Can endocranial volume be estimated accurately from external skull measurements in great-tailed grackles (*Quiscalus mexicanus*)?

Corina J Logan, Christin R Palmstrom

There is an increasing need to validate and collect data approximating brain size on individuals in the field to understand what evolutionary factors drive brain size variation within and across species. We investigated whether we could accurately estimate endocranial volume (a proxy for brain size), as measured by computerized tomography (CT) scans, using external skull measurements and/or by filling skulls with beads and pouring them out into a graduated cylinder for male and female great-tailed grackles. We found that while females had higher correlations than males, estimations of endocranial volume from external skull measurements or beads did not tightly correlate with CT volumes. We found no accuracy in the ability of external skull measures to predict CT volumes because the prediction intervals for most data points overlapped extensively. We conclude that we are unable to detect individual differences in endocranial volume using external skull measurements. These results emphasize the importance of validating and explicitly quantifying the predictive accuracy of brain size proxies for each species and each sex.

2 **Can endocranial volume be estimated accurately from external skull measurements in great-**

3 **tailed grackles (*Quiscalus mexicanus*)?**

4 Corina J. Logan*[1,2], Christin R. Palmstrom[3]

6 [1]SAGE Center for the Study of the Mind and [3]College of Creative Studies, University of California

7 Santa Barbara, Santa Barbara, California USA 93106

8 [2]Department of Zoology, University of Cambridge, Cambridge, CB2 3EJ, UK

10 *Corresponding author: Corina J. Logan, Department of Zoology, University of Cambridge, Downing

11 Street, Cambridge, CB2 3EJ, UK, +44 7868 357 294, cl417@cam.ac.uk

13 **Abstract**

14 There is an increasing need to validate and collect data approximating brain size on individuals in the

15 field to understand what evolutionary factors drive brain size variation within and across species. We

16 investigated whether we could accurately estimate endocranial volume (a proxy for brain size), as

17 measured by computerized tomography (CT) scans, using external skull measurements and/or by

18 filling skulls with beads and pouring them out into a graduated cylinder for male and female great-

19 tailed grackles. We found that while females had higher correlations than males, estimations of

20 endocranial volume from external skull measurements or beads did not tightly correlate with CT

21 volumes. We found no accuracy in the ability of external skull measures to predict CT volumes

22 because the prediction intervals for most data points overlapped extensively. We conclude that we are

23 unable to detect individual differences in endocranial volume using external skull measurements. These

24 results emphasize the importance of validating and explicitly quantifying the predictive accuracy of

25 brain size proxies for each species and each sex.

27 **Introduction**

28 While comparing relative brain sizes (corrected for body size) across species has led to a greater

29 understanding of the evolutionary factors correlated with brain size at a broad taxonomic scale (e.g.,

30 Iwaniuk & Nelson 2003, Sakai et al. 2011, Sol et al. 2005), little is known about the within species

31 causes and consequences of variation in brain sizes (see Gonda et al. 2013, Thornton & Lukas 2012).

32 Additionally, the accuracy of brain size proxies, which are frequently used in such comparisons, is not

33 often validated (Healy & Rowe 2007). Therefore, the accuracy of brain size estimations and how they

34 compare to estimations in other species is questionable (Healy & Rowe 2007). Intraspecific brain size

35 comparisons are rare perhaps due to the difficulty of obtaining data on a number of factors for the same

36 individuals (e.g., biometric measurements, reproductive success, dominance rank, position in the social

37 network, and cognitive abilities). Acquiring such data is key for understanding what contributes to the

38 evolution of brain size among individuals, as well as across species (Gonda et al. 2013, Logan &

39 Clutton-Brock 2013, Thornton & Lukas 2012).

40   We investigated whether endocranial volume, a proxy for brain size (Iwaniuk & Nelson 2002),

41 can be approximated using measurements of the external skull in great-tailed grackles (*Quiscalus*

42 *mexicanus*, JF Gmelin, 1788). Grackles are innovative, having many novel foraging techniques

43 (Overington et al. 2009, 2011), and have successfully expanded their geographical range by exploiting

44 new environments (Peer 2011). However, it is unclear how grackles achieve such feats. Relatively

45 large-brained birds such as New Caledonian crows solve novel problems using enhanced cognitive

46 abilities (e.g., Taylor et al. 2010, Logan et al. 2014), yet grackles have smaller relative brain sizes than

47 would be expected for their many innovations (Overington et al. 2009). One of us (Logan) is

48 conducting in depth investigations of great-tailed grackle cognition and innovation in the lab and field

49 to understand whether innovativeness indicates behavioral flexibility and what kinds of cognition

50   grackles use to solve novel problems. This study system provides a rare opportunity to examine

51   intraspecific differences in brain size, which could have implications for understanding range

52   expansions in the context of innovation and behavioral flexibility. Finding an accurate proxy for

53   endocranial volume would greatly ease the collection of data on brain sizes since external skull

54   measurements could be further validated to account for head measurements that can be taken on live

55   birds, thus allowing for correlations with a number of other factors on which data are gathered on this

56   species.

57   This study is intended as a first step in validating the accuracy of using head measurements

58   taken from live birds to predict their actual brain size, which would require two additional steps: 1)

59   validating the link between external skull measurements and external head measurements, and 2)

60   validating the link between endocranial volume and actual brain size. Regarding the latter validation,

61   there is reason to believe that endocranial volume accurately approximates actual brain volume in birds

62   because the meninges, the matter between the skull and the brain, are thinner than in mammals and the

63   shape of the brain case in birds tightly tracks the shape of the brain (Iwaniuk & Nelson 2002).

