

---

1                   **Supplementary information for:**

2                   **Unifying individual differences in personality,**

3                   **predictability, and plasticity: a practical guide**

---

5                   Data, code, and model objects to reproduce the results contained in this supplement are  
6                   available in this dedicated repository: <http://doi.org/10.17605/OSF.IO/V3QAX>

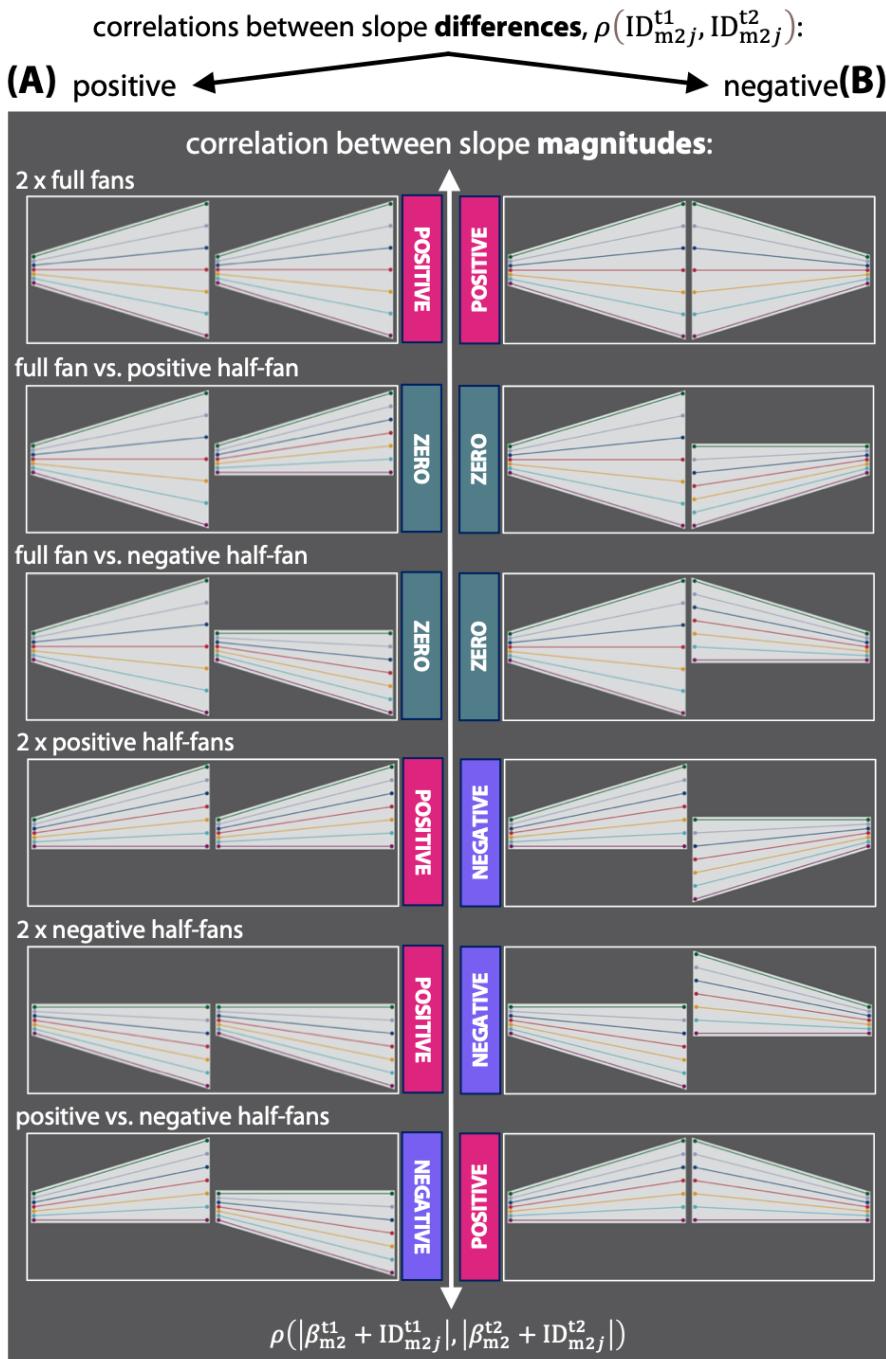
8 <b>SUPPLEMENT TO SECTION 2: INDIVIDUAL DIFFERENCES IN INTERCEPTS AND SLOPES</b>	3
9 <b>Figure S1</b>	3
10 <b>SUPPLEMENT TO SECTION 4: SUMMARY STATISTICS</b>	4
11                  Reversing non-linear transformations	4
12                  Coefficient of individual variation for the mean model	5
13 <b>4.3  THE COEFFICIENT OF VARIATION IN PREDICTABILITY</b>	6
14 <b>SUPPLEMENTARY WORKED EXAMPLE</b>	8
15 <b>Data collection</b>	8
16                  Analysis methods	9
17                  Data preparation	9
18                  Stan models	10
19                  Model checks and calculations	11
20 <b>Results and discussion: worked example</b>	12
21 <b>Figure S2</b>	13
22                  Individual differences in personalities	14
23                  Zebrafish behavioural syndromes	15
24                  Individual differences in predictability of behaviour	15
25                  Zebrafish personality-plasticity associations and plasticity syndromes	16
26 <b>Figure S3</b>	17
27                  Zebrafish predictabilities and predictability syndromes	18
28                  Zebrafish personality-predictability associations	18
29                  Zebrafish plasticity-predictability associations	19
30 <b>Figure S4</b>	20
31 <b>RESULTS TABLES</b>	21
32 <b>Supplementary Table S1</b>	21
33 <b>Supplementary Table S2</b>	24

34	<b>Supplementary Table S3</b>	25
35	<b>DIAGNOSTIC FIGURES</b>	27
36	<b>Diagnostic Figure SD01: Trace plots for bivariate double hierarchical generalized linear model</b>	27
37	<b>Diagnostic Figure SD02: R-hat statistics for univariate models with random intercepts but no random slopes</b>	28
38	<b>Diagnostic Figure SD03: R-hat statistics for univariate models with random intercepts and slopes</b>	29
39	<b>Diagnostic Figure SD04: R-hat statistics for bivariate models with random intercepts and slopes</b>	30
40	<b>Diagnostic Figure SD05: Distributions of residuals for aggression models</b>	31
41	<b>Diagnostic Figure SD06: Distributions of residuals for sociability models</b>	32
42	<b>Diagnostic Figure SD07: Distributions of residuals for neophilia models</b>	33
43	<b>Diagnostic Figure SD08: Distributions of residuals for boldness models</b>	34
44	<b>Diagnostic Figure SD09: Distributions of residuals for activity models</b>	35
45	<b>Diagnostic Figure SD10: Correlation between observed and predicted values for aggression models</b>	36
46	<b>Diagnostic Figure SD11: Correlation between observed and predicted values for sociability models</b>	38
47	<b>Diagnostic Figure SD12: Correlation between observed and predicted values for neophilia models</b>	39
48	<b>Diagnostic Figure SD13: Correlation between observed and predicted values for boldness models</b>	40
49	<b>Diagnostic Figure SD14: Correlation between observed and predicted values for activity models</b>	41
50	<b>SUPPLEMENTARY REFERENCES</b>	42

51

52      **Supplement to section 2: Individual differences in**  
 53      **intercepts and slopes**

Variation in plasticity syndromes, for pairs of traits,  
 when the rank-order of intercepts is maintained



54

55      **Figure S1**

56      Plasticity syndromes are influenced by simplified shapes of phenotypic plasticity (full fans,

negative half-fans, and positive half-fans, which either 'fan in' or 'fan out'; Fig. 2, Main Text). Boxes outlined in white depict reaction norms for two types of traits, where the order of individual averages (i.e. intercepts) is maintained across environments and across traits. Rows are arranged according to which fan shapes are paired together. In the centre of the figure, rotated text inside coloured boxes show variation in plasticity syndromes when measured with slope magnitudes (equation 16). Plasticity syndromes measured with slope magnitudes are always positive for a pair of full fans, always zero when a full fan is paired with a half-fan, and either positive or negative when half-fans are paired (depending on the direction of their slopes and whether reaction norms fan 'in' or 'out'). **(A)** Reaction norms in the left column all 'fan out'. When the effect of the environmental change on between-individual variance is identical for both traits (i.e. both sets of reaction norms 'fan out', or both sets of reaction norms 'fan in'), then plasticity syndromes measured with individual differences (equation 15) are always positive. **(B)** Reaction norms in the right column fan out for one trait, and fan in for the second trait. In this case, plasticity syndromes measured with individual differences (equation 15) are always negative.

73

## 74 Supplement to section 4: summary statistics

### 75 Reversing non-linear transformations

76 Non-linear transformations of the response variable are commonly used to improve the  
 77 normality of residuals. The assumption of normality of residuals is particularly important in  
 78 a DHGLM, because skewed residuals will inflate heterogeneity in within-individual  
 79 variances (i.e.  $\sigma_{ID_{v0,exp}}^2$ ), and therefore affect the estimation of both  $R_p$  and  $CV_{ID}$  (Yang et  
 80 al., 2011). Three transformations are commonly used to improve residual normality: (1) the  
 81 ln transformation; (2) the  $\log_{10}$  transformation; and (3) the square-root transformation. In  
 82 the main text we saw how to transform back from the ln-normal distribution (equations 36  
 83 and 37 in the main text). To convert back from  $\log_{10}$ , we can use the same formulas, after  
 84 first converting means and variances from the  $\log_{10}$  to the ln scale:

$$85 \ln(10)\mu_{\log_{10}(y)} = \mu_{y,exp} \quad \text{eqn 45}$$

$$86 \ln(10)^2\sigma_{\log_{10}(y)}^2 = \sigma_{y,exp}^2 \quad \text{eqn 46}$$

87  
 88 To reverse the square-root transformation, we can use an approximation from second  
 89 order Taylor expansions (also known as the delta method). The general formulas for  
 90 converting the mean and variance using the second order Taylor expansions are (Oehlert,  
 91 1992):

92  $\mu_y \approx f(\mu_{f^{-1}(y)}) + \frac{1}{2}\sigma_{f^{-1}(y)}^2 f'(\mu_{f^{-1}(y)})'',$  eqn 47

93  $\sigma_y^2 \approx \sigma_{f^{-1}(y)}^2 (f'(\mu_{f^{-1}(y)}))'^2 + \frac{1}{2}\sigma_{f^{-1}(y)}^4 (f''(\mu_{f^{-1}(y)}))'^2,$  eqn 48

94 where  $f$  and  $f^{-1}$  are the inverse functions of each other (i.e., transformation and back-  
 95 transformation), and  $f'$  and  $f''$  are the first and second-order derivatives, respectively. In  
 96 our case,  $f$  is taking square roots so that  $f(\mu_{f^{-1}(y)}) = \mu_{f^{-1}(y)}^2$ . Therefore,  $f'(\mu_{f^{-1}(y)})' =$   
 97  $2\mu_{f^{-1}(y)}$ , and  $f''(\mu_{f^{-1}(y)})'' = 2$ . By substituting these values into equation 45 and equation  
 98 46, we find:

99  $\mu_y \approx \mu_{y,sq}^2 + \sigma_{y,sq}^2,$  eqn 49

100  $\sigma_y^2 \approx 4\sigma_{y,sq}^2\mu_{y,sq}^2 + 2\sigma_{y,sq}^4.$  eqn 50

101 Importantly, formulas for reversing a non-linear transformation are only valid for the total  
 102 variance on the transformed scale. Therefore, to obtain the individual variance  
 103 component, we need to first back-transform the total phenotypic variance, and then  
 104 calculate the proportion of this variance that is due to the individual component  
 105 (assuming that the proportion of variance is preserved across scale types, as we did in the  
 106 main text for estimating  $\sigma_{ID_{v_0}}^2$  in equation 44).

## 107 Coefficient of individual variation for the mean model

108 If no transformations have been applied to the response variable,  $y$ , then we can take the  
 109 standard deviation directly from the between-individual variance component, such that  
 110  $CV_{IDm} = \frac{\sigma_{IDm}}{\beta_{pm0}}$ , where  $\sigma_{IDm}$  is described by equation 39 (main text), and the population  
 111 mean is calculated for an even sex ratio at the average age of the population ( $\beta_{pm0} =$   
 112  $\frac{2\beta_{m0} + \beta_{m1}}{2}$ ). In the worked example, all our response variables have been z-transformed  
 113 prior to analysis, so they need to be brought back to the observed scale as  $\mu_p = \beta_{pm0}\sigma_y +$   
 114  $\mu_y$  and  $\sigma_{ID\mu_p} = \sigma_{IDm}\sigma_y$  (thus returning  $y$  to the ratio scale, with a lower bound at zero).

115  
 116 In the below worked example, two of the response variables were square-root

transformed (a non-linear transformation) prior to being  $z$ -transformed. In this case, returning estimates to the ratio scale involves five steps: (1) the mean is brought back from the  $z$ -sq-scale to the sq-scale,  $\mu_{p,sq} = \beta_{pm0,sq}\sigma_{y,sq} + \mu_{y,sq}$ ; (2) the mean is brought back from the sq-scale, using the approximation from equation 47,  $\mu_p \approx \beta_{pm0,sq}^2 + \sigma_{p,sq}^2$ ; (3) the total phenotypic standard deviation ( $\sqrt{\sigma_p^2}$ , from equation 38) is brought back from the  $z$ -sq-scale to the sq-scale,  $\sigma_{p,sq} = \sigma_{y,sq}\sqrt{\sigma_{ID_{m,sq}}^2 + \sigma_{fixed_{m,sq}}^2 + \bar{\sigma}_{w,sq}^2}$ ; (4) the total phenotypic standard deviation is converted to a variance on the ratio scale, using the approximation from equation 48:  $\sigma_p^2 \approx 4\sigma_{p,sq}^2\mu_{p,sq} + 2\sigma_{p,sq}^4$ ; (5) we extract the individual variance component,  $\sigma_{ID_{\mu_p}}$ , by assuming the proportion of variance components are preserved across scales, such that:  $\sigma_{ID_{\mu_p}} = \sigma_p^2 \left( \frac{\sigma_{ID_{m,sq}}^2}{\sigma_{ID_{m,sq}}^2 + \sigma_{fixed_{m,sq}}^2 + \bar{\sigma}_{w,sq}^2} \right)$ .

### 4.3| THE COEFFICIENT OF VARIATION IN PREDICTABILITY

Cleasby et al. (2015) presents an analogous statistic to  $CV_{IDv}$ , called the 'coefficient of variation in predictability' ( $CV_{IDSd}$ ).  $CV_{IDSd}$  was introduced as a unitless measure, to facilitate comparative analyses of individual differences in predictability. There are two main differences between  $CV_{IDv}$  and  $CV_{IDSd}$ : (1) In Cleasby et al., the dispersion model uses residual standard deviations as the response variable, rather than residual variances (so they estimate  $CV_{IDSd}$  from  $\sigma_{ID_{sd}}$  and  $\bar{\sigma}_w$ , rather than  $CV_{IDv}$  from  $\sigma_{ID_v}$  and  $\bar{\sigma}_w^2$ ). Currently  $\sigma_{ID_{sd}}$  is the default when fitting a DHGLM in the R package '*brms*' [Bürkner, 2018]), and; (2) The formula for  $CV_{IDSd}$  (eqn 17 in Cleasby et al.) is applicable to a DHGLM that has a single variance component (i.e. only one random intercept, for individual identity. Variance due to fixed effects is ignored). In this simple case,  $CV_{IDSd} = \frac{\sigma_{ID_{sd}}}{\bar{\sigma}_w}$  is calculated by simply converting  $\sigma_{ID_{sd,exp}}$  and  $\bar{\sigma}_{w,exp}$  back from the ln-scale, using the conversion formulas presenting in the main text (equations 36 and 37):

$$CV_{IDSd} = \frac{\sqrt{\sigma_{ID_{sd}}^2}}{\bar{\sigma}_w} = \frac{\sqrt{(\exp(\sigma_{ID_{sd,exp}}^2) - 1)\exp(2\bar{\sigma}_{w,exp} + \sigma_{ID_{sd,exp}}^2)}}{\exp\left(\bar{\sigma}_{w,exp} + \frac{\sigma_{ID_{sd,exp}}^2}{2}\right)}.$$

This fraction simplifies to  $CV_{IDSd} = \sqrt{\exp(\sigma_{ID_{sd,exp}}^2) - 1}$ .

