Chimpanzees with positive welfare are happier, extraverted, and emotionally stable

Chimpanzees with Positive Welfare Are Happier, Extraverted, and Emotionally Stable

Running header: Chimpanzee Welfare and Happiness

L. M. Robinson[1,2,3]*, D. M. Altschul[1, 3]*, E. K. Wallace[3,4], Y. Úbeda[5,6], M. Llorente[5,7], Z. Machanda[8,9], K. E. Slocombe[3,4], M. C. Leach[10], N. K. Waran[2,11], and A. Weiss[1,3]

[1] Department of Psychology, School of Philosophy, Psychology and Language Sciences, University of Edinburgh, Edinburgh, UK, EH8 9JZ

[2] Jeanne Marchig International Centre for Animal Welfare Education, The University of Edinburgh, Roslin, UK, EH25 9RG

[3] Scottish Primate Research Group, UK

[4] Department of Psychology, University of York, York, UK, YO10 5DD

[5] Unitat de Recerca i Etologia, Fundació Mona, Riudellots de la Selva (Girona), Spain

[6] Facultat d'Educació i Psicologia, Universitat de Girona, Girona, Spain.

[7] IPRIM, Institut de Recerca i Estudis en Primatologia, Spain.

[8] Department of Human Evolutionary Biology, Harvard University, Cambridge, MA, USA, 02138

[9] Department of Anthropology, Tufts University, Medford, MA, USA, 02155

[10]School of Agriculture, Food & Rural Development, Newcastle University, Newcastle upon Tyne, UK, NE1 7RU

[11] Faculty of Education, Humanities and Health Science, Eastern Institute of Technology, Hawke's Bay, New Zealand.

* Authors contributed equally to this work.

Corresponding author: Lauren M. Robinson. Department of Psychology, School of Philosophy, Psychology and Language Sciences, The University of Edinburgh, 7 George Square, Edinburgh EH8 9JZ, United Kingdom. E-mail: l.robinson@ed.ac.uk

Abstract

Facilities housing captive animals are full of staff who, every day, interact with the animals under their care. The expertise and familiarity of staff can be used to monitor animal welfare by means of questionnaires. It was the goal of our study to examine the association between chimpanzee (*Pan troglodytes*) welfare, happiness, and personality. To these ends we collected two waves of welfare and subjective well-being ratings of 18 chimpanzees housed at the Edinburgh Zoo and one set of ratings of 13 chimpanzees housed at Fundació Mona. Ratings were made on a welfare questionnaire that included 12 items related to stress, psychological stimulation, and behavioural indicators of negative and positive welfare states, and a 4-item subjective well-being questionnaire. In addition, ratings were made on the 54-item Hominoid Personality Questionnaire and an abbreviated version of this scale consisting of 37 antonym pairs. We used generalizability theory to test whether welfare ratings generalized across items, raters, chimpanzees, and time. We then assessed the validity of the welfare and subjective well-being questionnaires by examining their associations with behaviour. Finally, we tested whether the welfare and subjective well-being ratings were associated with personality. Welfare ratings generalized across items, raters, chimpanzees, and time. Principal components analysis and regularized exploratory factor analysis indicated that ten welfare items and all four subjective well-being items formed a single dimension (welfareSWB). LASSO regression found that lower welfareSWB was associated with regurgitation, coprophagy, urophagy, and decreased proximity to nearest neighbour. A linear model that adjusted for age, sex, and facility, indicated that higher Extraversion and lower Neuroticism were related to higher welfareSWB. Welfare ratings were reliable and associated with subjective well-being and personality, demonstrating that staff ratings are a valid and potentially valuable tool for chimpanzee welfare assessment.

Keywords: Behaviour, chimpanzee, happiness, *Pan troglodytes*, personality, welfare

Chimpanzees (*Pan troglodytes*) with Positive Welfare Are Happier, Extraverted, and Emotionally Stable

1. Introduction

The chimpanzee fear grimace, which indicates panic but looks like a smile to humans, is a classic example of how humans can misinterpret animal behaviour and emotion in that, what appears to be the same behaviour in animals and humans, reflects different emotional states (Parr & Waller, 2006). This mismatch between animal signals and human interpretations of these signals may explain why some researchers find that observer ratings of welfare are not always associated with behaviours, including abnormal behaviours (Lesimple & Hausberger, 2014). However, animal care staff and researchers with years of education, training, and experience with their charges appear to be able to assess the personalities of these animals (Gosling, 2001; Vazire, Gosling, Dickey, & Schapiro, 2007) and also the animals' well-being (Diener & Chan, 2011; Gartner & Weiss, 2013). However, there is still debate about whether members of staff or other observers can reliably assess animal welfare (Meagher, 2009; Whitham & Wielebnowski, 2009).

Welfare is traditionally measured by coding behaviours believed to indicate negative or positive welfare, or by collecting physiological measures, such as blood cortisol. These approaches are, however, time consuming and/or expensive. Welfare questionnaires, on the other hand, can cover a broad set of states, and can be designed to be brief (Clarke, Pluske, & Fleming, 2016). These questionnaires can thus be used to rapidly and inexpensively assess welfare (Whitham & Wielebnowski, 2009). Ratings are also based on staff knowledge (Meagher, 2009), which is an underused resource. When welfare ratings are studied in combination with traits, such as personality, they can help further our understanding of why some animals do better in captivity than others.

