Associations between Manakin Display Complexity and Both Body and Brain Size Challenge Assumptions of Allometric Correction: A Response to Gutierrez-Ibanez et al. (2016)

Lainy B. Day^a Willow R. Lindsay^{a,b}

^aDepartment of Biology, University of Mississippi, Oxford, Miss., USA; ^bDepartment of Biological and Environmental Sciences, Gothenburg University, Gothenburg, Sweden

Essentially, all models are wrong, but some are useful.

Box and Draper, 1987

The consequence of sexual selection for variation in brain size across animal taxa remains controversial [Jacobs, 1996; Miller, 2000; Madden, 2001; Lefebvre and Sol, 2008] with evidence both for brain enlargement driven by complex cognition [Miller, 2000; Madden, 2001] and for brain reduction driven by competing investments in other sexually selected traits [García-Peña et al., 2013]. We published compelling evidence showing that sexual selection for acrobatic displays is related to the enlargement of brain size in manakins (Pipridae) [Lindsay et al., 2015]. In response, Gutierrez-Ibanez et al. [2016] delivered a commentary focused on arguments about allometric scaling [McCoy et al., 2006; Sol and Price, 2008; Weisbecker and Goswami, 2010; Berner, 2011; Smaers and Soligo, 2013] and phylogenetic independence in comparative analyses [Revell, 2009, 2010]. Reanalyzing our data with alternative methods, Gutierrez-Ibanez et al. [2016] suggest that absolute rather than relative brain size positively predicts display com-

plexity. This conclusion is conceptually similar to our own, and it little alters the broader impacts of our published findings; still, the general relevance of the issues warrants a response.

In their reanalysis, Gutierrez-Ibanez et al. [2016] perform a ‘residual analysis’, which regresses body weight on brain weight with phylogenetic generalized least squares (PGLS) to obtain size-corrected relative brain size, a method that is widely criticized [Newman, 1993; Smith, 1999; Darlington and Smulders, 2001; Garcia-Berthou, 2001; McCoy et al., 2006; Hoverman and Relyea, 2008; Revell, 2009; Berner, 2011]. An implicit assumption of relative brain size analysis is that selection is neutral regarding body size and, accordingly, that the portion of brain tissue needed to control bodily functions can be statistically removed [Deaner et al., 2007] – an improbable assumption for many sexually selected traits in sexually dimorphic species [García-Peña et al., 2013] and a particularly dubious assumption when the trait in question is a physically demanding display. Avoiding these pitfalls, we corrected for body size using general linear modeling (GLM) to obtain estimated marginal means

(EMMs) representing adjusted brain weight [Day et al., 2008, 2011] for each species [Lindsay et al., 2015]. Both the residuals analysis and our GLM were followed by a PGLS regression to examine the relationship between relative or adjusted brain weight and manakin display complexity.

Focusing their commentary on statistical issues, Gutierrez-Ibanez et al. [2016] neglect the truly interesting finding that both body weight and brain weight are positively associated with the complexity of courtship display in manakins, and overlook the fact that the strength of these variable interactions clearly contraindicate both residual analyses and GLM [Newman, 1993; Garcia-Berthou, 2001; Engqvist, 2005; McCoy et al., 2006; Dormann et al., 2013]. Thus, in relation to our data set [Lindsay et al., 2015], the statistical arguments and alternatives proposed by Gutierrez-Ibanez et al. [2016] are rendered moot; their methods do not provide a useful model of relative brain size in relationship to courtship complexity in manakins. The commentary highlights the difficulty in interpreting biological relationships between size-corrected morphological variables and a dependent variable in the many

data sets similar to ours [Lindsay et al., 2015] where all the variables (e.g. brain weight, body weight and complexity) are highly interrelated [Deacon, 1990; Deaner et al., 2007].