64   Additionally, it was previously observed in the genus *Quiscalus* that individual variation in endocranial

65   volume is particularly large when compared with variation in other bird species (*Quiscalus quiscala*;

66   Iwaniuk & Nelson 2002). Assuming great-tailed grackles have similar amounts of variation, our study

67   provides an opportunity to understand how much of this variation in endocranial volume is due to

68   variation in skull morphology (i.e., changes in skull length, width, and/or height specifically).

69   Great-tailed grackle body sizes are sexually dimorphic (Johnson et al. 2000), therefore we

70   expect sex differences in endocranial volumes and we investigate proxies for each sex independently.

71   We used endocranial volumes calculated from computerized tomography (CT) scans to represent

72   actual endocranial volumes since this measure is the most precise. The complete area of the inside of

73   the skull is accounted for in CT scans, while other methods do not cover the whole endocranial space

74 (Witmer et al. 2008, Knoll et al. 2012). We compared CT volumes to skull length, width, and height

75 measurements to determine whether the correlation between these two methods and the accuracy of

76 external measures in predicting CT volumes warrants their use as a proxy for endocranial volume. We

77 also evaluated the bead method of generating endocranial volume, where glass beads are poured into

78 the skull and then out into a graduated cylinder, to increase the value of our research by determining

79 whether this widely used method (e.g., Isler et al. 2008, Iwaniuk & Nelson 2002) also accurately

80 predicts actual endocranial volume as estimated by CT scans in this species.

81

82 **Methods**

83 *Specimens*

84 We collected data from February through September 2014, and in March 2015, on 40 great-tailed

85 grackle skulls (Supplemental Table S1), 20 female and 20 male (some analyses have 19 males because

86 on one of their skulls the bill was broken off, thus we could not acquire its skull length measurement),

87 obtained from the Museum of Southwestern Biology (n=24, Albuquerque, NM), the Ornithology

88 Division of the University of Kansas (KU) Biodiversity Institute (n=15, Lawrence, KS), and the Santa

89 Barbara Museum of Natural History (n=1, Santa Barbara, CA). Skulls of unknown age were aged by

90 Andy Johnson if they were from the Museum of Southwestern Biology or by us if they were from KU.

91 Skulls were aged using the percentage of ossification to classify each as adult (>7 months old; 100%

92 ossified unless it was collected in February-May because this would mark the start of that individual's

93 first breeding season after having hatched June-August in the previous year) or immature (<7 months

94 old; <100% ossified when collected September-December indicating it had hatched that year; del Hoyo

95 et al. 1992, Winker 2000, Pyle 1997).

96

97 *Collecting endocranial volume measurements*

98   Linear measurements: Linear measurements of skulls were collected by placing calipers in locations on

99   the skull that would also be accessible and measurable on a live bird in the field. We recorded skull

100 length from the base of the bill to the back of the skull along the occipital crest (Figure 1A), height

101 from the posterior edge of the foramen magnum (the posterior edge of the neck at the base of the skull

102 on a live bird) to the top of the skull along the frontal region (Figure 1B), and width at the widest part

103 of the brain case along the squamosal bones (Figure 1C). All measurements were taken to the nearest

104 0.1mm. Research on other species has found positive correlations between actual brain mass and brain

105 volume estimated from linear skull measurements by calculating the volume of an ellipsoid (barn

106 swallows: Møller 2010), and endocranial volume and the volume of a cube as estimated using head

107 width (zebra finches: Bonaparte et al. 2011). Therefore, we estimated endocranial volume using a

108 number of volumetric shapes and data transformations to determine which best correlated with

109 endocranial volumes from CT scans. The volumetric shapes included were: cube (Length x Width x

110 Height), sphere ($\frac{4}{3}\pi r^3$, where r=$\frac{1}{2}$L or $\frac{1}{2}$W or $\frac{1}{2}$H), ellipsoid ($\frac{4}{3}\pi abc$, where a=$\frac{1}{2}$L, b=$\frac{1}{2}$W, c=$\frac{1}{2}$H), and

111 cone/pyramid ($\frac{1}{3}$bh, where b=W, h=H). We included log, natural log, and exponential transformations

112 of the data, and also allowed polynomial terms.

113

114 CT scans: Skulls were CT scanned at the Pueblo Radiology Medical Group in Santa Barbara,

115 California using a Siemans 16-slice Somatom Sensation 16 (1mm slices, 100Kv, 150MAs, 380mm

116 FOV, soft tissue window, analyzed with bone algorithm on). Endocranial volume (cm^3) was calculated

117 using the DICOM viewer OsiriX v5.8.5 (32-bit, Pixmeo SARL, Switzerland; Supplemental Figures S1

118 and S2) for 1mm cross-sectional slices (regular) and for 1mm cross-sectional slices that were taken

119 with the CT scanner bed moved 0.5mm forward (offset), using the average endocranial volume (

120 $\frac{regular + offset}{2}$) in analyses. The offset was added to increase the precision of the endocranial volume

121 measurements since grackle craniums are small (approximately 20mm in length), resulting in about 20

122 slices per scan (one slice every 1mm). The offset allowed us to measure more area (one slice every

123 0.5mm) by increasing the number of slices to approximately 40 per skull.