In previous study (Robinson et al., 2016), we designed a 12-item welfare questionnaire for use in brown capuchins (*Sapajus apella*). For that study we collected observer ratings of personality and happiness, the latter being measured using the subjective well-being questionnaire (King & Landau, 2003), which was based on measure of human

happiness (Sandvik, Diener, & Seidlitz, 1993). Observers agreed on their ratings of welfare and that welfare and happiness were measuring the same construct in brown capuchins. Welfare and happiness were positively associated with brown capuchin Sociability, Assertiveness, and Attentiveness, and negatively associated with Neuroticism.

To extend these findings across primates, we sought to conduct a similar study with another nonhuman primate species. To these ends, chimpanzees (*Pan troglodytes*) are an ideal study species due to the high number that reside in zoos, sanctuaries, and research facilities, and the welfare challenges captivity may pose for them (Hosey, 2005). We also tested whether observer ratings of welfare are valid by examining their association with behavioural indicators of welfare states. Behaviours associated with positive welfare in chimpanzees include prosocial grooming (Martin, 2005) and play behaviour (Held & Špinka, 2011). Behaviours commonly associated with negative welfare include rocking and pacing (Jacobson, Ross, & Bloomsmith, 2016).

Previous research has shown that ratings of chimpanzee happiness and behavioural indicators of chimpanzee welfare are associated with personality. Chimpanzees with higher Dominance, Extraversion, Agreeableness, and Openness and lower Neuroticism have higher subjective well-being (King & Landau, 2003; Weiss et al., 2009). Chimpanzees higher in Neuroticism also perform more self-directed behaviours, such as scratching (Herrelko, Vick, & Buchanan-Smith, 2012). However, we do not yet know whether chimpanzee subjective well-being and welfare states, such as stress frequency and physical health, are related. Moreover, research is needed to determine whether and how chimpanzee personality is related to a more extensive set of welfare indicators.

To address these questions we collected ratings of welfare, subjective well-being, and personality in chimpanzees. Our first aim was to test the degree to which staff agreed in their ratings on an existing welfare questionnaire. Our second aim was to test whether there were associations between ratings on the welfare scale and ratings on the subjective well-being questionnaire. Our third aim was to test for associations between the ratings and observed

behaviours. Our fourth aim was to test for associations between welfare, subjective well-being, and personality.

2. Methods

2.1 Ethical approval

Our study was observational and non-invasive. Approval was sought and gained from the Edinburgh Zoo's Budongo Trail Research Committee on 3 July 2014, by the Edinburgh Biological Services Unit on 11 March 2014 (AWERB No: OS04-14), and by Fundació Mona's head researcher, Dr. Miquel Llorente, on 10 March, 2015.

2.2 Subjects

We studied 31 chimpanzees, of whom 17 were males that ranged in ge from 12.32 to 51.05 years (mean = 27.04 years ± SD = 10.06 years). Of these chimpanzees, 8 males and 11 females were housed at the Royal Zoological Society of Scotland's Edinburgh Zoo and 9 males and 4 females were housed at Fundació Mona in Girona, Spain. The Edinburgh Zoo chimpanzees lived in 1832m^2 enclosure with outdoor and indoor areas, each with complex climbing structures, and ropes (Schel et al., 2013). The Fundació Mona chimpanzees lived in a 5640m^2 naturalistic outdoor enclosure. The enclosure is divided in two areas (2420m^2 and 3220m^2) and contained natural grasses, wooden platforms, towers and ropes. Additional details on the Edinburgh Zoo chimpanzees and Fundació Mona are available in Schel et al. (2013) and Úbeda and Llorente (2015), respectively.

2.3 Instruments and observations

We used four questionnaires: the welfare questionnaire (Robinson et al., 2016), the subjective well-being questionnaire (King & Landau, 2003), the Five Factor Model Questionnaire (Úbeda & Llorente, 2015), and the Hominoid Personality Questionnaire (Weiss et al., 2009). Úbeda and Llorente translated the welfare and subjective well-being

questionnaires into Spanish[1] and put them into an Excel spreadsheet. We also used existing behavioural data.

2.3.1 Welfare questionnaire

The welfare questionnaire is comprised of three sections. In the first section, raters are instructed to answer questions about themselves and to explain the signs they use as indicators of positive and negative welfare, for example, "What signs (both physically and behaviourally) do you use as an indication that an animal has positive welfare?" The second section consists of a 12-item welfare questionnaire that is based on the factors---social relationships, mental stimulation, physical health, stress, and control---that McMillan (2005) identified as important welfare domains. The questionnaire design was also influenced by descriptions of animal stress, coping, and physical health (see Broom & Johnson, 1993; Broom, 1991, 2007). The wording of the ninth and tenth question was influenced by the quality of life scale (Table 3 in Green & Mellor, 2011). Printed instructions asked raters to answer each question using a five-point scale, which includes answers that range from "very bad" to "very good" (e.g. Boissy et al., 2007; Yeates & Main, 2008). The third section asks for demographic details on animals, including their date of birth and sex.

2.3.2 Subjective well-being questionnaire

The chimpanzee subjective well-being questionnaire is comprised of four questions relating to the animal's ability to achieve their goals, the animal's satisfaction with social relationships, the amount of time the animal is happy, and how happy the rater would be if they were that animal for a week (King & Landau, 2003). For each item, raters are instructed to indicate on a Likert scale ranging from 1 "Displays either total absence or negligible amounts of the trait or state" to 7 "Displays extremely large amounts of the trait" the standing of a particular animal.

2.3.3 Personality questionnaires

[1] The English version of the welfare questionnaire can be found in the supplementary materials and the original questionnaire can be found in Robinson et al. (2016; Appendix A).

