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

Corina J. Logan*1,2, Christin R. Palmstrom3

1SAGE Center for the Study of the Mind and 3College of Creative Studies, University of California Santa Barbara, Santa Barbara, California USA 93106

2Department of Zoology, University of Cambridge, Cambridge, CB2 3EJ, UK

*Corresponding author: Corina J. Logan, Department of Zoology, University of Cambridge, Downing Street, Cambridge, CB2 3EJ, UK, +44 7868 357 294, cl417@cam.ac.uk

Abstract

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.
Introduction

While comparing relative brain sizes (corrected for body size) across species has led to a greater understanding of the evolutionary factors correlated with brain size at a broad taxonomic scale (e.g., Iwaniuk & Nelson 2003, Sakai et al. 2011, Sol et al. 2005), little is known about the within species causes and consequences of variation in brain sizes (see Gonda et al. 2013, Thornton & Lukas 2012). Additionally, the accuracy of brain size proxies, which are frequently used in such comparisons, is not often validated (Healy & Rowe 2007). Therefore, the accuracy of brain size estimations and how they compare to estimations in other species is questionable (Healy & Rowe 2007). Intraspecific brain size comparisons are rare perhaps due to the difficulty of obtaining data on a number of factors for the same individuals (e.g., biometric measurements, reproductive success, dominance rank, position in the social network, and cognitive abilities). Acquiring such data is key for understanding what contributes to the evolution of brain size among individuals, as well as across species (Gonda et al. 2013, Logan & Clutton-Brock 2013, Thornton & Lukas 2012).

We investigated whether endocranial volume, a proxy for brain size (Iwaniuk & Nelson 2002), can be approximated using measurements of the external skull in great-tailed grackles (*Quiscalus mexicanus*, JF Gmelin, 1788). Grackles are innovative, having many novel foraging techniques (Overington et al. 2009, 2011), and have successfully expanded their geographical range by exploiting new environments (Peer 2011). However, it is unclear how grackles achieve such feats. Relatively large-brained birds such as New Caledonian crows solve novel problems using enhanced cognitive abilities (e.g., Taylor et al. 2010, Logan et al. 2014), yet grackles have smaller relative brain sizes than would be expected for their many innovations (Overington et al. 2009). One of us (Logan) is conducting in depth investigations of great-tailed grackle cognition and innovation in the lab and field to understand whether innovativeness indicates behavioral flexibility and what kinds of cognition
grackles use to solve novel problems. This study system provides a rare opportunity to examine intraspecific differences in brain size, which could have implications for understanding range expansions in the context of innovation and behavioral flexibility. Finding an accurate proxy for endocranial volume would greatly ease the collection of data on brain sizes since external skull measurements could be further validated to account for head measurements that can be taken on live birds, thus allowing for correlations with a number of other factors on which data are gathered on this species.

This study is intended as a first step in validating the accuracy of using head measurements taken from live birds to predict their actual brain size, which would require two additional steps: 1) validating the link between external skull measurements and external head measurements, and 2) validating the link between endocranial volume and actual brain size. Regarding the latter validation, there is reason to believe that endocranial volume accurately approximates actual brain volume in birds because the meninges, the matter between the skull and the brain, are thinner than in mammals and the shape of the brain case in birds tightly tracks the shape of the brain (Iwaniuk & Nelson 2002). Additionally, it was previously observed in the genus *Quiscalus* that individual variation in endocranial volume is particularly large when compared with variation in other bird species (*Quicalus quiscalus*; Iwaniuk & Nelson 2002). Assuming great-tailed grackles have similar amounts of variation, our study provides an opportunity to understand how much of this variation in endocranial volume is due to variation in skull morphology (i.e., changes in skull length, width, and/or height specifically).

Great-tailed grackle body sizes are sexually dimorphic (Johnson et al. 2000), therefore we expect sex differences in endocranial volumes and we investigate proxies for each sex independently. We used endocranial volumes calculated from computerized tomography (CT) scans to represent actual endocranial volumes since this measure is the most precise. The complete area of the inside of the skull is accounted for in CT scans, while other methods do not cover the whole endocranial space
We compared CT volumes to skull length, width, and height measurements to determine whether the correlation between these two methods and the accuracy of external measures in predicting CT volumes warrants their use as a proxy for endocranial volume. We also evaluated the bead method of generating endocranial volume, where glass beads are poured into the skull and then out into a graduated cylinder, to increase the value of our research by determining whether this widely used method (e.g., Isler et al. 2008, Iwaniuk & Nelson 2002) also accurately predicts actual endocranial volume as estimated by CT scans in this species.