Courtship Complexity Is Related to Brain Weight and Body Weight

In a re-examination of our data preceding the commentary [Gutierrez-Ibanez et al., 2016] (see suppl. material at www.karger.com/doi/10.1159/000446341), we discovered that both body weight and absolute brain weight are positively related to display complexity, findings which are in keeping with the concluding statements made by Gutierrez-Ibanez et al. [2016]. This unexpected relationship between courtship complexity and body weight is intriguing in light of the muscularly demanding, acrobatic nature of courtship display in manakins [Day et al., 2007; Fusani et al., 2007a, b; Schlinger et al., 2008a, b; Fuxjager et al., 2015; Lindsay et al., 2015]. Gutierrez-Ibanez et al. [2016] show that there is a positive relationship between absolute courtship complexity and both brain weight and body weight. Based on lambda (λ) values of 0 or 1, respectively, they judge brain weight as tentatively significant ($p = 0.023$ or 0.085). Body weight ($p = 0.052$ or 0.093) is deemed nonsignificant because they assume that Pagel's λ , which adjusts analyses for lack of phylogenetic independence among species, is closer to 1 (Brownian motion assumed) than to 0 (phylogenetic independence). Our own reanalysis, using maximum likelihood to calculate the appropriate λ between 0 and 1 [Symonds and Blomberg, 2014], yields slightly lower p values. We find significant relationships for body weight and display complexity ($p = 0.039$) and for brain weight and complexity ($p = 0.016$). Differences between our values and those produced by Gutierrez-Ibanez et al. [2016] may have occurred if they did not log-transform brain and body weight to improve the fit to normality [Pérez-Barbería et al., 2007; Swanson et al., 2012; Symonds et al., 2014]. In any case, we would argue that $p = 0.052$ suggests a biologically significant relationship. Additionally, we analyzed the effects of increasing λ in intervals of 0.2; the brain weight relationship remains significant for $\lambda = 0.2, 0.4$ and 0.8 , and the body weight relationship remains significant for $\lambda = 0.2$ and 0.4 . For brain weight, the maximum likelihood λ is

not significantly different from $\lambda = 0$ ($p = 1.0$), but it is significantly different from $\lambda = 1$ ($p = 0.046$). For body weight, λ is not significantly different from $\lambda = 0$ ($p = 1.0$), but trends towards significance for $\lambda = 1$ ($p = 0.056$). Similar maximum likelihood profiles are not uncommon, occurring in 24% of traits [Freckleton et al., 2002]. Importantly, Gutierrez-Ibanez et al. [2016] showed no significant difference in their PGLS models using $\lambda = 0$ compared to a $\lambda = 1$, unequivocally demonstrating that assuming $\lambda = 1$ does not model our data better than assuming $\lambda = 0$. Thus, it is unreasonable to arbitrarily adopt a $\lambda = 1$ [Shultz and Dunbar, 2010]; and our use of a maximum likelihood to estimate λ (Caper software package in R; Orme [2013]) is supported.

Despite these clear results, Gutierrez-Ibanez et al. [2016] argue that a strong phylogenetic signal (high λ) is parsimoniously assumed for all their analyses of our data because a phylogenetic signal was found for 92% of 121 traits (e.g. tail length) examined in published studies [Blomberg et al., 2003]. However, trait λ s are not related to λ s for a regression of these traits, as λ from the regression relates to residual errors of the regression [Symonds and Blomberg, 2014]. Thus, data on trait λ s cannot be used to make assumptions about regression λ s, as Gutierrez-Ibanez et al. [2016] imply. While we acknowledge that traits examined in ecology often show a phylogenetic signal (60% of traits in Freckleton et al. [2002]), this is hardly a reason to assume a phylogenetic signal for traits in our data set. Whether a particular trait in a given lineage shows a phylogenetic signal is an empirical question and is influenced by a number of factors, including the trait examined, the tree structure and the index used [Blomberg et al., 2003; Münkemüller et al., 2012]. λ appears to be the most appropriate index for our data as it is less sensitive to small species numbers (<20) than three other commonly used signal indices [Münkemüller et al., 2012] and allows for the use of maximum likelihood estimates of λ that will optimize models regardless of power, thereby negating concerns about low power with small sample sizes [Freckleton et al., 2002]. Additionally, given the robust association of complexity with brain and body weight across a wide range of λ s, concerns about the accuracy of predicting λ are obviated. Still, the effect of species number on λ could be analyzed empirically for a larger subset of the 56 or so species of

manakins (vs. our $n = 13$). Body weight and other morphological variables for most manakin species are easily found in museum databases, and display complexity measures can be obtained from the literature for the majority of species, as outlined in Lindsay et al. [2015].