Ten of the Fundació Mona chimps were rated on the chimpanzee Five Factor Model questionnaire (Úbeda & Llorente, 2015). This questionnaire was based on King and Figueredo's (1997) Chimpanzee Personality Questionnaire. The Five Factor Model questionnaire includes 38 Spanish antonym pairs (Úbeda & Llorente, 2015). Each chimpanzee is rated on a seven-point scale where one word is assigned a "1" and its antonym is assigned a "7". An example item is "Sumiso [Submissive] (1) vs. Dominante [Dominant] (7)".

The 18 Edinburgh chimpanzees were rated on the Hominoid Personality Questionnaire, which consists of 54 items and is an expansion of the Chimpanzee Personality Questionnaire (Weiss et al., 2009). Each item consists of an adjective followed by one to three descriptive sentences, e.g., "**GENTLE:** Subject responds to others in an easy-going, kind, and considerate manner. Subject is not rough or threatening." Each chimpanzee is rated on a seven-point Likert scale from 1 "Displays either total absence or negligible amounts of the trait." to 7 "Displays extremely large amounts of the trait."

2.3.4 Behavioural observations

Behavioural data has been collected on the Edinburgh Zoo chimpanzees by researchers since 2012. Researchers collect data on variable days and at variable times between 09:00 and 17:00. Observations occur over 30-minute periods and are focussed on one area of the enclosure, with researchers performing three 10-minute focal observations on different chimpanzees, with an instantaneous group scan sample in between each focal. This creates the following observation schedule: group scan, 10-minute focal, group scan, 10-minute focal, group scan, final 10-minute focal, final group scan.

At these scan points, the identity and activity of each chimpanzee who was visible in the observed area was recorded. Each chimpanzee was recorded as resting, travelling (walking, running), climbing, foraging (moving around, looking for food), eating, playing, autogrooming, displaying, receiving or giving grooming, mutual grooming, fighting, or "other". Chimpanzees not visible were recorded as "out of sight". Operational definitions for these behaviours are available in Supplementary Table 1. Data collected from focal

observations were primarily binary, that is, the behaviour did or did not occur during the focal sample period. During the 10-minute focal observations, a point sample of the focal animal's nearest neighbour (identity and distance in metres) was taken at the start of the focal period. Zero-one sampling of the occurrence of the following behaviours was then recorded for the duration of the 10-minute focal sample: displaced by other, displace other, regurgitation, coprophagy, urophagy, allogrooming, autogrooming, eating, auto-hairplucking, hairplucked by other, and hairpluck other.

2.4. Data collection

The welfare and subjective well-being ratings of all the chimpanzees were collected at two time points. The first wave of data collection (Time 1) occurred between August and November of 2014 and only included the 18 Edinburgh Zoo chimpanzees. Ratings were made by two researchers and four keepers who were familiar with, i.e. able to differentiate and describe, the chimpanzees (mean = 3.0 raters per subject). The second wave of data collection (Time 2) occurred between April and June of 2016 and included the 18 Edinburgh Zoo chimpanzees and the 13 Fundació Mona chimpanzees. These 31 chimpanzees were rated by 15 researchers and 4 keepers who were familiar with the individual animals (mean = 7.61 raters per subject). One researcher rated the Edinburgh chimpanzees in both waves of data collection. All other raters only rated chimpanzees in a single wave.

Between March and July of 2015 three researchers familiar with the Edinburgh Zoo chimpanzees rated all 18 individuals on the 54-item Hominoid Personality Questionnaire (mean = 2 raters per subject) (Smith, 2015). As part of Úbeda and Llorente's 2015 study, 14 researchers, 12 keepers, and 2 volunteers familiar with the Fundació Mona chimpanzees rated 10 individuals on the 38-item Five Factor Model questionnaire (mean = 28 raters per subject).

For the 18 chimpanzees rated in Time 1, there were 3 missing responses to welfare questionnaire items out of 660 possible responses and 1 missing subjective well-being data point out of 220 possible responses. For the 31 chimpanzees rated in Time 2 there were 33 missing responses to welfare questionnaire items out of 2844 possible responses. For all

questionnaires, missing data were replaced with the overall mean for that item (Downey & King, 1998).

2.5 Data analysis

Unless otherwise stated, statistical analyses were performed using R version 3.1.1 (R Development Core Team, 2014).

2.5.1 Generalizability analysis of welfare and subjective well-being ratings

Because our data on welfare and subjective well-being contained four sources of variation or *facets*: Rater, Chimpanzee, Time, and Item, we used the framework of generalizability theory, of which interrater reliabilities are a special case, to assess reliability (Webb, Shavelson, & Haertel, 2006). For our study, we carried out a generalizability study on the facets and their interactions. We evaluated every combination of facet interaction using maximum likelihood estimation and linear mixed models, with the R package 'lmer'.

2.5.2 Interrater reliabilities of the personality items

The reliability of chimpanzee personality items has been described (King & Figueredo, 1997; Weiss et al., 2009), and as we only measured personality at one time point, it was not possible to evaluate reliability across time. However, to examine the reliability of personality items between raters, we used two intraclass correlations or *ICC*s. The first, *ICC*(3,1), indicates the reliability of individual ratings, and the second, *ICC*(3,*k*), indicates the reliability of mean ratings across *k* raters (Shrout & Fleiss, 1979).