To synthesise estimates of the coefficient of individual variation from different studies, we can convert  $CV_{IDv}$  to  $CV_{IDsd}$ , and vice versa, using basic properties of logarithms and variances. The two dispersion models are related to each other as  $\ln(\sigma_{e_{ij}}) = \frac{1}{2} \ln(\sigma_{e_{ij}}^2)$ . Therefore, we can convert the mean standard deviation into the mean variance by multiplying it by two:  $2\bar{\sigma}_{w,\exp} = \bar{\sigma}_w^2$ . Similarly, a variance estimated from the  $\ln$  standard deviation model can be converted to the  $\ln$  variance model as:  $4\sigma_{ID_{sd,\exp}}^2 = \sigma_{ID_{v,\exp}}^2$ . By modifying the main text equations 38 ( $\bar{\sigma}_w^2$ ), 42 ( $\sigma_{\bar{\sigma}_w}^2$ ), and 44 ( $\sigma_{ID_{v0}}^2$ ), and assuming the same model structure used throughout this supplement, we present the following conversion formulas for  $CV_{IDv}$  to  $CV_{IDsd}$ :

$$CV_{IDsd} = \frac{\sqrt{\sigma_{ID_{sd}}^2}}{\bar{\sigma}_w}, \quad \text{eqn 51}$$

$$\sigma_{\bar{\sigma}_w}^2 = \left( \exp\left( \frac{\sigma_{ID_{v0,\exp}}^2 + \sigma_{fixed_{v,\exp}}^2}{4} \right) - 1 \right) \exp\left( \beta_{pv0} + \frac{\sigma_{ID_{v0,\exp}}^2 + \sigma_{fixed_{v,\exp}}^2}{4} \right), \quad \text{eqn 52}$$

$$\bar{\sigma}_w = \exp\left( \frac{\beta_{pv0,\exp}}{2} + \frac{\sigma_{ID_{v0,\exp}}^2 + \sigma_{fixed_{v,\exp}}^2}{8} \right), \quad \text{eqn 53}$$

$$\sigma_{ID_{sd}}^2 = \sigma_{ID_{sd0}}^2 = \sigma_{\bar{\sigma}_w}^2 \left( \frac{\sigma_{ID_{v0,\exp}}^2}{\sigma_{ID_{v0,\exp}}^2 + \sigma_{fixed_{v,\exp}}^2} \right), \quad \text{eqn 54}$$

where once again we are converting a variance component by assuming the preservation

of proportionality, such that  $\frac{\sigma_{ID_{v0,\exp}}^2}{\sigma_{ID_{v0,\exp}}^2 + \sigma_{fixed_{v,\exp}}^2} = \frac{\sigma_{ID_{sd0,\exp}}^2}{\sigma_{ID_{sd0,\exp}}^2 + \sigma_{fixed_{sd,\exp}}^2}$ .

158

159 For converting  $CV_{IDsd}$  to  $CV_{IDv}$ , we multiply, rather than divide, the mean and variance  
160 from the dispersion model by 2 and 4, respectively, so that:

$$CV_{IDv} = \frac{\sigma_{IDv}}{\bar{\sigma}_w}, \quad \text{eqn 55}$$

$$\bar{\sigma}_w^2 = \exp\left( 2\beta_{psd0,\exp} + 2\left( \sigma_{ID_{sd0,\exp}}^2 + \sigma_{fixed_{sd,\exp}}^2 \right) \right), \quad \text{eqn 56}$$

$$\sigma_{\bar{\sigma}_w}^2 = \left( \exp\left( 4\left[ \sigma_{ID_{sd0,\exp}}^2 + \sigma_{fixed_{sd,\exp}}^2 \right] \right) - 1 \right) \exp\left( 4\left[ \beta_{psd0,\exp} + \sigma_{ID_{sd0,\exp}}^2 + \sigma_{fixed_{sd,\exp}}^2 \right] \right), \quad \text{eqn 57}$$

$$\sigma_{IDv}^2 = \sigma_{ID_{v0}}^2 = \sigma_{\bar{\sigma}_w}^2 \left( \frac{\sigma_{ID_{sd0,\exp}}^2}{\sigma_{ID_{sd0,\exp}}^2 + \sigma_{fixed_{sd,\exp}}^2} \right). \quad \text{eqn 58}$$

166

167

## 168 Supplementary Worked Example

### 169 Data collection

170 To demonstrate the models and calculations described here and in the main text, we use  
171 behaviour data from a laboratory population of 248 adult zebrafish (*Danio rerio*) following  
172 protocols developed by Fangmeier et al. (2018).

173

174 We conducted behavioural trials over six months, starting when zebrafish were around  
175 four months old, measuring each individual up to 12 times (with an average of two weeks  
176 between trials). Briefly, in each trial we recorded the movements of a single zebrafish in a  
177 rectangular tank (40cm long, 20cm wide), using an overhead camera. The sides of the tank  
178 were opaque, but the ends were transparent, so that the fish could see computer tablets  
179 through both ends: one screen was always blank (the control), and one displayed a  
180 stimulus intermittently.

181

182 During a trial, to elicit behavioural responses from the focal zebrafish, we played four  
183 different stimulus videos lasting three minutes each: (1) a video of a single aggressive  
184 zebrafish (eliciting aggression); (2) a video of a zebrafish shoal (eliciting sociability); (3) a  
185 novel object animation (eliciting neophobia, i.e. aversion to a novel object); (4) a predator  
186 animation (eliciting a risk response). Between each stimulus period, the stimulus screen  
187 was blank to match the control. The order in which the four stimuli were presented was  
188 pseudo-randomised, and we had 24 versions of each stimulus video so that zebrafish did  
189 not see the same video twice. Each trial lasted 34 minutes and, after an acclimation period  
190 (3 minutes), consisted of four repetitions of the following sequence: pre-stimulus (3  
191 minutes), stimulus (3 minutes), post-stimulus (1 minute). At the end of this sequence there  
192 was a post-assay period (3 minutes).

193

194 From each behavioural trial, we analysed five behavioural traits (measured using  
195 EthoVision XT 11.5 tracking software): (1) aggression: time spent within 5 cm of the  
196 aggressor stimulus; (2) sociability: time spent within 5 cm of the shoal stimulus;  
197 (3) neophilia: time spent within 10 cm of the novel object stimulus; (4) boldness: time  
198 spent within 10 cm of the predator stimulus; (5) activity: the average distance travelled

199 without stimulus present, taken from the four pre-stimulus periods and one post-assay  
200 period.

201

## 202 **Analysis methods**

203 Data and code to reproduce results from this worked example are available from  
204 <http://doi.org/10.17605/OSF.IO/V3QAX> (O'Dea et al., 2020). Models were fit in a Bayesian  
205 statistical program, Stan (Carpenter et al., 2017), accessed through the *R* environment (R  
206 Core Team 2020) using the '*stan*' function from the '*RStan*' package (v. 2.21.2; Stan  
207 Development Team, 2020). To reduce computing time, we ran Stan models with an  
208 external computer cluster ('Katana' at the University of New South Wales), which was  
209 loaded with a module for compiling C++ tool chains ('gcc/7.5.0'), and *R* v. 4.0.2-gcc7 (a  
210 version of *R* including the GUN complier collection, GCC). We reproduced a subset of  
211 models in the popular *R* package '*brms*' (v. 2.15.0; Bürkner, 2018), which is a relatively user-  
212 friendly interface to Stan, to demonstrate conversions between the log variance and log  
213 standard deviation dispersion models.

## 214 **Data preparation**

215 We performed initial data checks by plotting the distribution of residuals from intercept-  
216 only mixed models, with a random intercept for each individual (we ran these models  
217 using the *lmer* function in the *lme4* package, v. 1.1-18-1; Bates et al., 2015). Neophilia and  
218 boldness were strongly positively skewed, due to most fish spending little to no time  
219 within 10 cm of the stimulus (reflecting aversive behavioural responses to the 'risky'  
220 stimulus zone). We therefore analysed neophilia and boldness on the square-root scale,  
221 which greatly improved the normality of residuals. In addition, the five response variables  
222 were z-transformed prior to analysis (to put them all in standard deviation units, with a  
223 mean of zero, making it easier to re-use the same model code for different data). To help  
224 with biological interpretation of model estimates, the unscaled descriptive statistics were  
225 used for back-transformation from the z-scale (back-transformations are described in the  
226 proceeding section, and shown in Fig. S2, below). Next, we created data lists to feed into  
227 each of the six models described below (under 'stan models'). We ran each of the  
228 univariate models five times (i.e. for the five different response variables), and both  
229 bivariate models were each run ten times (for each possible combination of response

230 variables).

231 **Stan models**

232 For illustrative purposes, we specified six different models: (1) random intercept in mean  
233 model only (personality; single hierarchical, as in equations 1-4); (2) random intercept and  
234 random slope in mean model (personality and plasticity; single hierarchical, as in  
235 equations 5-7); (3) bivariate model, random intercept and random slope in mean model  
236 only (personality and plasticity syndromes and associations; single hierarchical, as in  
237 equations 10-13); (4) random intercept in both the mean and dispersion models  
238 (personality and predictability; double hierarchical); (5) random intercept in both the  
239 mean and dispersion models, random slope in the mean model only (personality,  
240 plasticity, and predictability; double hierarchical, as in equations 17-20); (6) bivariate  
241 model, random intercept in both the mean and dispersion models, random slope in the  
242 mean model only (personality, plasticity, and predictability syndromes and associations;  
243 double hierarchical, as in equations 25-30). Estimates of repeatability ( $R_{Pm}$  and  $R_{Pv}$ ) and  
244 the coefficient of individual variation ( $CV_{IDm}$  and  $CV_{IDv}$ ) were made using the formulas  
245 presented in the main text (with small modifications for single hierarchical models,  
246 available in the supplementary script 'Step 3 - Process Models.R').

247

248 Using the 'stan' function from the 'RStan' package requires two arguments: data for the  
249 model, stored as a list, and an external text file of instructions for how Stan should analyse  
250 that data, saved with the '.stan' file extension. The stan code is written as six structured  
251 'chunks', each contained within curly brackets. First, the 'data' chunk defines the raw  
252 variables that were specified in the data list (e.g. sample sizes, vectors, and design  
253 matrices). Second, the 'transformed data' chunk performs transformations of the raw data  
254 necessary to run the model. Third, the 'parameters' chunk defines parameters that will be  
255 estimated in the model (e.g. intercept, residual variance). Fourth, 'transformed parameters'  
256 are used to estimate each of the individual differences (i.e. BLUPs). Fifth, the 'model' chunk  
257 defines the model, priors, and likelihoods. Sixth, and finally, the 'generated quantities'  
258 chunk can be used to calculate parameters from the results of the model. In all models, we  
259 included a random intercept for each individual, a categorical fixed effect for sex of the  
260 individual (female = 0; male = 1), and a continuous fixed effect for the age of the individual  
261 (z-transformed).

262

263 The likelihood functions for all models were based on a normal distribution. For priors of  
264 population parameters, we used the log of the student *t* distribution with degrees of  
265 freedom = 3, location (mean) = 0, and scale (variance) = 10. The priors of individual  
266 differences were based on a normal distribution, with a mean of zero and variance of 1.  
267 Covariances were modelled using Cholesky factors of correlation matrices, specifying an  
268 LKJ (Lewandowski-Kurowicka-Joe) correlation distribution (uniform probability between -  
269 1 and 1). All models were fit using 3 chains, each with 6,000 iterations and a warm-up of  
270 2,000 iterations.

271

## 272 **Model checks and calculations**

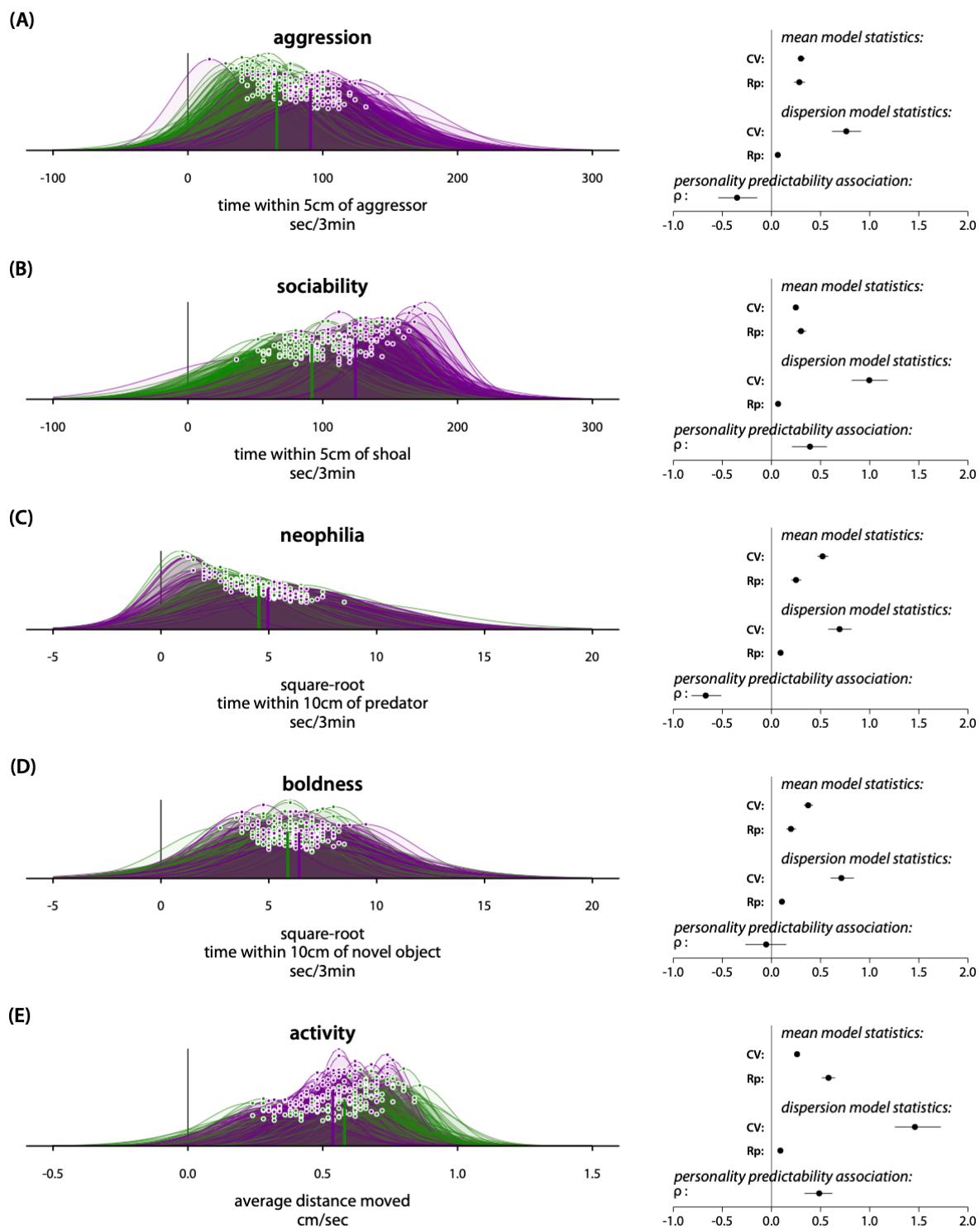
273 Estimates of repeatability and the coefficient of individual variation were made using the  
274 formulas presented in the main text and above (with small modifications for single  
275 hierarchical models, available in the supplementary script 'Step 3 - Process Models.R'). We  
276 performed four types of model checks. First, for the key parameter estimates for each  
277 model, we visually assessed mixing and convergence of all Markov chains with trace-plots  
278 (by plotting iteration on the x-axis and the value of the parameter on the y-axis). Second,  
279 we quantified the consistency of the chains with the split R-hat statistic, which was below  
280 1.01 for the key parameters from almost all models, and the effective sample sizes were  
281 above 1,000. Third, we calculated residuals for each model (based on posterior means) and  
282 inspected histogram plots to confirm residuals were approximately normally distributed.  
283 Fourth, for the response values ('observed'), we calculated corresponding model  
284 predictions ('predicted') and squared the Pearson correlation between observed and  
285 predicted values as an indicator of model fit ( $R^2$ ).