**Methods**

**Specimens**

We collected data from February through September 2014, and in March 2015, on 40 great-tailed grackle skulls (Supplemental Table S1), 20 female and 20 male (some analyses have 19 males because on one of their skulls the bill was broken off, thus we could not acquire its skull length measurement), obtained from the Museum of Southwestern Biology (n=24, Albuquerque, NM), the Ornithology Division of the University of Kansas (KU) Biodiversity Institute (n=15, Lawrence, KS), and the Santa Barbara Museum of Natural History (n=1, Santa Barbara, CA). Skulls of unknown age were aged by Andy Johnson if they were from the Museum of Southwestern Biology or by us if they were from KU. Skulls were aged using the percentage of ossification to classify each as adult (>7 months old; 100% ossified unless it was collected in February-May because this would mark the start of that individual’s first breeding season after having hatched June-August in the previous year) or immature (<7 months old; <100% ossified when collected September-December indicating it had hatched that year; del Hoyo et al. 1992, Winker 2000, Pyle 1997).

**Collecting endocranial volume measurements**
Linear measurements: Linear measurements of skulls were collected by placing calipers in locations on the skull that would also be accessible and measurable on a live bird in the field. We recorded skull length from the base of the bill to the back of the skull along the occipital crest (Figure 1A), height from the posterior edge of the foramen magnum (the posterior edge of the neck at the base of the skull on a live bird) to the top of the skull along the frontal region (Figure 1B), and width at the widest part of the brain case along the squamosal bones (Figure 1C). All measurements were taken to the nearest 0.1mm. Research on other species has found positive correlations between actual brain mass and brain volume estimated from linear skull measurements by calculating the volume of an ellipsoid (barn swallows: Möller 2010), and endocranial volume and the volume of a cube as estimated using head width (zebra finches: Bonaparte et al. 2011). Therefore, we estimated endocranial volume using a number of volumetric shapes and data transformations to determine which best correlated with endocranial volumes from CT scans. The volumetric shapes included were: cube (Length x Width x 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 cone/pyramid \( \frac{1}{3} bh \), where \( b = W \), \( h = H \). We included log, natural log, and exponential transformations of the data, and also allowed polynomial terms.

CT scans: Skulls were CT scanned at the Pueblo Radiology Medical Group in Santa Barbara, California using a Siemans 16-slice Somatom Sensation 16 (1mm slices, 100Kv, 150MAs, 380mm FOV, soft tissue window, analyzed with bone algorithm on). Endocranial volume (cm\(^3\)) was calculated using the DICOM viewer OsiriX v5.8.5 (32-bit, Pixmeo SARL, Switzerland; Supplemental Figures S1 and S2) for 1mm cross-sectional slices (regular) and for 1mm cross-sectional slices that were taken with the CT scanner bed moved 0.5mm forward (offset), using the average endocranial volume (
in analyses. The offset was added to increase the precision of the endocranial volume measurements since grackle crania are small (approximately 20mm in length), resulting in about 20 slices per scan (one slice every 1mm). The offset allowed us to measure more area (one slice every 0.5mm) by increasing the number of slices to approximately 40 per skull.

Beads: Endocranial volume was measured by pouring 1mm diameter glass beads (BioSpec Products, catalog number 11079110) into the cranium through the foramen magnum until full. The skull was repeatedly shaken to settle the beads and then filled again until the beads reached the posterior foramen magnum without falling out (Figure 2). The volume was calculated by pouring the beads out of the skull and into a graduated cylinder (5ml in 0.1ml graduations, World Precision Instruments, Inc., catalog number CG-0160; note that 1ml=1cm³). In cross-species comparisons, there is mixed evidence about whether pouring the beads into a graduated cylinder introduces error when compared with pouring the beads onto a scale and converting their mass into volume (4% difference: Miller 1997, 0% difference: Isler et al. 2008). The skulls in this study were measured with an average error of 2% (i.e., there was a 2% difference in volume between two sets of volume measurements carried out on a subset of 12 skulls), which is small in comparison to the variance between skulls (intra-class correlation coefficient=0.94; Hutcheon et al. 2010). Therefore, the error should not affect the power to detect a correlation with the more precise CT method (intra-class correlation coefficient=0.98) of measuring skulls.