2.5.3 Principal component analyses and regularized exploratory factor analyses

To create a single item score for each chimpanzee, we took the mean of the items for the welfare and subjective well-being item across raters. If these scores were consistent across Time 1 and Time 2, then we aggregated across both time points. We performed three principal component analyses (PCA) and three regularized exploratory factor analyses (REFA) (Jung & Lee, 2011). The first included the welfare items and the subjective well-being items. The second of each of these analyses included the welfare items. The third included the subjective well-being items. To determine the number of components to extract we used parallel analyses using the paran function (Dinno & Dinno, 2010) and inspected the

scree plots. We used the principle function (Revelle, 2011) to perform the PCAs. We conducted the REFAs in MATLAB (R20014b) using code provided by Sunho Jung.

We then created unit-weighted component scores (Gorsuch, 1983) based on the three structures delineated by the PCAs and REFAs. To do so, items with loadings less than |0.4| were assigned a weight of zero, items with loadings that were greater than or equal to 0.4 were assigned a weight of +1, and items with loadings that were less than or equal to -0.4 were assigned a weight of -1. If an item had loadings greater than |0.4| onto two components or factors, we assigned the item to the component or factor with the highest loading.

2.5.4 Personality component scores

Because the chimpanzees at Edinburgh Zoo and Fundació Mona were assessed on related questionnaires that had different answer formats, we matched items by their names and descriptions to create comparable component scores. For example, the antonym pair Triste [sad]/Alegre[cheerful] was matched with the adjective depressed. Only one antonym pair, Desconfiado [mistrustful]/Confiado [trustful], did not match a Hominoid Personality Questionnaire item. This antonym pair was therefore not used in our analyses (for a list of matched personality items see Supplementary Table 2). After matching the personality items, using the published structure in Weiss et al. (2009), we calculated unit-weighted scores for Dominance, Extraversion, Conscientiousness, Agreeableness, Neuroticism, and Openness.

2.5.5 Regression of welfare component and observed behaviour

In exploratory analyses, we examined the focal and scan data during two time frames; of the three months before and of the three months after each of the two welfare and subjective well-being rating time points. At Time 1 our models were based on 253 focal observations (mean per chimpanzee ± SD = 14.06 ± 2.69), from July to December, 2014. At Time 2, our models were based on 105 focal observations (mean per chimpanzee ± SD = 5.83 ± 1.42), taken between January and June, 2015. While the amount of scan data was larger (5471 records at Time 1, 2553 at Time 2), each record contained only a single behaviour for each chimpanzee, whereas within our focal data, each row contained many different bits of information, one bit for each behaviour, e.g. aggression or grooming.

For regression modelling we tested for associations between component scores representing welfare and observed behaviours. The sparsity of the behavioural data was problematic for ordinary least squares regression. We therefore we used least absolute shrinkage and selection operator (LASSO) regression for variable selection as it is robust to the effects of sparse data (Zou & Hastie, 2005). LASSO regression operates by iteratively increasing the penalty against each estimate in the regression, and will eventually reduce an estimate to zero and remove it from the model if it does not make a significant contribution to model fit. LASSO can thus be used as a computationally objective alternative to model selection functions like the Akaike Information Criteria and log-likelihood difference tests. LASSO models also produce accurate parameter estimates like other linear models. We used the 'glmnet' package (Friedman, Hastie, Simon, Tibshirani, & Hastie, 2015) to run the LASSO regressions.

2.5.6 Correlation of welfare, subjective well-being, and personality components

For the 28 chimpanzees rated on personality, component scores based on the welfare and subjective well-being structures were correlated with the six chimpanzee personality dimensions using Spearman-rank correlations. The welfare and subjective well-being components were also correlated using the same method.

2.5.7 Linear models of welfare, subjective well-being, and personality dimensions

We fit linear models using the lm function. Our dependent variables were the component score or scores based on the PCA and REFA of the welfare and subjective well-being items. The predictors included sex, age, and facility, which served as covariates, and the six personality component scores. Age and the personality component scores were mean centred and divided by two standard deviations to make their effect sizes comparable to those for facility and sex (Gelman, 2008). The dependent variables were converted into *z*-scores. We used the MuMIn package (Barton, 2015) to calculate the marginal r^2, which indicates the amount of variance in each model.

As the personality questionnaires used to rate the chimpanzees at Edinburgh Zoo and Fundació Mona differed, we ran additional linear models excluding the ten Fundació Mona

chimpanzees. If we found similar effects, both in terms of size and direction, then this would suggest that the results were unaffected by the use of different questionnaires.

3. Results

3.1 Generalizability theory and reliability

The variance components of the generalizability study are shown in Table 1. The majority of the variance (excluding the residual) in responses to the questionnaire is captured by the Item facet, which captures the variation between answers to different questions. Other facets that captured more than 0% of the variance include Subject and Subject × Item, both of which reflected the variation in welfare we would expect to see across a group of chimpanzees with differing mental and physical health.

--- Insert Table 1 About Here ---

The generalizability coefficients in Table 2 represent the reliability of the individual items, i.e., how consistently they are used, across different chimpanzees, different raters, and different study waves. All but three of these coefficients are strong (> 0.8) and all coefficients are greater than 0.6. In small sample animal studies, these estimates are within reason (Figueredo, Cox, & Rhine, 1995) and alone, do not suggest that any items should be removed from further analysis.

The overall generalizability of the instrument was $E\hat{\rho}^2 = 0.92$, and so was well above all minimum acceptable standards (0.7 to 0.8) (Hernandez-Lloreda & Colmenares, 2006), indicating that the relative (or between-individual) reliability is strong. The overall dependability ($\Phi = 0.82$) also indicated that the absolute (or within-individual) reliability was strong.

All 37 personality items were found to be reliable (Supplementary Table 3). The *ICC*(3,1) estimates ranged from 0.11 to 0.59 with a mean of 0.36 and the *ICC*(3,*k*) estimates ranged from 0.59 to 0.94 with a mean of 0.84.