286

287 **Results and discussion: worked example**

288 Full results are available in the Results Tables S1, S2, and S3, and a selection of diagnostic  
289 plots are shown below in Diagnostic Figures SD01-SD14. Statistical significance was  
290 determined from 95% credible intervals not crossing zero. Descriptive statistics ( $\mu \pm \sigma$ ) for  
291 the five behavioural traits were: (1) aggression =  $79.3 \pm 49.3$  seconds spent within 5 cm of  
292 aggressor video; (2) sociability =  $109.0 \pm 54.1$  seconds spent within 5 cm of shoal video; (3)  
293 neophilia (square-root transformed) =  $4.75 \pm 3.48$  seconds spent within 10 cm of novel  
294 object animation; (4) boldness (square-root transformed) =  $6.16 \pm 3.18$  seconds spent  
295 within 10 cm of predator animation; (5) activity =  $0.559 \pm 0.217$  cm moved in 1 second.

296



297

298 **Figure S2**

299 Individual differences, summary statistics, and sex differences in five behaviours,  
 300 estimated from univariate double hierarchical linear models, with a random intercept in  
 301 mean and dispersion (i.e. personality and predictability), and a random slope for the effect  
 302 of age (i.e. plasticity). **Left-hand panels** show simulated distributions for the behaviour of

303  $N_{ID} = 248$  individuals. Females = **purple**; males = **green**. Purple and green vertical lines  
 304 show the population averages for female and male zebrafish, respectively. Filled points  
 305 indicate the mean and peak of each individual's distribution; taller peaks indicate a  
 306 narrower distribution, and therefore a smaller variance (i.e. greater predictability). Each  
 307 individual's distribution is simulated from the posterior mean estimate of their phenotypic  
 308 mean and variance, which were back-transformed from the z-scale to the raw scale  
 309 (within-individual variances were first converted back from the ln- scale by taking the  
 310 exponent). **Right-hand panels** show summary statistics (posterior means and 95%  
 311 credible intervals) for the magnitude of individual differences in both the mean and  
 312 dispersion models ( $CV$  = coefficient of individual variation;  $R_p$  = repeatability; see main  
 313 text), and the personality predictability association for each behaviour (i.e. the sign-  
 314 reversed correlation between individual intercepts in the mean and dispersion model;  
 315 equation 22).

316

### 317 Individual differences in personalities

318 The average behaviour of individuals varied substantially, as can be seen by the spread of  
 319 distributions in Fig. S2. From the coefficient of individual variation, relative to the  
 320 population mean, neophilia and boldness showed the most individual differences in  
 321 average behaviours ( $CV_{IDm} \approx 0.5$  and 0.4, respectively). Note that neophilia was positively  
 322 skewed, even after the square-root transformation, which likely over-estimated  $CV_{IDm}$  (Fig.  
 323 S2C; the importance of normality for DHGLMs is discussed in the main text). While  
 324 aggression, sociability, and activity had a similar magnitude of individual differences  
 325 ( $CV_{IDm} \approx 0.3$ ), those differences were the most consistent for activity ( $R_{pm} \approx 0.6$ ; Fig. S2E).  
 326 The consistencies of individual differences were similar for sociability, aggression,  
 327 neophilia, and boldness ( $R_{pm} \approx 0.2-0.3$ ). Estimates of repeatability from the single  
 328 hierarchical models were around 5-10% lower than those from DHGLMs, suggesting that  
 329 when homoscedasticity is assumed, the total estimate of residual variance can be inflated,  
 330 and therefore underestimate repeatability.

331

332 Males were significantly more active, less aggressive and social, and less bold and  
 333 neophilic than females (Fig. S3; Table S1). Previous studies on zebrafish have also found  
 334 male zebrafish to be more active (e.g. Dereje et al., 2012; Moretz et al., 2007; Mustafa et al.,

335 2019), but males are also generally reported to be bolder than females (e.g. Dahlbom et  
336 al., 2011; Dereje et al., 2012; Kern et al., 2016). However, estimates of sex differences can  
337 depend on both the type of behavioural assays, and the zebrafish strain (Mustafa et al.,  
338 2019). Often measures of 'boldness' are conflated with activity levels (e.g. distanced  
339 moved in an open arena). Here, it is conceivable that female zebrafish, who were less  
340 active than males, happened to settle close to the stimulus screen, although across  
341 individuals more activity was correlated with more time near the predator stimulus (e.g.  
342 positive activity-boldness correlation).

343

#### 344 **Zebrafish behavioural syndromes**

345 While zebrafish studies have found mixed results for a total (i.e. un-partitioned)  
346 phenotypic correlation between aggression and boldness (Martins & Bhat, 2014; Norton &  
347 Bally-Cuif, 2012), the among-individual correlations we estimated here provide evidence  
348 for proactive behavioural syndromes. Statistically significant correlations between mean  
349 individual intercepts (equation 14) were all in the positive direction (Fig. S4). Strong  
350 correlations existed between neophilia and boldness, and between aggression and  
351 sociability. Moderate correlations were seen between neophilia and aggression, neophilia  
352 and sociability, and aggression and boldness. In addition to the moderate positive  
353 correlation between activity and boldness, there was a weaker negative correlation  
354 between activity and sociability (albeit this negative correlation was not statistically  
355 significant). Activity was not significantly correlated with aggression or neophilia.

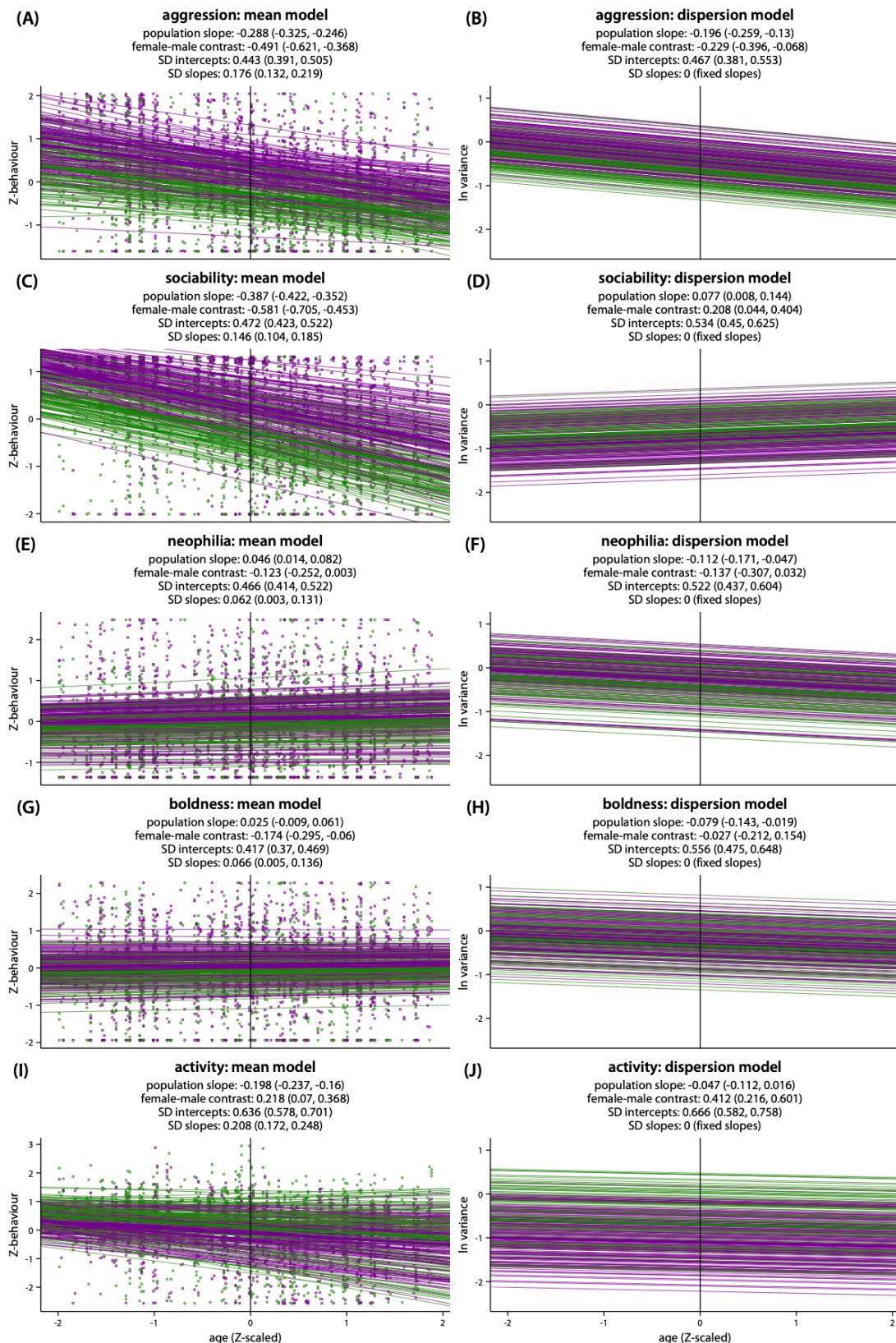
#### 356 **Individual differences in predictability of behaviour**

357 All measured behaviours showed considerable heteroscedasticity, which is depicted in  
358 Fig. S2 as variability in the heights of individual peaks. The magnitude of individual  
359 differences in predictability was most pronounced in activity ( $CV_{IDv} \approx 1.25$ ) and least  
360 pronounced for aggression, neophilia, and boldness ( $CV_{IDv} \approx 0.6$ ) (Table S2). For  
361 comparisons to studies that model individual differences in residual standard deviations,  
362 rather than variances, this translates into a range of  $CV_{IDsd}$  from 0.3 to 0.5 (the main text  
363 and supplementary code contain conversion formulas), which is in the range of estimates  
364 seen over shorter timespans for guppies and zebrafish (Mitchell & Biro, 2017; Mitchell et  
365 al., 2016). As with mean differences, the behaviour with the greatest magnitude of

366 individual differences in predictability did not have the greatest *consistency* of differences,  
367 as measured by repeatability. Individual differences in predictability were most consistent  
368 for boldness ( $R_{pv} \approx 0.11$ ), although the other behaviours were not far behind ( $R_{pv} \approx 0.06 -$   
369 0.09).  
370

### 371 **Zebrafish personality-plasticity associations and plasticity syndromes**

372 Over the course of the experiment zebrafish became, on average, less active, aggressive,  
373 and sociable, but tended to become more neophilic and bold (Fig. S3). The plastic effect of  
374 age could be driven by habituation (Biro, 2012). Plasticity in sociability (Fig. S3C) was the  
375 most integrated with other individual differences. Individuals with more active and more  
376 sociable personalities were less plastic (moderate negative personality-plasticity  
377 associations; equation 9, Fig. 3D), which could provide some support for predictions by  
378 Dubois (2019) (i.e. under the assumption that proactive individuals are less capable of  
379 accurately assessing their environment, due to the higher cognitive demands of  
380 proactivity, a negative correlation is predicted between proactive personalities and  
381 adaptive plasticity). However, the magnitudes of plasticity in aggression, neophilia, and  
382 boldness were unrelated to individual personalities (Fig. S4, green diagonal: A,F,J,M,O). For  
383 plasticity syndromes (equation 16, Fig. 3B), individuals who were more plastic in sociability  
384 also tended to be more plastic in aggression, and less plastic in activity (moderately  
385 positive and negative plasticity syndromes, respectively; Fig. S4, blue boxes). No other  
386 combinations of traits showed statistically significant plasticity syndromes. Within the  
387 same system, therefore, we found all possible outcomes for plasticity syndromes —  
388 position, negative, and null relationships — suggesting individual differences in plasticity  
389 are underpinned by multiple mechanisms.



390

**Figure S3**

391 Individual differences in personality, plasticity, and predictability. **Purple** = female  
 392 zebrafish, and **green** = male zebrafish. In left-hand panels, points show the raw values of  
 393

zebrafish behaviour across age. Population estimates for male and female zebrafish are shown as thick green and purple lines, respectively. **Personality:** the left-hand column shows individual differences in intercepts (i.e. the point at which the sloped lines cross the vertical line at zero). **Plasticity:** the left-hand column shows individual differences in slopes (i.e. the change in average behaviour with zebrafish age). **Predictability:** the right-hand column shows individual intercepts for within-individual variances, and the population-level change in predictability with time.

401

## 402 **Zebrafish predictabilities and predictability syndromes**

Male zebrafish were significantly less predictable than female zebrafish in activity and sociability (i.e. greater within-individual variance), and more predictable in aggression (i.e. smaller within-individual variance) (Fig. S3, right column; Results Table S1). There were no statistically significant sex differences in predictability for neophilia (Fig. S3 F) or boldness (Fig. S3H). Predictability tended to increase with age for aggression (Fig. S3B), neophilia (Fig. S3F), and boldness (Fig. S3H), as seen by negative slopes for age against within-individual variance. In contrast, sociability became less predictable with time (Fig. S3D), and there was no change in the predictability of activity with time (Fig. S3J). From bivariate models, we found moderate-to-strong positive predictability syndromes for all combinations of traits (see blue boxes in Fig. S4, derived from equation 31; see also Fig. 3C and Table S3). In contrast to plasticity, therefore, predictability was strongly integrated among behavioural traits. Positive predictability syndromes could be indicative of mechanistic constraints, which limit the independent evolution of individual's predictability across traits, or correlated selective pressures, such that individuals are selected to be more or less predictable for a suite of related traits. Quantitative genetic methods can be incorporated into studies of predictability to determine the relative contribution of environmental and genetic effects on individual differences in predictability (e.g. Martin et al., 2017; Prentice et al., 2020).

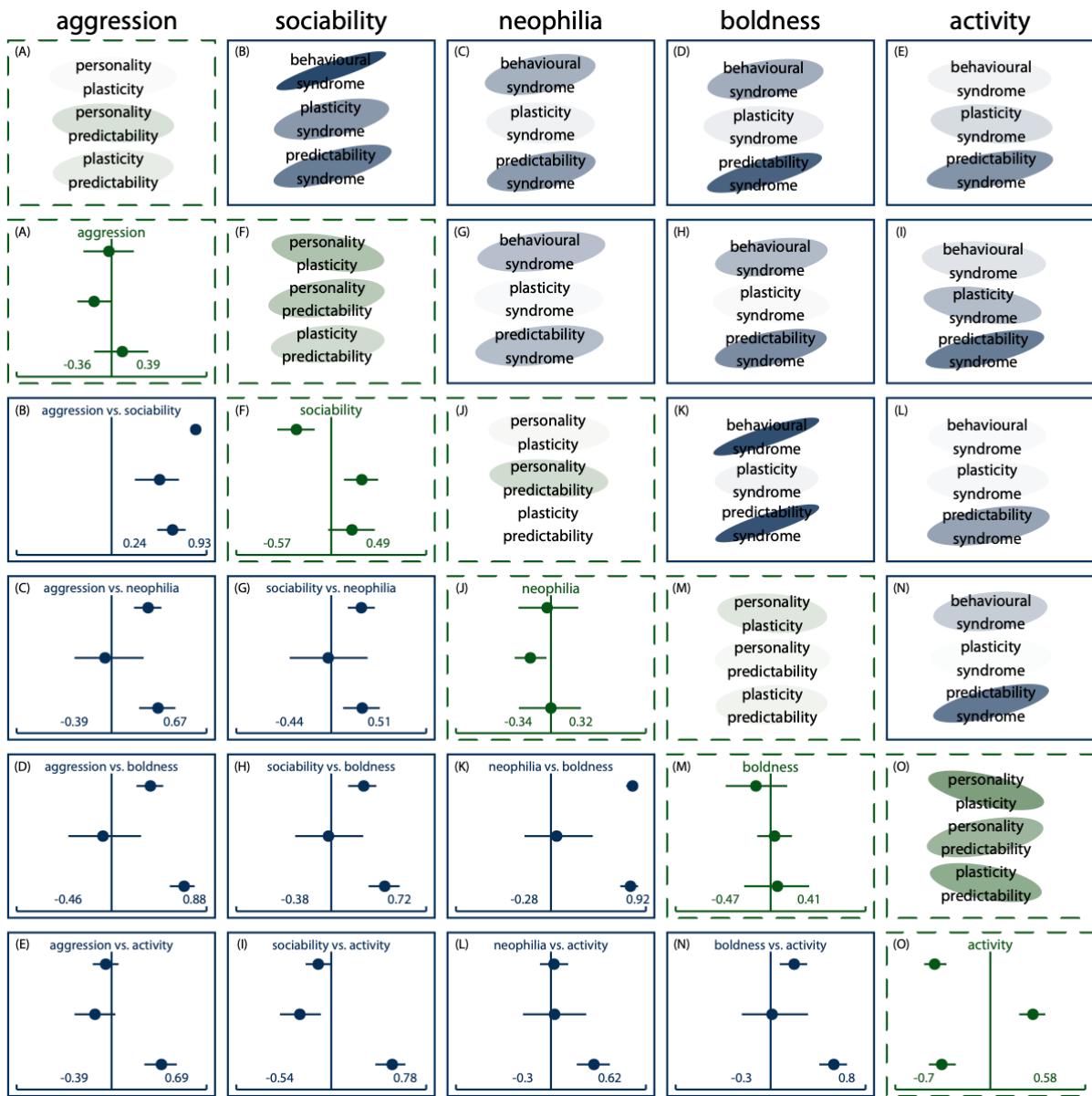
## 421 **Zebrafish personality-predictability associations**

We found both moderately positive, weakly negative, negative, and null personality-predictability associations (Fig. 3E; Fig. S4, green diagonal; equation 22). More active and sociable zebrafish also tended to be more predictable, as measured by a narrow

425 distribution in behaviours. In contrast, individuals who were more aggressive and  
426 neophilic tended to be less predictable in their behaviour. Boldness did not show a  
427 significant personality-predictability association (Fig. S4M). Therefore, there was no clear  
428 trend for riskier personality types to also be less predictable, or for more extreme  
429 personalities to show greater precision around their mean.

