes at the beginning of behavior data collection are shown in Table S1.

Data collection

1) Trait ratings

All adult females and males of the four groups were rated with the 54 item Hominoid Personality Questionnaire (HPQ; King & Figueredo, 1997; Weiss et al., 2009), which had previously been successfully used to assess personality in different macaque species, including Assamese macaques (Adams et al., 2015). Each adjective item is defined within the context of general behaviors common to primates. For example, 'fearful' is defined as "Subject reacts excessively to real or imagined threats by displaying behaviors such as screaming, grimacing, running away or other signs of anxiety or distress." Adjective items are rated on a 7-point scale with 1 defined as "Displays either total absence or negligible amounts of the trait" and 7 defined as "Displays extremely large amounts of the trait." (Weiss et al., 2009). Ratings were done in an office away from the animals, and observers were instructed to base their ratings on overall impressions of the individual's behavior, and not to discuss the questionnaires with each other. The Thai project members used a Thai language version of the questionnaire that was back-translated to ensure accuracy. Within the study period from April 2014 to March 2016, four Thai field assistants, AE and another PhD student, all familiar with the subjects for 6 months to 7 years in April 2014, rated the animals twice. Eighty-one individuals were rated in the first year, all adult males in March 2015, and all adult females and sub-adult males in September 2015. Of these 81, 74 individuals were

rated again in March 2016. Animals were rated by the same observers who collected the behavior data within the study period.

One of the groups (ASM) had been rated earlier (2009) by eight observers (2 of them rated again in this study) familiar with the animals from focal animal data collection (60 adult and juvenile males and females) (Adams et al., 2015). Of these 60 individuals, 26 had emigrated into a non-study group or died by the beginning of this study and were included here to extend our data set to ensure a higher observations-to-variables ratio.

2) Behavior codings

We collected 4,628 hours of focal animal observations (Altmann, 1974) from 24 adult males (mean per subject = 193h; range = 86h–284h) of the four study groups. Focal animals were included in the study if they were present longer than three months within one year of the two-year study period. Individuals were followed for 40 minutes with continuous recording of all approaches and departures within 1.5m of the focal animal and all affiliative and agonistic social interactions with directionality and the identities of interaction partners, and several solitary behaviors (e.g., scratching, yawning, display; supplement Table S2). Activity of the focal animal was recorded instantaneously at 2-minutes intervals. Every 10min, we recorded the identities of all individuals within a 5m sphere around the focal animal. An effort was made to equally distribute observation time across individuals and time of the day. This study includes an extended data set of socially motivated behaviors, such as affiliative triadic male-infant-male interactions (Kalbitz et al., 2017; Paul, Kuester, & Arnemann, 1996).

Data analysis

All statistical analyses were run with R (version 3.4.2; R Core Team, 2017).

1) Trait ratings

We reduced the data to only those raters who used the entire 7-point scale, and those items that were rated by more than half of the remaining reliable observers, and those items that showed normal distribution. We calculated interrater-reliability, a measure of consistency across raters, and test-retest reliability, examining consistency over time, to assess reliable adjective items that would then be included in a factor analysis.

For each year (2009, 2015, 2016), data were visually inspected via scatterplots, histograms and Q-Q plots to examine the rating performance of observers and the distribution of each adjective item. Three observers had to be excluded from further analyses due to incorrect use of the coding protocol. Specifically, in 2009 three of eight observers and in 2015 and 2016 one of six observers (one of the two, who rated in all 3 years) did not use the complete 7-point scale, and used "1" ("Displays either total absence or negligible amounts of the trait") instead of "4" as mean rating value. The item "autistic" was excluded since 8 of the remaining reliable 9 observers did not rate it, leaving 53 items for further analysis.

Interrater-reliability was measured as Intraclass Correlation Coefficients (*ICC*; Shrout & Fleiss, 1979). *ICC*s assess reliability by comparing item variability across all observers. *ICC* quantifies the similarity of single trait ratings of each adjective for every individual among all observers (3,1), or mean trait ratings of each adjective for every individual among all observers (3,k). A good level of agreement for primate studies has been set at $ICC(3,k) = 0.48$ (Freeman & Gosling, 2010). A meaningful item selection is based on positive values when confidence intervals do not include zero, and often a cut-off criterion of $ICC(3,k) > 0.4$ is applied (Cicchetti, 1994).

Temporal stability of trait ratings was assessed with Pearson correlations (corrected for multiple testing by the false discovery rate, hereafter fdr) of mean adjective items, averaged across observers, for each individual from one rating to the next. High retest reliability measures for the trait ratings of 74 individuals present in 2015 and 2016 ($N_{adjective\ items} = 53$; $M = 0.66$; range: 0.3–0.92) support data aggregation. Thus, we assessed the overall personality structure from trait ratings of 107 subjects, a combination of mean ratings of 74 individuals present in 2016, plus seven individuals present in 2015 (which emigrated or died later), and additional 26 individuals that were present in the study group only in 2009. Only temporally stable adjective items, i.e. significant positive correlations > 0, were subjected to a factor analysis.

After data reduction based on reliability analyses, two measures of sampling adequacy were applied to check for moderate inter-correlations and hence factorability of variables. The Kaiser-Mayer-Olkin–(KMO) index compares values of correlations between items and those of partial correlations to check if at least two or three variables correlate with each other ("KMO" function in "psych" package; Revelle, 2018). The Bartlett's test of sphericity compares the observed correlation matrix to the identity matrix with no correlation and is significant when it deviates from identity ("bartlett.test" function in "REdaS" package; Maier, 2015).

The number of factors to be retained for factor analysis was determined using the "fa.parallel" function ("psych" package) and the "paran" function ("paran" package; Dinno, 2012) to perform a scree test (Cattell, 1966) with parallel analysis (Horn, 1965). Exploratory factor analysis was done with the "fa" function of the "psych" package. Principle axis factoring with promax rotation was applied to attain an optimal simple structure that

maximize high loadings on one factor for each variable. Oblique factor rotation allows for inter-correlations of factors, which is more adequate for trait ratings by observers as well as behavior codings (Uher & Visalberghi, 2016). For interpretation, salient factor loadings were considered to be ≥ |0.4|, and items with salient cross loadings were included in the factor on which they had the highest loading (Field, Miles, & Field, 2012). A clean factor structure is achieved by item loadings above |0.30| with preferably no cross loadings, and no factors with less than three items (Costello & Osborne, 2005; Freeman et al., 2013). Behavioral personality research, however, does often accept dimensions with less than three salient items (Koski, 2011b; Manson & Perry, 2013; Pritchard, Sheeran, Gabriel, Li, & Wagner, 2014; Seyfarth, Silk, & Cheney, 2012; Sussman, Ha, Bentson, & Crockett, 2013). Internal consistency for each retained personality dimension was calculated as Cronbach's Alpha on mean ratings of adjective items across raters for all 107 subjects ("alpha" function in "psych" package). Alpha ranges from zero to one, with higher values indicating greater internal consistency.

For further analyses, factor scores for each dimension were extracted with the "factor.scores" function ("psych" package) using the regression ("Thurstone") method. Rank order stability of the personality structures from one year to the next (March 2015 and 2016) was assessed with the "factor.congruence" function ("psych" package), comparing the factor loadings of respective dimensions. Confidence intervals (CI) were calculated with the "boot.data" function, we applied 1000 iterations and sampled 60 out of 74 individuals with no replacement. Further, to demonstrate that the convergent structure relies on individual stability, individual factor scores were correlated for every dimension applying fdr corrections for multiple testing. Due to a rather small sample size, which exaggerates the

294 effect of single individuals on the analysis output, the same variables and individuals were
295 submitted to factor analysis for both years.

296 The data sets for trait rating and behavior coding were imbalanced. TR was conducted for
297 male and female infants, juveniles and adults to increase the number of observations above
298 that of items in the factor analysis. Observational data for behavior coding were collected
299 only on adult males though. To assess whether this imbalance affected comparisons of
300 constructs derived from both methods we used a parceling approach for item reduction in the
301 TR data set (Little, Cunnigham, Shahar, Widaman, 2002). We ran separate factor analyses for
302 males and females, and compared the resulting constructs (further details in SEM). The
303 resulting separate four factor solutions for males (n=59) and females (n=48) had high factor
304 congruence of factor loadings (M=0.93, range: 0.90-0.98) suggesting they were similar or
305 equal (Lorenzo-Seva & ten Berge, 2006). We ran further tests to assess whether the trait-rated
306 factors from the males only (n=24) structure differed from the structure derived from the full
307 data set on males and females (n=107) and found high factor congruence in factor loadings
308 again (M=0.98, range: 0.96-0.99). All further analyses were run on the TR construct built
309 from the full data set.

310 *2) Behavior codings*

311 Variables for behavioral coding were derived from a long ethogram provided in the
312 supplement (Table S2). To arrive at our final set of variables for behavioral coding, we first
313 omitted variables that could not be measured reliably either because the behavior was shown
314 very rarely by the adult male subjects (‘yawning’; ‘treeshake’; ‘submission’; ‘reconciliation’;
315 ‘rejection’; ‘dominance’; ‘interference’; ‘ignorance’; ‘playface’) or because of problems with
316 data collection; the variable ‘fidgeting’ which should have captured how often an individual

317 changed between different activities, was omitted because we were unable to record these
318 changes in the continuous protocol and our instantaneous recording interval was too long to
319 capture its effects. In an effort to reduce the overall number of variables, we collapsed those
320 variables that had been defined per age-sex class into just one variable for all partners. We
321 then assessed for each study period separately whether those behaviors that comprised an
322 aggregated variable were positively correlated amongst each other and retained only those
323 that were. For 'friendly behavior' this meant that 'MIMI' was added, but 'peer', 'play',
324 'kiss', 'genital touch', 'mount', 'present' were omitted from the aggregate. We had planned
325 to aggregate aggressive behaviors into 'threats' and 'aggression'; we tested the validity of
326 these constructs by PCA and found three instead of two principle components that we named
327 'mild aggression', 'overt aggression' and 'physical aggression'. The aggressive behaviors
328 'open mouth', 'point', 'pretend grab', 'bite', and 'chase' were omitted, because they were not
329 consistently across years correlated to behaviors from one of the three aggregated aggression
330 variables. New aggregated variables were built from correlated behaviors that individually
331 occurred at rather low rates; 'auto-grooming' and 'scratching' into 'self-directed'; 'body-
332 contact' and 'grooming' into 'contact time' and 'contact diversity'; 'involvement', 'policing'
333 and 'agonistic support' aggregated into 'agonistic involvement'; the low frequency variables
334 mentioned above were omitted because they did not fit into any aggregate variable. The
335 aggregate variable 'vocalization' was dropped, because components were moved into other
336 aggregates; 'growl' was included in 'overt aggression'. One variable 'being left' was omitted
337 to further reduce the number of variables.