124

125 Beads: Endocranial volume was measured by pouring 1mm diameter glass beads (BioSpec Products,

126 catalog number 11079110) into the cranium through the foramen magnum until full. The skull was

127 repeatedly shaken to settle the beads and then filled again until the beads reached the posterior foramen

128 magnum without falling out (Figure 2). The volume was calculated by pouring the beads out of the

129 skull and into a graduated cylinder (5ml in 0.1ml graduations, World Precision Instruments, Inc.,

130 catalog number CG-0160; note that 1ml=1cm^3). In cross-species comparisons, there is mixed evidence

131 about whether pouring the beads into a graduated cylinder introduces error when compared with

132 pouring the beads onto a scale and converting their mass into volume (4% difference: Miller 1997, 0%

133 difference: Isler et al. 2008). The skulls in this study were measured with an average error of 2% (i.e.,

134 there was a 2% difference in volume between two sets of volume measurements carried out on a subset

135 of 12 skulls), which is small in comparison to the variance between skulls (intra-class correlation

136 coefficient=0.94; Hutcheon et al. 2010). Therefore, the error should not affect the power to detect a

137 correlation with the more precise CT method (intra-class correlation coefficient=0.98) of measuring

138 skulls

139

140 Data availability: The data from skull measurements and intraobserver reliability measurements are

141 available at the KNB Data Repository (Logan & Palmstrom 2015).

142

143   *Statistical analyses*

144   The female and male data (analyzed separately) were normally distributed (Anderson Darling

145   normality test: females: skull height p=0.27, length p=0.30, width p=0.86; males: skull height p=0.35,

146   length p=0.63, width p=0.38). We defined statistical significance as $p < 0.05$ throughout the paper.

147   Two sets of linear, bead, and CT scan measurements were taken on a subset of skulls on different days

148   by Palmstom to quantify the random measurement error (intraobserver reliability). We used intra-class

149   correlation coefficients (ICC) to determine the precision of our estimates using the equation in Figure 2

150   in Hutcheon and colleagues (2010): $true\ slope \times \frac{variance\ (true\ X\ values)}{variance\ (true\ X\ values)\ +\ variance\ (random\ error)}$ (we

151   assumed that variance in the observed values was equal to the variance in the true values). This ICC is

152   a measure of consistency, not agreement since it does not include rater effects (Auerbach et al. 2004).

153   We used generalized linear models (GLMs) to determine how well linear and bead

154   measurements correlated with volumes from CT scans, while examining whether the age of the

155   individual at death and the year the skull was collected improved the model fit. GLMs were carried out

156   in R v3.1.2 (R Core Team 2014) using the MCMCglmm function (MCMCglmm package; Hadfield

157   2010), while applying the dredge function (MuMIn package; Barton 2012) to select the best-fitting

158   model using the Akaike weight (Akaike 1981, Burnham & Anderson 2002). We considered the best-

159   fitting model to be strongly supported and reliable if its Akaike weight was ≥ 0.9 as suggested by

160   Burnham and Anderson (2002) since this would indicate that the likelihood of the model given the data

161   is very high. Female and male data were analyzed in separate models. Full models included 1)

162   endocranial volumes from CT scans as the response variable with the following explanatory variables:

163   volume of a cube or sphere or ellipsoid or cone + age and the interaction with year collected; or 2)

164   endocranial volumes from CT scans as the response variable with the following explanatory variables:

165   skull length + skull width + skull height + age and the interaction with year collected. GLMs were

166 conducted on the best-fitting model for each sex to explore whether the adjusted coefficient of

167 determination (adjusted r^2) improved by transforming the endocranial volume proxy (explanatory

168 variable) in the following ways: squared, cubed, quadratic, exponential, square root, log, log base 10,

169 and a polynomial with a degree of two or three. Of these, the model with the highest adjusted r^2 was

170 chosen as the final best-fitting model for that sex and included in the results below. Interpretations of

171 correlation strengths were taken from Taylor (1990): ≤ 0.35=weak correlation, 0.36-0.67=moderate,

172 0.68-0.89=high, 0.90-1.00=very high. We set the minimum criteria for a correlation of sufficient

173 strength such that it might be predictive at $r^2 \geq 0.88$, which is equivalent to Pearson's r set to

174 alpha=0.05 or $r \geq 0.95$, adjusted for the random measurement error in the response variable (CT

175 measurements), which has an intra-class correlation coefficient of 0.99 (0.95*0.99=0.94, 0.94^2=0.88;

176 Hutcheon et al. 2010).

177   Since we want to predict CT volumes from linear measures, we validated whether this was

178 possible by generating prediction intervals with 95% confidence levels. We applied the predict

179 function in the MCMCglmm package to the best-fitting model for each sex and evaluated whether

180 fitted values (predicted CT volumes) had credible intervals small enough such that there was little to no

181 overlap with other fitted values, thus allowing the discrimination of individual differences.

182

183 **Results**

184 *Intraobserver reliability*

185 There was very high consistency (precision) between the two sets of CT and bead volume

186 measurements, but no consistency for linear volume (LWH) measurements when sexes were analyzed

187 together and separately (Table 1). There was high (sexes combined) to very high (sexes analyzed

188 separately) consistency when comparing the two sets of skull width measurements, high consistency

189 for skull length (when sexes were separate and combined), and moderate (males and sexes combined)

190 to very high (females) consistency for skull height (Table 1).

191

192 *Correlations between methods*

193 None of the models using linear measurements to explain variation in CT volumes were likely given

194 the data, as indicated by the low Akaike weights of the best-fitting models ($<$ 0.90; Table 2; Burnham

195 & Anderson 2002). Regardless, we used the best-fitting models to examine these relationships further.