430 **Zebrafish plasticity-predictability associations**

431 Only activity and sociability showed an association between plasticity and predictability  
432 (equation 24; Fig. 3F). This pattern is similar to the personality-plasticity associations,  
433 except in this case sociability and activity showed correlations in opposite directions.  
434 Individuals who were more predictable in activity were less plastic (i.e. had a smaller  
435 change in activity as they aged), whereas individuals who were more predictable in  
436 sociability showed a greater change in sociability as they aged. The remaining behaviours  
437 (aggression, neophilia, and boldness) did not show significant plasticity-predictability  
438 associations (Fig. S4; Results Table S3). Given that we found strong phenotypic integration  
439 of predictability (in positive predictability syndromes), but variable or null integration of  
440 plasticity (in positive, negative, and null predictability syndromes), the variable  
441 relationships between plasticity and predictability were likely driven by independence of  
442 plasticity between traits, rather than flexibility in predictability.



443

444 **Figure S4**

445 Six types of correlations between and within personality, plasticity, and predictability,  
 446 estimated from bivariate double hierarchical linear models on five behavioural traits.  
 447 Diagonal boxes, in green, show within-trait associations (averaged across the four  
 448 bivariate models featuring each focal trait). Off-diagonal boxes, in blue, show between-  
 449 trait correlations. In the **upper triangle**, each ellipse represents a correlation; ellipses  
 450 tilted to the right represent positive correlations (i.e.  $r > 0$ ), and ellipses tilted to the left  
 451 represent negative correlations (i.e.  $r < 0$ ). Narrower ellipses, with more opaque shading,  
 452 represent stronger correlations. Ellipses with no tilt, and more transparent colouration,  
 453 represent little or no correlation (i.e.  $r = 0$ ). The **lower triangle** shows posterior mean

454 estimates and 95% credible intervals for each correlation (bound between -1 and 0), with  
 455 the range of lower and upper estimates shown on the x-axis. Personality-plasticity  
 456 associations and plasticity syndromes were calculated from slope magnitudes (the  
 457 absolute value of individual slopes; equations 9 and 16). Equivalent correlations for slope  
 458 deviations (equations 8 and 15) are available from the supplementary code.

## 459 Results Tables

### 460 Supplementary Table S1

461 *Parameter estimates from univariate models with and without predictability.*

462 Sample sizes for all models are  $N_i = 2,766$  observations, taken from  $N_j = 248$  individuals.  
 463 Models were either 'SHGLM' (Single Hierarchical Generalized Linear Model: assumes  
 464 homoscedasticity of residual variances), or 'DHGLM' (Double Hierarchical Generalized  
 465 Linear Model: models heteroscedasticity). Estimates with 95% credible intervals crossing  
 466 zero are shown in red font. 'Residual standard deviation' of the mean model for a DHGLM  
 467 is  $\sqrt{\sigma_w^2}$ , from equation 38 in the main text. The five response variables were analysed on  
 468 the z-scale (with a mean of zero and standard deviation of one). The un-scaled variables  
 469 are described by the following statistics (mean  $\pm$  SD): activity (average distance travelled):  
 470  $0.559 \pm 0.217$  cm/sec; aggression (time spent within 5 cm of aggressor stimulus):  $79.3 \pm$   
 471  $49.3$  sec; sociability (time spent within 5 cm of social stimulus):  $109.0 \pm 54.1$  sec; neophilia  
 472 (square-root of time spent within 10cm of novel object stimulus):  $4.75 \pm 3.48$  sec; boldness  
 473 (square-root of time spent within 10cm of predator stimulus)  $6.16 \pm 3.18$  sec.

<b>Hierarchy</b>	<b>Coefficient</b>	<b>Behaviour</b>	<b>Model</b>	model with random intercept			model with random intercept & slope		
				<b>Mean</b>	<b>Cl.lb</b>	<b>Cl.ub</b>	<b>Mean</b>	<b>Cl.lb</b>	<b>Cl.ub</b>
mean model	intercept	activity	SHGLM	-0.104	-0.223	0.016	-0.121	-0.239	-0.011
			DHGLM	-0.114	-0.230	-0.002	-0.109	-0.215	-0.008
	aggression		SHGLM	0.242	0.155	0.327	0.238	0.156	0.315
			DHGLM	0.235	0.146	0.320	0.228	0.138	0.317
	neophilia		SHGLM	0.060	-0.034	0.147	0.059	-0.032	0.146
			DHGLM	0.061	-0.026	0.152	0.063	-0.029	0.150
	boldness		SHGLM	0.077	-0.014	0.157	0.079	0.002	0.162
			DHGLM	0.076	-0.013	0.162	0.080	-0.006	0.165

	sociability	SHGLM	0.287	0.195	0.373	0.286	0.200	0.382
		DHGLM	0.288	0.198	0.380	0.278	0.191	0.368
female-male contrast	activity	SHGLM	0.205	0.032	0.371	0.240	0.087	0.395
		DHGLM	0.225	0.052	0.390	0.220	0.069	0.365
	aggression	SHGLM	-0.513	-0.642	-0.379	-0.504	-0.633	-0.389
		DHGLM	-0.498	-0.629	-0.360	-0.490	-0.612	-0.357
	neophilia	SHGLM	-0.132	-0.252	0.002	-0.137	-0.267	-0.016
		DHGLM	-0.126	-0.260	0.005	-0.135	-0.267	0.006
	boldness	SHGLM	-0.166	-0.287	-0.037	-0.170	-0.283	-0.052
		DHGLM	-0.164	-0.287	-0.034	-0.172	-0.296	-0.046
	sociability	SHGLM	-0.594	-0.721	-0.464	-0.597	-0.739	-0.473
		DHGLM	-0.601	-0.730	-0.464	-0.584	-0.719	-0.453
age slope	activity	SHGLM	-0.207	-0.234	-0.180	-0.206	-0.243	-0.168
		DHGLM	-0.135	-0.162	-0.111	-0.198	-0.236	-0.159
	aggression	SHGLM	-0.308	-0.339	-0.276	-0.306	-0.345	-0.265
		DHGLM	-0.298	-0.327	-0.266	-0.288	-0.324	-0.249
	neophilia	SHGLM	0.049	0.015	0.084	0.049	0.013	0.084
		DHGLM	0.040	0.010	0.073	0.045	0.012	0.079
	boldness	SHGLM	<b>0.024</b>	<b>-0.010</b>	<b>0.058</b>	<b>0.024</b>	<b>-0.014</b>	<b>0.062</b>
		DHGLM	<b>0.025</b>	<b>-0.008</b>	<b>0.057</b>	<b>0.025</b>	<b>-0.010</b>	<b>0.060</b>
	sociability	SHGLM	-0.375	-0.404	-0.346	-0.372	-0.406	-0.339
		DHGLM	-0.391	-0.418	-0.363	-0.387	-0.419	-0.354
random intercept standard deviation	activity	SHGLM	0.643	0.581	0.717	0.650	0.586	0.715
		DHGLM	0.635	0.574	0.705	0.634	0.575	0.703
	aggression	SHGLM	0.434	0.385	0.488	0.438	0.386	0.493
		DHGLM	0.441	0.392	0.496	0.442	0.391	0.496
	neophilia	SHGLM	0.442	0.391	0.497	0.442	0.392	0.502
		DHGLM	0.467	0.417	0.523	0.465	0.411	0.521
	boldness	SHGLM	0.396	0.346	0.448	0.395	0.346	0.446
		DHGLM	0.412	0.365	0.463	0.416	0.366	0.466
	sociability	SHGLM	0.472	0.424	0.524	0.470	0.421	0.522
		DHGLM	0.484	0.432	0.542	0.473	0.421	0.529
random slope standard deviation	activity	SHGLM				0.227	0.189	0.264
		DHGLM				0.208	0.172	0.248
	aggression	SHGLM				0.188	0.136	0.238
		DHGLM				0.177	0.130	0.220
	neophilia	SHGLM				0.076	0.007	0.150
		DHGLM				0.061	0.003	0.134

	boldness	SHGLM				0.046	0.001	0.119	
		DHGLM				0.064	0.003	0.138	
	sociability	SHGLM				0.141	0.083	0.188	
		DHGLM				0.147	0.105	0.191	
residual standard deviation	activity	SHGLM	0.740	0.720	0.761	0.703	0.681	0.725	
		DHGLM	0.572	0.511	0.645	0.518	0.459	0.579	
	aggression	SHGLM	0.822	0.800	0.843	0.801	0.778	0.824	
		DHGLM	0.690	0.635	0.754	0.658	0.600	0.722	
	neophilia	SHGLM	0.895	0.872	0.921	0.892	0.867	0.917	
		DHGLM	0.818	0.743	0.903	0.812	0.742	0.897	
	boldness	SHGLM	0.916	0.889	0.945	0.915	0.890	0.940	
		DHGLM	0.855	0.777	0.943	0.851	0.765	0.942	
	sociability	SHGLM	0.764	0.744	0.785	0.751	0.729	0.773	
		DHGLM	0.595	0.545	0.653	0.582	0.527	0.646	
dispersion model	intercept	activity	DHGLM	-1.007	-1.136	-0.871	-1.110	-1.252	-0.982
		aggression	DHGLM	-0.412	-0.532	-0.300	-0.445	-0.563	-0.333
	neophilia	DHGLM	-0.277	-0.403	-0.153	-0.281	-0.405	-0.167	
	boldness	DHGLM	-0.299	-0.411	-0.177	-0.311	-0.436	-0.185	
	sociability	DHGLM	-0.732	-0.849	-0.616	-0.802	-0.925	-0.669	
	female-male contrast	activity	DHGLM	0.424	0.221	0.620	0.409	0.221	0.592
		aggression	DHGLM	-0.173	-0.347	0.001	-0.224	-0.386	-0.059
	neophilia	DHGLM	<b>-0.144</b>	<b>-0.308</b>	<b>0.035</b>	<b>-0.149</b>	<b>-0.318</b>	<b>0.030</b>	
	boldness	DHGLM	<b>-0.034</b>	<b>-0.205</b>	<b>0.158</b>	<b>-0.029</b>	<b>-0.205</b>	<b>0.153</b>	
	sociability	DHGLM	<b>0.161</b>	<b>-0.013</b>	<b>0.336</b>	0.210	0.037	0.392	
age slope	activity	DHGLM	<b>-0.065</b>	<b>-0.131</b>	<b>0.007</b>	<b>-0.047</b>	<b>-0.116</b>	<b>0.020</b>	
	aggression	DHGLM	-0.220	-0.282	-0.156	-0.199	-0.265	-0.133	
	neophilia	DHGLM	-0.109	-0.168	-0.049	-0.111	-0.172	-0.051	
	boldness	DHGLM	-0.086	-0.140	-0.025	-0.081	-0.144	-0.017	
	sociability	DHGLM	<b>0.067</b>	<b>0.002</b>	<b>0.136</b>	0.076	0.006	0.146	

474 **Supplementary Table S2**475 *Repeatability and the coefficient of individual variation, from univariate models with and*  
476 *without predictability.*477 Equation numbers in the first column refer to Section 4 'Summary statistics for meta-  
478 analysis' in the main text. Further descriptions are as given in the description of Table S1.

Parameter	Interpretation	Behaviour	Model	model with random intercept			model with random intercept & slope		
				Mean	CI.lb	CI.ub	Mean	CI.lb	CI.ub
Repeatability (mean) (eqn 32)	proportion of phenotypic variance due to differences between individuals in the same environment	activity	SHGLM	0.406	0.356	0.460	0.461	0.412	0.513
		aggression	DHGLM	0.528	0.457	0.593	0.581	0.515	0.651
		neophilia	SHGLM	0.185	0.151	0.223	0.222	0.183	0.265
		boldness	DHGLM	0.237	0.188	0.292	0.284	0.233	0.340
		sociability	SHGLM	0.195	0.158	0.235	0.202	0.163	0.247
		activity	DHGLM	0.245	0.195	0.303	0.250	0.196	0.303
Coefficient of variation (mean) (eqn 34)	magnitude of variability in average behaviour	boldness	SHGLM	0.156	0.123	0.194	0.159	0.123	0.196
		aggression	DHGLM	0.188	0.141	0.237	0.197	0.150	0.249
		neophilia	SHGLM	0.216	0.180	0.257	0.234	0.197	0.279
		sociability	DHGLM	0.283	0.233	0.341	0.301	0.247	0.359
		activity	SHGLM	0.250	0.225	0.279	0.267	0.241	0.297
		aggression	DHGLM	0.247	0.222	0.274	0.259	0.235	0.288
Repeatability (variance) (eqn 33)	degree to which individual differences in predictability are consistent through time	neophilia	SHGLM	0.272	0.239	0.307	0.299	0.264	0.336
		boldness	DHGLM	0.277	0.245	0.311	0.300	0.268	0.335
		aggression	SHGLM	0.475	0.425	0.531	0.485	0.429	0.542
		boldness	DHGLM	0.515	0.460	0.576	0.520	0.462	0.576
		neophilia	SHGLM	0.343	0.301	0.388	0.346	0.304	0.391
		sociability	DHGLM	0.364	0.321	0.408	0.372	0.331	0.415
Coefficient of variation (predictability - variance) (eqn 35)	magnitude of variability in residual variance	neophilia	SHGLM	0.236	0.211	0.264	0.246	0.220	0.275
		boldness	DHGLM	0.242	0.215	0.272	0.248	0.221	0.278
		activity	DHGLM	0.098	0.073	0.127	0.090	0.065	0.118
		aggression	DHGLM	0.060	0.042	0.081	0.064	0.043	0.088
		neophilia	DHGLM	0.093	0.071	0.115	0.092	0.071	0.115
		boldness	DHGLM	0.105	0.082	0.130	0.107	0.082	0.131

479

480 **Supplementary Table S3**

481 *Six types of between-individual correlations from bivariate personality, plasticity, and predictability*  
 482 *models.*

483 The table is organised as a correlation matrix, with rows showing a between-individual  
 484 correlation and its 95% credible interval (in parentheses). Green boxes – on the diagonal – show  
 485 in within-trait correlations (averaged across the four bivariate models featuring the focal trait).  
 486 For each green box the top row shows a personality plasticity association (the correlation  
 487 between individual intercepts and absolute slopes from the mean model), the middle row  
 488 shows a personality predictability association (the correlation between individual intercepts in  
 489 the mean and sign-reversed individual intercepts in the dispersion model), and the bottom row  
 490 shows a plasticity predictability association (the correlation between absolute individual slopes  
 491 in the mean and sign-reversed individual intercepts in dispersion). Blue boxes – the off-diagonal  
 492 positions – show between-trait correlations for the three individual differences from each  
 493 bivariate model. The top rows of blue boxes show behavioural syndromes (the correlation  
 494 between random intercepts in the mean), the middle rows show plasticity syndromes (the  
 495 correlation between absolute random slopes in the mean), and the bottom rows show  
 496 predictability syndromes (the correlation between random intercepts in dispersion, i.e. residual  
 497 variance). Rows shown in red font indicate 95% credible intervals that cross zero.