Data availability: The data from skull measurements and intraobserver reliability measurements are available at the KNB Data Repository (Logan & Palmstrom 2015).
Statistical analyses

The female and male data (analyzed separately) were normally distributed (Anderson Darling normality test: females: skull height p=0.27, length p=0.30, width p=0.86; males: skull height p=0.35, length p=0.63, width p=0.38). We defined statistical significance as p < 0.05 throughout the paper.

Two sets of linear, bead, and CT scan measurements were taken on a subset of skulls on different days by Palmstom to quantify the random measurement error (intraobserver reliability). We used intra-class correlation coefficients (ICC) to determine the precision of our estimates using the equation in Figure 2 in Hutcheon and colleagues (2010):

\[
\text{true slope} \times \frac{\text{variance (true X values)}}{\text{variance (true X values)} + \text{variance (random error)}}
\]

(we assumed that variance in the observed values was equal to the variance in the true values). This ICC is a measure of consistency, not agreement since it does not include rater effects (Auerbach et al. 2004).

We used generalized linear models (GLMs) to determine how well linear and bead measurements correlated with volumes from CT scans, while examining whether the age of the individual at death and the year the skull was collected improved the model fit. GLMs were carried out in R v3.1.2 (R Core Team 2014) using the MCMCglmm function (MCMCglmm package; Hadfield 2010), while applying the dredge function (MuMIn package; Barton 2012) to select the best-fitting model using the Akaike weight (Akaike 1981, Burnham & Anderson 2002). We considered the best-fitting model to be strongly supported and reliable if its Akaike weight was ≥ 0.9 as suggested by Burnham and Anderson (2002) since this would indicate that the likelihood of the model given the data is very high. Female and male data were analyzed in separate models. Full models included 1) endocranial volumes from CT scans as the response variable with the following explanatory variables: volume of a cube or sphere or ellipsoid or cone + age and the interaction with year collected; or 2) endocranial volumes from CT scans as the response variable with the following explanatory variables: skull length + skull width + skull height + age and the interaction with year collected. GLMs were
conducted on the best-fitting model for each sex to explore whether the adjusted coefficient of
determination (adjusted $r^2$) improved by transforming the endocranial volume proxy (explanatory
variable) in the following ways: squared, cubed, quadratic, exponential, square root, log, log base 10,
and a polynomial with a degree of two or three. Of these, the model with the highest adjusted $r^2$ was
chosen as the final best-fitting model for that sex and included in the results below. Interpretations of
correlation strengths were taken from Taylor (1990): ≤ 0.35=weak correlation, 0.36-0.67=moderate,
0.68-0.89=high, 0.90-1.00=very high. We set the minimum criteria for a correlation of sufficient
strength such that it might be predictive at $r^2 \geq 0.88$, which is equivalent to Pearson’s $r$ set to
alpha=0.05 or $r \geq 0.95$, adjusted for the random measurement error in the response variable (CT
measurements), which has an intra-class correlation coefficient of 0.99 (0.95*0.99=0.94, 0.94²=0.88;
Hutcheon et al. 2010).

Since we want to predict CT volumes from linear measures, we validated whether this was
possible by generating prediction intervals with 95% confidence levels. We applied the predict
function in the MCMCglmm package to the best-fitting model for each sex and evaluated whether
fitted values (predicted CT volumes) had credible intervals small enough such that there was little to no
overlap with other fitted values, thus allowing the discrimination of individual differences.

Results

Intraobserver reliability

There was very high consistency (precision) between the two sets of CT and bead volume
measurements, but no consistency for linear volume (LWH) measurements when sexes were analyzed
together and separately (Table 1). There was high (sexes combined) to very high (sexes analyzed
separately) consistency when comparing the two sets of skull width measurements, high consistency
for skull length (when sexes were separate and combined), and moderate (males and sexes combined) to very high (females) consistency for skull height (Table 1).

Correlations between methods
None of the models using linear measurements to explain variation in CT volumes were likely given the data, as indicated by the low Akaike weights of the best-fitting models ($< 0.90$; Table 2; Burnham & Anderson 2002). Regardless, we used the best-fitting models to examine these relationships further. The volume of a sphere was the best-fitting shape for both sexes (the radius was based on skull width for males and skull height for females). The best-fitting female model showed a positive relationship between CT volumes and volumes from using the skull height as the radius for a sphere, volumes were larger for immatures than for adults, and volumes slightly decreased over the years collected (Table 2, model 1; Figure 3A). The best-fitting male model showed a positive correlation between CT volumes and volumes using a quadratic polynomial of the skull width as the radius for a sphere, volumes were slightly larger for immatures than for adults, and volumes decreased slightly over the years collected (Table 2, model 2; Figure 3B). Transformations of the explanatory volume variables or substituting volume for individual linear measurements (length, width, height, or some combination of these) did not improve the adjusted $r^2$ for females. Volumes from CT scans were moderately (males) to highly (females) correlated with spherical volumes from linear measurements (Table 2).