196 The volume of a sphere was the best-fitting shape for both sexes (the radius was based on skull width

197 for males and skull height for females). The best-fitting female model showed a positive relationship

198 between CT volumes and volumes from using the skull height as the radius for a sphere, volumes were

199 larger for immatures than for adults, and volumes slightly decreased over the years collected (Table 2,

200 model 1; Figure 3A). The best-fitting male model showed a positive correlation between CT volumes

201 and volumes using a quadratic polynomial of the skull width as the radius for a sphere, volumes were

202 slightly larger for immatures than for adults, and volumes decreased slightly over the years collected

203 (Table 2, model 2; Figure 3B). Transformations of the explanatory volume variables or substituting

204 volume for individual linear measurements (length, width, height, or some combination of these) did

205 not improve the adjusted r^2 for females. Volumes from CT scans were moderately (males) to highly

206 (females) correlated with spherical volumes from linear measurements (Table 2).

207   None of the models using bead volumes to explain variation in CT volumes were strongly

208 supported given the data as indicated by their low Akaike weights for the best-fitting models ($<$ 0.90;

209 Table 2; Burnham & Anderson 2002). Nonetheless, the best-fitting female model showed that

210 endocranial volumes decreased slightly over time (Table 2, model 3; Figure 4A), while the best-fitting

211 male model included age, with immatures having smaller volumes than adults (Table 2, model 4;

212 Figure 4B). Volumes from CT scans were highly positively correlated with volumes from the bead

213 method for both sexes (Table 2).

214 None of the correlations between CT volumes and linear measures met our minimum criteria (r^2

215 ≥ 0.88) for a strong enough relationship such that it might predict endocranial volumes from linear

216 measurements of skulls. Since we want to predict CT volumes from linear measures, we determined

217 whether this was possible by generating prediction intervals for the best-fitting female and male

218 models for the linear measurements (models 1 and 2) and bead method (models 3 and 4; Table 2). We

219 found that the lower and upper limits of the 95% confidence intervals of the predicted values for both

220 sexes show extensive overlap such that individual differences would not be able to be resolved if a

221 new, unvalidated data point was obtained (Table 3).

222

223 *Comparing method means*

224 Endocranial volume means were significantly different from each other when comparing across

225 methods (mean±standard deviation: females: CT 2.29cm^3±0.20, sphere 32459.1mm^3±4344.7, bead

226 2.60ml±0.28; males: CT 2.54cm^3±0.15, sphere 59292.1mm^3±2360.4, bead 2.91ml±0.21; Welch two

227 sample t-test: females: CT x Sphere t=31, p<0.0001, df=19; sphere x bead t=-31, p<0.0001, df=19;

228 bead x CT t=4, p=0.0003, df=34; males: CT x sphere t=96, p<0.0001, df=19; sphere x bead t=-96,

229 p<0.0001, df=19; bead x CT t=6, p<0.0001, df=35).

230

231 **Discussion**

232 While female great-tailed grackle endocranial volumes from linear measurements were highly

233 correlated with volumes from CT scans, which we consider a more accurate proxy for brain size than

234 bead volume (Witmer et al. 2008, Knoll et al. 2012), the correlation did not meet our criteria of having

235 a coefficient of determination (r^2) equal to or greater than 0.88 – a level of correlation that might be

236 strong enough to allow for the resolution of individual differences in endocranial volumes. This

237 correlation was only moderate in males, which is likely due to the sexual dimorphism in this species.

238 Our sample includes individuals from a range of populations in which the extent of sexual dimorphism

239 might vary. If, in some populations, there has been strong selection for males to increase in body size,

240 then the skeletal measures will not reflect brain size accurately because, in many instances, skeletal

241 size changes faster than brain size as has been found in primates (Montgomery 2011). Perhaps

242 additional biometric measurements would explain more of the variation in their endocranial volumes

243 from CT scans, however we only had access to skulls for most of the specimens and therefore could

244 not test this hypothesis.

245   We were more interested in whether a given value of an external skull measurement could

246 accurately predict actual endocranial volume from CT scans, rather than setting a subjective criterion

247 about how high r^2 should be, especially given the extensive debate around the latter approach (e.g.,

248 Legates & McCabe Jr. 1999, Müller & Büttner 1994). In particular, r^2 "[…] describes the proportion of

249 the *total* variance in the observed data that can be explained by the model" (Legates & McCabe Jr.

250 1999, p. 233, emphasis added) and thus does not allow one to investigate differences in the variance of

251 individual data points. Our predictive analyses showed that prediction intervals for new data points

252 overlapped to such a degree (within 95% credible intervals) that it was not possible to distinguish

253 among individuals, as we would need to when collecting linear measurements on new individuals in

254 the field. We must conclude that external skull measurements are not accurate enough to estimate

255 endocranial volume in great-tailed grackles.

256   Predictive analyses are crucial for determining the accuracy of predicting individual data points

257 by a particular method and should be applied extensively in future research, rather than relying solely

258 on correlation coefficients (r) or coefficients of determination (r^2). The omission of such an analysis

259   leaves data uninterpretable for its purported use of discerning intraspecific differences in a

260   morphological feature. Additionally, we caution against using a proxy validated in one species as

261   evidence that the same proxy will apply to other species (e.g., great tits: Dreyer 2012). Until

262   intraspecific validations of brain size proxies using skull or head measurements have been validated

263   across species, we cannot assume that what works (or not) for one species will work (or not) for

264   another.

265   The bead method was highly correlated with CT volumes in both sexes, however, it also did not

266   meet our minimum criteria ($r^2 \geq 0.88$), and prediction intervals extensively overlapped for individual

267   data points. Great-tailed grackles and common grackles are among the species with the largest ranges

268   in endocranial volumes (as measured using the bead method) when compared with most other species

269   in Iwaniuk & Nelson's (2002) study on 81 bird species. Both grackle species had large standard

270   deviations when compared with mean volumes (common grackles: mean ± SD=2.59ml ± 0.37, n=10,

271   Iwaniuk & Nelson 2002; great-tailed grackles: female 2.60ml ± 0.28, n=20, male 2.91ml ± 0.21, n=20,

272   this study). It appears that grackle endocranial volumes are more variable than those of many other

273   species. This is not likely due to variation in skull morphology since we did not find a perfect

274   correlation between endocranial volume and external skull measurements. Therefore, we caution

275   against using external skull measurements to estimate endocranial volume without proper validation.