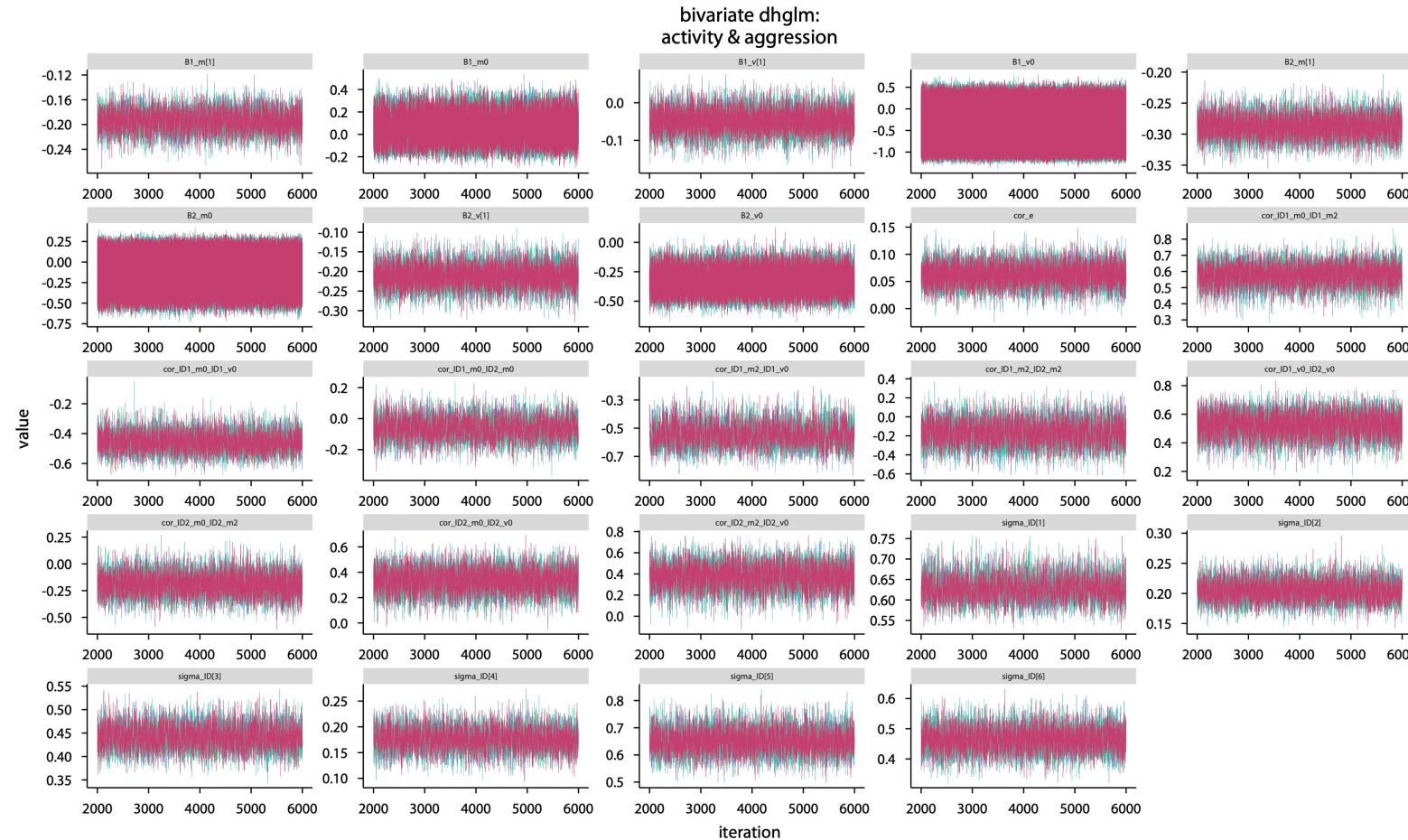
aggression	sociability	neophilia	boldness	activity
-0.024 (-0.311, 0.247)	0.888 (0.833, 0.933)	0.383 (0.240, 0.514)	0.415 (0.276, 0.541)	-0.069 (-0.207, 0.079)
-0.184 (-0.002, -0.364)	0.512 (0.254, 0.725)	-0.065 (-0.411, 0.36)	-0.083 (-0.425, 0.283)	-0.178 (-0.398, 0.026)
0.116 (-0.194, 0.415)	0.645 (0.481, 0.783)	0.484 (0.304, 0.664)	0.759 (0.615, 0.871)	0.528 (0.343, 0.702)
	-0.369 (-0.576, -0.167)	0.314 (0.163, 0.447)	0.345 (0.203, 0.490)	-0.132 (-0.268, -0.004)
	0.321 (0.132, 0.502)	-0.038 (-0.458, 0.349)	-0.021 (-0.360, 0.296)	-0.335 (-0.528, -0.124)
	0.220 (-0.030, 0.462)	0.324 (0.125, 0.503)	0.565 (0.376, 0.729)	0.638 (0.482, 0.771)
		-0.043 (-0.373, 0.283)	0.862 (0.803, 0.916)	0.036 (-0.117, 0.177)
		-0.218 (-0.040, -0.381)	0.063 (-0.330, 0.474)	0.041 (-0.274, 0.387)
		0.005 (-0.349, 0.347)	0.843 (0.731, 0.927)	0.454 (0.279, 0.616)

	-0.157 (-0.497, 0.171)	0.249 (0.095, 0.388)
	0.039 (-0.144, 0.233)	0.007 (-0.327, 0.369)
	0.075 (-0.289, 0.423)	0.665 (0.517, 0.797)
<b>Legend</b>		
<b>diagonal</b>	<b>off-diagonal</b>	
personality plasticity association		-0.584
personality predictability association		(-0.697, -0.454)
plasticity predictability association		0.448
behavioural syndrome		(0.303, 0.58)
plasticity syndrome		-0.511
predictability syndrome		(-0.363, -0.651)

498

499

## Diagnostic Figures



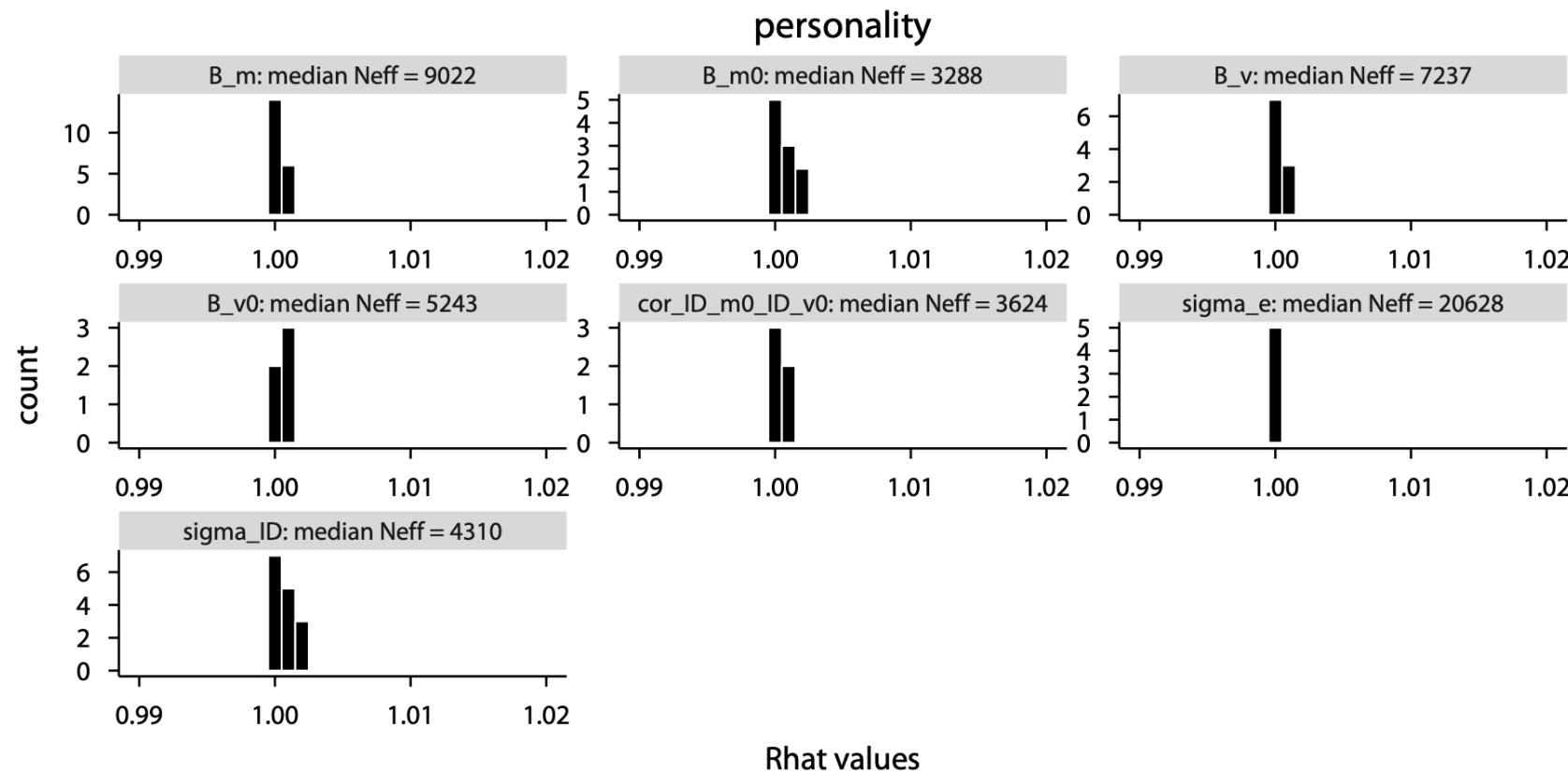
500

### Diagnostic Figure SD01: Trace plots for bivariate double hierarchical generalized linear model

502 Trace plots are shown for one model – activity and aggression. Each panel shows the trace plot for a different parameter, where the chain

iteration on the x-axis is plotted against its observed value on the y-axis. In a trace plot we are looking for all chains to behave similarly, with random scatter around a stable mean value producing a 'hairy caterpillar' appearance. Deviation from this look (e.g. divergent chains, moving mean values, or uneven scatter) indicate problems with chain mixing or convergence. Equivalent figures for the remaining 39 models look very similar (and can be reproduced with the supplementary models and code).

507

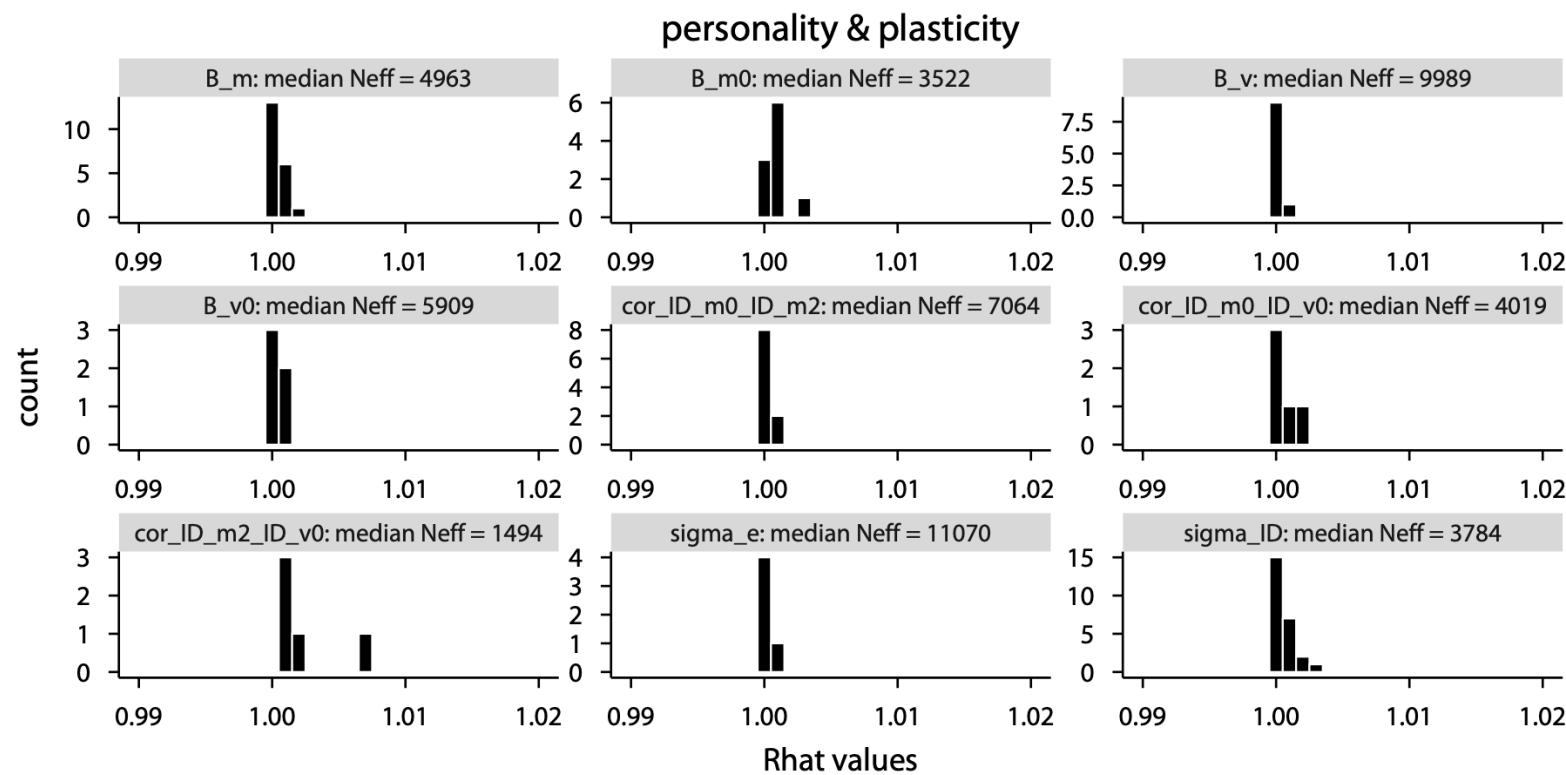


508

#### 509 Diagnostic Figure SD02: R-hat statistics for univariate models with random intercepts but no random slopes

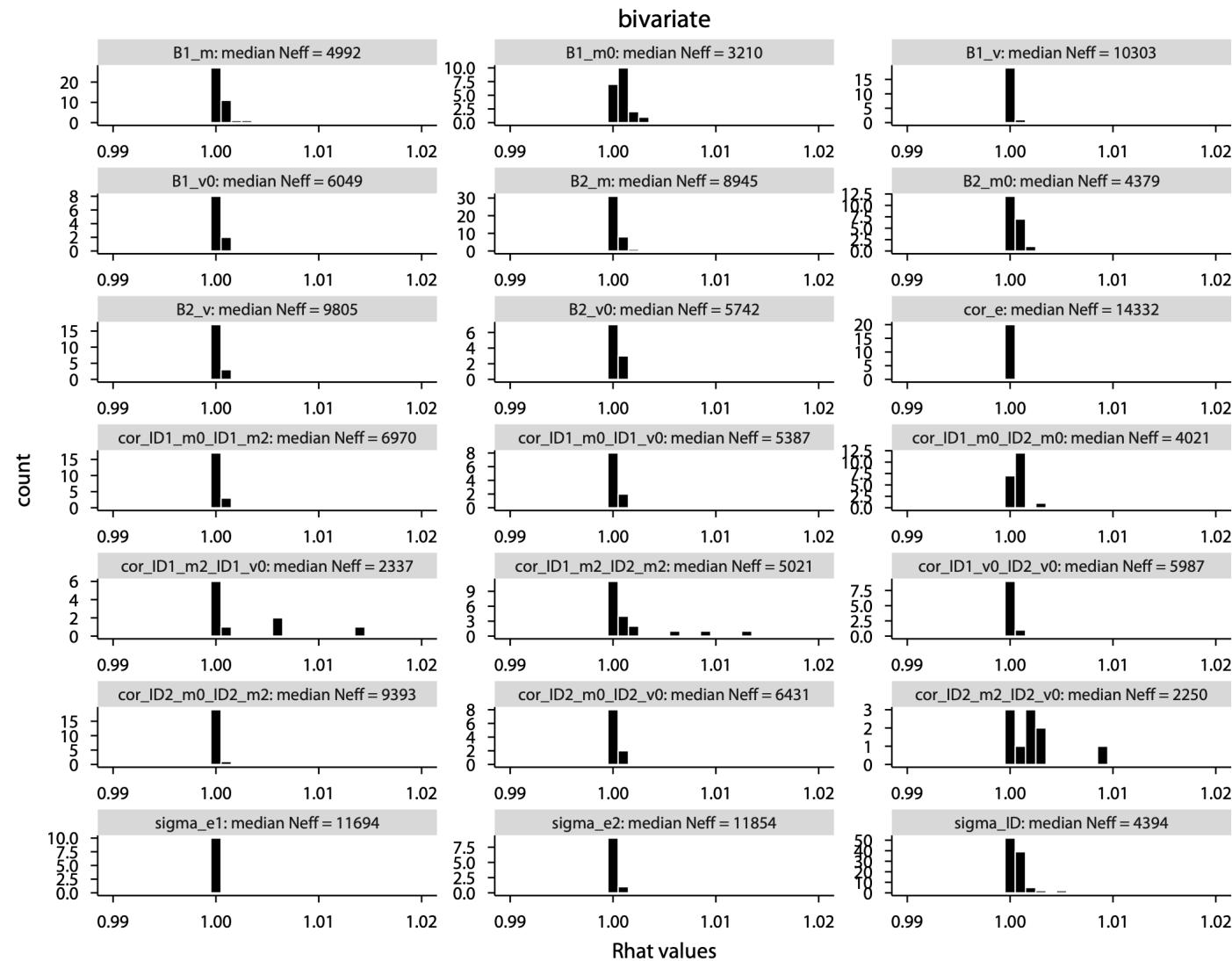
510 R-hat values – also known as the potential scale reduction factor or the Gelman-Rubin Diagnostic – take the ratio of variance within chains to