338 . .

After this first data reduction, 20 behavior variables were defined and extracted from the focal animal observations (supplement Table S7). Most of the behavior variables were calculated as rate per hour, corrected by the individual observation time. Other variables were calculated as proportions, across all partners in social measures. To assess the personality construct, variables were aggregated from the overall data collection period to maximize construct stability, which is compromised by low frequencies (Freeman et al., 2013). Reliability was assessed with Pearson correlations of mean values from one year to the next, and only the 18 stable out of the 20 variables were used (Table S7). The two years of data collection were split in half, to have one year of behavior codings before the trait ratings in March 2015 and 2016, respectively. In the first period, continuous data collection started half a year later for six adult males due to age-class change in October 2014. All behavior variables were visually inspected via histograms and Q-Q plots to examine their distribution pattern. "Display" was log transformed for the single years and square root transformed for the both years together. Variables were z-transformed before submitted to further analyses.

After data reduction, behavior variables were treated as described for the trait ratings (see above) to evaluate their suitability for factor analysis, except that further methods were considered to determine the number of factors to retain for factor analysis (besides scree test with parallel analysis using "fa.parallel" and "paran" function), to search for similar results across several criteria to overcome data insecurities (Freeman et al., 2013; Preacher & MacCallum, 2003). Additional methods were the Very Simple Structure Criterion (VSS; Revelle & Rocklin, 1979) and Velicer's Minimum Average Partial test (MAP; Velicer, 1976), both using the "vss" function ("psych" package), as well as Kaiser's rule (Kaiser, 1960) and scree tests with parallel analysis using the "nScree" function ("nFactors" package; Raiche, 2010).

Exploratory factor analysis and further steps were implemented as described above for the rating analysis, except that maximum likelihood method with oblimin rotation was applied, and that the CI's for factor congruence were calculated based on a sample of 19 out of 22 individuals (bootstrapping required subsampling).

3) Construct validity

Convergent construct validity examines the correlation between different measures of the same construct (John & Soto, 2007). Convergent construct validity between personality constructs, derived from trait ratings and behavior codings, was assessed in two ways. First, individual factor scores of 24 adult males were correlated (Pearson's *r*) for every dimension (Freeman et al., 2013; Garai, Weiss, Arnaud, & Furuichi, 2016b). Second, percentage bend correlation (Wilcox, 1994) was used to measure the relation between the behavior data from 24 adult males and the overall personality scores from questionnaire ratings from 107 individuals (Iwanicki & Lehmann, 2015; Morton et al., 2013; Tkaczynski et al., 2018). Z-transformed behavior variables were correlated via the "pbcor" function ("WRS2" package; Mair, Schoenbrodt, & Wilcox, 2017) with factor scores of every trait rating personality dimension.

Results

Trait ratings (TR)

Eight adjective items were excluded because of low interrater reliability, leaving 45 items with ICC_{mean} = 0.6; range: 0.4–0.87 (Table S8). The remaining 45 adjective items were submitted to measures of sampling adequacy and met the criteria. The overall KMO – index

was higher than 0.5 (KMO = 0.87; range: 0.57–0.92) and the Bartlett's test for homogeneity of variances was significant (Bartlett's K-squared (44) = 360.17; $p < 2.2e\text{-}16$).

The visual inspection of the scree plot with parallel analysis suggested retaining four factors to explain most of the variability in the data. Two items (thoughtless and unperceptive) had loadings < |0.4| on all factors and were excluded. Factor analysis with the remaining 43 items yielded a similar construct. Four dimensions were extracted: $Opportunism_{TR}$, $Confidence_{TR}$, $Activity_{TR}$ and $Friendliness_{TR}$ (Table 1). $Opportunism_{TR}$ comprises items such as aggressive, bullying, irritable, impulsive, not gentle and not stable. $Confidence_{TR}$ includes dominant, not a follower, not vulnerable, not timid and not submissive. $Activity_{TR}$ is described by being curious, active, playful, not depressed and not lazy. $Friendliness_{TR}$ is characterized with affectionate, sociable, friendly and not solitary. All four factors together explain 72% of the item variance (Table 1). The communalities are the sum of the squared factor loadings for a given variable and may be interpreted as a reliability indicator (Field, 2000). No item loadings were below |0.4|, there were relatively few cross loadings, and mean item communality h^2 was rather high (M = 0.72; range: 0.39–0.93), with only three items below 0.5 (clumsy, individualistic and intelligent), indicating robust personality dimensions (Costello & Osborne, 2005; de Winter, Dodou, & Wieringa, 2009).

Insert Table 1 about here

Correlations among personality dimensions, as computed with the "fa" function (M = |0.16|; range: |0.01–0.44|), were generally weak, apart from a moderate correlation between

Opportunism$_{TR}$ and Activity$_{TR}$. Internal consistencies for the factor scores (*Cronbach's alpha*$_{mean}$ = 0.92; range: 0.88–0.98) were high..

Rank order stability of personality dimensions from March 2015 to March 2016 was confirmed, with similar results for the congruence of factor loadings Φ (M = 0.95; range: 0.91–0.98) and the correlations of factor scores *r* (M = 0.81; range: 0.72–0.87; Table 2).

Insert Table 2

Behavior codings (BC)

Test-retest reliabilities *r* of behavior variables were rather low ($N_{behavior\ variables}$ = 20; *M* = 0.39; range: -0.16–0.69; Table S7). Two variables, grooming symmetry and overt aggression, did not meet the criteria for temporal stability, i.e. positive correlations > 0, and were excluded from further analyses.

The remaining 18 behavior variables were submitted to measures of sampling adequacy. The overall KMO – index for the total observation period was slightly lower than 0.5 (KMO = 0.46; range: 0.20–0.76) and the Bartlett's test for homogeneity of variances was significant (Bartlett's K-squared(17) = 1014.8; p < 2.2e-16). Low KMO values might result from the rather small observations-to–variables ratio and these variables were not rejected.

The visual inspection of the scree plot and parallel analysis ("psych" and "paran" package) suggested to retain three factors to explain most of the variability in the data. The VSS suggested four and the MAP test five factors to retain ("psych" package). The "nFactors"

428 package revealed for Kaiser's rule five factors, and for scree test and parallel analysis four
429 factors to retain.

430

431 ----------

432 Insert Table 3 about here

433 ----------

434

435 Due to these inconsistencies, we ran two factor analyses, extracting three and four factors,
436 respectively. The 4-factor solution, with the dimensions $Gregariousness_{BC}$, $Aggressiveness_{BC}$,
437 $Sociabiliy_{BC}$ and $Vigilance_{BC}$, (Table 3) yielded more robust factors as described below
438 (Costello & Osborne, 2005), though less robust than in the trait rating analysis, with seven of
439 18 variables showing item communalities below 0.5. The 4-factor solution had higher mean
440 communality values (4-factor solution: M(h^2) = 0.62; range: 0.3–1 and 3-factor solution:
441 M(h^2) = 0.55; range: 0.2–0.95), as well as a higher proportion of variance explained (4-factor
442 solution: 62% and 3-factor solution: 55%). The variable "display" had no reliable loading (<
443 |0.4|) in the 4-factor solution and cross loadings of |0.4| (in two dimensions) in the 3-factor
444 solution, reflecting its instability. In addition, the fourth factor was mainly marked by the
445 variable "vigilant", with a very high factor loading (0.99) and communality (h^2 = 1), but the
446 lowest communality (h^2 = 0.2) and a weak factor loading (-0.43) in the 3-factor solution. This
447 emphasizes that "vigilant" represents a separate factor, though it is rather unstable, with only
448 one additional variable "not active" loading on it, probably, due to the lack of more non-
449 social variables in the data. In other personality studies, vigilance also groups with negatively

loaded activity behaviors, and positively with other behaviors, not captured in our analysis, e.g., provisioning (Iwanicki & Lehmann, 2015).

As with the rating analysis, there were also moderate intercorrelations *r* among dimensions (M = |0.16|; range: |0.02–0.35|), speaking against varimax factor rotation. Internal consistency of the personality dimensions was fair to excellent (M(*Cronbach's alpha*) = 0.77; range: 0.62–0.92).. Alpha for Vigilance_{BC} was rather low at 0.62, but comparable with other studies where personality dimensions include less than three variables (e.g., Manson & Perry, 2013).

To assess rank order stability, the three factor solution was applied for the single years, even though there was a Heywood case (i.e. factor loading > 1.0) for "quitting" in the second year. There was no clear result for factor determination, considering different methods and the criteria for construct robustness, as mentioned above. This ambiguity underlines the need to aggregate the data (two-year observation period) to overcome the shortcoming of low frequencies in behavior codings, in order to enable interpretation of the personality structure.

Rank order stability of personality dimensions from one year to the next was moderate, and revealed some differences between the congruence of factor loadings Φ (M = 0.69; range: 0.37–0.92) and the correlations of factor scores *r* (M = 0.40; range: 0.24 – 0.57; Table 4).

Insert Table 4 about here

Construct validity TR with BC

Significant positive correlations of individual factor scores were found between personality dimensions $Friendliness_{TR}$ and $Gregariousness_{BC}$ ($r = 0.69$; $p < 0.001$), as well as between $Sociability_{BC}$ and $Activity_{TR}$ ($r = 0.63$; $p = 0.001$), and a low negative correlation between $Friendliness_{TR}$ and $Sociability_{BC}$ ($r = -0.41$; $p = 0.046$) (Table 5.; Figure 1).