276   To infer differences in brain size among individuals of the same species, and of the same sex,

277   there must be a high degree of accuracy to have the ability to detect actual individual differences

278   (Legates & McCabe Jr. 1999, Logan & Clutton-Brock 2013). Our results highlight the need to validate

279   brain size and/or endocranial volume proxies and their predictive power for each species under

280   investigation, and for each sex if they are sexually dimorphic. It is unfortunate that there is not an

281   easier, more accurate way to approximate brain size in the field where we have the potential to

282 understand how evolutionary factors drive brain size variation within species. However, this study

283 accentuates the importance of knowing how accurate brain size measures and proxies are when

284 including such data in analyses.

285

286 **Acknowledgements**

287 We thank Scott Grafton at the University of California Santa Barbara (UCSB) and Lawrence Harter

288 and Chuck Scudelari at the Pueblo Radiology Medical Group for providing access to the CT scanner;

289 Mario Mendoza at UCSB for scan planning; Sabrina Cruz and Kerri Moore at the Pueblo Radiology

290 Medical Group for scanning; Steve Rothstein for advice and sponsoring CP's grant; and Mireia Beas-

291 Moix and the Cheadle Center for Biodiversity and Ecological Restoration at UCSB for hosting the

292 skull loans, Dieter Lukas for analysis assistance, and Jay Taylor for feedback on a previous version of

293 this manuscript. We are grateful for skull loans from the Museum of Southwestern Biology (and to

294 Andy Johnson for aging skulls), Krista Fahy and the Santa Barbara Museum of Natural History, and

295 the Ornithology Division of the KU Biodiversity Institute.

296

297 **References**

298 Akaike H. 1981. Likelihood of a model and information criteria. *Journal Econometrics* 16: 3-14.

299 Auerbach C, La Porte HH, Caputo RK. 2004. Evidence-based practice manual: research and outcome

300   measures in health and human services [eds. Roberts AR, Yeager KR]. Oxford University Press

301 Barton K. 2012. MuMIn: Multi-model inference. R package version 1.7.7. Available: http://CRAN.R-

302   project.org/package=MuMIn. Accessed 3 Jan 2015.

303 Bonaparte KM, Riffle-Yokoi C, Burley NT (2011) Getting a Head Start: Diet, Sub-Adult Growth, and

304   Associative Learning in a Seed-Eating Passerine. PLoS ONE 6(9): e23775.

305   Burnham KP, Anderson DR. 2002. Model selection and multimodel inference: a practical information-

306   theoretic approach. New York: Springer.

307   del Hoyo J, Elliot A, Christie DA. 1992. Handbook of the Birds of the World. Barcelona: Lynx

308   Edicions.

309   Dreyer, S. 2012. Cranium size and fitness measures in a wild population of great tits (*Parus major*).

310   Masters thesis, University of Oslo.

311   Gonda A, Herczeg G, Merilä J. 2013. Evolutionary ecology of intraspecific brain size variation: a

312   review. *Ecology and Evolution* 3:2751-2764.

313   Hadfield JD. 2010. MCMC Methods for Multi-Response Generalized Linear Mixed Models: The

314   MCMCglmm   R   Package.   *Journal   of   Statistical   Software*   33:1-22.

315   http://www.jstatsoft.org/v33/i02/. Accessed 3 Jan 2015.

316   Healy SD, Rowe C. 2007. A critique of comparative studies of brain size. *Proceedings of the Royal*

317   *Society B* 274:453-464.

318   Hutcheon JA, Chiolero A, Hanley JA. 2010. Random measurement error and regression dilution bias.

319   *BMJ* 340:c2289.

320   Isler K, Christopher Kirk E, Miller J, Albrecht GA, Gelvin BR, Martin RD. 2008. Endocranial volumes

321   of primate species: scaling analyses using a comprehensive and reliable data set. *Journal of*

322   *Human Evolution* 55:967-978.

323   Iwaniuk AN, Nelson JE. 2002. Can endocranial volume be used as an estimate of brain size in birds?

324   *Canadian Journal of Zoology* 80:16-23.

325   Iwaniuk AN, Nelson JE. 2003. Developmental differences are correlated with relative brain size in

326   birds: a comparative analysis. *Canadian Journal of Zoology* 81:1913-1928.

327   Johnson K, DuVal E, Kielt M, Hughes C. 2000. Male mating strategies and the mating system of great-

328   tailed grackles. *Behavioral Ecology* 11:132-141.

329 Knoll F, Witmer LM, Ortega F, Ridgely RC, Schwarz-Wings D. 2012. The braincase of the basal
330   sauropod dinosaur Spinophorosaurus and 3D reconstructions of the cranial endocast and inner
331   ear. *PLOS ONE* 7:e30060.

332 Legates DR, McCabe Jr. GJ. 1999. Evaluating the use of "goodness-of-fit" measures in hydrologic and
333   hydroclimatic model validation. *Water Resources Research* 35:233-241.

334 Logan CJ, Clutton-Brock TH. 2013. Validating methods for estimating endocranial volume in
335   individual red deer (*Cervus elaphus*). *Behavioral Processes* 92:143-146.