511 the variance pooled across chains. These are diagnostic statistics to indicate how similarly chains are behaving. When all chains are at  
 512 equilibrium the R-hat value will be exactly 1 (the most common value for models run in this worked example). Values above 1.01 indicate the  
 513 chains have not converged, and the sampling period should be increased (Vehtari et al., 2019). Increasing the number of iterations will  
 514 increase the effective sample size; the median effective sample size is shown in the facet label for each panel, and all are above 1,000  
 515 (minimum of 3,288, maximum of 20,628). R-hat values should be used in conjunction with diagnostic plots, such as the trace-plots shown  
 516 above.



518 **Diagnostic Figure SD03: R-hat statistics for univariate models with random intercepts and slopes**

519 Range of median effective sample sizes: 1494 – 11,070.

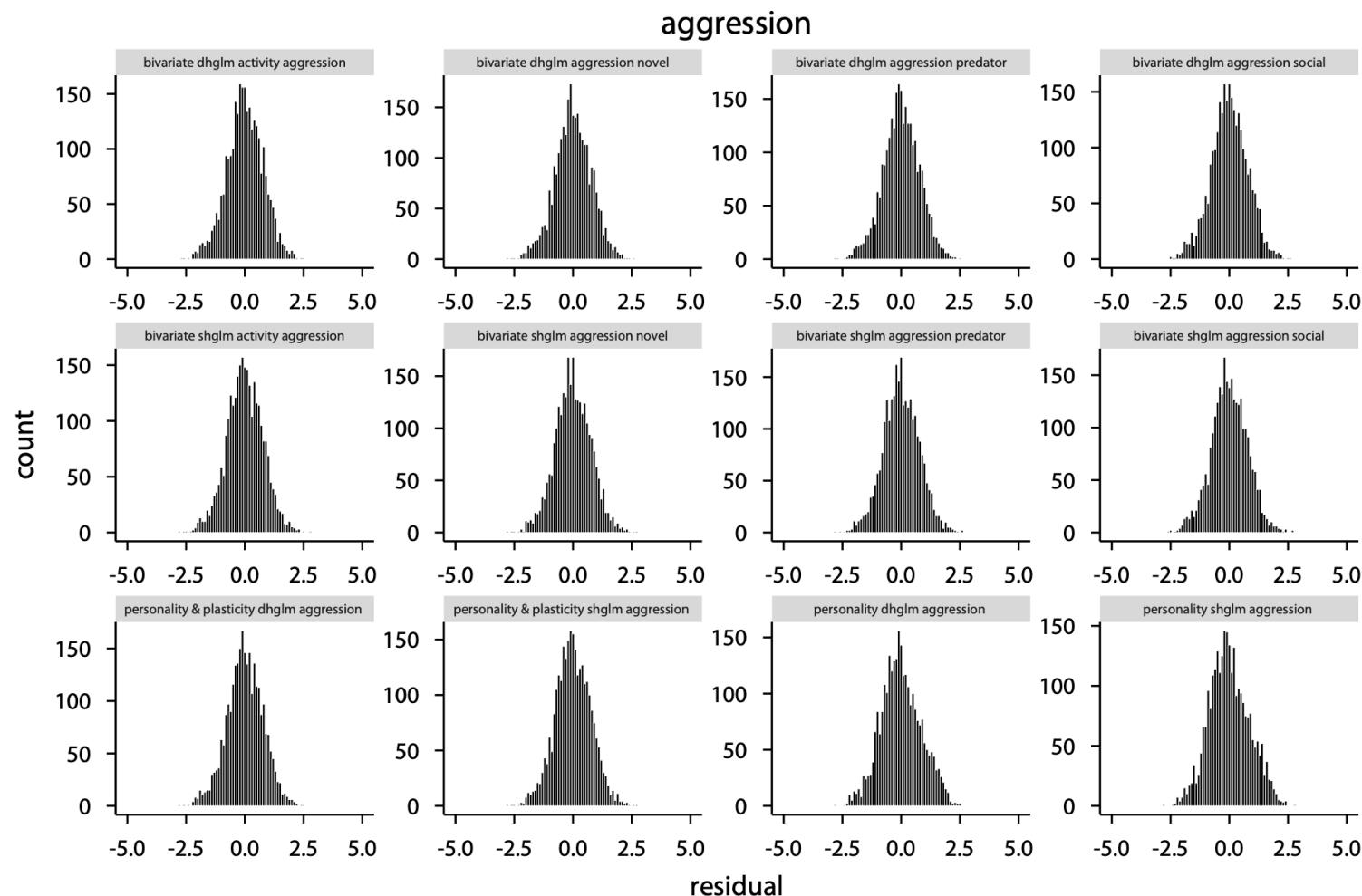


520

521 **Diagnostic Figure SD04: R-hat statistics for bivariate models with random intercepts and slopes**

522 Range of median effective sample sizes: 2,250 – 11,854.

523

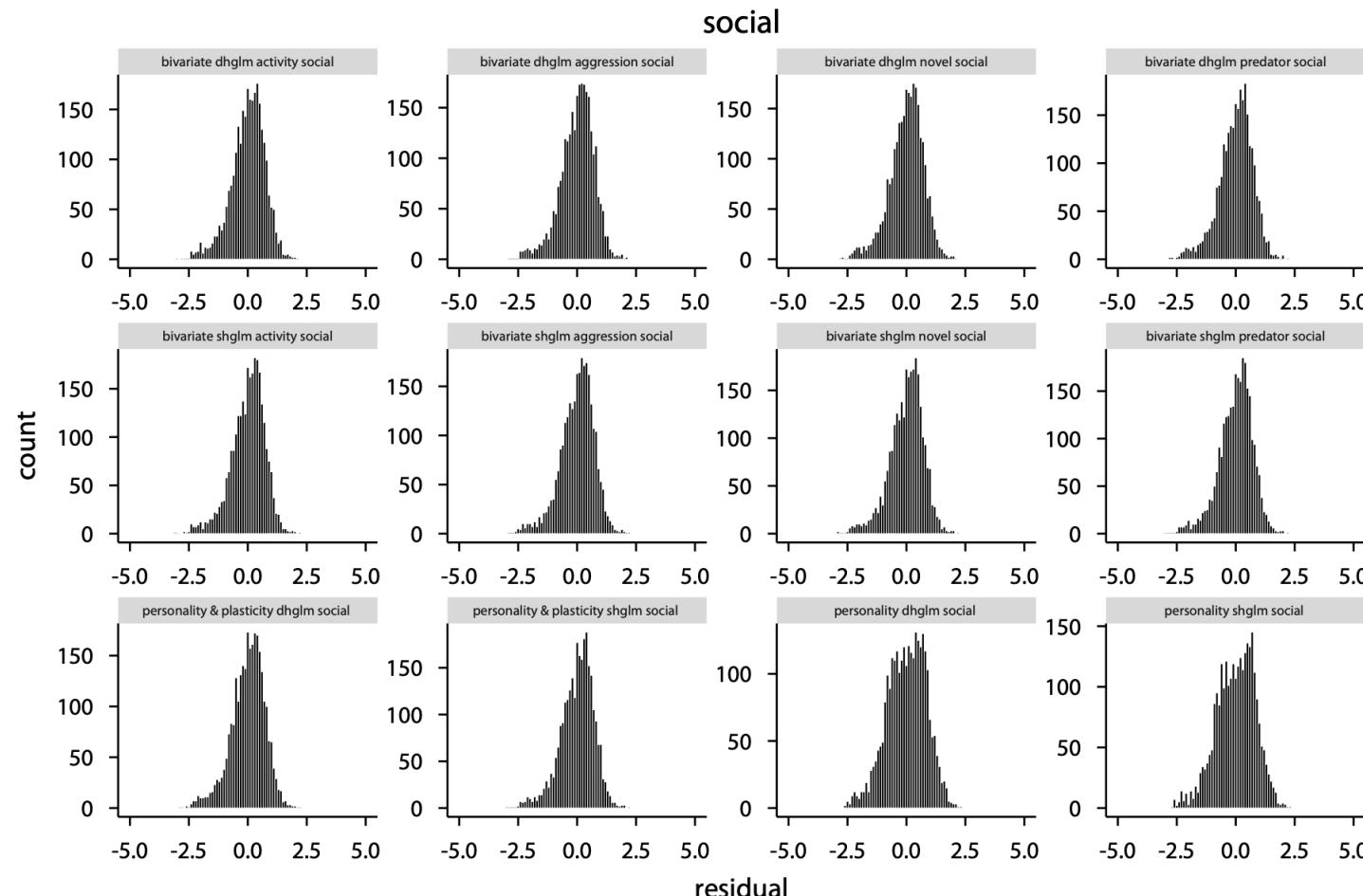


524

### 525 Diagnostic Figure SD05: Distributions of residuals for aggression models

526 Residuals were calculated as the difference between observed values (i.e. the response variable) and the values predicted by each model,  
 527 based on the posterior mean estimates for each parameter (including all fixed and random effects). As the likelihood function for all models

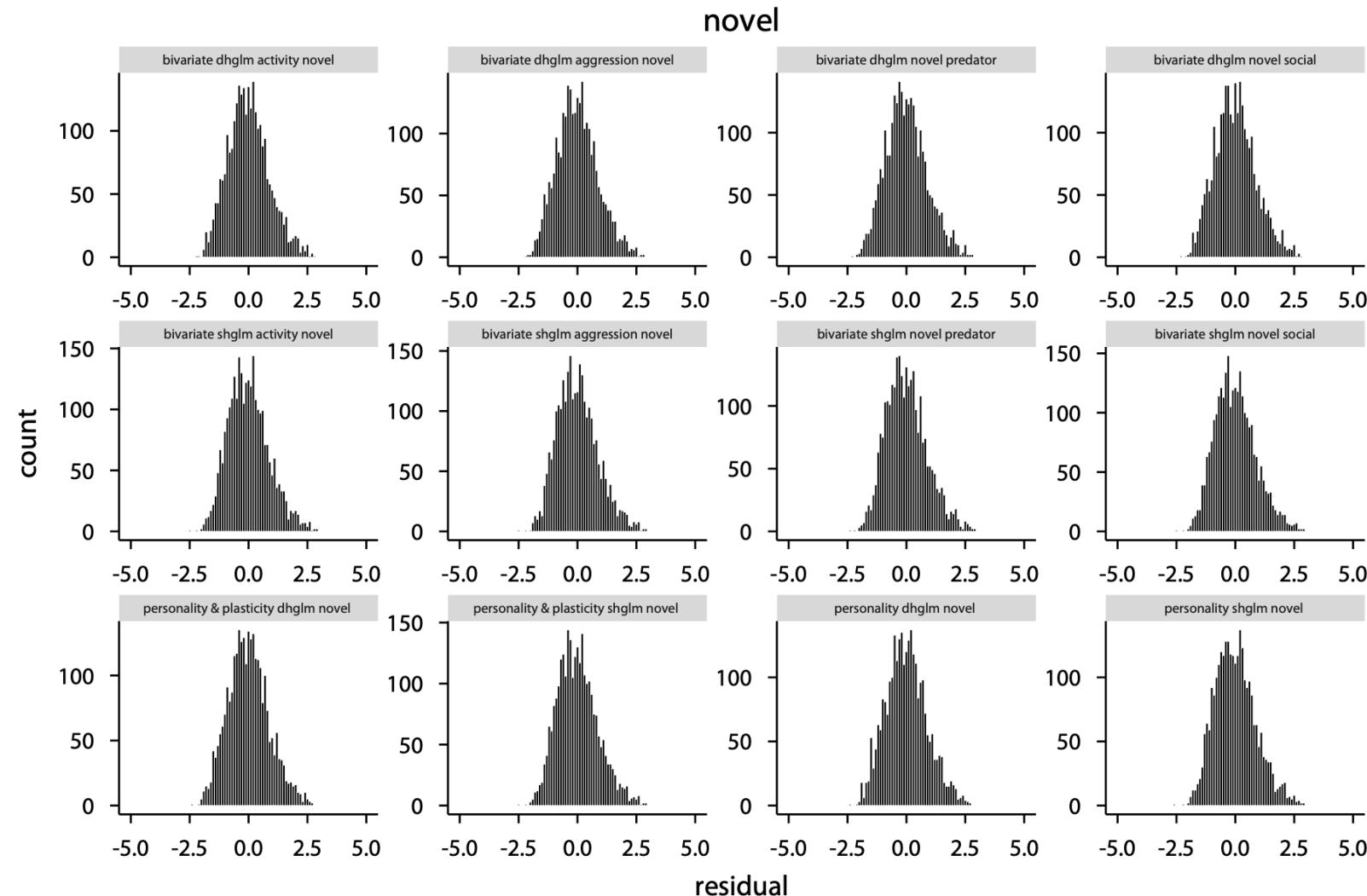
528 was based on a normal distribution, the residual distributions should look normally distributed (i.e. give the appearance of a bell curve). If  
 529 deviations from normality cannot be adequately improved with non-linear transformations, then a generalised model with a non-gaussian  
 530 likelihood function might be necessary (e.g., Poisson regression).