Insert Table 5 about here

Insert Figure 1 about here

Regarding the relations between behavior variables many of which were integrated and personality dimensions based on trait ratings (Table 6), "mild aggression" correlated with $Opportunism_{TR}$ (ρ percentage bent, hereafter $\rho_{pb} = 0.41$; $p = 0.049$). "Display" ($\rho_{pb} = 0.59$; $p = 0.003$), "vigilant" ($\rho_{pb} = -0.43$; $p = 0.035$) and "self-directed" ($\rho_{pb} = -0.56$; $p = 0.005$) correlated with $Confidence_{TR}$. "Friendly behavior" ($\rho_{pb} = 0.55$; $p = 0.006$) and "peripheral" ($\rho_{pb} = 0.43$; $p = 0.037$) correlated with $Activity_{TR}$. Nine out of 18 variables correlated with $Friendliness_{TR}$: "active" ($\rho_{pb} = -0.44$; $p = 0.031$), "alone" ($\rho_{pb} = -0.66$; $p = 0.001$), "display" ($\rho_{pb} = 0.44$; $p = 0.033$), "friendly approach" ($\rho_{pb} = 0.60$; $p = 0.002$), "leaving" ($\rho_{pb} = 0.63$; $p = 0.001$), "neighbor diversity" ($\rho_{pb} = 0.65$; $p = 0.001$), "physical aggression" ($\rho_{pb} = 0.50$; $p = 0.012$), "peripheral" ($\rho_{pb} = -0.45$; $p = 0.026$) and "tolerance" ($\rho_{pb} = 0.60$; $p = 0.002$).

---

492 Insert Table 6 about here

493 ----------

494

495 **Discussion**

496 This study was not designed as a validation study comparing personality structures derived
497 with different methods or with the goal to compare results with the published literature. An
498 elaborate method comparison is hampered by the imbalance in our data sets for the two
499 methods and also would have required deriving a priori predictions about which factors from
500 each method would be congruent (Koski et al., 2017; Tkaczynski et al., 2018). A method
501 comparison would also have benefitted from including rating of adjectives that had been
502 derived from the species' behavior repertoire instead of applying constructs derived from
503 lexical accounts in humans (Uher et al., 2008). A validation study would assess which
504 constructs derived by one method are congruent with a construct derived by the second
505 method and thereby arrive at a possibly reduced set of constructs that have been validated.
506 Accepting that to some extent each method will grasp slightly different aspects of personality,
507 the aim of this study was to build a comprehensive personality structure by also retaining
508 dimensions that were not congruent across methods and thereby extending the structure space
509 beyond what could be achieved by a single method. Before we describe this structure, we will
510 relate the personality structures and dimensions developed here with published work on the
511 same and other macaque species and other nonhuman primates to establish how robust and
512 reproducible they are across studies using slightly or vastly different methods.

513 **Reproducibility of trait ratings and construct validity**

514 A previous study on Assamese macaque personality (Adams et al., 2015) did not employ item
515 selection and extracted five personality dimensions from HPQ ratings (labelled as
516 $Confidence_{TR}$, $Activity_{TR}$, $Openness_{TR}$, $Friendliness_{TR}$ and $Opportunism_{TR}$) whereas our
517 rating analysis revealed four dimensions: $Opportunism_{TR}$, $Confidence_{TR}$, $Activity_{TR}$ and
518 $Friendliness_{TR}$ which we named the same when replicated. Our $Opportunism_{TR}$ domain
519 comprised all adjectives representing $Opportunism_{TR}$, as well as 4 out of 10 items of
520 $Openness_{TR}$ in Adams et al. Behavior codings from adult males also resulted in a four-factor
521 structure, with $Gregariousness_{BC}$, $Aggressiveness_{BC}$, $Sociability_{BC}$ and $Vigilance_{BC}$, that has
522 considerable overlap with the personality structure from trait rating and does reflect previous
523 findings in other macaque studies (Neumann et al., 2013). $Friendliness_{TR}$ was convergent
524 with $Gregariousness_{BC}$ and inverse $Sociability_{BC}$, whereas $Activity_{TR}$ correlated with
525 $Sociability_{BC.}$ In addition, $Opportunism_{TR}$ was weakly associated with $Aggressiveness_{BC}$.
526 $Confidence_{TR}$ replicated very well, but showed no association with the behavior personality
527 dimensions.

528 The trait rating dimension $Friendliness_{TR}$ confirmed previous findings (Adams et al., 2015;
529 Konečná et al., 2008; Sussman et al., 2013; Weiss, Adams, Widdig, & Gerald, 2011): the
530 rather strong link between $Friendliness_{TR}$ and $Gregariousness_{BC}$ is explained by all behavior
531 variables belonging to $Gregariousness_{BC}$ having significant correlations with $Friendliness_{TR}$.
532 More specifically, adjectives, like “sympathetic” (kind towards others), “sociable” (seeks and
533 enjoys company) and “not solitary” (spends considerable time alone) correlate with behaviors
534 like “friendly approach” or “tolerance”, “neighbor diversity” and “not alone”.
535 $Gregariousness_{BC}$ represents spatial proximity aspects that have been suggested to measure
536 social integration, i.e., close proximity to and higher rates of affiliation with diverse partners.
537 The evidence for a $Gregariousness_{BC}$ dimension is mixed for macaques. There are similar

538 domains with the focus on close and distant proximity in Barbary, crested, and Rhesus
539 macaques (Capitanio, 1999; Neumann et al., 2013; Tkaczynski et al., 2018), or domains with
540 a mixture of behaviors belonging to our $Gregariousness_{BC}$ and $Sociability_{BC}$ in Barbary,
541 longtailed, lion-tail, and Tibetan macaques (Pritchard et al., 2014; Rouff, Sussman, & Strube,
542 2005; Tkaczynski et al., 2018; Uher, Werner, et al., 2013).

543 The negative correlation between $Friendliness_{TR}$ and $Sociability_{BC}$, while being weak, still
544 seems unexpected. $Friendliness_{TR}$ does not seem to cover the behavior personality dimension
545 $Sociability_{BC}$, since four out of the five variables with positive factor loadings on
546 $Sociability_{BC}$ correlated negatively with $Friendliness_{TR}$. $Sociability_{BC}$ seems to have two
547 facets: first, it was primarily described by high rates of short-term affiliation (such as
548 "embrace"), with a factor loading of 0.99 (mean loadings of other variables < 0.5) and a
549 negative correlation with $Friendliness_{TR}$, and thus may reflect an aspect of social integration,
550 not included in $Friendliness_{TR}$. Second, $Sociability_{BC}$ seems to express variation in social
551 bonding, i.e. the tendency to focus affiliation on a few close partners (Ostner & Schülke,
552 2014). Individuals with stronger bonds have a higher "contact time" (time in body contact
553 and grooming) with less social partners leading to a negative correlation of "contact
554 diversity" with $Friendliness_{TR}$. Also, "contact time", as a measure for "affectionate"
555 (closeness and frequent grooming with others), was not correlated with $Friendliness_{TR}$, as
556 high scores may result from a few long or from many short periods of contact. In addition,
557 rates of "friendly behavior" do not need to be high for strongly bonded individuals, since they
558 might engage in prolonged but less frequent social interactions with their bonded partners.
559 Crested and Rhesus macaques exhibit a similar behavior $Sociability_{BC}$ dimension with high
560 grooming and affiliation rates as well as diverse grooming and affiliation (Neumann et al.,
561 2013; von Borell, Kulik, & Widdig, 2016). Interestingly, "peripheral" (staying outside the

562 group center) included in Sociability$_{BC}$ underlines that there is no need for individuals to stay
563 close to the core of the group to be socially integrated. In Barbary and Crested macaques
564 being "central" loads on a dimension most similar to this study's Gregariousness$_{BC}$
565 (Neumann et al., 2013; Tkaczynski et al., 2018).

566 Opportunism$_{TR}$ completely comprises adjectives referring to Adam's and colleagues
567 Opportunism$_{TR}$, as well as four out of ten adjectives belonging to their Openness$_{TR}$
568 dimension, i.e. "excitable", "impulsive", "erratic", "distractible" (Adams et al., 2015). The
569 remaining five 'Openness$_{TR}$ adjectives' (e.g., "thoughtless", "innovative") were excluded
570 from our analysis and "individualistic" loaded on Confidence$_{TR}$. It was unexpected that
571 Opportunism$_{TR}$ and Aggressiveness$_{BC}$ were not strongly correlated. This weak relationship
572 was likely driven by the behaviors "mild aggression" and "quitting", because observers may
573 have perceived more frequent behaviors as more important ("mild aggression" was 5 times
574 more frequent than "physical aggression"). Similarly, in chimpanzees, aggression related trait
575 ratings converged strongly with "threat", but only weakly with the lower frequency "attack"
576 behavior measures (Vazire et al., 2007). "Physical aggression" and "leaving" correlated
577 positively with Friendliness$_{TR}$, which is similar to a positive association of "contact
578 aggression" and the socio-positive Extraversion$_{TR}$ found in chimpanzees (Freeman et al.,
579 2013). As "physical aggression" is often expressed in more serious conflicts which in
580 chimpanzees and Assamese macaques often turn polyadic, an individual acting aggressively
581 may at the same time be supportive and helpful in a coalitionary context. "Helpful" is
582 included in Friendliness$_{TR}$, possibly causing the positive correlation between "physical
583 aggression" and Friendliness$_{TR}$. In addition, there is a general link between affiliative and
584 aggressive interactions among partners due to the increasing probability of conflict with
585 increasing time spent in close proximity to others (Silk et al., 2010; Widdig, Nürnberg,

586 Krawczak, Streich, & Bercovitch, 2002). In line with this, in macaque personality studies
587 affiliative and aggressive behaviors are frequently correlated as are social behavior and
588 aggressiveness-related rating dimensions (Capitanio, 1999; Rouff et al., 2005; Tkaczynski et
589 al., 2018; Uher, Werner, et al., 2013). So far, rating Opportunism$_{TR}$ domains are described for
590 Assamese (this study), crested (Adams et al., 2015), pigtail and longtailed (Sussman et al.
591 2013, Uher, Werner, et al., 2013) as well as Tibetan macaques (Pritchard et al., 2014),
592 whereas Confidence$_{TR}$ or Dominance$_{TR}$ dimensions are found in Rhesus (Adams et al., 2015;
593 Capitanio, 1999, Weiss et al., 2011) and Confidence$_{TR}$ and Opportunism$_{TR}$ or Excitability$_{TR}$
594 in Barbary macaques (Adams et al. 2015; Konečná et al., 2012; Tkaczynski et al., 2018).