--- Insert Table 2 About Here ---

3.2 PCA and REFA of welfare and subjective well-being items

The parallel analysis and scree plot revealed that the four subjective well-being items with the 12 welfare items all 16 items loaded onto a single component (Table 3, left panel), which we named welfareSWB. We also found a single component structure when we performed separate PCAs and REFA with the 12 welfare items (Table 3, middle panel) and four subjective well-being items (Table 3, right panel). The REFA revealed that the welfare items relating to the physical health and stress frequency did not have loadings that exceeded |0.4|. We therefore did not include these items in our component scores.

--- Insert Table 3 About Here ---

3.3 Regression of welfare and subjective well-being components and observed behaviour

Every instance of each behaviour on the focal and scan ethograms that could be related to welfare was included in our models. These focal behaviours were regurgitation, coprophagy, urophagy, allogrooming, autogrooming, autoplucking, plucked by other, pluck other, and distance to nearest neighbour. The scan sample behaviours included were resting, playing, autogrooming, aggression (displaying), receiving or giving grooming, mutual grooming, and fighting. Out of sight and "other" were included as control variables. Individual recordings of these behaviours were regressed onto the welfareSWB component using LASSO regression. We created two models for focal behaviours (one for each time point) and two models for scan behaviours (one for each time point).

The LASSO models predicting welfareSWB ratings from focal behaviour are shown in Table 4. In the first model, negative behaviours, such as regurgitation and coprophagy, were retained. The more instances of these behaviours observed, the lower the individual's welfareSWB score. The other behaviours included were social, e.g. grooming, distance to nearest neighbour, which were related to lower welfareSWB.

--- Insert Table 4 About Here ---

The smaller coefficients in the first model could have been retained as a result of noise. This bears out in the second regression, which is modelled with fewer observations, and does not retain grooming or plucking with the same sign in both models. The second regression is thus useful for determining which variables consistently predict welfare.

Regurgitation, coprophagy, and urophagy were retained in the second model, with the same signs and similar effect sizes. The only other variable to be retained was nearest neighbour distance, which also had a negative and small effect size.

Effect sizes for the LASSO models of the scan samples were smaller and less consistent between time points (Table 5). The scan models retained resting, playing, and "out of sight", with the same sign in both models. Chimpanzees who were resting were not engaged in other behaviours, most of which were positively associated with welfareSWB.

--- Insert Table 5 About Here ---

3.4 Correlation of welfare, subjective well-being, welfareSWB, and personality

The welfare and subjective well-being components were highly correlated (r_s = 0.95, 95% CI = 0.90 to 0.98). Higher Openness and lower Neuroticism were associated with higher welfare, subjective well-being, and welfareSWB; higher Extraversion was associated with higher subjective well-being and welfareSWB (Table 6).

--- Insert Table 6 About Here ---

3.5 Linear models of welfareSWB and personality dimensions

In the model of welfareSWB predicted by the personality dimensions we found that chimpanzees who were higher in Extraversion and lower in Neuroticism were higher in welfareSWB (Table 7). This held true for models predicting welfare but there was no association between personality and subjective well-being (Supplementary Table 4). There were no significant effects of age, sex, or facility in any model predicting the welfare, SWB, or welfareSWB.

--- Insert Table 7 About Here ---

As a sensitivity check, we ran the same models, but only included the 18 Edinburgh Zoo chimpanzees. We found that personality was not significantly associated with the subjective well-being, welfare, or welfareSWB, but that the directions of the effects were unchanged. The only exceptions to this were for Agreeableness in all models and the association between sex and subjective well-being. In both cases, the direction of the effect

flipped from positive to negative (Supplementary Table 5). This suggests that we did not find significant associations due to the reduced sample size and low statistical power.

4. Discussion

We found that staff familiar with captive chimpanzees agreed on ratings of welfare and happiness. These ratings were highly correlated and comprised a single dimension, welfareSWB. Higher welfareSWB scores were associated with behavioural indicators of positive and negative welfare states, a finding that was replicated across two time points. Higher welfareSWB scores were associated with higher Extraversion and lower Neuroticism. These results are consistent with previous studies of brown capuchin monkeys (Robinson et al., 2016) and happiness in nonhuman primates (Schaefer & Steklis, 2014; Weiss et al., 2009).

Higher welfareSWB was associated with behaviours, including, but not limited to, regurgitation, urophagy, and coprophagy, as well as with lower frequencies of resting and greater distance from conspecifics. The negative association between resting and welfareSWB likely reflects a lack of independence between observations; when a chimpanzee is resting, he or she is not doing something more enjoyable, for example, engaging in social behaviours.

We expected that chimpanzees who played more would be rated as higher in welfare, but the opposite was true. This may be because, although play is often considered a sign of positive welfare, it can be higher in stressful situations (Held & Špinka, 2011). When we examined our data, we found that there were few observations of play and that these instances were observed in a select group of females at Edinburgh Zoo. It may be that individuals who spend more time playing alone may lack the social standing to engage in prosocial grooming and other social activities.

Welfare ratings were stable over time and the variation across time that we did find was not easily attributable to any particular source. It may be that that there were no events strong enough to influence welfare, such as the introduction of a new group member (Brent, Kessel, & Barrera, 1997). It may be worth evaluating welfare ratings before and after such a major change to see how the chimpanzees and corresponding ratings are affected. An alternative explanation for this finding is that our welfare and subjective well-being

instruments are not sensitive enough to detect small changes in welfare or subjective well-being over short periods of time. To test this future studies could include additional indicators of welfare, such as self-injury and cortisol levels. As there is now work validating animal welfare and personality questionnaires, future studies may want to examine how ratings are influenced by characteristics of the observers, such as their gender, experience, and understanding of animal welfare.