None of the models using bead volumes to explain variation in CT volumes were strongly supported given the data as indicated by their low Akaike weights for the best-fitting models ($< 0.90$; Table 2; Burnham & Anderson 2002). Nonetheless, the best-fitting female model showed that endocranial volumes decreased slightly over time (Table 2, model 3; Figure 4A), while the best-fitting male model included age, with immatures having smaller volumes than adults (Table 2, model 4;
Figure 4B). Volumes from CT scans were highly positively correlated with volumes from the bead method for both sexes (Table 2).

None of the correlations between CT volumes and linear measures met our minimum criteria ($r^2 \geq 0.88$) for a strong enough relationship such that it might predict endocranial volumes from linear measurements of skulls. Since we want to predict CT volumes from linear measures, we determined whether this was possible by generating prediction intervals for the best-fitting female and male models for the linear measurements (models 1 and 2) and bead method (models 3 and 4; Table 2). We found that the lower and upper limits of the 95% confidence intervals of the predicted values for both sexes show extensive overlap such that individual differences would not be able to be resolved if a new, unvalidated data point was obtained (Table 3).

Comparing method means

Endocranial volume means were significantly different from each other when comparing across methods (mean±standard deviation: females: CT 2.29cm$^3$±0.20, sphere 32459.1mm$^3$±4344.7, bead 2.60ml±0.28; males: CT 2.54cm$^3$±0.15, sphere 59292.1mm$^3$±2360.4, bead 2.91ml±0.21; Welch two sample t-test: females: CT x Sphere $t=31$, $p<0.0001$, df=19; sphere x bead $t=-31$, $p<0.0001$, df=19; bead x CT $t=4$, $p=0.0003$, df=34; males: CT x sphere $t=96$, $p<0.0001$, df=19; sphere x bead $t=-96$, $p<0.0001$, df=19; bead x CT $t=6$, $p<0.0001$, df=35).

Discussion

While female great-tailed grackle endocranial volumes from linear measurements were highly correlated with volumes from CT scans, which we consider a more accurate proxy for brain size than bead volume (Witmer et al. 2008, Knoll et al. 2012), the correlation did not meet our criteria of having
a coefficient of determination \( (r^2) \) equal to or greater than 0.88 – a level of correlation that might be strong enough to allow for the resolution of individual differences in endocranial volumes. This correlation was only moderate in males, which is likely due to the sexual dimorphism in this species. Our sample includes individuals from a range of populations in which the extent of sexual dimorphism might vary. If, in some populations, there has been strong selection for males to increase in body size, then the skeletal measures will not reflect brain size accurately because, in many instances, skeletal size changes faster than brain size as has been found in primates (Montgomery 2011). Perhaps additional biometric measurements would explain more of the variation in their endocranial volumes from CT scans, however we only had access to skulls for most of the specimens and therefore could not test this hypothesis.

We were more interested in whether a given value of an external skull measurement could accurately predict actual endocranial volume from CT scans, rather than setting a subjective criterion about how high \( r^2 \) should be, especially given the extensive debate around the latter approach (e.g., Legates & McCabe Jr. 1999, Müller & Büttner 1994). In particular, \( r^2 \) “[…] describes the proportion of the total variance in the observed data that can be explained by the model” (Legates & McCabe Jr. 1999, p. 233, emphasis added) and thus does not allow one to investigate differences in the variance of individual data points. Our predictive analyses showed that prediction intervals for new data points overlapped to such a degree (within 95% credible intervals) that it was not possible to distinguish among individuals, as we would need to when collecting linear measurements on new individuals in the field. We must conclude that external skull measurements are not accurate enough to estimate endocranial volume in great-tailed grackles.