336 Logan CJ, Jelbert SA, Breen AJ, Gray RD, Taylor AH. 2014. Modifications to the Aesop's Fable
337   paradigm change performances in New Caledonian crows. *PLOS ONE* 9(7):e103049.

338 Logan C, Palmstrom C. 2015. Skull measurements from 42 great-tailed grackles (*Quiscalus*
339   *mexicanus*) from Central and North America. The Knowledge Network for Biocomplexity
340   (KNB). doi: 10.5063/F1668B3W. https://knb.ecoinformatics.org/#view/doi:10.5063/F1668B3W

341 Miller JMA. 1997. A hierarchical analysis of primate brain size and body size: patterns of
342   morphometric variation. Ph.D. dissertation, University of Southern California, Los Angeles.

343 Møller, A. P. (2010). Brain size, head size and behaviour of a passerine bird. *Journal of evolutionary*
344   *biology*, *23*(3), 625-635.

345 Müller R, Büttner P. 1994. A critical discussion of intraclass correlation coefficients. *Statistics in*
346   *Medicine* 13:2465-2476.

347 Montgomery SH. 2011. The primate brain: evolutionary history and genetics. PhD dissertation,
348   University of Cambridge.

349 Overington, S. E., Morand-Ferron, J., Boogert, N. J., & Lefebvre, L. (2009). Technical innovations
350   drive the relationship between innovativeness and residual brain size in birds. *Animal Behaviour*,
351   *78*(4), 1001-1010.

352 Overington, S. E., Griffin, A. S., Sol, D., & Lefebvre, L. (2011). Are innovative species ecological

353   generalists? A test in North American birds. *Behavioral Ecology* 22:1286-1293.

354 Peer BD. 2011. Invasion of the Emperor's grackle. *Ardeola* 58:405-409.

355 Pyle P, Howell SN, DeSante DF. 1997. *Identification guide to North American birds* (pp. 421-422).

356   Bolinas, CA: Slate Creek Press.

357 R Core Team. 2014. R: A Language and Environment for Statistical Computing. R Foundation for

358   Statistical Computing, Vienna, Austria. http://www.R-project.org. Accessed 3 Jan 2015.

359 Sakai ST, Arsznov BM, Lundrigan B, Holekamp KE. 2011. Brain size and social complexity: a

360   computed tomography study in hyaenidae. *Brain, Behavior and Evolution* 77:91-104.

361 Sol D, Duncan RP, Blackburn TM, Cassey P, Lefebvre L. 2005. Big brains, enhanced cognition, and

362   response of birds to novel environments. *Proceedings of the National Academy of Sciences of the*

363   *United States of America* 102:5460-5465.

364 Taylor, A. H., Elliffe, D., Hunt, G. R., & Gray, R. D. (2010). Complex cognition and behavioural

365   innovation in New Caledonian crows. *Proceedings of the Royal Society of London B: Biological*

366   *Sciences*, *277*(1694), 2637-2643.

367 Thornton A, Lukas D. 2012. Individual variation in cognitive performance: developmental and

368   evolutionary perspectives. *Philosophical Transactions of the Royal Society B* 367:2773-2783.

369 Winker K. 2000. Obtaining, preserving, and preparing bird specimens. *Journal of Field Ornithology*

370   71:250-297.

371 Witmer LM, Ridgely RC, Dufeau DL, Semones MC. 2008. Using CT to peer into the past: 3D

372   visualization of the brain and ear regions of birds, crocodiles, and nonavian dinosaurs. In

373   *Anatomical Imaging* (pp. 67-87). Springer Japan.