531

532 **Diagnostic Figure SD06: Distributions of residuals for sociability models**

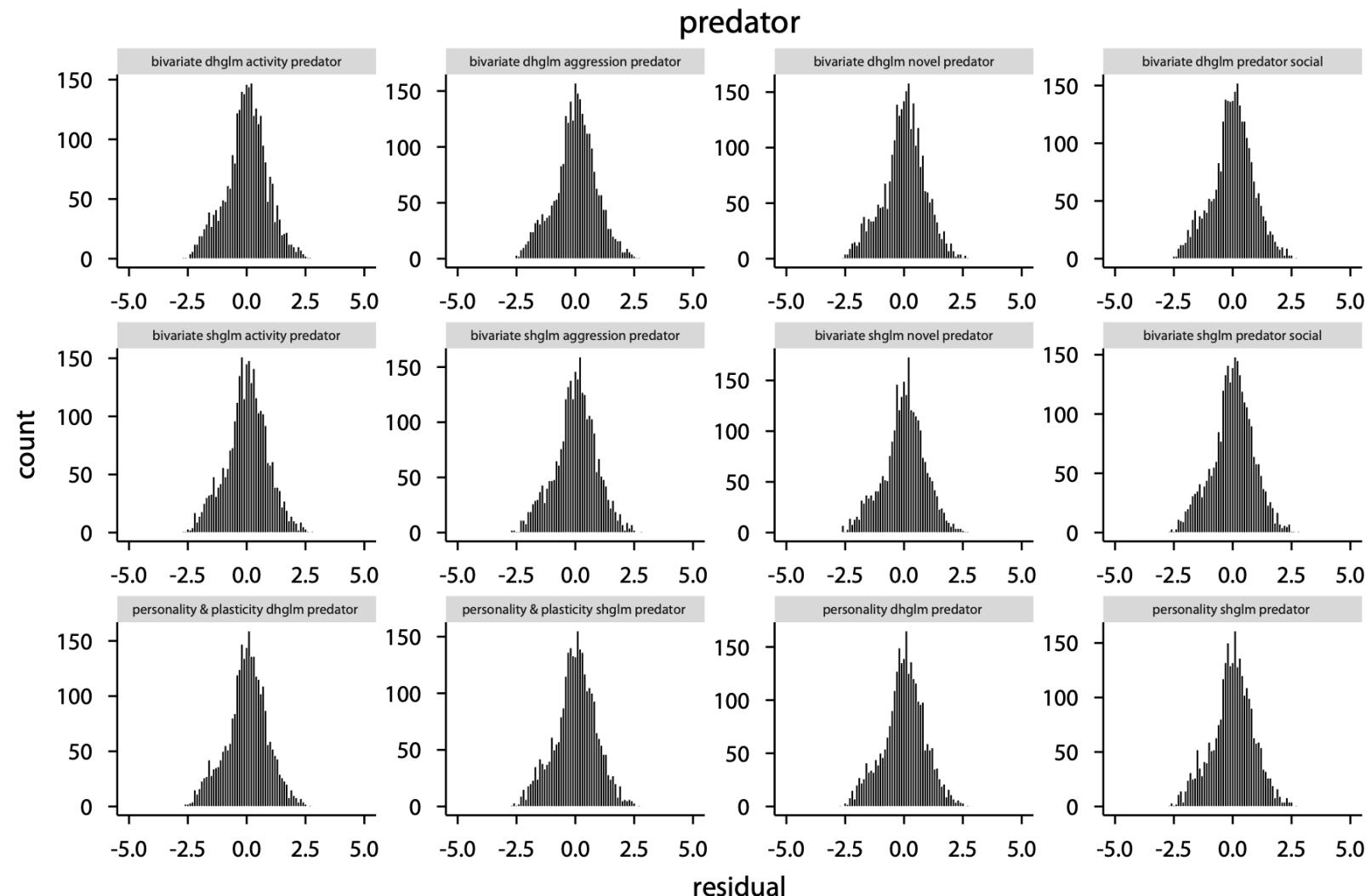
533



534

535 **Diagnostic Figure SD07: Distributions of residuals for neophilia models**

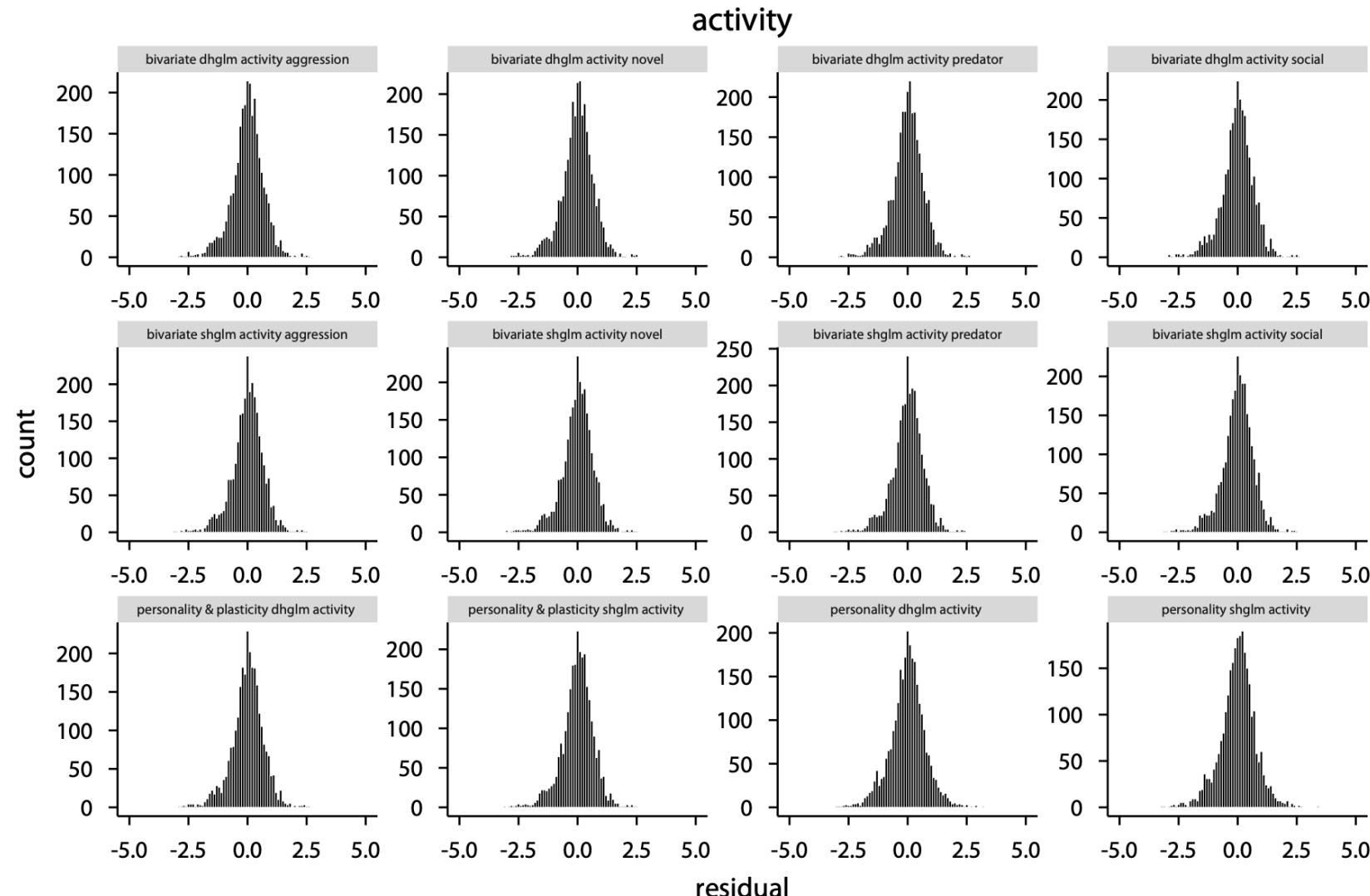
536



537

538 **Diagnostic Figure SD08: Distributions of residuals for boldness models**

539

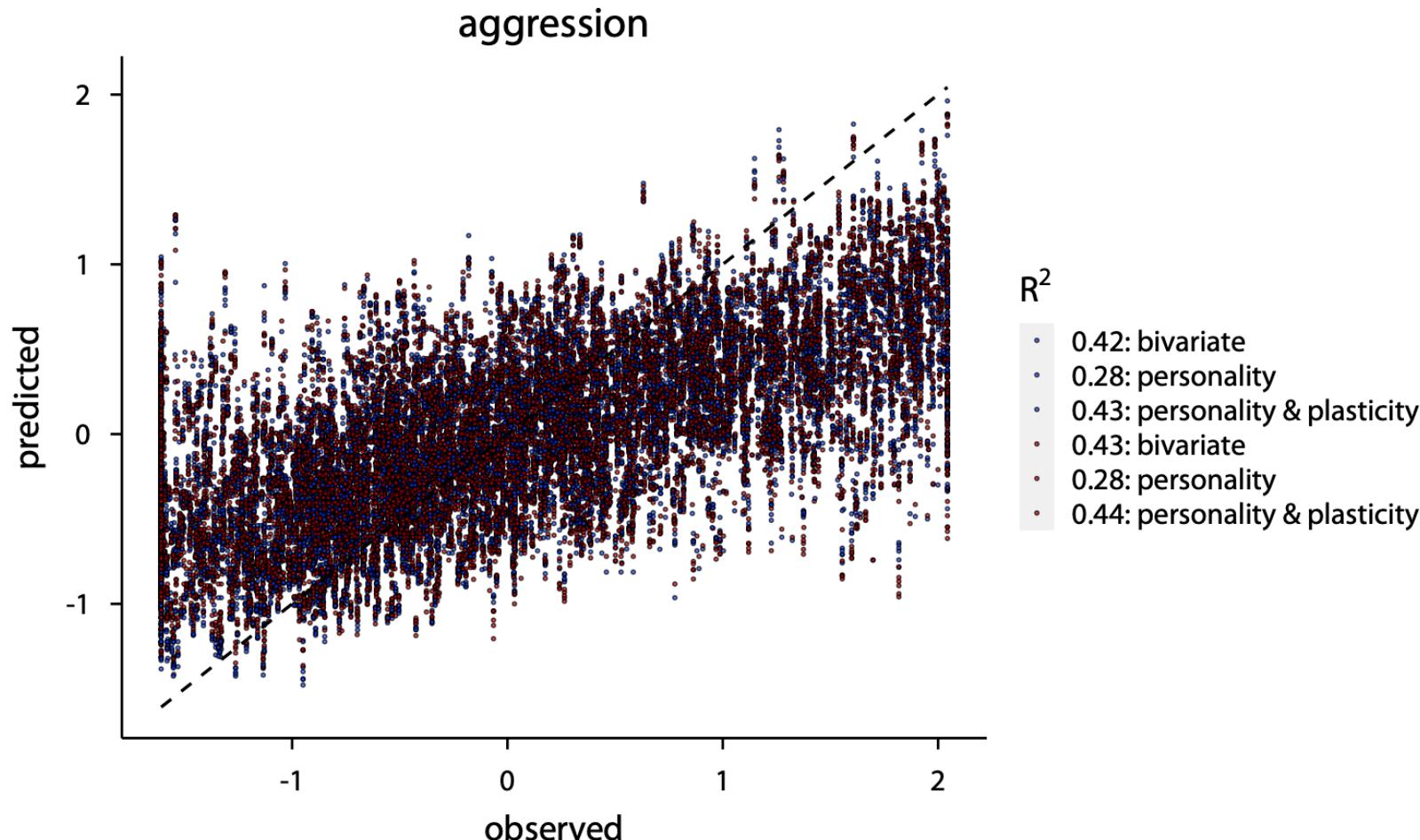


540

541 **Diagnostic Figure SD09: Distributions of residuals for activity models**

542

543 In the following Diagnostic Figures SD10-SD14: Blue points represent double hierarchical models (i.e. predictability models),  
544 and red points represent single hierarchical models.

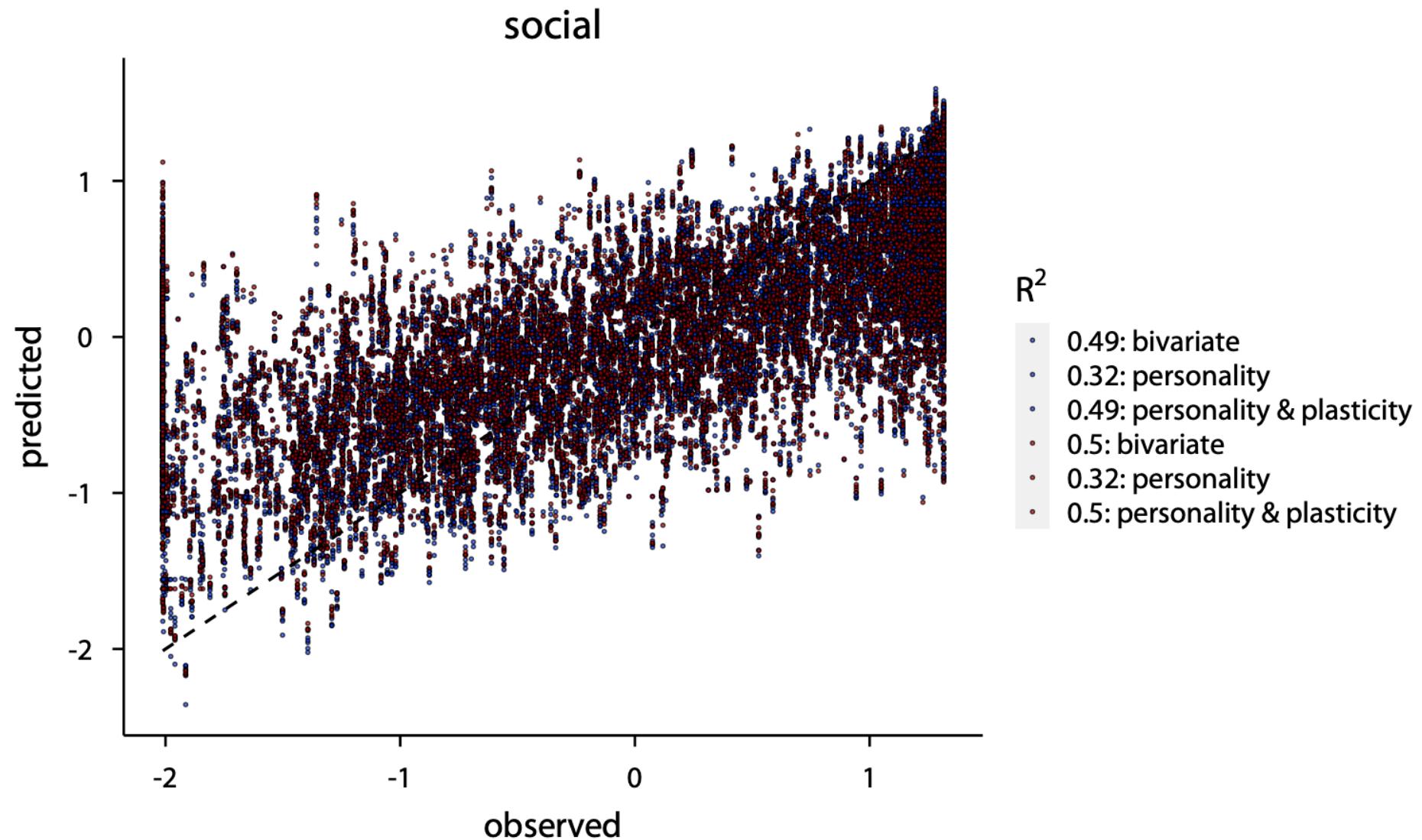


545

546 **Diagnostic Figure SD10: Correlation between observed and predicted values for aggression models**

547 Observed values (on the x-axis) are the response variable values fed into each model. Predicted values (on the y-axis) are the values