Residual and Absolute Brain Weight Are Correlated

Gutierrez-Ibanez et al. [2016] questioned our use of a nonphylogenetically informed GLM and the use of EMMs from this GLM in a secondary analysis. We completely agree that trait variables should be adjusted for phylogenetic dependence [Revell, 2009]. We chose to use a GLM because of known problems with residual analyses and the recognized advantages of GLM. GLM allows you to include potentially meaningful intraspecific variation in your model [Revell and Reynolds, 2012; Garamszegi, 2014]. Using GLM also allows a simple test for violation of the assumption of heterogeneity of slopes rather than assuming homogeneity and using species means, as the residuals analysis does. Lastly, phylogenetically informed models with similar properties were rather novel at the time [Revell and Reynolds, 2012]. Reporting λ values for our traits would have been more theoretically inclusive, but in light of the values we have now demonstrated for λ , adding phylogenetic adjustments would have been unlikely to alter the EMMs substantively.

Using GLM, brain size EMMs were adjusted for body weight [Lindsay et al., 2015]. EMMs were little altered from absolute brain weight [Gutierrez-Ibanez et al., 2016] because of the strong relationship between body weight and brain weight. EMMs are not arithmetic means [Gutierrez-Ibanez et al., 2016] but are predicted from the linear regression of the dependent variable (e.g. brain weight) and the covariate (e.g. body weight) using the weighted within-groups (e.g. species) regression coefficient for the slope and subtracting from each species the family mean for the covariate [Goodnight and Harvey, 1978; Searle et al., 1980; Field, 2013]. Thus, EMMs are adjusted brain weights [Goodnight and Harvey, 1978; Searle et al., 1980; Cotton et al., 2004; McCoy et al., 2006; Hoverman and Relyea, 2008; Berner, 2011] and effectively adjust for allometry [as per Cotton et al., 2004; Hoverman and Relyea, 2008; Berner,

2011]. However, assumptions of the GLM must be met in order for estimated means to be biologically meaningful.

We quantitatively met all of the modeling assumptions for GLM, but consider that qualitatively interpreting the statistics would have given a broader understanding of our data. Our nonsignificant interaction term between species and body weight met the model assumptions, but the low $p = 0.056$ and the partial $\eta^2 = 0.48$ suggest species identity is not independent from body weight. Because body weight is observed rather than experimentally obtained, a strong interaction does not necessarily violate model assumptions, but reminds us that EMMs are calculated at a value of body weight that may not be ideally representative of each species [Grace-Martin, 2012]. While it is possible to adjust for species by body size interactions when they exist [Engqvist, 2005], this is unrealistic with our sample sizes. Similarly, within-species sample sizes make a lack of slope heterogeneity ($p = 0.1$) difficult to interpret, which means that the EMM regression coefficient may not best represent the slope for all species. The same inherent data structure that produces little adjustment of brain weight for a body weight covariate also leads to residual analyses results that are incomprehensible.

Conflation of Body and Brain Weight

As noted by Gutierrez-Ibanez et al. [2016], using size-corrected residuals in a secondary analysis has been widely criticized for a variety of reasons related to assumptions about measurement error and the homogeneity of group slopes [Newman, 1993; Smith, 1999; Darlington and Smulders, 2001; Garcia-Berthou, 2001; Freckleton, 2002]. Some of these problems with residuals analysis can be corrected (as with reduced major axis regression [Shultz and Dunbar, 2010]), but it has become commonplace to acknowledge the flaws of residuals analyses and use this method anyway (as in Gutierrez-Ibanez et al. [2016] and citations within), seemingly for a historical lack of an alternative, readily available, phylogenetically informed model. However, residuals analysis is not just statistically flawed; when applied to our data set, the results cannot be meaningfully interpreted.