1

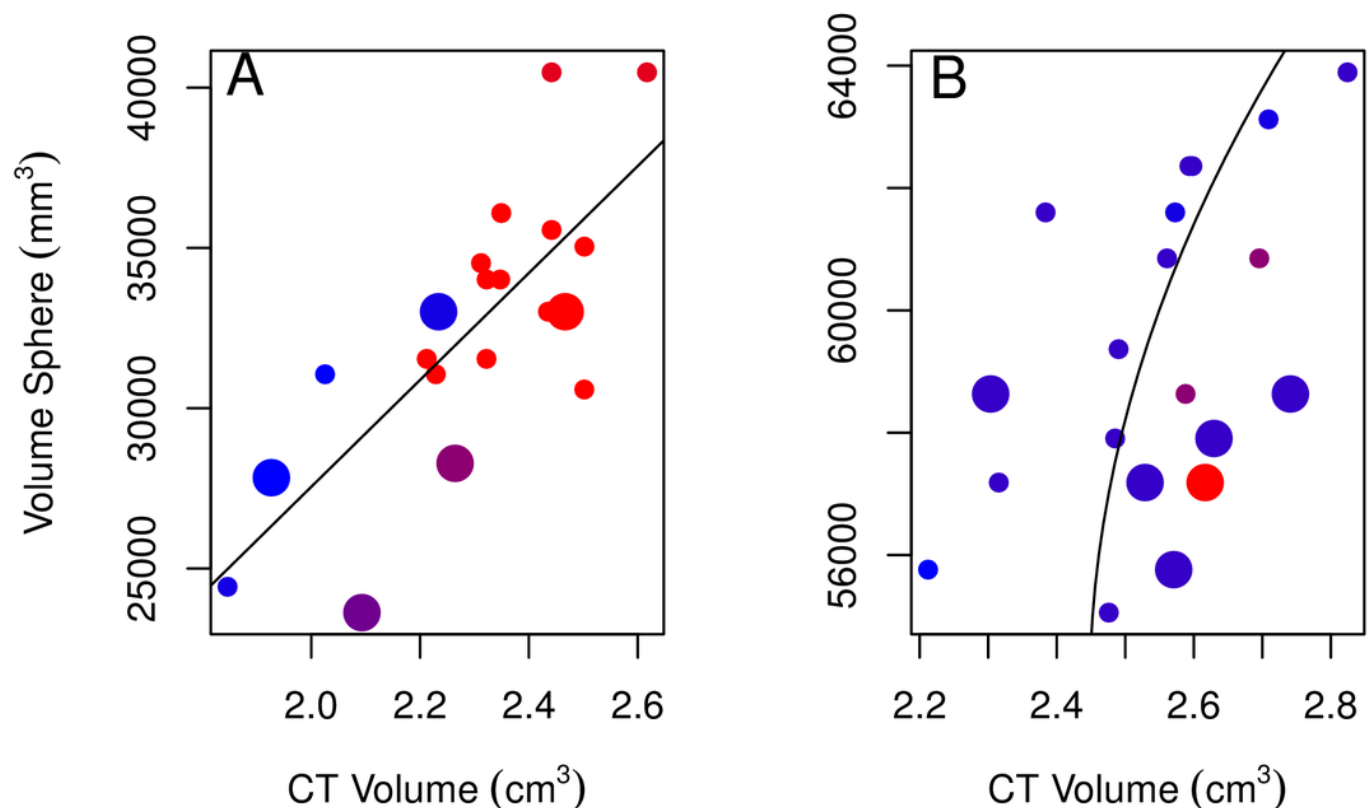
Skull length, height, and width

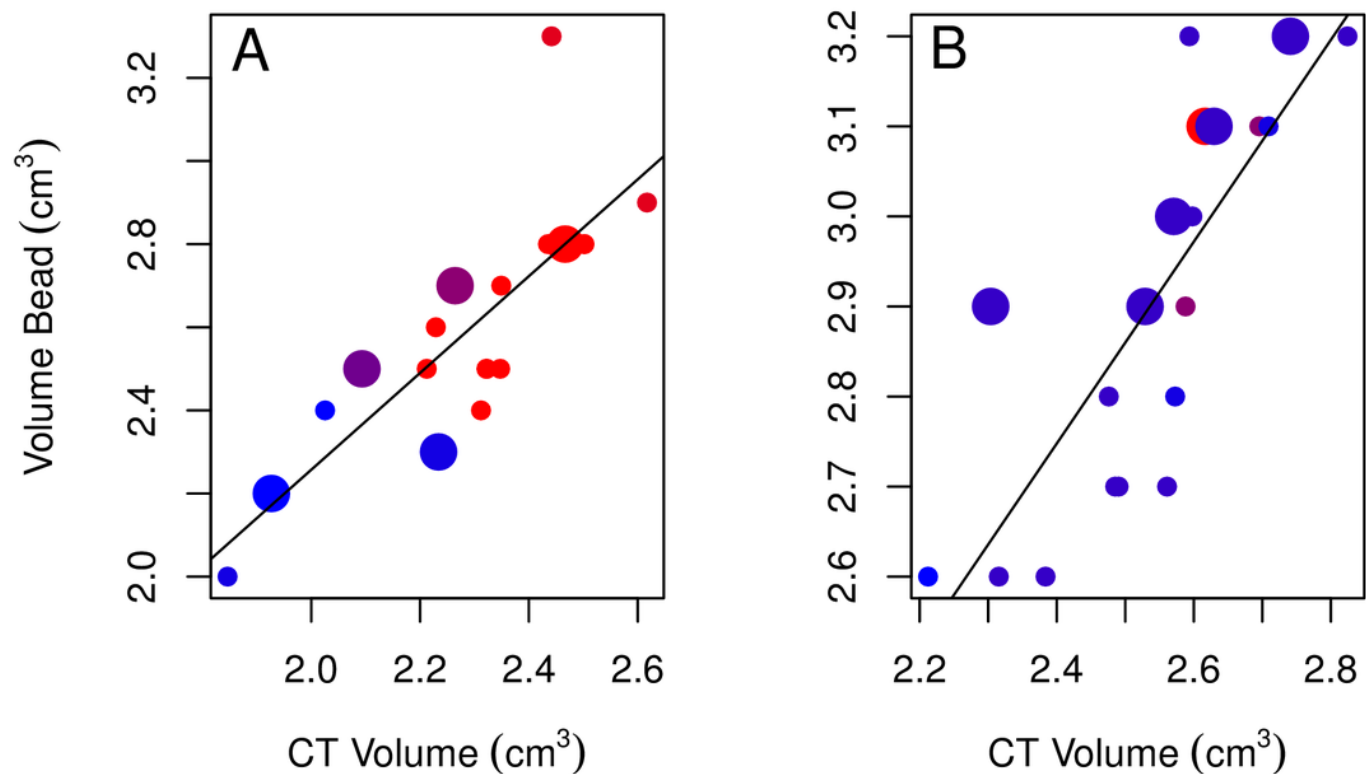
Measuring skull length (A), skull height (B), replicating the height that can be measured on live birds, and skull width (C) at the widest part of the braincase

Bead method

Skull holes are plugged with cotton and then the cranium is filled with glass beads

Plots of the volumes of spheres and volumes calculated from CT scans for females and males

Correlations between CT volumes and the volume of a sphere as calculated from linear measurements for female (A) and male (B) adults (small circles) and immatures (large circles), with the year the skull was collected represented by a red-blue spectrum (earlier years are redder and recent years are bluer). Note that regression lines only reflect the relationship between spherical volume and CT volume and do not correct for age or year (factors in the best-fitting model for both sexes) as in the GLMs. Skulls were aged as described in the methods.

