Supplementary Information for

Facial Width-to-Height Ratio is Associated with Agonistic and Affiliative Dominance in Bonobos (*Pan paniscus*)

J. S. Martin, N. Staes, A. Weiss, J. M. G. Stevens, & A. V. Jaeggi

Correspondence should be addressed to Jordan S. Martin
Email: jsm.primatology@gmail.com

Table of Contents

Supplementary Methods and Results

Subjects and measures

A random sampling procedure was used to reduce researcher bias during facial measurement. NS removed identifiable information from photographs and randomly selected a subset of subjects from each zoo with available facial photographs, which were subsequently measured by JSM and a research assistant, neither of whom took part in collection of the behavioral and psychometric data. As described in the main text, appropriate photos with neutral expressions and forward-facing orientation were subsequently selected and measured for these subjects. fWHRs were then paired back with the remaining individual data after these measurements were completed.

Organizational effects of androgen exposure on behavior and facial morphology could plausibly occur from the prenatal period until sexual maturity. We therefore sought to focus our analysis on sexually mature bonobos. Previous research on captive bonobos suggests that the onset of puberty is likely to occur from approximately 6-10 years of age, with the sharpest increase in urinary testosterone around 8-9 years of age for males and an earlier but more gradual increase in females [1]. We therefore excluded three 7 year old subjects from our final dataset who we could not confidently classify as sexually mature. This resulted in a final sample of 38 individuals across five social groups. Demographic data on the resultant sample is provided below (**Table S1**).

Table S1. Sample demographics.

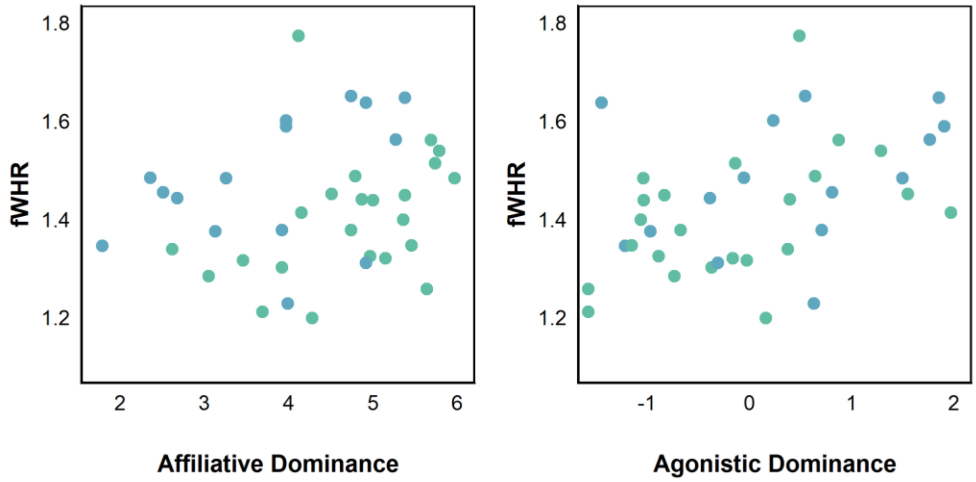
Zoo	*n*	# Males	# Females	Average age (range)
Apenheul	5	2	3	19.6 (13-34)
Frankfurt	7	2	5	25.6 (11-62)
Planckendael	5	3	2	17.4 (10-27)
Twycross	8	3	5	22 (10-36)
Wilhelma	7	2	5	28.1 (11-48)
Wuppertal	6	3	3	28.3 (12-49)

Footnote. Age in listed in years.

In the 22 subjects with available body weight measures, moderate to strong associations were also observed between sex and weight ($r_{Biserial} = 0.82$) and fWHR and weight ($r = 0.36$). In our full sample, fWHR and sex exhibit a similarly sized association ($r_{Biserial} = 0.36$). While the relationship between sex and fWHR may be mediated by body weight, as suggested by our primary regression model (**M1**; see below), testosterone is also a known cause of individual differences in body size [2]. It therefore remains unclear whether organizational androgen effects may be a latent common cause of these associations. The statistically uncertain sex effect reported in the main text, after conditioning on body weight, should therefore be cautiously interpreted.