Chimpanzees who were higher in Extraversion and lower in Neuroticism were rated as being happier and having higher welfare. Similar associations have been found in humans (DeNeve & Cooper, 1998; Steel, Schmidt, & Shultz, 2008), gorillas (Schaefer & Steklis, 2014), orangutans (Weiss, King, & Perkins, 2006), and rhesus macaques (Weiss, Adams, Widdig, & Gerald, 2011). This consistency of this finding across the primate order suggests that emotional stability and sociability are pillars of primate happiness. This finding may be unsurprising given the social nature of primate species and the influence that emotional stability has on health (Chapman, Roberts, & Duberstein, 2011; Goodwin & Stein, 2003). Further research to determine the degree to which this association is across primates and other orders of animal.

The association between Extraversion and welfare is consistent with the literature demonstrating the importance of social relationships in nonhuman primates (reviewed by McCowan et al., 2016 and Olsson & Westlund, 2007) and the association between social relationships and well-being in humans (Jose, Ryan, & Pryor, 2012; Pavot, Diener, & Fujita, 1990). In nonhuman primates, the presence of a social partner and the quality of social relationships are associated with reduced risk of self-injury (Lutz, Well, & Novak, 2003; Rommeck, Anderson, Heagerty, Cameron, & McCowan, 2009). Furthermore, primates with higher quality relationships exhibit lower levels of stress (Wittig et al., 2008) and reduced parasite loads (Akinyi et al., 2013). The association between higher welfare and happiness and lower Neuroticism matches a 2012 study of the Edinburgh Zoo chimpanzees, which found an inverse association between Neuroticism and self-directed behaviours, including self-rubbing (Herrelko et al., 2012).

Our results deviated some from previous findings. For one, Weiss et al. (2009) found that higher Dominance, Agreeableness, and Openness were associated with higher subjective well-being. Similarly, King and Landau (2003) found that higher subjective well-being was associated with higher Dominance, Extraversion, and Dependability (since renamed Conscientiousness). In the present study, higher Extraversion and lower Neuroticism were only associated with higher welfare and welfareSWB. The reason for this may be that we lacked the statistical power to detect these associations. This could be tested by extending our data collection to include additional facilities, thus increasing our sample size and statistical power. Even with a limited sample size we found significant associations between welfare, welfareSWB, and Extraversion, and between lower Neuroticism. This suggests that these personality dimensions play an important role in chimpanzee welfare.

Overall, these findings suggest that questionnaire may be useful for improving animal welfare. For example, facilities housing chimpanzees can use these questionnaires to identify chimpanzees that are lower in Extraversion and higher in Neuroticism and behaviourally monitor the welfare of these individuals more carefully. These facilities may also be able to use the welfare questionnaire to track welfare over time to provide interventions when welfare is compromised. Whitham and Wielebnowski (2009) outline automated ways to do so, such as online questionnaires that notify keepers when ratings change drastically. Finally, these questionnaires can be used to identify chimpanzees at low risk, so that time resources dedicated to enrichment can be effectively allocated.

More than ever, animal welfare is at centre stage. In fact, in 2015, when Francis Collins, director of United States National Institutes of Health, announced the removal of chimpanzees from biomedical research, he cited the need to track the chimpanzees' welfare when they were moved to sanctuaries (Collins, 2015). However, he did not mention how welfare would be assessed. Questionnaires offer a quick and easy-to-use way to monitor chimpanzee welfare and can be integrated into common husbandry practice and applied during major changes in the circumstances or chimpanzees and other nonhuman primates. As

we demonstrated here, the reliability and validity of welfare ratings makes it worthwhile to expand this line of research to include other species.

Acknowledgements

We thank the Budongo Trail staff at the Edinburgh Zoo, the staff at Fundació Mona, Emma Tecwyn, and Vanessa Wilson for filling out welfare questionnaires. We would further like to thank Conor Smith for performing data formatting and for sharing personality ratings he collected for his MSc dissertation. Thank you to Sunho Jung for providing the MATLAB code.

References

Akinyi, M. Y., Tung, J., Jeneby, M., Patel, N. B., Altmann, J., & Alberts, S. C. (2013). Role of grooming in reducing tick load in wild baboons (Papio cynocephalus). *Animal Behaviour*, *85*, 559–568. http://doi.org/10.1016/j.anbehav.2012.12.012

Barton, K. (2015). MuMIn: Multi-model inference. R package version 1.9.13. *Version*, *1*, 18. http://doi.org/citeulike:11961261

Boissy, A., Manteuffel, G., Jensen, M. B., Moe, R. O., Spruijt, B., Keeling, L. J., … Aubert, A. (2007). Assessment of positive emotions in animals to improve their welfare. *Physiology & Behavior*, *92*(3), 375–397. http://doi.org/10.1016/j.physbeh.2007.02.003

Brent, L., Kessel, a. L., & Barrera, H. (1997). Evaluation of introduction procedures in captive chimpanzees. *Zoo Biology*, *16*(4), 335–342. http://doi.org/10.1002/(SICI)1098-2361(1997)16:4<335::AID-ZOO5>3.0.CO;2-B

Broom, D. M. (1991). Animal welfare: concepts and measurement. *Journal of Animal Science*, *69*, 4167–4175. http://doi.org//1991.69104167x

Broom, D. M. (2007). Quality of life means welfare: How is it related to other concepts and assessed? *Animal Welfare*, *16*(S), 45–53.