Predictive analyses are crucial for determining the accuracy of predicting individual data points by a particular method and should be applied extensively in future research, rather than relying solely on correlation coefficients \( (r) \) or coefficients of determination \( (r^2) \). The omission of such an analysis
leaves data uninterpretable for its purported use of discerning intraspecific differences in a morphological feature. Additionally, we caution against using a proxy validated in one species as evidence that the same proxy will apply to other species (e.g., great tits: Dreyer 2012). Until intraspecific validations of brain size proxies using skull or head measurements have been validated across species, we cannot assume that what works (or not) for one species will work (or not) for another.

The bead method was highly correlated with CT volumes in both sexes, however, it also did not meet our minimum criteria ($r^2 \geq 0.88$), and prediction intervals extensively overlapped for individual data points. Great-tailed grackles and common grackles are among the species with the largest ranges in endocranial volumes (as measured using the bead method) when compared with most other species in Iwaniuk & Nelson’s (2002) study on 81 bird species. Both grackle species had large standard deviations when compared with mean volumes (common grackles: mean ± SD=2.59ml ± 0.37, n=10, Iwaniuk & Nelson 2002; great-tailed grackles: female 2.60ml ± 0.28, n=20, male 2.91ml ± 0.21, n=20, this study). It appears that grackle endocranial volumes are more variable than those of many other species. This is not likely due to variation in skull morphology since we did not find a perfect correlation between endocranial volume and external skull measurements. Therefore, we caution against using external skull measurements to estimate endocranial volume without proper validation.

To infer differences in brain size among individuals of the same species, and of the same sex, there must be a high degree of accuracy to have the ability to detect actual individual differences (Legates & McCabe Jr. 1999, Logan & Clutton-Brock 2013). Our results highlight the need to validate brain size and/or endocranial volume proxies and their predictive power for each species under investigation, and for each sex if they are sexually dimorphic. It is unfortunate that there is not an easier, more accurate way to approximate brain size in the field where we have the potential to
understand how evolutionary factors drive brain size variation within species. However, this study accentuates the importance of knowing how accurate brain size measures and proxies are when including such data in analyses.

Acknowledgements

We thank Scott Grafton at the University of California Santa Barbara (UCSB) and Lawrence Harter and Chuck Scudelari at the Pueblo Radiology Medical Group for providing access to the CT scanner; Mario Mendoza at UCSB for scan planning; Sabrina Cruz and Kerri Moore at the Pueblo Radiology Medical Group for scanning; Steve Rothstein for advice and sponsoring CP’s grant; and Mireia Beas-Moix and the Cheadle Center for Biodiversity and Ecological Restoration at UCSB for hosting the skull loans, Dieter Lukas for analysis assistance, and Jay Taylor for feedback on a previous version of this manuscript. We are grateful for skull loans from the Museum of Southwestern Biology (and to Andy Johnson for aging skulls), Krista Fahy and the Santa Barbara Museum of Natural History, and the Ornithology Division of the KU Biodiversity Institute.

References

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

Auerbach C, La Porte HH, Caputo RK. 2004. Evidence-based practice manual: research and outcome measures in health and human services [eds. Roberts AR, Yeager KR]. Oxford University Press

Barton K. 2012. MuMIn: Multi-model inference. R package version 1.7.7. Available: http://CRAN.R-project.org/package=MuMIn. Accessed 3 Jan 2015.

Bonaparte KM, Riffle-Yokoi C, Burley NT (2011) Getting a Head Start: Diet, Sub-Adult Growth, and Associative Learning in a Seed-Eating Passerine. *PLoS ONE* 6(9): e23775.
Burnham KP, Anderson DR. 2002. Model selection and multimodel inference: a practical information-theoretic approach. New York: Springer.

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

Dreyer, S. 2012. Cranium size and fitness measures in a wild population of great tits (Parus major). Masters thesis, University of Oslo.

Gonda A, Herczeg G, Merilä J. 2013. Evolutionary ecology of intraspecific brain size variation: a review. Ecology and Evolution 3:2751-2764.

Hadfield JD. 2010. MCMC Methods for Multi-Response Generalized Linear Mixed Models: The MCMCglmm R Package. Journal of Statistical Software 33:1-22. http://www.jstatsoft.org/v33/i02/. Accessed 3 Jan 2015.

Healy SD, Rowe C. 2007. A critique of comparative studies of brain size. Proceedings of the Royal Society B 274:453-464.

Hutcheon JA, Chiolero A, Hanley JA. 2010. Random measurement error and regression dilution bias. BMJ 340:c2289.