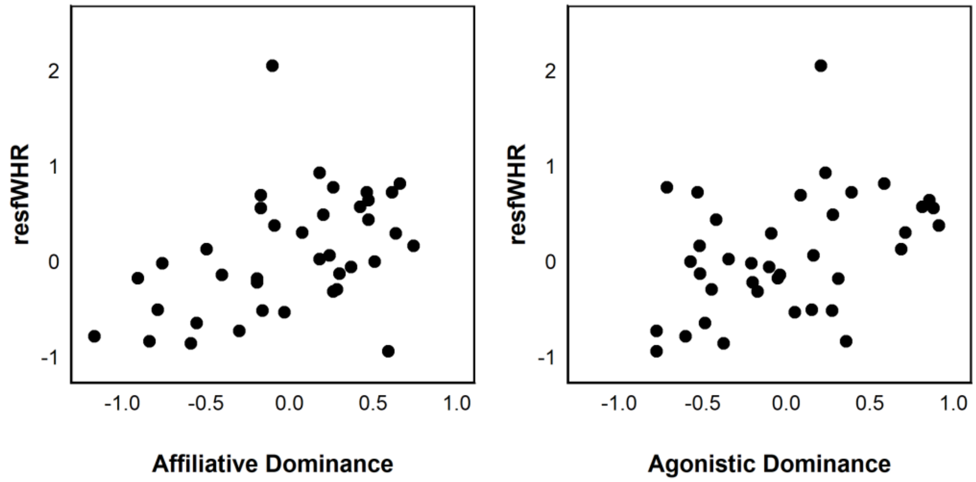
Scatterplots of our raw data provide initial support for the association between fWHR and both affiliative and agonistic dominance (**Fig S1**), but also suggest that the strength of affiliative dominance in particular is enhanced by controlling for sex, age and body weight. Consistent with this interpretation, a clearer affiliative dominance association is observed with fWHR residuals after controlling for these factors (**Fig S2**).

Fig S1. Scatterplots of fWHR and the social dominance measures.



Footnote. Datapoints are colored separately for females (green) and males (blue). Social dominance measures are shown on the original data scale.

Fig S2. Scatterplots of fWHR residuals and the social dominance measures.



Footnote. Social dominance measures are standardized to 2 SD. resfWHR = residual fWHR controlling for age, sex, and body weight.

It is important to emphasize that our agonistic dominance measure was analyzed using within-group deviations rather than absolute scores. By centering individual scores within zoos, we effectively accounted for differential opportunities for agonistic encounters across zoos. This is necessary because the raw David's scores used as a measure of agonistic dominance are contingent upon the sample size within each zoo. As further described below, we did not find support for further zoo-specific effects in a random slopes model (**M11** below).

Statistical Analysis

We estimated Bayesian linear measurement error models for all analyses using the R package 'brms' [3], which interfaces with the Stan statistical programming language [4]. As noted in the main text, we employed a fully Bayesian approach to statistical estimation and inference. Therefore, rather than relying upon null hypothesis tests and arbitrary designations of statistical significance, we used multiple sources of information to summarize and draw inferences from our posterior model estimates [5]. The R Code and dataset for this manuscript have been provided as additional supplementary material and can be used to replicate all analyses described below.

We examined the association between fWHR and measures of affiliative and agonistic dominance, while accounting for error in the measurement of fWHR across photos. In addition to these covariates, we also included fixed effects for years of age and sex in all models. We found that inclusion of random zoo-specific intercepts did not account for a meaningful degree of variance in fWHR ($\tilde{\sigma}^2$ = 0.03 [MAD = 0.04]) and reduced the efficiency of MCMC model convergence. We therefore excluded this term from our statistical models.

Our first model (**M0**) excluded information on body weight to assess potential sexual dimorphism in fWHR irrespective of body size. We therefore estimated the following formal model structure conditional on the average fWHR measurement for subject *i* using Hamiltonian Markov Chain Monte Carlo.

Model 0 (M0). Main effects without body weight covariate.

$$\mathrm{fWHR}_{\mathrm{EST},i} \sim \mathrm{Normal}(\mu_i, \sigma)$$

$$\mu_i = \alpha + \beta_{\mathrm{AssR}} + \beta_{\mathrm{wgDS}} + \beta_{\mathrm{Age}} + \beta_{\mathrm{Sex}}$$

$$\mathrm{fWHR}_{\mathrm{OBS},i} \sim \mathrm{Normal}\left(\mathrm{fWHR}_{\mathrm{EST},i}, \mathrm{fWHR}_{\mathrm{SD},i}\right)$$