Broom, D. M., & Johnson, K. G. (1993). *Stress and animal welfare*. *Animal Welfare*. London: Chapman and Hall.

Chapman, B. P., Roberts, B., & Duberstein, P. (2011). Personality and longevity: knowns, unknowns, and implications for public health and personalized medicine. *Journal of Aging Research*. http://doi.org/10.4061/2011/759170

Clarke, T., Pluske, J. R., & Fleming, P. A. (2016). Are observer ratings influenced by prescription? A comparison of Free Choice Profiling and Fixed List methods of Qualitative Behavioural Assessment. *Applied Animal Behaviour Science*, *177*, 77–83. http://doi.org/10.1016/j.applanim.2016.01.022

Collins, F. (2015, November 18). NIH Will No Longer Support Biomedical Research on Chimpanzees. *National Institutes of Health*.

DeNeve, K. M., & Cooper, H. (1998). The happy personality: a meta-analysis of 137 personality traits and subjective well-being. *Psychological Bulletin*, *124*(2), 197–229. http://doi.org/10.1037/0033-2909.124.2.197

Diener, E., & Chan, M. Y. (2011). Happy people live longer: Subjective well-being contributes to health and longevity. *Applied Psychology: Health and Well-Being*, *3*(1), 1–43. http://doi.org/10.1111/j.1758-0854.2010.01045.x

Dinno, A., & Dinno, M. (2010). R package "paran."

Downey, R. G., & King, C. (1998). Missing data in Likert ratings: A comparison of replacement methods. *The Journal of General Psychology*, *125*(2), 175–91. http://doi.org/10.1080/00221309809595542

Figueredo, A. J., Cox, R. L., & Rhine, R. J. (1995). A Generalizability Analysis of Subjective Personality Assessments in the Stumptail Macaque and the Zebra Finch. *Multivariate Behavioral Research*, *30*(2), 167–197. http://doi.org/10.1207/s15327906mbr3002_3

Friedman, A. J., Hastie, T., Simon, N., Tibshirani, R., & Hastie, M. T. (2015). Lasso and Elastic-Net Regularized Generalized Linear Models. *Available Online at Https://cran.r-Project.org/web/packages/glmnet/glmnet.pdf. (Verified 29 July. 2015).*

Gartner, M. C., & Weiss, A. (2013). Scottish wildcat (Felis silvestris grampia) personality and subjective well-being: Implications for captive management. *Applied Animal Behaviour Science*, *147*(3–4), 261–267. http://doi.org/10.1016/j.applanim.2012.11.002

Gelman, A. (2008). Scaling regression inputs by dividing by two standard deviations. *Statistics in Medicine*, *27*(15), 2865–2873. http://doi.org/10.1002/sim.3107

Goodwin, R. D., & Stein, M. B. (2003). Peptic ulcer disease and neuroticism in the United States adult population. *Psychother Psychosom*, *72*(1), 10–15.

Gorsuch, R. L. (1983). *Factor Analysis* (2nd ed.). Hillsdale, NJ: Erlbaum.

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

Green, T. C., & Mellor, D. J. (2011). Extending ideas about animal welfare assessment to include "quality of life" and related concepts. *New Zealand Veterinary Journal*, *59*(6), 263–271. http://doi.org/10.1080/00480169.2011.610283

Held, S. D. E., & Špinka, M. (2011). Animal play and animal welfare. *Animal Behaviour*, *81*(5), 891–899. http://doi.org/10.1016/j.anbehav.2011.01.007

Hernandez-Lloreda, M. V., & Colmenares, F. (2006). The utility of generalizability theory in the study of animal behaviour. *Animal Behaviour*. http://doi.org/10.1016/j.anbehav.2005.04.023

Herrelko, E. S., Vick, S. J., & Buchanan-Smith, H. M. (2012). Cognitive research in zoo-housed chimpanzees: Influence of personality and impact on welfare. *American Journal of Primatology*, *74*(9), 828–840. http://doi.org/10.1002/ajp.22036

Hosey, G. R. (2005). How does the zoo environment affect the behaviour of captive primates? *Applied Animal Behaviour Science*, *90*(2), 107–129. http://doi.org/10.1016/j.applanim.2004.08.015

Jacobson, S. L., Ross, S. R., & Bloomsmith, M. A. (2016). Characterizing abnormal behavior in a large population of zoo-housed chimpanzees: prevalence and potential influencing factors. *PeerJ*, *4*, e2225. http://doi.org/10.7717/peerj.2225

Jose, P. E., Ryan, N., & Pryor, J. (2012). Does Social Connectedness Promote a Greater Sense of Well-Being in Adolescence Over Time? *Journal of Research on Adolescence*, *22*(2), 235–251. http://doi.org/10.1111/j.1532-7795.2012.00783.x

King, J. E., & Figueredo, A. (1997). The Five-Factor Model plus Dominance in Chimpanzee Personality. *Journal of Research in Personality*, *271*(31), 257–271.

King, J. E., & Landau, V. I. (2003). Can chimpanzee (Pan troglodytes) happiness be estimated by human raters? *Journal of Research in Personality*, *37*(1), 1–15. http://doi.org/10.1016/S0092-6566(02)00527-5

Lesimple, C., & Hausberger, M. (2014). How accurate are we at assessing others' well-being? The example of welfare assessment in horses. *Frontiers in Psychology*, *5*(JAN). http://doi.org/10.3389/fpsyg.2014.00021

Lutz, C. K., Well, A., & Novak, M. a. (2003). Stereotypic and self-injurious behavior in rhesus macaques: A survey and retrospective analysis of environment and early experience. *American Journal of Primatology*, *60*(1), 1–15. http://doi.org/10.1002/ajp.10075

Martin, J. E. (2005). The effects of rearing conditions on grooming and play behaviour in captive chimpanzees. *Animal Welfare*, *14*(2), 125–133.