595 Confidence$_{TR}$ is almost equivalent to the dimension found previously (Adams et al., 2015).
596 Confidence$_{TR}$ was weakly related to Vigilance$_{BC}$ as a statistical trend, given the negative
597 correlation of "vigilant" behavior with Confidence$_{TR}$. However, the two dimensions seem not
598 convergent, since other candidate behaviors for Confidence$_{TR}$ did not correlate with
599 "vigilant". For instance, "display", presumably a dominance behavior (Freemann et al., 2013;
600 Uher, Werner, et al., 2013; Tkaczynski et al., 2018), showed no reliable loading on any
601 behavior personality dimension, but a prominent positive correlation with Confidence$_{TR}$. In
602 addition, "self-directed" behavior correlated negatively with Confidence$_{TR}$, suggesting it is a
603 measure of anxiety (Brent et al., 2014; Koski, 2011b; Neumann et al., 2013; Iwanicki &
604 Lehmann, 2015; Tkaczynski et al., 2018), but here it only loaded negatively, although
605 weakly, on Sociability$_{BC}$. There was a lack of behavior variables supporting Confidence$_{TR}$
606 adjectives like "dominant" and "submissive", because adult Assamese macaques rarely show
607 unprovoked submissive behavior towards other individuals. Similarly, a study on wild
608 Barbary macaques did not find behavior variables to be correlated with their Confidence$_{TR}$
609 dimension, which is a mixture of items belonging to our Confidence$_{TR}$ and Opportunism$_{TR}$

610 (Tkaczynski et al., 2018). Confidence$_{TR}$ was correlated with dominance rank though in
611 Hanuman langurs and with rank stability in Barbary macaques (Konečná et al., 2008;
612 Konečná, Weiss, Lhota, & Wallner, 2012). Taken together, a fifth behavior dimension
613 replicating Confidence$_{TR}$ may be possible.

614 In contrast to the reverse correlation of "vigilant" and "active" in this study, a positive
615 correlation of "vigilance" with "playfulness" and "physical activity" was found in longtailed
616 macaques (Uher et al., 2013) and play behavior in brown capuchins (Morton et al., 2013). In
617 these studies, vigilance was measured as social attentiveness, monitoring the activities of
618 other group members. Further, "vigilant" loaded on the Anxiety dimension derived from
619 playback experiments in wild Barbary macaques (Tkaczynski et al., 2018). However, it was
620 only used as playback reaction measure and not as a neutral variable in the every-day
621 behavior coding. A behavior study with free-ranging Rhesus macaques (von Borell et al.,
622 2016) described a Fearfulness$_{BC}$ dimension, resembling some of the behavior variables of this
623 study loading on Gregariousness$_{BC}$ (i.e. friendly approach, proximity, resting) grouped
624 together with "submissive". If our "vigilance" represented an anxiety measure, it would most
625 likely group with other potential anxiety behaviors, but instead it represented its own
626 dimension. Thus, the "vigilance" variable in this study reflected a general attentiveness to the
627 surrounding environment, social or ecological.

628 Activity$_{TR}$ shared the most descriptive adjectives with highest loadings, e.g., "active"
629 (considerable time moving or engaging in energetic behavior) and "not lazy" and "playful",
630 with earlier findings (Adams et al., 2015). Additionally, our Activity$_{TR}$ included "curious",
631 like in other macaque studies (Barbary, Konecná et al., 2012; longtailed, Uher, Werner, et al,
632 2013; Tibetan macaques, Pritchard et al., 2014), but the Activity$_{TR}$ dimension was not found

633 across all macaque rating studies. $Activity_{TR}$ was not correlated with "active" behavior, but
634 with "friendly behavior" and "peripheral" instead, leading to a strong correlation between
635 $Activity_{TR}$ and $Sociability_{BC}$. $Activity_{TR}$ may thus be mainly perceived as socially active and
636 spending a greater amount of time outside the group center, yet not necessarily alone.
637 However, there was no convergence between $Activity_{TR}$ and $Sociability_{BC}$, because the
638 correlation was only driven by one social behavior (out of six behaviors loading on
639 $Sociability_{BC}$); it was not correlated with "active" behavior, despite the high loadings of
640 "active" and "not lazy" on $Activity_{TR}$. Generally, rating and behavior Activity dimensions are
641 related to social behaviors in primates (Konečná et al., 2008; Tkaczynski et al., 2018; Uher,
642 Werner, et al., 2013; Vazire et al., 2007), but further analyses are needed to better understand
643 these links.

644 The inconsistencies identified in this descriptive comparison between our and previous
645 findings for Assamese macaques (Adams et al., 2015) could be due to sensitivity of the
646 personality structure to rather small sample sizes, leading to small observations-to-variables
647 ratios for the factor analyses (trait ratings with 107 individuals and 43 adjective items). If we
648 assume moderately plastic personality, the construct might change due to new animals being
649 rated in 2015 and 2016, and 19 of the 34 animals already rated in 2009 having matured into
650 adulthood (Dingemanse & Wolf, 2013). Inconsistencies may additionally be due to
651 differences in method; the original study (Adams et al., 2015) neither exclude unreliable
652 raters, nor items with low interrater reliability, while we exclude half of the items included in
653 the original $Openness_{TR}$ dimension which was found in four out of five species studied. Some
654 studies found similar Openness dimensions as well (e.g., Konečná et al. 2012; Uher, Werner,
655 et al., 2013), but others did not (e.g., Capitanio, 1999; Tkaczynski et al., 2018). Differences in
656 method may thus explain the main discrepancies between ours and the previous study on the

same population. Another replication study, using the same trait rating method on different populations of common marmosets also revealed to some extent differences in the personality structures, even adding a fifth dimension (Koski et al., 2017).

Comparison with other multi-method primate studies

In the following we provide a brief review of seven macaque and eleven other primate studies that employed a multi-method approaches (macaques: Capitanio 1999; Pritchard et al., 2014; Rouff et al., 2005; Sussman et al., 2013; Tkaczynski et al., 2018; Uher, Werner, et al., 2013; this study; other nonhuman primates: Carter, Marshall, Heinsohn, & Cowlishaw, 2012; Eckardt et al., 2015; Freeman et al., 2013; Garai et al., 2016; Iwanicki & Lehmann, 2015; Konečná et al., 2008; Morton et al., 2013; Pederson, King, & Landau, 2005; Uher & Asendorpf, 2008; Uher & Visalberghi, 2016; Vazire et al., 2007). To date, trait rating studies identified a unique macaque $Friendliness_{TR}$ dimension described with the same adjectives in all species, whereas other primate studies often define two discriminant social domains, mostly named $Extraversion_{TR}$ and $Agreeableness_{TR}$. In nearly all of the reviewed multi-method studies social rating domains were best validated, revealing positive correlations with social behaviors or behavior domains, generally resembling a mixed pattern of behaviors reflecting social integration (e.g., proximity) and bonding (e.g., grooming skew). $Opportunism_{TR}$ and $Dominance_{TR}$ related dimensions (or $Confidence_{TR}$ as in most macaques) were validated in half of the studies. $Dominance_{TR}$ and $Opportunism_{TR}$ are not seen as interchangeable, because $Dominance_{TR}$, described with dominant and confident adjectives, is indeed often correlated with dominant and aggressive behaviors. In contrast, $Opportunism_{TR}$ is correlated with dominance behavior only, if the two aspects are integrated in one Dominant-Competitive-Aggressive dimension. Most studies describe either $Opportunism_{TR}$

680 (macaques) or Dominance$_{TR}$ (other species) and one third made both domains part of the
681 personality profile. Other dimensions, as Excitability$_{TR+BC}$, Emotionality$_{TR+BC}$ or
682 Boldness$_{TR+BC}$ are less reliably measured and validated. Half of the studies in non-macaque
683 species, and one in macaques, define an Openness$_{TR+BC}$ domain, consistently loaded with
684 "curious", which in all cases correlated with "playful" behavior.

685 Disparities between methods measuring the same construct, as found in the majority of
686 studies, demand multi-method approaches to check for validity in every species under study
687 to further ensure reproducibility. We should not assume validity if only shown once, since
688 most personality dimensions have not been reliably validated in the nonhuman literature (but
689 see discussion for nun-human primates in Koski et al., 2017). Particularly studies on wild
690 primates with low sample sizes and limited observational data (partly due to lower behavior
691 frequencies compared to captivity) need to be replicated. Replication studies on the same
692 subjects, groups and populations could build up on each other enabling longer observation
693 periods, and hence larger sample sizes, which may allow including behaviors with low
694 frequencies. Eventually, male, female and juvenile behavior data could be integrated.
695 Subsequent studies could focus on a single dimension, which might require experimental
696 settings (e.g., boldness) which are difficult to conduct in the wild (e.g., playback experiments
697 in demanding habitats). Altogether, this may lead to more complex and fine grained
698 personality structures of nonhuman animals. In addition, mixed findings in personality
699 profiles between species also call for replication studies to understand the socio-ecological
700 relevance of similar behaviors in different species.

701 **A comprehensive personality structure for male Assamese macaques**

We propose two solutions for how to combine the 8 dimensions from trait rating and behavior coding. Congruence between $Friendliness_{TR}$ and $Gregariousness_{BC}$ as well as between $Activity_{TR}$ and $Sociability_{BC}$ make them largely redundant. Based on the considerations laid out above about the role of behavior frequency we also consider $Opportunism_{TR}$ and $Aggressiveness_{BC}$ to reflect a similar dimension of inter-individual variation in behavior. Despite the correlation between factor loadings on $Confidence_{TR}$ and $Vigilance_{BC}$ being the same as between $Opportunism_{TR}$ and $Aggressiveness_{BC}$ dimensions, we considered congruence to be low between the former. It follows that the full 5-factor structure is either built from all trait rating dimensions plus $Vigilance_{BC}$ or from all behavior coding dimensions plus $Confidence_{TR}$. Both solutions have their strengths and weaknesses.