Isler K, Christopher Kirk E, Miller J, Albrecht GA, Gelvin BR, Martin RD. 2008. Endocranial volumes of primate species: scaling analyses using a comprehensive and reliable data set. Journal of Human Evolution 55:967-978.

Iwaniuk AN, Nelson JE. 2002. Can endocranial volume be used as an estimate of brain size in birds? Canadian Journal of Zoology 80:16-23.

Iwaniuk AN, Nelson JE. 2003. Developmental differences are correlated with relative brain size in birds: a comparative analysis. Canadian Journal of Zoology 81:1913-1928.

Johnson K, DuVal E, Kielt M, Hughes C. 2000. Male mating strategies and the mating system of great-tailed grackles. Behavioral Ecology 11:132-141.
Knoll F, Witmer LM, Ortega F, Ridgely RC, Schwarz-Wings D. 2012. The braincase of the basal sauropod dinosaur Spinophorosaurus and 3D reconstructions of the cranial endocast and inner ear. *PLOS ONE* 7:e30060.

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

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

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

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

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

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

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

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

Overington, S. E., Morand-Ferron, J., Boogert, N. J., & Lefebvre, L. (2009). Technical innovations drive the relationship between innovativeness and residual brain size in birds. *Animal Behaviour*, 78(4), 1001-1010.
Overington, S. E., Griffin, A. S., Sol, D., & Lefebvre, L. (2011). Are innovative species ecological generalists? A test in North American birds. *Behavioral Ecology* 22:1286-1293.

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

Pyle P, Howell SN, DeSante DF. 1997. *Identification guide to North American birds* (pp. 421-422). Bolinas, CA: Slate Creek Press.

R Core Team. 2014. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org. Accessed 3 Jan 2015.

Sakai ST, Arsznov BM, Lundrigan B, Holekamp KE. 2011. Brain size and social complexity: a computed tomography study in hyaenidae. *Brain, Behavior and Evolution* 77:91-104.

Sol D, Duncan RP, Blackburn TM, Cassey P, Lefebvre L. 2005. Big brains, enhanced cognition, and response of birds to novel environments. *Proceedings of the National Academy of Sciences of the United States of America* 102:5460-5465.

Taylor, A. H., Elliffe, D., Hunt, G. R., & Gray, R. D. (2010). Complex cognition and behavioural innovation in New Caledonian crows. *Proceedings of the Royal Society of London B: Biological Sciences*, 277(1694), 2637-2643.

Thornton A, Lukas D. 2012. Individual variation in cognitive performance: developmental and evolutionary perspectives. *Philosophical Transactions of the Royal Society B* 367:2773-2783.

Winker K. 2000. Obtaining, preserving, and preparing bird specimens. *Journal of Field Ornithology* 71:250-297.

Witmer LM, Ridgely RC, Dufeau DL, Semones MC. 2008. Using CT to peer into the past: 3D visualization of the brain and ear regions of birds, crocodiles, and nonavian dinosaurs. In *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.
Table 1 (on next page)

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 | Male | Both | Female | Male | Both |
| Volume CT       | 2.30 ± 0.15   | 2.56 ± 0.06   | 2.43 ± 0.17 | 2.38 ± 0.17 | 2.48 ± 0.08 | 2.38 ± 0.16 | 0.99 | 0.009 | 0.95 |
| Volume LWH      | 13814 ± 141   | 15605 ± 127   | 14650 ± 141 | 13779 ± 123 | 15320 ± 78 | 14672 ± 121 | 0.99 | 0.003 | 0.95 |
| Volume bead     | 2.72 ± 0.15   | 2.93 ± 0.15   | 2.88 ± 0.21 | 2.68 ± 0.19 | 2.91 ± 0.13 | 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 | 30.24 ± 0.7 | 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 | 24.01 ± 0.3 | 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 | 21.09 ± 0.5 | 20.68 ± 0.76 | 0.91 | 0.51  | 0.65 |
Table 2 (on next page)

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 0.47 x VolumeSphere + 0.19 x VolumeSphere² + 0.12 x Age – 0.003 x Year + 5.25 |
|         | Male   | 2     | 0.26          | 0.39          | 0.02   | 0.47 x VolumeSphere + 0.19 x VolumeSphere² + 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 0.66 x VolumeBead – 0.09 x Age + 0.66 |
|         | Male   | 4     | 0.45          | 0.68          | <0.0001| 0.66 x VolumeBead – 0.09 x Age + 0.66 |
Table 3 (on next page)

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

|        | 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    |