$$\alpha, \beta \sim \mathrm{Normal}(0,2)$$

$$\sigma \sim \mathrm{Half-Cauchy}(0,2)$$

Here, the expected subject-specific fWHR μ_i is represented as a function of the population-level intercept α and population-level/fixed effects β for Assertiveness scores of affiliative dominance (AssR), within-group David's scores of agonistic dominance (wgDS), age, and sex. We account for measurement error in fWHR measurements by parameterizing observed fWHR measurements $\mathrm{fWHR}_{\mathrm{OBS},i}$ as arising from a normal distribution characterized by unknown mean parameter $\mathrm{fWHR}_{\mathrm{EST},i}$ and the standard deviation $\mathrm{fWHR}_{\mathrm{SD},i}$ of fWHR measurements for each subject. This structure effectively accounts for uncertainty in our response variable while estimating the regression parameters, and vice versa [5]. The expected measurement error for subjects with multiple photographs was assigned to 3 subjects with single photographs. Please note that we simplify specification of model priors to represent shared priors over fixed effects (α, β), and we also suppress observed covariate values to ease interpretation, so that terms such as β_{AssR} implicitly denote $\beta_{\text{AssR}}\text{AssR}_i$.

For our primary analysis (**M1**), we then included body weight as an additional covariate to assess whether links between fWHR, sex, and social dominance were independent of body size. Recent body weight measures were only available for a subset of our sample, and we therefore used a Bayesian imputation procedure to avoid an appreciable loss of information and statistical power. We used an inclusive predictive model for estimating unmeasured body weights, incorporating all main effect terms in the primary regression model, so as to reduce systematic error and better approximate data missing completely at random (MCAR) [6]. We therefore estimated the following model conditional on our dataset

Model 1 (M1). Main effects with body weight covariate.

$$\text{fWHR}_{\text{EST},i} \sim \text{Normal}(\mu_i, \sigma)$$

$$\mu_i = \alpha + \beta_{\text{AssR}} + \beta_{\text{wgDS}} + \beta_{\text{Age}} + \beta_{\text{Sex}} + \beta_{\text{Weight}}$$

$$\text{fWHR}_{\text{OBS},i} \sim \text{Normal}\left(\text{fWHR}_{\text{EST},i}, \text{fWHR}_{\text{SD},i}\right)$$

$$\text{Weight}_i \sim \text{Normal}\left(\nu_i, \sigma_{\text{Weight}}\right)$$

$$\nu_i = \alpha_{\text{Weight}} + \gamma_{\text{AssR\_}\nu} + \gamma_{\text{wgDS\_}\nu} + \gamma_{\text{Age\_}\nu} + \gamma_{\text{Sex\_}\nu} + \gamma_{\text{fWHR\_}\nu}$$

$$\alpha, \beta, \gamma \sim \text{Normal}(0,2)$$

$$\sigma \sim \text{Half} - \text{Cauchy}(0,2)$$

Here, missing values for body weight are imputed using the regression function defined for the subject-specific expectation ν_i, with random predictive uncertainty σ_{Weight}. Fixed effect terms in this predictive imputation model are noted by γ, rather than the β notation for fixed effects in the main fWHR model, to aid interpretation.

Cohen's f^2 [7] were calculated as suggested by Selya and colleagues [8] to provide a standardized metric of local effect size

$$\text{f}^2 = \frac{\text{R}^2_{\text{AB}} - \text{R}^2_{\text{A}}}{1 - \text{R}^2_{\text{AB}}}$$

Here R^2_{AB} is the variance explained by a model containing the parameter of interest B, and R^2_A is the variance explained by a model of all other parameters A excluding B. An estimated f^2 can be negative as the sampled posterior of R^2_{AB} may be smaller than R^2_A. For ease of interpretation, we report negative values as 0 to denote that no support was found for a relative increase in the explanatory power of the model.

Additional interaction effect models.

For comparison with previous research on capuchins, we also estimated additional interaction models with sex-specific effects for affiliative (**M2;** see R Code for further details) and agonistic dominance (**M3**), as well as an interaction between these dominance measures (**M4**). Given that associations between personality and dominance rank have been found to vary across the lifespan (e.g., [9]), we also fit supplementary exploratory models estimating interactions between age and affiliative (**M5**) and agonistic dominance (**M6**), as well as age by sex interactions with affiliative (**M7**) and agonistic dominance (**M8**). No clear interaction effects were observed across models. In addition to the absence of sex-specific interactions reported in the main text, we also did not find support for age interaction effects with affiliative ($\tilde{\beta}$ = 0.03 [0.37], 90% CI [-0.57, 0.66], $p_{>0}$ = 0.54, $\widetilde{f^2}$ = 0) or agonistic dominance ($\tilde{\beta}$ = 0.01 [0.35], 90% CI [-0.57, 0.58], $p_{>0}$ = 0.51, $\widetilde{f^2}$ = 0.01). Sex-specific age interactions were also not present for affiliative ($\tilde{\beta}$ = 0.14 [0.88], 90% CI [-1.28, 1.60], $p_{>0}$ = 0.56, $\widetilde{f^2}$= 0) or agonistic dominance ($\tilde{\beta}$ = -0.20 [0.94], 90% CI [-1.75, 1.36], $p_{<0}$ = 0.59, $\widetilde{f^2}$ = 0.02).