We demonstrated the importance of the examination of rating performances of observers, as well as reliability analyses for item selection. However, even if a strong agreement among observers can ensure that trait ratings are not purely idiosyncratic interpretations, all observers may be biased in the same way (Freeman et al. 2011; Koski et al 2011a; Uher 2008, Uher & Asendorpf 2008). This aspect of trait ratings deserves continued attention. For instance, an individual which is mainly staying in the periphery of the group and spotted less often will more likely be rated as unsocial, although it is extremely social when joining the group. In statistical terms trait ratings have to be preferred over behavior coding results here because the former were much more robust.

Advantages of behavior coding include that behaviors are directly perceivable and measurable and therefore more objective (Freeman et al., 2011) so that more subtle variation in sociability can be detected. So far, trait rating studies in macaques identified a single Friendliness dimension (Adams et al., 2015), whereas behavior coding studies frequently

found two distinct prosocial personality domains (Capitanio, 1999; Neumann et al., 2013; Tkaczynski et al., 2018; this study). Researchers in animal and human personality studies stress that social personality dimensions can only be identified in the context of social relationships especially in species with complex societies (Cooper, 2002; Koski, 2011b; Reis, Collins, & Berscheid, 2000). This is highlighted by the fact that affinitive behavior loaded on a different factor than actual affiliative interaction variables because individuals may tolerate and be tolerated sometimes without implications for friendly contact behaviors.

One disadvantage of the behavior coding data is that measurement reliability for individual differences may be low for rare behaviors that may, however, bebut biologically very important such as support in agonistic conflicts. Observers may be better able to represent variation in rare behaviors in their ratings.

We conclude that two solutions with five factors best describe the personality structure of the male Assamese macaques studied here. One 5-factor structure is built from all trait rating dimensions plus Vigilance_{BC} and the other from all behavior coding dimensions plus Confidence_{TR}. Both solutions have their strengths and weaknesses. We encourage the use of integrative approaches including trait ratings, behavioral codings, and experiments embracing the Transdisciplinary philosophy-of-science paradigm (Uher, 2018), which aims to broaden the horizon in personality research in a transdisciplinary way.

REFERENCES

Adams, M., Bonaventura, M., Ostner, J., Schulke, O., De Marco, A., Thierry, B., … Weiss, A. (2015). Personality Structure and Social Style in Macaques. *Journal of Personality and Social Psychology*, *109*(2), 338–353. https://doi.org/10.1037/pspp0000041

Altmann, J. (1974). Observational study of behavior: sampling methods. *Behaviour*, *49*, 227–267. https://doi.org/http://dx.doi.org/10.1163/156853974X00534.

Borries, C., Larney, E., Kreetiyutanont, K., & Koenig, A. (2002). The diurnal primate community in a dry evergreen forest in Phu Khieo Wildlife Sanctuary, Northeast Thailand, *50*(1), 75–88.

Brent, L. J. N., Semple, S., MacLarnon, A., Ruiz-Lambides, A., Gonzalez-Martinez, J., & Platt, M. L. (2014). Personality Traits in Rhesus Macaques (Macaca mulatta) Are Heritable but Do Not Predict Reproductive Output. *International Journal of Primatology*, *35*(1), 188–209. https://doi.org/10.1007/s10764-013-9724-6

Brommer, J. E., & Class, S. A. (2017). Personality from the Perspective of Behavioral Ecology. In J. Vonk, A. Weiss, & S. A. Kuczaj (Eds.), *Personality in Nonhuman Animals* (pp. 73–107). New York, NY: Springer, Cham.

Capitanio, J. P. (1999). Personality dimensions in adult male rhesus macaques: Prediction of behaviors across time and situation. *American Journal of Primatology*, *47*(4), 299–320. https://doi.org/10.1002/(sici)1098-2345(1999)47:4<299::aid-ajp3>3.0.co;2-p

Carere, C., & Maestripieri, D. (2013). *Animal Personalities: Behavior, Physiology, and Evolution*. Chicago: The University of Chicago Press.

Carter, A. J., Marshall, H. H., Heinsohn, R., & Cowlishaw, G. (2012). Evaluating animal personalities: do observer assessments and experimental tests measure the same thing? *Behavioral Ecology and Sociobiology*, *66*(1), 153–160. https://doi.org/10.1007/s00265-011-1263-6

768 Cattell, R. B. (1966). The scree test for the number of factors. *Multivariate Behavioral Research*, *1*(2),
769 245–276. https://doi.org/10.1207/s15327906mbr0102_10
770 Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and
771 standardized assessment instruments in psychology. *Psychological Assessment*, *6*(4).
772 doi:1040-3590/94/53.00
773 Cooper, M. L. (2002). Personality and Close Relationships: Embedding People in Important Social
774 Contexts. *Journal of Personality*, *70*(6), 757–782. https://doi.org/10.1111/1467-6494.05023
775 Costello, A. B., & Osborne, J. W. (2005). Best Practices in Exploratory Factor Analysis: Four
776 Recommendations for Getting the Most From Your Analysis. *Practical Assessment, Research*
777 *& Evaluation*, *10*(7), 1–9. http://pareonline.net/getvn.asp?v=10&n=7
778 Curley, J. P. (2016). *compete: Analyzing Social Hierarchies*. Retrieved from https://CRAN.R-
779 project.org/package=compete
780 Dammhahn, M. (2012). Are personality differences in a small iteroparous mammal maintained by a
781 life-history trade-off? *Proceedings of the Royal Society B: Biological Sciences*, 1-7.
782 https://doi.org/10.1098/rspb.2012.0212
783 Deary, I. J., Weiss, A., & Batty, G. D. (2010). Intelligence and Personality as Predictors of Illness and
784 Death: How Researchers in Differential Psychology and Chronic Disease Epidemiology Are
785 Collaborating to Understand and Address Health Inequalities. *Psychological Science in the*
786 *Public Interest*, *11*(2), 53–79. https://doi.org/10.1177/1529100610387081
787 de Winter, J. C. F., Dodou, D., & Wieringa, P. A. (2009). Exploratory Factor Analysis With Small
788 Sample Sizes. *Multivariate Behavioral Research*, *44*(2), 147–181.
789 https://doi.org/10.1080/00273170902794206
790 Digman, J. M. (1990). Personality Structure: Emergence of the Five-Factor Model. *Annual Review of*
791 *Psychology*, *41*(1), 417–440. https://doi.org/10.1146/annurev.ps.41.020190.002221

792 Dingemanse, N. J., Kazem, A. J. N., Réale, D., & Wright, J. (2010). Behavioural reaction norms: animal
793 personality meets individual plasticity. *Trends in Ecology & Evolution*, *25*(2), 81–89.
794 https://doi.org/10.1016/j.tree.2009.07.013
795 Dingemanse, N. J., & Wolf, M. (2013). Between-individual differences in behavioural plasticity within
796 populations: causes and consequences. *Animal Behaviour*, *85*(5), 1031–1039.
797 https://doi.org/10.1016/j.anbehav.2012.12.032
798 Dinno, A. (2012). *paran: Horn's Test of Principal Components/Factors*. Retrieved from
799 https://CRAN.R-project.org/package=paran
800 Eckardt, W., Steklis, H. D., Steklis, N. G., Fletcher, A. W., Stoinski, T. S., & Weiss, A. (2015). Personality
801 dimensions and their behavioral correlates in wild Virunga mountain gorillas (Gorilla beringei
802 beringei). *Journal of Comparative Psychology*, *129*(1), 26–41.
803 https://doi.org/10.1037/a0038370
804 Field, A. (2000). *Discovering statistics using SPSS for Windows: Advanced techniques for beginners*
805 *(Introducing Statistical Methods series)*. Sage publications.
806 Field, A., Miles, J., & Field, Z. (2012). *Discovering statistics using R*. Sage publications.
807 Freeman, H. D., Brosnan, S. F., Hopper, L. M., Lambeth, S. P., Schapiro, S. J., & Gosling, S. D. (2013).
808 Developing a comprehensive and comparative questionnaire for measuring personality in
809 chimpanzees using a simultaneous top-down/bottom-up design. *American Journal of*
810 *Primatology*, *75*(10), 1042–1053. https://doi.org/10.1002/ajp.22168
811 Freeman, H. D., & Gosling, S. D. (2010). Personality in Nonhuman Primates: A Review and Evaluation
812 of Past Research. *American Journal of Primatology*, *72*(8), 653–671.
813 https://doi.org/10.1002/ajp.20833
814 Freeman, H. D., Gosling, S. D., & Schapiro, S. J. (2011). Comparison of Methods for Assessing
815 Personality in Nonhuman Primates. In A. Weiss, J. E. King, & L. Murray (Eds.), *Personality and*
816 *Temperament in Nonhuman Primates* (pp. 17–40). New York, NY: Springer.

Garai, C., Weiss, A., Arnaud, C., & Furuichi, T. (2016a). Personality in wild bonobos (Pan paniscus). *American Journal of Primatology*, *78*(11), 1178–1189. https://doi.org/10.1002/ajp.22573

Gosling, S. D. (2001). From mice to men: What can we learn about personality from animal research? *Psychological Bulletin*, *127*(1), 45–86. https://doi.org/10.1037//0033-2909.127.1.45

Haunhorst, C.B., Schülke, O., Ostner, J. (2016). Opposite-sex social bonding in wild Assamese macaques *American Journal of Primatology 78,* 872-883 doi:10.1002/ajp.22554

Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika*, *30*(2), 179–185. https://doi.org/10.1007/BF02289447

Iwanicki, S., & Lehmann, J. (2015). Behavioral and trait rating assessments of personality in common marmosets (Callithrix jacchus). *Journal of Comparative Psychology*, *129*(3), 205–217. https://doi.org/10.1037/a0039318

John, O. P., & Soto, C. J. (2007). The importance of being valid: Reliability and the process of construct validation. In R. W. Robins, C. R. Fraley, & R. F. Krueger (Eds.), *Handbook of research methods in personality psychology* (pp. 461–494). New York, NY: Guilford Press.