McCowan, B., Beisner, B. A., Bliss-Moreau, E., Vandeleest, J. J., Jin, J., Hannibal, D., & Hsieh, F. (2016). Connections matter: Social networks and lifespan health in primate translational models. *Frontiers in Psychology*, *7*(433), 1–11.

McMillan, F. (2005). Mental wellness: The concept of quality of life in animals. In *Mental Health and Well-Being in Animals*.

Meagher, R. K. (2009). Observer ratings: Validity and value as a tool for animal welfare research. *Applied Animal Behaviour Science*, *119*(1), 1–14. http://doi.org/10.1016/j.applanim.2009.02.026

Olsson, I. A. S., & Westlund, K. (2007). More than numbers matter: The effect of social factors on behaviour and welfare of laboratory rodents and non-human primates. *Applied Animal Behaviour Science*, *103*(3–4), 229–254. http://doi.org/10.1016/j.applanim.2006.05.022

Parr, L. A., & Waller, B. M. (2006). Understanding chimpanzee facial expression: insights into the evolution of communication. *Social Cognitive and Affective Neuroscience*. http://doi.org/10.1093/scan/nsl031

Pavot, W., Diener, E., & Fujita, F. (1990). Extraversion and happiness. *Personality and Individual Differences*, *11*(12), 1299–1306. http://doi.org/10.1016/0191-8869(90)90157-M

R Development Core Team. (2014). R Development Core Team. *R: A Language and Environment for Statistical Computing*.

Revelle, M. W. (2011). Package ' psych. *October*, 1–250.

Robinson, L. M., Waran, N. K., Leach, M. C., Morton, F. B., Paukner, A., Lonsdorf, E., … Weiss, A. (2016). Happiness is positive welfare in brown capuchins (Sapajus apella). *Applied Animal Behaviour Science*, *181*, 145–151. http://doi.org/10.1016/j.applanim.2016.05.029

Rommeck, I., Anderson, K., Heagerty, A., Cameron, A., & McCowan, B. (2009). Risk factors and remediation of self-injurious and self-abuse behavior in rhesus macaques. *Journal of Applied Animal Welfare Science : JAAWS*, *12*(1), 61–72. http://doi.org/10.1080/10888700802536798

Sandvik, E., Diener, E., & Seidlitz, L. (1993). Subjective Well-Being - the Convergence and Stability of Self-Report and Non-Self-Report Measures. *Journal of Personality*, *61*(3), 318–342. http://doi.org/10.1111/j.1467-6494.1993.tb00283.x

Schaefer, S. A., & Steklis, H. D. (2014). Personality and subjective well-being in captive male western lowland gorillas living in bachelor groups. *American Journal of Primatology*, *76*(9), 879–889. http://doi.org/10.1002/ajp.22275

Schel, A. M., Rawlings, B., Claidière, N., Wilke, C., Wathan, J., Richardson, J., … Slocombe, K. (2013). Network Analysis of Social Changes in a Captive Chimpanzee Community Following the Successful Integration of Two Adult Groups. *American Journal of Primatology*, *75*, 254–266. http://doi.org/10.1002/ajp.22101

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: uses in assessing rater reliability. *Psychological Bulletin*, *86*(2), 420–428. http://doi.org/10.1037/0033-2909.86.2.420

Smith, C. (2015). *Correlations of Measures of Autistic-like Traits in Chimpanzees (Pan troglodytes)*. University of Edinburgh.

Steel, P., Schmidt, J., & Shultz, J. (2008). Refining the relationship between personality and subjective well-being. *Psychological Bulletin*, *134*(1), 138–161. http://doi.org/10.1037/0033-2909.134.1.138

Úbeda, Y., & Llorente, M. (2015). Personality in sanctuary-housed chimpanzees: A comparative approach of psychobiological and penta-factorial human models. *Evolutionary Psychology*, *13*(1), 182–196.

Vazire, S., Gosling, S. D., Dickey, A. S., & Schapiro, S. J. (2007). Measuring personality in nonhuman animals. In *Handbook of research methods in personality psychology* (pp. 190–206). http://doi.org/10.1002/zoo.20379

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. http://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*, 283–292. http://doi.org/10.1002/ajp.20649

Weiss, A., King, J. E., & Perkins, L. (2006). Personality and subjective well-being in orangutans (Pongo pygmaeus and Pongo abelii). *Journal of Personality and Social Psychology*, *90*, 501–511. http://doi.org/10.1037/0022-3514.90.3.501

Whitham, J. C., & Wielebnowski, N. (2009). Animal-based welfare monitoring: using keeper ratings as an assessment tool. *Zoo Biology*, *28*(6), 545–560. http://doi.org/10.1002/zoo.20281

Wittig, R. M., Crockford, C., Lehmann, J., Whitten, P. L., Seyfarth, R. M., & Cheney, D. L. (2008). Focused grooming networks and stress alleviation in wild female baboons. *Hormones and Behavior*, *54*(1), 170–7. http://doi.org/10.1016/j.yhbeh.2008.02.009

Yeates, J. W., & Main, D. C. J. (2008). Assessment of positive welfare: a review. *The Veterinary Journal*, *175*(3), 293–300.

Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, *67*(2), 301–320. http://doi.org/10.1111/j.1467-9868.2005.00503.x