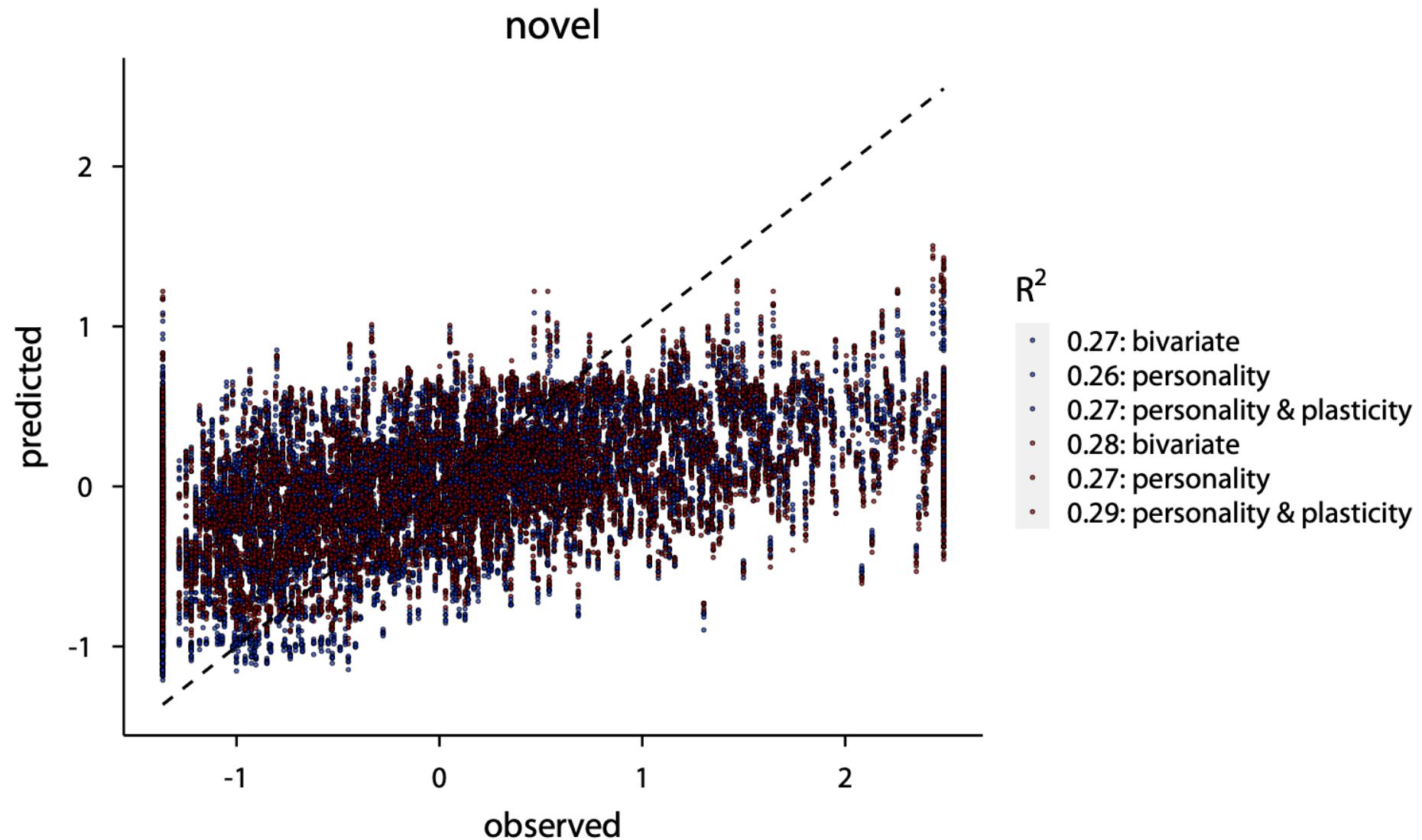
548 predicted by each model, based on posterior mean estimates for each parameter (including all fixed and random effects). The squared  
549 Pearson correlation between observed and predicted values gives an indication of model fit, with values closer to 1 indicating a tighter  
550 relationship between observed values and model predictions. The dashed diagonal line indicates a 1:1 relationship between observed and  
551 predicted values. In this and the remaining diagnostic plots, we can see that model estimates for each behaviour are over-estimated for  
552 observed values below the mean, and under-estimated for observed values above the mean.



553

554 **Diagnostic Figure SD11: Correlation between observed and predicted values for sociability models**

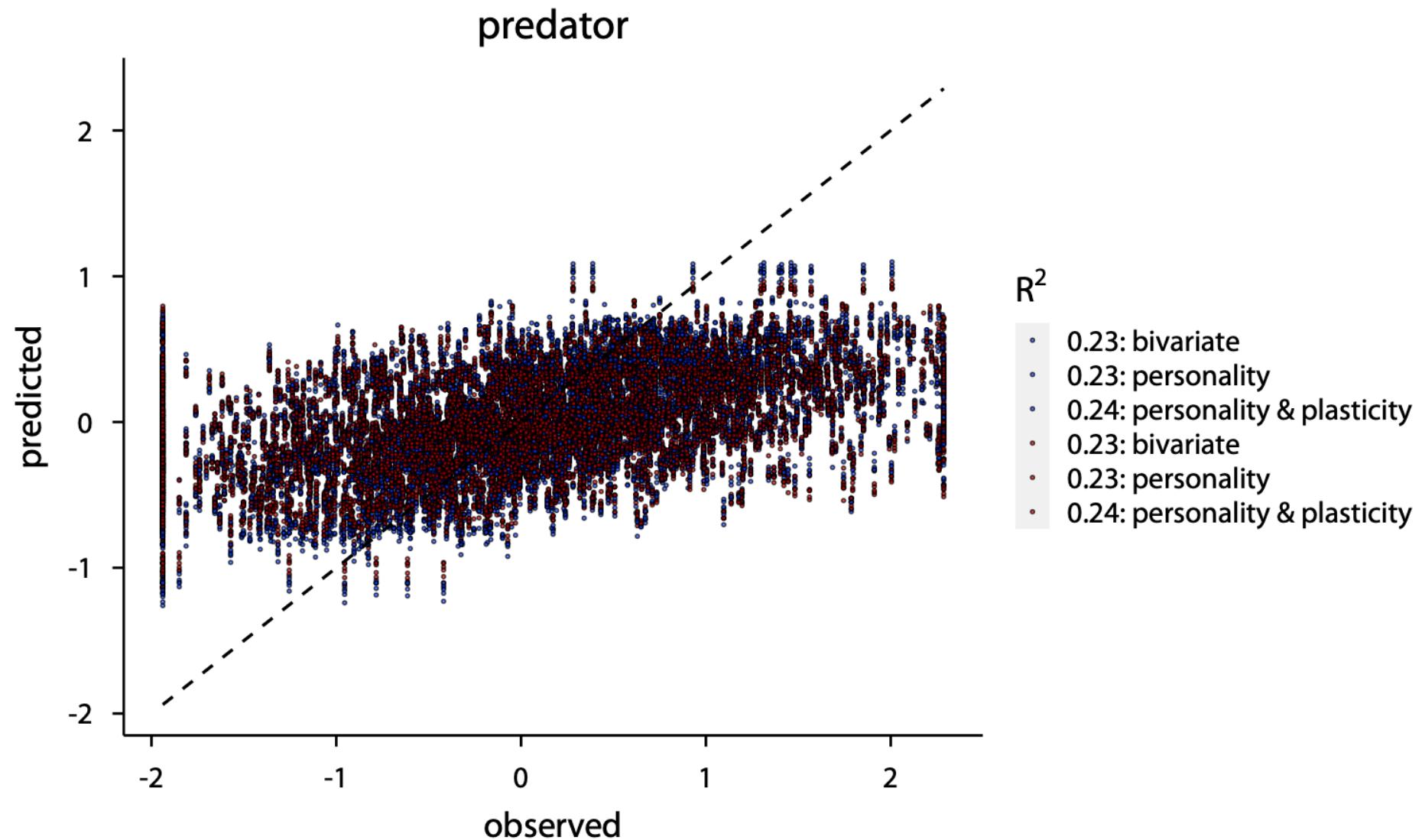
555



556

557 Diagnostic Figure SD12: Correlation between observed and predicted values for neophilia models

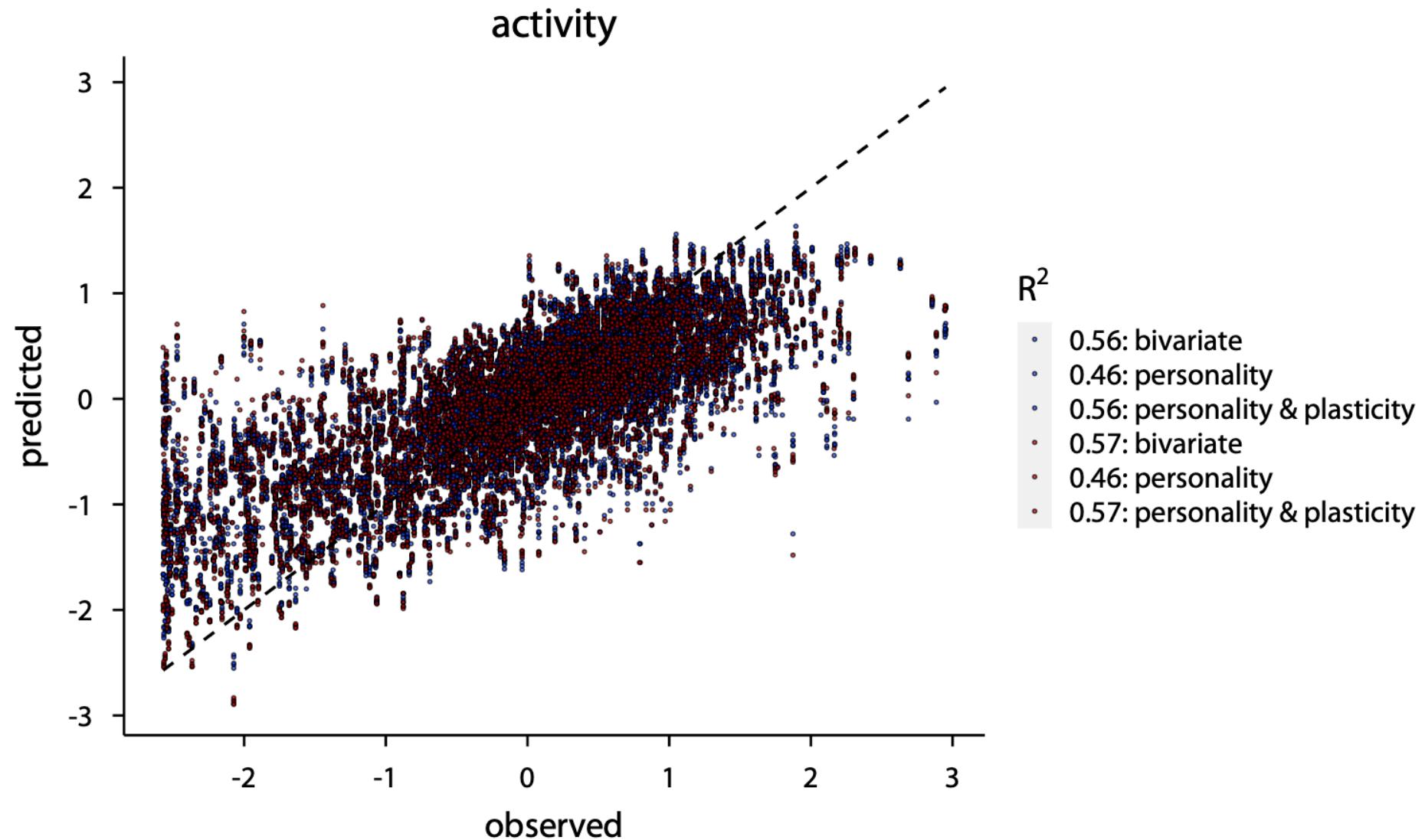
558



559

560 Diagnostic Figure SD13: Correlation between observed and predicted values for boldness models

561



562

563 Diagnostic Figure SD14: Correlation between observed and predicted values for activity models

## 564 Supplementary References

- 565 Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using  
 566 *lme4*. *Journal of Statistical Software*, 67(1). <http://doi.org/10.18637/jss.v067.i01>
- 567 Biro, P. A. (2012). Do rapid assays predict repeatability in labile (behavioural) traits? *Animal*  
 568 *Behaviour*, 83(5), 1295–1300. <http://doi.org/10.1016/j.anbehav.2012.01.036>
- 569 Bürkner, P (2018). "Advanced Bayesian Multilevel Modeling with the R Package brms." *The R*  
 570 *Journal*, 10(1), 395–411. <http://doi.org/10.32614/RJ-2018-017>
- 571 Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., et al. (2017). Stan:  
 572 A probabilistic programming language. *Journal of Statistical Software*, 76(1).  
 573 <http://doi.org/10.18637/jss.v076.i01>
- 574 Cleasby, I. R., Nakagawa, S., & Schielzeth, H. (2015). Quantifying the predictability of behaviour:  
 575 Statistical approaches for the study of between-individual variation in the within-individual  
 576 variance. *Methods in Ecology and Evolution*, 6(1), 27–37. <http://doi.org/10.1111/2041-210X.12281>
- 577 Dahlbom, S. J., Lagman, D., Lundstedt-Enkel, K., Sundström, L. F., & Winberg, S. (2011). Boldness  
 578 predicts social status in zebrafish (*Danio rerio*). *PLoS One*, 6(8), e23565.  
 579 <http://doi.org/10.1371/journal.pone.0023565>
- 580 Dereje, S., Sawyer, S., Oxendine, S. E., Zhou, L., Kezios, Z. D., Wong, R. Y., et al. (2012). Comparing  
 581 behavioral responses across multiple assays of stress and anxiety in zebrafish (*Danio rerio*).  
 582 *Behaviour*, 149(10-12), 1205–1240. <http://doi.org/10.1163/1568539X-00003018>
- 583 Fangmeier, M. L., Noble, D. W. A., O'Dea, R. E., Usui, T., Lagisz, M., Hesselson, D., & Nakagawa, S.  
 584 (2018). Computer animation technology in behavioral sciences: A sequential, automatic, and  
 585 high-throughput approach to quantifying personality in zebrafish (*Danio rerio*). *Zebrafish*,  
 586 zeb.2017.1532–4. <http://doi.org/10.1089/zeb.2017.1532>
- 587 Kern, E. M. A., Robinson, D., Gass, E., Godwin, J., & Langerhans, R. B. (2016). Correlated evolution  
 588 of personality, morphology and performance. *Animal Behaviour*, 117, 79–86.  
 589 <http://doi.org/10.1016/j.anbehav.2016.04.007>
- 590 Martins, E. P., & Bhat, A. (2014). Population-level personalities in zebrafish: Aggression-boldness  
 591 across but not within populations. *Behavioral Ecology*, 25(2), 368–373.  
 592 <http://doi.org/10.1093/beheco/aru007>

- 594 Mitchell, D. J., & Biro, P. A. (2017). Is behavioural plasticity consistent across different  
595 environmental gradients and through time? *Proceedings of the Royal Society B: Biological  
596 Sciences*, 284(1860), 20170893. <http://doi.org/10.1098/rspb.2017.0893>
- 597 Mitchell, D. J., Fanson, B. G., Beckmann, C., & Biro, P. A. (2016). Towards powerful experimental  
598 and statistical approaches to study intraindividual variability in labile traits. *Royal Society  
599 Open Science*, 3(10), 160352–10. <http://doi.org/10.1098/rsos.160352>
- 600 Moretz, J. A., Martins, E. P., & Robison, B. D. (2007). Behavioral syndromes and the evolution of  
601 correlated behavior in zebrafish. *Behavioral Ecology*, 18(3), 556–562.  
602 <http://doi.org/10.1093/beheco/arm011>
- 603 Mustafa, A., Roman, E., & Winberg, S. (2019). Boldness in male and female zebrafish (*Danio rerio*)  
604 is dependent on strain and test. *Frontiers in Behavioral Neuroscience*, 13, 20150509.  
605 <http://doi.org/10.3389/fnbeh.2019.00248>
- 606 Norton, W. H. J., & Bally-Cuif, L. (2012). Unravelling the proximate causes of the aggression-  
607 boldness behavioural syndrome in zebrafish. *Behaviour*, 149(10-12), 1063–1079.  
608 <http://doi.org/10.1163/1568539X-00003012>
- 609 O'Dea, R.E., Noble, D., & Nakagawa, S. (2020). Worked Example: Personality, Plasticity, and  
610 Predictability. *Open Science Framework*. <http://doi.org/10.17605/OSF.IO/V3QAX>
- 611 Oehlert, G. W. (1992). A note on the Delta method. *The American Statistician*, 46(1), 27–29.  
612 <http://doi.org/10.1080/00031305.1992.10475842>
- 613 Prentice, P. M., Houslay, T. M., Martin, J. G. A., & Wilson, A. J. (2020). Genetic variance for  
614 behavioural “predictability” of stress response. *Journal of Evolutionary Biology*, 39(03), 473.  
615 <http://doi.org/10.1111/jeb.13601>
- 616 R Core Team. (2020). *R*: A language and environment for statistical computing. Vienna, Austria: R  
617 Foundation for Statistical Computing. Retrieved from <http://www.R-project.org>.
- 618 Stan Development Team. (2020). *RStan*: the *R* interface to Stan. R package. Retrieved from  
619 <http://mc-stan.org/>
- 620 Vehtari, A., Gelman, A., Simpson, D., Carpenter, B., & Bürkner, P.C. (2019). Rank-normalization,  
621 folding, and localization: An improved R-hat for assessing convergence of MCMC. *arXiv  
622 preprint*, 1903.08008. <http://doi.org/10.1214/20-BA1221>
- 623 Yang, Y. E., Christensen, O. F., & Sorensen, D. (2011). Analysis of a genetically structured variance  
624 heterogeneity model using the Box–Cox transformation. *Genetics Research*, 93(1), 33–46.  
625 <http://doi.org/10.1017/S0016672310000418>