John, O. P., & Srivastava, S. (1999). The Big Five trait taxonomy: History, measurement, and theoretical perspectives. In L. A. Pervin & O. P. John (Eds.), *Handbook of personality: Theory and research* (Vol. 2, pp. 102–138). New York, NY: Guilford Press.

Kaiser, H. F. (1960). The Application of Electronic Computers to Factor Analysis. *Educational and Psychological Measurement*, *20*(1), 141–151. https://doi.org/10.1177/001316446002000116

Kalbitz, J., Schülke, O., & Ostner, J. (2017). Triadic male-infant-male interaction serves in bond maintenance in male Assamese macaques. *PLOS ONE*, *12*(10), e0183981. https://doi.org/10.1371/journal.pone.0183981

King, J. E., & Figueredo, A. J. (1997). The Five-Factor Model plus Dominance in Chimpanzee Personality. *Journal of Research in Personality*, *31*(2), 257–271. https://doi.org/10.1006/jrpe.1997.2179

842 Konečná, M., Lhota, S., Weiss, A., Urbánek, T., Adamová, T., & Pluháček, J. (2008). Personality in free-
843 ranging Hanuman langur (Semnopithecus entellus) males: Subjective ratings and recorded
844 behavior. *Journal of Comparative Psychology*, *122*(4), 379–389.
845 http://dx.doi.org/10.1037/a0012625
846 Konečná, M., Weiss, A., Lhota, S., & Wallner, B. (2012). Personality in Barbary macaques (Macaca
847 sylvanus): Temporal stability and social rank. *Journal of Research in Personality*, *46*(5), 581–
848 590. https://doi.org/10.1016/j.jrp.2012.06.004
849 Koski, S. E. (2011a). How to Measure Animal Personality and Why Does It Matter? Integrating the
850 Psychological and Biological Approaches to Animal Personality. In M. Inoue-Murayama, S.
851 Kawamura, & A. Weiss (Eds.), *From Genes to Animal Behavior* (pp. 115–136). Tokyo:
852 Springer, Japan.
853 Koski, S. E. (2011b). Social personality traits in chimpanzees: temporal stability and structure of
854 behaviourally assessed personality traits in three captive populations. *Behavioral Ecology*
855 *and Sociobiology*, *65*(11), 2161–2174. https://doi.org/10.1007/s00265-011-1224-0
856 Koski, S. E. (2014). Broader horizons for animal personality research. *Ecology and Evolution*, *2*(70), 1–
857 6. https://doi.org/10.3389/fevo.2014.00070
858 Koski, S. E., Buchanan-Smith, H. M., Ash, H., Burkart, J. M., Bugnyar, T., & Weiss, A. (2017). Common
859 marmoset (Callithrix jacchus) personality. *Journal of Comparative Psychology*, *131*(4), 326-
860 336. http://dx.doi.org/10.1037/com0000089
861 Lilley, M. K., Kuczaj, S. A., & Yeater, D. B. (2017). Individual Differences in Nonhuman Animals:
862 Examining Boredom, Curiosity, and Creativity. In J. Vonk, A. Weiss, & S. A. Kuczaj (Eds.),
863 *Personality in Nonhuman Animals* (pp. 257–275). New York, NY: Springer, Cham.
864 Little T. D., Cunningham W. A., Shahar, G., & Widaman K. F. (2002). To parcel or not to parcel:
865 exploring the question, weighing the merits. *Structural Equation Modeling 9(2)*, 151-173.

Lorenzo-Seva, U., & ten Berge, J. M. F. (2006). Tucker's congrunece coefficient as a meaningful index of factor similarity. *Methodology 2(2)*, 57-64.

Maier, M. J. (2015). *REdaS: Companion Package to the Book "R: Einführung durch angewandte Statistik."* Retrieved from https://CRAN.R-project.org/package=REdaS

Mair, P., Schoenbrodt, F., & Wilcox. (2017). *WRS2: Wilcox robust estimation and testing*. Retrieved from https://CRAN.R-project.org/package=WRS2

Manson, J. H., & Perry, S. (2013). Personality Structure, Sex Differences, and Temporal Change and Stability in Wild White-Faced Capuchins (Cebus capucinus). *Journal of Comparative Psychology*, *127*(3), 299–311. https://doi.org/10.1037/a0031316

Martin, P., & Bateson, P. (1993). *Measuring Behaviour*. Cambridge: Cambridge University Press.

Massen, J. J. M., Antonides, A., Arnold, A.-M. 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*(9), 947–958. https://doi.org/10.1002/ajp.22159

Massen, J. J. M., & Koski, S. E. (2014). Chimps of a feather sit together: chimpanzee friendships are based on homophily in personality. *Evolution and Human Behavior*, *35*(1), 1–8. https://doi.org/10.1016/j.evolhumbehav.2013.08.008

Mehta, P. H., & Gosling, S. D. (2008). Bridging human and animal research: A comparative approach to studies of personality and health. *Brain, Behavior, and Immunity*, *22*(5), 651–661. https://doi.org/10.1016/j.bbi.2008.01.008

Minge, C., Berghänel, A., Schülke, O., Ostner, J. (2016). Patterns and consequences of male–infant relationships in wild assamese macaques (*Macaca assamensis*). *International Journal of Primatology 37,* 350-370 doi:10.1007/s10764-016-9904-2

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*(3), 282–298. https://doi.org/10.1037/a0031723

Neumann, C., Agil, M., Widdig, A., Engelhardt, A., & Kalueff, A. (2013). Personality of Wild Male Crested Macaques (Macaca nigra). *PLoS ONE*, *8*(8), e69383. https://doi.org/10.1371/journal.pone.0069383

Ostner, J., Heistermann, M., & Schülke, O. (2008). Dominance, aggression and physiological stress in wild male Assamese macaques (Macaca assamensis). *Hormones and Behavior*, *54*(5), 613–619. https://doi.org/10.1016/j.yhbeh.2008.05.020

Ostner, J., & Schülke, O. (2014). The evolution of social bonds in primate males. *Behaviour, 151*(7), 871 – 906. https://doi.org/10.1163/1568539X-00003191

Paul, A., Kuester, J., & Arnemann, J. (1996). The sociobiology of male-infant interactions in Barbary macaques, Macaca sylvanus. *Animal Behaviour*, *51*(1), 155–170. https://doi.org/10.1006/anbe.1996.0013

Pederson, A. K., King, J. E., & Landau, V. I. (2005). Chimpanzee (Pan troglodytes) personality predicts behavior. *Journal of Research in Personality*, *39*(5), 534–549. https://doi.org/10.1016/j.jrp.2004.07.002

Penke, L., & Jokela, M. (2016). The evolutionary genetics of personality revisited. *Evolutionary Psychology*, *7*, 104–109. https://doi.org/10.1016/j.copsyc.2015.08.021

Preacher, K. J., & MacCallum, R. C. (2003). Repairing Tom Swift's Electric Factor Analysis Machine. *Understanding Statistics*, *2*(1), 13–43. https://doi.org/10.1207/S15328031US0201_02

Pritchard, A. J., Sheeran, L. K., Gabriel, K. I., Li, J.-H., & Wagner, R. S. (2014). Behaviors that predict personality components in adult free-ranging Tibetan macaques Macaca thibetana. *Current Zoology*, *60*(3), 362–372. https://doi.org/10.1093/czoolo/60.3.362

915 Raiche, G. (2010). *nFactors: an R package for parallel analysis and non graphical solutions to the*
916 *Cattell scree test*. Retrieved from https://CRAN.R-project.org/package=nFactors
917 Réale, D., Reader, S. M., Sol, D., McDougall, P. T., & Dingemanse, N. J. (2007). Integrating animal
918 temperament within ecology and evolution. *Biological Reviews*, *82*(2), 291–318.
919 https://doi.org/10.1111/j.1469-185X.2007.00010.x
920 Reis, H. T., Collins, W. A., & Berscheid, E. (2000). The relationship context of human behavior and
921 development. *Psychological Bulletin*, *126*(6), 844–872. 10.1037//0033-2909.126.6.844
922 Revelle, W. (2018). *psych: Procedures for Psychological, Psychometric, and Personality Research*.
923 Retrieved from https://CRAN.R-project.org/package=psych
924 Revelle, W., & Rocklin, T. (1979). Very simple structure: An alternative procedure for estimating the
925 optimal number of interpretable factors. *Multivariate Behavioral Research*, *14*(4), 403–414.
926 https://doi.org/10.1207/s15327906mbr1404_2
927 Rouff, J. H., Sussman, R. W., & Strube, M. J. (2005). Personality traits in captive lion-tailed macaques
928 (Macaca silenus). *American Journal of Primatology*, *67*(2), 177–198.
929 https://doi.org/10.1002/ajp.20176
930 Schülke, O., Bhagavatula, J., Vigilant, L., Ostner, J. (2010). Social bonds enhance reproductive success
931 in male macaques. *Current Biology 20*, 2207-2210 doi:10.1016/j.cub.2010.10.058
932 Seyfarth, R. M., & Cheney, D. L. (2013). Affiliation, empathy, and the origins of theory of mind.
933 *Proceedings of the National Academy of Sciences of the United States of America*, *110*,
934 10349–10356. https://doi.org/10.1073/pnas.1301223110
935 Seyfarth, R. M., Silk, J. B., & Cheney, D. L. (2012). Variation in personality and fitness in wild female
936 baboons. *Proceedings of the National Academy of Sciences*, *109*(42), 16980–16985.
937 https://doi.org/10.1073/pnas.1210780109
938 Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability.
939 *Psychological Bulletin*, *86*, 420–428. http://dx.doi.org/10.1037/0033-2909.86.2.420