By definition, residual analysis removes the correlation between body weight and the residual brain weights [Gutierrez-

Ibanez et al., 2016], thereby obfuscating any effects of body weight on courtship complexity, resulting in residual relative brain weights that are confounded by such interactions. Thus, relative brain weight is obtained under the implicit assumption that there is no relationship between body weight and the outcome variable (e.g. complexity [Deaner et al., 2007]). Gutierrez-Ibanez et al. [2016] showed that brain weight and body weight are similarly related to complexity and were thus aware of the strong associations among these variables. Under these statistical conditions, residuals analysis produces far too conservative results irrespective of sample size, and makes the true contributions of brain size and body size to complexity impossible to untangle [Darlington and Smulders, 2001; Garcia-Berthou, 2001; Freckleton, 2002]. The stronger the linear association between brain weight and body size, the more biased is relative brain weight as residuals from the estimated regression line will show little variance from this shared variable relationship.

Models and Meanings for Bodies and Brains

For our data set, residual relative brain weights are overcorrected and EMM-adjusted brain weights may under adjust brain weight. The brain weight, body weight and display complexity relationships that contribute to these biases are unlikely to be mitigated by using other PGLS models (not be confused with GLM, as in Gutierrez-Ibanez et al. [2016]), for which issues of collinearity are also a problem [Dale et al., 2015].

If collinearity is not a problem, the use of EMMs in a secondary PGLS to test the relationship between a size-adjusted variable and another variable (e.g. courtship complexity) is tenable, although it is uncommon [but see McCoy et al., 2006; Hoverman, and Relyea, 2008; Fuxjager et al., 2015]. Gutierrez-Ibanez et al. [2016] have not provided conceptual reasons to undermine this approach and do not perform a technical evaluation of this statistical technique [as in Sayol et al., 2016]. An empirical evaluation is beyond the scope of our commentary response. However, a similarity in the results of GLM and residuals analysis based on individual brain weights suggests that both methods could be reasonable approaches for comparative analyses.

Testing for proof-of-concept rather than conducting a final phylogenetically informed model, we have found that cerebellar volume, corrected for whole brain volume (minus cerebellum volume) and then regressed on display complexity using PGLS, is significantly associated with complexity at $p = 0.05$ (GLM) and $p = 0.06$ (residuals analysis) (online suppl. material; Pano et al., [2015]). Similarly, for another brain region, these values were $p = 0.05$ and $p = 0.08$.

GLM and residuals calculated for individuals before means are taken reduce – but do not eliminate – the relationship between the species means for adjusted cerebellum volume and the size correction variable, namely brain weight minus cerebellar volume, and thus are not measures of brain region ‘relative’ to body weight according to the definition of Gutierrez-Ibanez et al. [2016]. It seems essential to their argument that allometric corrections mathematically decouple brain weight from body weight as species-level residual analyses do. However, the authors contradict themselves when they accept the use of PGLS models for size correction when at least some of these models would not remove the unique variance of body weight and/or brain weight on a third variable [Pérez-Barbería et al., 2007; Swanson et al., 2012; Smaers, and Soligo, 2013; Symonds et al., 2014].

Thus, we suggest that Gutierrez-Ibanez et al. [2016] have not actually shown that relative brain size is not related to courtship complexity in manakins, but rather have demonstrated that a simple residual analysis cannot model this relationship for our data. Our manakin data set now includes morphometrics for multiple brain regions that each have some correlation with one another and with body weight. It will be important to use phylogenetically informed multivariate analyses, likely in conjunction with a step-wise or sequential regression to obtain reduced models [Vincze et al., 2015], in order to identify the biologically meaningful variables that contribute to courtship complexity. Preliminary analyses suggest that body weight may better account for display complexity than regional brain volumes or whole brain volume, and that the tarsus length cubed, a common measure used to estimate body size, is not associated with courtship complexity. If these relationships hold, we would suggest that body mass is related to complexity as a consequence of coordinated motor function being necessary for the performance of court-