It is possible that such age by sex interactions for social dominance are non-linear across the lifespan, particularly for male bonobos. We therefore further explored non-linear sex by age interactions for affiliative (**M9**) and agonistic dominance (**M10**) using tensor product smoothing [10]. Given the difficulty of directly interpreting non-linear regression coefficients, we used the Watanabe-Akaike information criterion (WAIC) to conduct a fully Bayesian model comparison [11] between the main effects model (**M1**) and these more complex non-linear interaction models. As with other information criteria such as AIC or BIC, smaller values indicate greater relative model quality and expected predictive validity, such that $\text{WAIC}_{\textbf{Model A}} - \text{WAIC}_{\textbf{Model B}} \leq -2$ provides minimal support for selection of the more complex Model A. Consistent with the aforementioned results, we found that allowing for non-linear interaction effects did not meaningfully enhance the quality of our models and their expected predictive validity ($\text{WAIC}_{\textbf{M9}} - \text{WAIC}_{\textbf{M1}}$ = 4.87 [SE = 9.33]; $\text{WAIC}_{\textbf{M10}} - \text{WAIC}_{\textbf{M1}}$ = 5.85 [SE = 6.47]).

Finally, although we used within-zoo centering on David's scores, thus controlling for differential opportunities for agonistic encounters among zoos, it is possible that other unmeasured zoo-specific effects could still confound our main results. We therefore also estimated a supplementary model (**M11**) examining whether random zoo-specific slopes between social dominance and fWHR enhanced model quality. In support of our main effects model (**M1**), we found that adding parameters for zoo-specific slopes reduced the expected predictive validity of our model ($\text{WAIC}_{\textbf{M11}} - \text{WAIC}_{\textbf{M1}}$ = 4.89 [SE = 2.26]).

Our data therefore do not provide support for more complex relationships between social dominance and fWHR than are described in our main effects model. For these reasons, we relied on **M1** for drawing statistical inferences. Nonetheless, it is important to emphasize that our data provide only modest statistical power for detecting interaction and random slope effects, which would be more effectively examined in larger samples.

Supplementary References

1. Behringer, V., Deschner, T., Deimel, C., Stevens, J. M., & Hohmann, G. (2014). Age-related changes in urinary testosterone levels suggest differences in puberty onset and divergent life history strategies in bonobos and chimpanzees. *Hormones and Behavior*, *66*, 525-533.
2. Veldhuis, J. D., Roemmich, J. N., Richmond, E. J., Rogol, A. D., Lovejoy, J. C., Sheffield-Moore, M., ... & Bowers, C. Y. (2004). Endocrine control of body composition in infancy, childhood, and puberty. *Endocrine Reviews*, *26*, 114-146.
3. Bürkner, P. C. (2017). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software, 80*, 1-28.
4. Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., ... & Riddell, A. (2017). Stan: A probabilistic programming language. *Journal of Statistical Software, 76*. DOI: 10.18637/jss.v076.i01
5. McElreath, R. (2016) *Statistical Rethinking: A Bayesian course with examples in R and Stan*. Boca Raton: CRC Press.
6. Collins, L. M, Schafer, J. L., & Kam, C. M. (2001). A comparison of inclusive and restrictive strategies in modern missing data procedures. *Psychological Methods*, *6*, 330–351.
7. Cohen, J. E. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Lawrence Erlbaum Associates.
8. Selya, A. S., Rose, J. S., Dierker, L. C., Hedeker, D., & Mermelstein, R. J. (2012). A practical guide to calculating Cohen's f2, a measure of local effect size, from PROC MIXED. *Frontiers in Psychology*, *3*, 111.
9. Altschul, D. M., Hopkins, W. D., Herrelko, E. S., Inoue-Murayama, M., Matsuzawa, T., King, J. E., ... & Weiss, A. (2018). Personality links with lifespan in chimpanzees. *eLife*, *7*, e33781.
10. Wood, S. N., Scheipl, F., & Faraway, J. J. (2012). Straightforward intermediate rank tensor product smoothing in mixed models. *Statistics and Computing*, *23,* 341–360.
11. Gelman, A., Hwang, J., & Vehtari, A. (2014). Understanding predictive information criteria for Bayesian models. *Statistics and Computing*, *24,* 997-1016.