940 Silk, J. B., Beehner, J. C., Bergman, T. J., Crockford, C., Engh, A. L., Moscovice, L. R., … Cheney, D. L.
941 (2010). Female chacma baboons form strong, equitable, and enduring social bonds.
942 *Behavioral Ecology and Sociobiology*, *64*(11), 1733–1747.
943 https://doi.org/10.1007/s00265-010-0986-0
944 Sussman, A. F., Ha, J. C., Bentson, K. L., & Crockett, C. M. (2013). Temperament in Rhesus, Long-
945 Tailed, and Pigtailed Macaques Varies by Species and Sex. *American Journal of Primatology*,
946 *75*(4), 303–313. https://doi.org/10.1002/ajp.22104
947 Thierry, B., Singh, M., & Kaumanns, W. (2004). *Macaques Societies - A model for the Study of Social*
948 *Organization*. Cambridge: Cambridge University Press.
949 Tkaczynski, P., Ross, C., MacLarnon, A.M., Mouna, M., Majolo, B., Lehmann, J. (2018) Measuring
950 personality in the field: an in situ comparison of personality quantification methods in wild
951 Barbary Macaques (*Macaca sylvanus*). *Journal of Comparative Psychology,*
952 http://dx.doi.org/10.1037/com0000163
953 Uher, J. (2008). Comparative personality research: Methodological approaches. *European Journal of*
954 *Personality*, *22*(5), 427–455. https://doi.org/10.1002/per.680
955 Uher, J. (2008). Three Methodological Core Issues of Comparative Personality Research. *European*
956 *Journal of Personality*, *22*, 475–496
957 Uher, J. (2011). Personality in Nonhuman Primates: What Can We Learn from Human Personality
958 Psychology? In A. Weiss, J. E. King, & L. Murray (Eds.), *Personality and Temperament in*
959 *Nonhuman Primates* (pp. 41–76). New York, NY: Springer.
960 Uher J (2015). Conceiving "personality": Psychologist's challenges and basic fundamentals of the
961 Transdisciplinary Philosophy-of-Science Paradigm for Research on Individuals Integrative
962 *Psychological and Behavioral Science* 49, 398-458 doi:10.1007/s12124-014-9283-1
963 Uher J, Visalberghi E (2016). Observations versus assessments of personality: A five-method multi-
964 species study reveals numerous biases in ratings and methodological limitations of

965 standardised assessments *Journal of Research in Personality 61*, 61-79
966 doi:https://doi.org/10.1016/j.jrp.2016.02.003
967 Uher, J. (2018). Taxonomic models of individual differences: a guide to transdisciplinary approaches.
968 *Philosophical Transactions of the Royal Society B: Biological Sciences*, *373*(1744).
969 https://doi.org/10.1098/rstb.2017.0171
970 Uher, J., Addessi, E., & Visalberghi, E. (2013). Contextualised behavioural measurements of
971 personality differences obtained in behavioural tests and social observations in adult
972 capuchin monkeys (Cebus apella). *Journal of Research in Personality*, *47*(4), 427–444.
973 https://doi.org/10.1016/j.jrp.2013.01.013
974 Uher, J., & Asendorpf, J. B. (2008). Personality assessment in the Great Apes: Comparing ecologically
975 valid behavior measures, behavior ratings, and adjective ratings. *Journal of Research in*
976 *Personality*, *42*(4), 821–838. https://doi.org/10.1016/j.jrp.2007.10.004
977 Uher, J., Asendorpf, J.B., & Call, J. (2008) Personality in the behaviour of great apes: temporal
978 stability, cross-situational consistency and coherence in response. *Animal Behaviour, 75*, 99-
979 112
980 Uher, J., & Visalberghi, E. (2016). Observations versus assessments of personality: A five-method
981 multi-species study reveals numerous biases in ratings and methodological limitations of
982 standardised assessments. *Journal of Research in Personality*, *61*, 61–79.
983 https://doi.org/10.1016/j.jrp.2016.02.003
984 Uher, J., Werner, C. S., & Gosselt, K. (2013). From observations of individual behaviour to social
985 representations of personality: Developmental pathways, attribution biases, and limitations
986 of questionnaire methods. *Journal of Research in Personality*, *47*(5), 647–667.
987 https://doi.org/10.1016/j.jrp.2013.03.006

van Aken, M. A. G., & Asendorpf, J. B. (2018). Personality and peer relationships. In K. H. Rubin, W. M. Bukowski, & B. Laursen (Eds.), *Handbook of peer interactions, relationships, and groups* (pp. 159–176). New York, NY: Guilford Press.

Vazire, S., Gosling, S. D., Dickey, A. S., & Schapiro, S. J. (2007). *Measuring Personality in Nonhuman Animals*. New York, NY: Guilford Press.

Velicer, W. F. (1976). Determining the number of components from the matrix of partial correlations. *Psychometrika*, *41*(3), 321–327. https://doi.org/10.1007/BF02293557

von Borell, C., Kulik, L., & Widdig, A. (2016). Growing into the self: the development of personality in rhesus macaques. *Animal Behaviour*, *122*, 183–195. http://dx.doi.org/10.1016/j.anbehav.2016.10.013

Weiss, A. (2017). Exploring factor space (and other adventures) with the Hominoid Personality Questionnaire. In J. Vonk, A. Weiss, & S. A. Kuczaj (Eds.), *Personality in Nonhuman Animals* (pp. 73–107). New York, NY: Springer

Weiss, A., Adams, M. J., Widdig, A., & Gerald, M. S. (2011). Rhesus Macaques (Macaca mulatta) as Living Fossils of Hominoid Personality and Subjective Well-Being. *Journal of Comparative Psychology*, *125*(1), 72–83. https://doi.org/10.1037/a0021187

Weiss, A., Inoue-Murayama, M., Hong, K.-W., Inoue, E., Udono, T., Ochiai, T., … King, J. E. (2009). Assessing chimpanzee personality and subjective well-being in Japan. *American Journal of Primatology*, *71*(4), 283–292. https://doi.org/10.1002/ajp.20649

Widdig, A., Nürnberg, P., Krawczak, M., Streich, W. J., & Bercovitch, F. (2002). Affiliation and aggression among adult female rhesus macaques: a genetic analysis of paternal cohorts. *Behaviour*, *139*(2), 371–391. https://doi.org/10.1163/156853902760102717

Wilcox, R. R. (1994). The percentage bend correlation coefficient. *Psychometrika*, *59*(4), 601–616. https://doi.org/10.1007/BF02294395

1012 Wolf, M., & Weissing, F. J. (2012). Animal personalities: consequences for ecology and evolution.
1013 *Trends in Ecology & Evolution*, *27*(8), 452–461. https://doi.org/10.1016/j.tree.2012.05.001
1014
1015

1016 *Table 1: Personality structure after promax rotation with factor loadings and item*

1017 *communalities (h^2) derived from trait ratings (TR).*

Item	Opportunism$_{TR}$	Confidence$_{TR}$	Activity$_{TR}$	Friendliness$_{TR}$	h^2
bullying	**0.96**	0.21	-0.12	-0.02	0.92
aggressive	**0.93**	0.28	-0.1	-0.01	0.92
irritable	**0.93**	0	0.01	-0.09	0.88
greedy	**0.92**	0.21	-0.38	0.15	0.77
jealous	**0.92**	0.01	-0.14	0.10	0.75
manipulative	**0.87**	0.24	-0.17	0.09	0.77
defiant	**0.81**	0.09	0.12	0.02	0.81
excitable	**0.78**	-0.31	0.15	-0.05	0.77
impulsive	**0.72**	0.12	0.27	-0.01	0.82
erratic	**0.71**	-0.17	0.25	-0.21	0.76
distractable	**0.55**	**(-0.49)**	0.16	-0.05	0.58
persistent	**0.49**	**(0.44)**	0.20	0	0.63
cool	**-0.46**	**(0.42)**	-0.36	0.13	0.63
unemotional	**-0.62**	0.12	-0.39	-0.03	0.76
stable	**-0.66**	0.16	-0.36	0.15	0.79
gentle	**-0.77**	-0.17	-0.06	0.41	0.80
dominant	0.30	**0.82**	-0.19	0.05	0.81
decisive	0.20	**0.79**	-0.16	0.03	0.70
independent	-0.07	**0.68**	-0.32	**(-0.42)**	0.65
protective	0.17	**0.61**	-0.21	**(0.47)**	0.75
intelligent	0.10	**0.60**	-0.13	0.05	0.39
individualistic	0	**0.53**	-0.08	**(-0.49)**	0.41
clumsy	0.36	**-0.53**	-0.52	0.22	0.44
quitting	0.18	**-0.71**	-0.37	0.01	0.60
anxious	0.18	**-0.74**	0.05	-0.21	0.66
timid	-0.24	**-0.78**	0.11	-0.15	0.77
fearful	0.25	**-0.83**	-0.04	0.01	0.68
submissive	-0.23	**-0.86**	-0.04	0.10	0.82

vulnerable	-0.11	**-0.88**	-0.1	0.03	0.83
follower	0.03	**-0.91**	0.01	0.30	0.78
playful	-0.17	-0.17	**0.84**	0.09	0.65
active	0.09	0.01	**0.79**	-0.06	0.68
curious	0.32	-0.06	**0.61**	0.15	0.73
inquisitive	0.31	-0.04	**0.61**	0.08	0.68
depressed	-0.07	**(-0.44)**	**-0.48**	-0.29	0.71
cautious	0.19	**(-0.42)**	**-0.58**	-0.09	0.50
lazy	-0.16	0.13	**-0.73**	-0.03	0.68
affectionate	-0.10	-0.10	0.01	**0.91**	0.81
sympathetic	-0.31	-0.11	-0.02	**0.81**	0.71
friendly	**(-0.46)**	-0.14	-0.05	**0.77**	0.77
sociable	0.22	0.11	0.14	**0.76**	0.79
helpful	0.12	0.30	0.16	**0.66**	0.74
solitary	-0.15	-0.24	-0.22	**-0.64**	0.72
Variance explained	27%	22%	11%	12%	

Note. Salient factor loadings ≥ |0.4| are shown in bold.

Table 2: *Rank-order stability of personality dimensions from trait ratings (TR) in 2015 and 2016 with factor congruence of factor loadings (first number in cell) and Pearson correlations of extracted factor scores (second number in cell).*

	2016			
2015	**Opportunism$_{TR}$**	**Confidence$_{TR}$**	**Friendliness$_{TR}$**	**Activity$_{TR}$**
Opportunism$_{TR}$	**0.97 / 0.87**	-0.04 / -0.20	-0.03 / -0.04	-0.18 / **-0.39**
Confidence$_{TR}$	-0.03 / -0.12	**0.98 / 0.79**	-0.17 / **-0.29**	0.05 / 0.07
Friendliness$_{TR}$	-0.02 / -0.07	-0.11 / -0.17	**0.94 / 0.72**	-0.02 / **-0.29**
Activity$_{TR}$	-0.20 / **-0.46**	0.01 / 0.14	-0.01 / -0.13	0.91 / **0.84**

Note. $N_{adjectives} = 41$. Values with 90% confidence intervals excluding zero, shown in bold. $N_{individuals} = 74$. Significant values after correcting for fdr shown in bold.