ship display [Fusani et al., 2007b, 2014; Schlinger et al., 2013; Fuxjager et al., 2012a, b, 2015]. We would also speculate that body mass relationships with complexity are linked to muscle hypertrophy and the increased bone density that is associated with certain display elements [Schlinger et al., 2013; Friscia et al., 2016]. Measures of body size (not mass), like the tarsus length cubed, would be less likely to relate to courtship display. Lack of bone length associations would fit with our impression that manakins with more complex displays do not 'look' bigger. Undoubtedly, some forms of advanced learning, including knowledge of local spatial relationships [Coccon et al., 2012], are involved in courtship complexity. However, demands on the peripheral systems of manakins may override the importance of cognitive factors in driving the evolution of complex display, given the highly postural, ballistic movements required of those species with the most complex displays. Such a result would not mean that the brain was unimportant for display, but rather point to the importance of musculoskeletal and neuromuscular systems in the courtship of manakins.

The primary alternative interpretation of our data, that absolute but not relative brain mass influences courtship display [Gutierrez-Ibanez et al., 2016], is supported by researchers who have found that ab-

solute measures of brain or total neuron numbers are more relevant to predicting cognitive and behavioral complexity than is relative brain size [Deaner et al., 2007; Azevedo et al., 2009; Herculanou-Houzel, 2012]. What these studies and our data highlight is a general problem with analyses that seek to remove the effects of body size allometry based on the long-held [Jerison, 1973; Fox and Wilczynski, 1986; Deacon, 1997; Deaner et al., 2000] but unsupported [Deaner et al., 2000; Deaner et al., 2007] assumption that we can isolate the portion of the nervous system performing functions related to control of bigger bodies from those portions needed to control cognitive and other functions in these bigger bodies by removing body size statistically. While we obviously need to adjust for body size when asking questions about mental capacity between certain groups, such as men and women or mice and men, we are now recognizing that many standard correction methods (residuals analysis, EQ) over adjust and obscure true relationships between neural capacity and performance measures. For instance, in non-human primates at least some portion of brain mass that is removed by the EQ or residuals analysis must contribute to cognition because this removal reduces the ability to predict cognitive skills [Deaner et al., 2007]. In manakins, where one would

predict that a good deal of the brain would serve functions related to the neuromuscular system, it seems obvious from a theoretical standpoint that removing body size-dependent somatic control elements would also eliminate portions of the brain that are important for courtship display. Thus, relative brain weight cannot appropriately account for neural capacity in manakins.

With the increasing demands by journals and funding agencies to publish raw data, peers will have greater ease in reanalyzing and evaluating published findings. This greatly promotes transparency and provides excellent resources for meta-analyses and large-scale comparative analyses. With this wealth of data also comes the responsibility to focus on modeling with the goal of best explaining the relevant biological phenomena and avoidance of emphasis on cutoff p values of 0.05 [Siegfried, 2015] determined by particular statistical approaches. Favored statistical methods change based on zeitgeist as well as improvements in modeling and computing. We appreciate the opportunity to respond to the commentary of Gutierrez-Ibanez et al. [2016] as it has highlighted improvements in phylogenetically informed modeling that will be necessary to address relationships between brain morphometrics and courtship complexity in manakins and other species in the future.