Plots of the bead volumes and CT volumes for females and males

Correlations between CT volumes and bead volumes for female (A) and male (B) adults
(small circles) and immatures (large circles), with the year the skull was collected
represented by a red-blue spectrum (earlier years are redder and recent years are bluer).
Note that regression lines only reflect the relationship between bead volume and CT volume
and do not correct for age (in the best-fitting male model) or year (in the best-fitting female
model) as in the GLMs. Skulls were aged as described in the methods.

Intraobserver reliability

Intraobserver reliability summary data and results (measurements: mean±standard deviation; Both: data from both sexes combined; ICC: intra-class correlation)

Variable	Measurement 1			Measurement 2						ICC		
	Female	Male	Both	Female	n	Male	n	Both		Female	Male	Both
Volume CT	2.30 ± 0.15	2.56 ± 0.06	2.43 ± 0.17	2.38 ± 0.17	4	2.48 ± 0.08	4	2.38 ± 0.16		0.99	0.99	0.95
Volume LWH	13814 ± 1415	15605 ± 1276	14650 ± 1417	13779 ± 1233	5	15320 ± 785	7	14672 ± 1218		0.009	0.003	0.003
Volume bead	2.72 ± 0.15	2.93 ± 0.15	2.88 ± 0.21	2.68 ± 0.19	5	2.91 ± 0.13	7	2.84 ± 0.21		0.92	0.94	0.94
Skull Length	29.04 ± 0.98	30.06 ± 0.81	29.72 ± 0.99	29.40 ± 1.27	5	30.24 ± 0.78	7	29.94 ± 1.00		0.83	0.76	0.82
Skull Width	23.83 ± 0.86	23.99 ± 0.36	23.83 ± 0.67	23.39 ± 0.90	5	24.01 ± 0.39	7	23.81 ± 0.69		0.92	0.92	0.88
Skull Height	20.24 ± 0.87	20.89 ± 1.44	20.65 ± 1.18	19.96 ± 0.48	5	21.09 ± 0.54	7	20.68 ± 0.76		0.91	0.51	0.65

4

Model results

Model results for the best-fitting female and male models

Method	Sex	Model	Akaike weight	Adjusted r^2	p	y =
CT-sphere	Female	1	0.60	0.80	<0.0001	0.00002 x VolumeSphere + 1.10 x Age – 0.007 x Year + 1.44
	Male	2	0.26	0.39	0.02	0.47 x VolumeSphere + 0.19 x VolumeSphere^2 + 0.12 x Age – 0.003 x Year + 5.25
CT-bead	Female	3	0.74	0.77	<0.0001	0.37 x VolumeBead – 0.0005 x Year + 11.24
	Male	4	0.45	0.68	<0.0001	0.66 x VolumeBead – 0.09 x Age + 0.66

Prediction analysis results

Predicted CT volume (fitted value) and the predicted intervals in which these new data points would occur with 95% credible intervals based on inputs from linear measures or the bead method in the best-fitting female and male models for each method

| Linear Measurements | | | | | | Bead Method | | | | | |
| Males | | | Females | | | Males | | | Females | | |
Fitted value	Lower	Upper	Fitted value	Lower	Upper	Fitted value	Lower	Upper	Fitted value	Lower	Upper
2.72	2.37	3.07	2.23	2.03	2.42	2.61	2.42	2.80	2.30	2.07	2.48
2.41	2.17	2.73	2.08	1.86	2.29	2.43	2.23	2.61	2.21	1.99	2.42
2.40	2.12	2.69	1.87	1.61	2.07	2.37	2.18	2.55	1.90	1.63	2.11
2.60	2.33	2.92	1.99	1.78	2.22	2.70	2.52	2.90	2.02	1.78	2.26
2.42	2.10	2.70	2.20	2.02	2.39	2.50	2.34	2.68	2.05	1.81	2.30
2.52	2.22	2.83	2.30	2.10	2.51	2.55	2.35	2.75	2.42	2.22	2.62
2.77	2.45	3.07	2.43	2.23	2.60	2.76	2.54	2.94	2.39	2.16	2.59
2.67	2.41	2.94	2.39	2.20	2.58	2.70	2.50	2.89	2.28	2.07	2.52
2.35	2.06	2.64	2.32	2.14	2.53	2.37	2.17	2.56	2.31	2.11	2.54
2.55	2.30	2.84	2.51	2.30	2.71	2.50	2.31	2.67	2.57	2.31	2.81
2.46	2.21	2.72	2.51	2.30	2.71	2.43	2.25	2.63	2.43	2.20	2.64
2.64	2.40	2.92	2.39	2.19	2.59	2.63	2.45	2.82	2.32	2.10	2.53
2.53	2.24	2.77	2.48	2.27	2.70	2.44	2.26	2.63	2.46	2.20	2.70
2.58	2.31	2.85	2.40	2.21	2.60	2.37	2.18	2.57	2.32	2.11	2.52
2.55	2.31	2.85	2.32	2.13	2.52	2.47	2.28	2.67	2.31	2.11	2.56
2.64	2.39	2.94	2.41	2.23	2.60	2.76	2.57	2.97	2.43	2.20	2.64
2.54	2.29	2.82	2.36	2.16	2.55	2.61	2.40	2.80	2.43	2.22	2.66
2.55	2.24	2.80	2.42	2.21	2.60	2.67	2.49	2.87	2.43	2.20	2.65
2.52	2.26	2.80	2.32	2.13	2.52	2.48	2.27	2.68	2.36	2.15	2.57
2.51	2.24	2.80	2.02	1.82	2.13	2.57	2.38	2.74	1.98	1.74	2.18