Table 3: *Personality structure derived from behavior codings (BC) after oblimin rotation with factor loadings and item communalities (h^2).*

Variable	**Gregariousness$_{BC}$**	**Aggressiveness$_{BC}$**	**Sociability$_{BC}$**	**Vigilance$_{BC}$**	h^2
neighbor diversity	**0.91**	0.08	-0.14	-0.06	0.95
tolerance	**0.88**	0.02	0.13	0.12	0.78
friendly approach	**0.84**	0.10	-0.10	0.01	0.80
active	**-0.52**	0.08	-0.04	**(-0.46)**	0.46
alone	**-0.99**	0.06	-0.05	0.06	0.94
quitting	0.04	**0.85**	0.10	-0.12	0.77
mild aggression	-0.14	**0.73**	0.13	0.16	0.49
agonistic involvement	-0.02	**0.67**	0.03	-0.08	0.45
physical aggression	0.27	**0.60**	-0.14	0.16	0.61
leaving	0.21	**0.48**	-0.24	0.01	0.43
friendly behavior	-0.01	0.14	**0.99**	-0.02	1.00
contact diversity	0.03	-0.21	**0.56**	0.11	0.34
peripheral	-0.34	-0.24	**0.49**	-0.06	0.55
initiation	-0.17	-0.05	**0.42**	-0.28	0.38
contact time	0.38	-0.19	**0.42**	-0.10	0.30
self-directed	-0.35	0.01	**-0.44**	0.24	0.38
vigilant	-0.03	0.00	-0.03	**0.99**	1.00
display	0.39	0.38	-0.06	-0.24	0.46
Variance explained	25%	15%	13%	9%	

Table 4: *Rank-order stability of personality dimensions from behavior codings (BC) in 04/14–03/15 and 04/15–03/16 with factor congruence of loadings (first number in cell) and Pearson correlation of extracted factor scores (second number in cell).*

	2015–2016		
2014–2015	**Gregariousness$_{BC}$**	**Sociability$_{BC}$**	**Quitting$_{BC}$[a]**
Gregariousness$_{BC}$	**0.92 / 0.57**	0.01 / -0.27	0.38 / 0.04
Sociability$_{BC}$	-0.01 / 0.01	0.78 / *0.40*	0.02 / 0.25
Mild aggression$_{BC}$[a]	0.18 / -0.01	-0.25 / -0.24	0.37 / 0.24

Note. $N_{behaviors} = 18$. Significant value, i.e. 95% confidence interval excluding zero, shown in bold. $N_{individuals} = 22$. Significant values after controlling for fdr shown in bold and trends in italics.

[a]3rd factor named after behavior variables with highest loading on that factor, resembling, but not equivalent to, Aggressiveness$_{BC}$ in the 4-factor solution of the aggregated data.

Table 5: *Construct validity as Pearson correlations of individual factor scores derived from trait ratings (TR) and behavior codings (BC).*

	Behavior Coding			
Trait Rating	**Gregariousness$_{BC}$**	**Aggressiveness$_{BC}$**	**Sociability$_{BC}$**	**Vigilance$_{BC}$**
Opportunism$_{TR}$	-0.26	*0.36*	0.06	-0.08
Confidence$_{TR}$	0.30	0.29	0.07	*-0.36*
Friendliness$_{TR}$	**0.69**	*0.35*	**-0.41**	0.20
Activity$_{TR}$	-0.13	0.14	**0.63**	-0.05

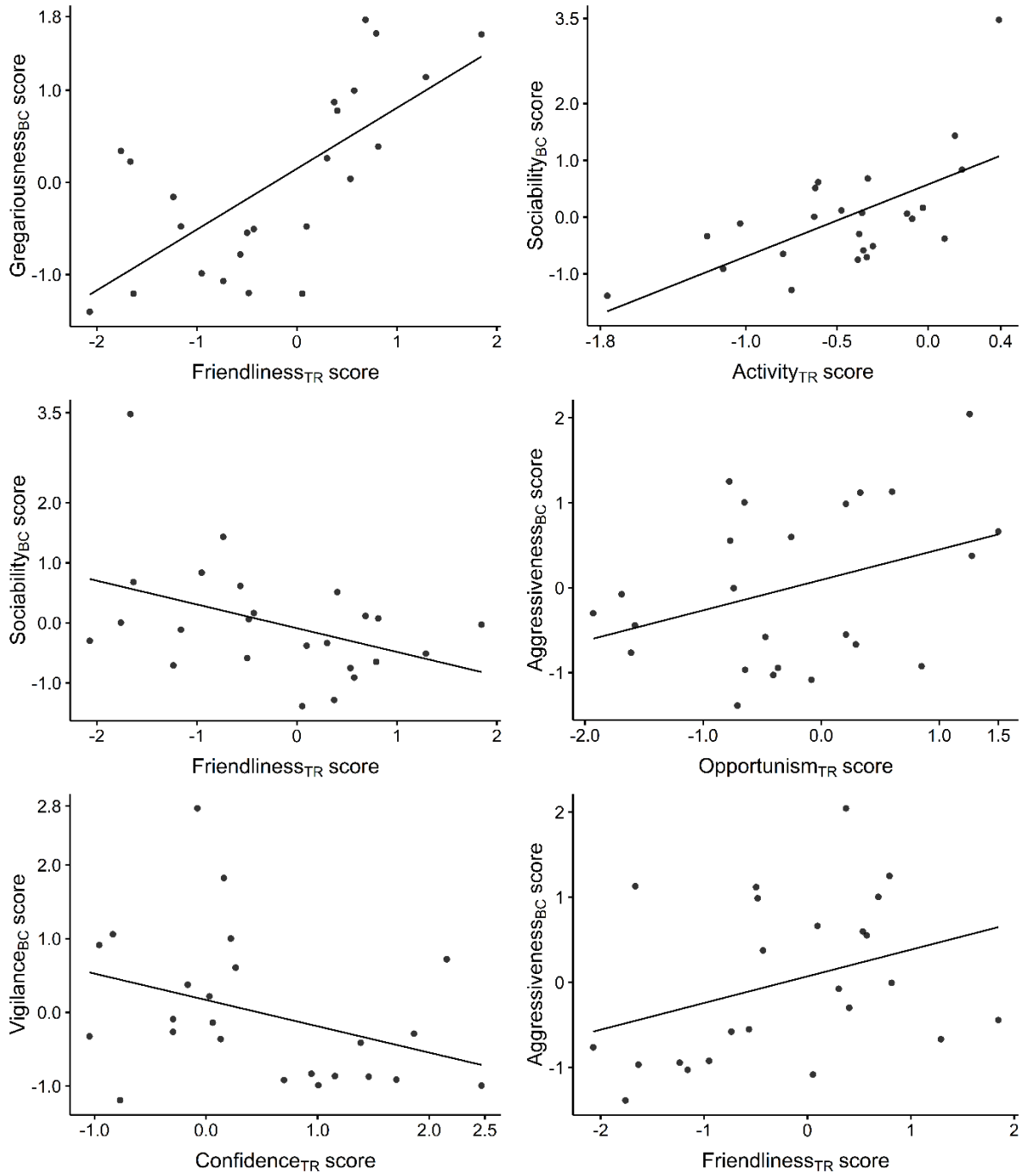
Note. Bold type significant after controlling for fdr, italic type statistical trend.

1043 Table 6: *Correlation coefficients (ρ_{pb}) of behavior variables with personality factor scores from trait ratings (TR).*

	Personality dimension$_{TR}$							
Behavior variable	**Opportunism$_{TR}$**	***p***	**Confidence$_{TR}$**	***p***	**Activity$_{TR}$**	***p***	**Friendliness$_{TR}$**	***p***
active	0.34	0.11	0.02	0.92	0.13	0.06	**-0.44**	0.03
agonistic involvement	0.07	0.75	0.32	0.13	-0.15	0.48	0.14	0.52
alone	0.32	0.13	-0.34	0.10	0.24	0.27	**-0.66**	0.00
contact diversity	-0.21	0.32	0.01	0.96	0.10	0.63	-0.40	0.05
contact time	-0.40	0.06	0.29	0.17	-0.01	0.95	-0.03	0.88
display	-0.23	0.28	**0.59**	0.00	-0.18	0.40	**0.44**	0.03
friendly approach	-0.02	0.92	0.08	0.71	-0.25	0.24	**0.60**	0.00
friendly behavior	0.06	0.79	0.01	0.96	**0.55**	0.01	-0.34	0.11
initiation	0.27	0.21	-0.24	0.25	0.19	0.37	-0.34	0.11
leaving	0.15	0.49	0.22	0.31	0.14	0.52	**0.63**	0.00
mild aggression	**0.41**	0.05	-0.11	0.61	0.08	0.72	0.10	0.64
neighbor diversity	-0.20	0.35	0.22	0.29	-0.33	0.12	**0.65**	0.00
physical aggression	0.05	0.82	0.17	0.43	-0.19	0.39	**0.50**	0.01
quitting	0.34	0.10	0.23	0.28	0.17	0.42	0.32	0.13
self-directed	0.09	0.68	**-0.56**	0.01	-0.20	0.36	-0.04	0.85
peripheral	0.15	0.48	-0.10	0.66	0.43	0.04	**-0.45**	0.03

tolerance	-0.34	0.11	0.15	0.49	-0.32	0.13	**0.60**	0.00
vigilant	-0.04	0.85	**-0.43**	0.04	-0.13	0.55	0.06	0.79

Note. Significant values after fdr correction shown in bold.



1046

1047 *Figure 1*. Congruence in factor scores of 24 males between constructs from behavior coding

1048 BC and trait rating TR. For statistics see table 5.

1049