References

- Azevedo FC, Carvalho LRB, Grinberg LT, Farfel JM, Ferretti REL, Leite REP, Filho, WJ, Lent R, Herculanou-Houzel S (2009): Equal numbers of neuronal and nonneuronal cells make the human brain an isometrically scaled-up primate brain. *J Comp Neurol* 513:532–541.
- Berner D (2011): Size correction in biology: how reliable are approaches based on (common) principal component analysis? *Oecologia* 166:961–971.
- Blomberg SP, Garland T, Ives AR (2003): Testing for phylogenetic signal in comparative data: behavioral traits are more labile. *Evolution* 57: 717–745.
- Coccon F, Schlinger BA, Fusani L (2012): Male golden-collared manakins *Manacus vitellinus* do not adapt their courtship display to spatial alteration of their court. *Ibis (Lond 1859)* 154: 173–176.
- Cotton S, Fowler K, Pomiankowski A (2004): Condition dependence of sexual ornament size and variation in the stalk-eyed fly *Cyrtodiopsis Dalmanni* (Diptera: Diopsidae). *Evolution (NY)* 58:1038–1046.
- Dale S, Liffield JT, Rowe M (2015): Commonness and ecology, but not bigger brains, predict urban living in birds. *BMC Ecol* 15:12.
- Darlington RB, Smulders TV (2001): Problems with residual analysis. *Anim Behav* 62:599–602.
- Day LB, Fusani L, Hernandez E, Billo TJ, Sheldon KS, Wise PM, Schlinger BA (2007): Testosterone and its effects on courtship in golden-collared manakins (*Manacus vitellinus*): seasonal, sex, and age differences. *Horm Behav* 51:69–76.
- Day LB, Fusani L, Kim C, Schlinger B (2011): Sexually dimorphic neural phenotypes in golden-collared manakins (*Manacus vitellinus*). *Brain Behav Evol* 77:206–218.
- Day LB, Guerra M, Schlinger BA, Rothstein SI (2008): Sex differences in the effects of captivity on hippocampus size in brown-headed cowbirds (*Molothrus ater obscurus*). *Behav Neurosci* 122:527–534.
- Deacon TW (1990): Fallacies of progression in theories of brain-size evolution. *Int J Primatol* 11:193–236.
- Deacon TW (1997): *The Symbolic Species: The Co-Evolution of Language and the Brain*. New York, W. W. Norton & Company.
- Deaner RO, Isler K, Burkart J, van Schaik C (2007): Overall brain size, and not encephalization quotient, best predicts cognitive ability across non-human primate. *Brain Behav Evol* 70:115–124.
- Deaner RO, Nunn CL, van Schaik CP (2000): Comparative tests of primate cognition: different scaling methods produce different results. *Brain Behav Evol* 55:44–52.
- Dormann CF, Elith J, Bacher S, Buchmann C, Carl G, Carré G, Marquéz JRG, Gruber B, Laffourcade B, Leitão PJ, Münkemüller T, McClean C, Osborne PE, Reineking B, Schröder B, Skidmore AK, Zurell D, Lautenbach S (2013): Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. *Ecography* 36:27–46.
- Engqvist L (2005): The mistreatment of covariate interaction terms in linear model analyses of behavioural and evolutionary ecology studies. *Anim Behav* 70:967–971.

- Field AP (2013): Discovering statistics using IBM SPSS statistics: and sex and drugs and rock 'n' roll. Los Angeles, Sage. <http://umiss.idm.oclc.org/login?url=http://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,url,uid&db=catt02590a&AN=olemiss.b6700521&site=eds-live&scope=site>.
- Fox JH, Wilczynski W (1986): Allometry of major CNS divisions: towards a reevaluation of somatic brain-body scaling. *Brain Behav Evol* 28:157–169.
- Freckleton RP (2002): On the misuse of residuals in ecology: regression of residuals vs. multiple regression. *J Anim Ecol* 71:542–545.
- Freckleton RP, Harvey PH, Pagel M (2002): Phylogenetic analysis and comparative data: a test and review of evidence. *Am Nat* 160:712–726.
- Frischia A, Sanin GD, Lindsay WR, Day LB, Schlinger BA, Tan J, Fuxjager MJ (2016): Adaptive evolution of a derived radius morphology in manakins (Aves, Pipridae) to support acrobatic display behavior. *J Morphol*. Epub ahead of print.
- Fusani L, Barske J, Day LD, Fuxjager MJ, Schlinger BA (2014): Physiological control of elaborate male courtship: female choice for neuromuscular systems. *Neurosci Biobehav Rev* 46: 534–546.
- Fusani L, Day LB, Canoine V, Reinemann D, Hernandez E, Schlinger BA (2007a): Androgen and the elaborate courtship behavior of a tropical lekking bird. *Horm Behav* 51:62–68. <http://www.sciencedirect.com/science/article/pii/S0018506X06002546>.
- Fusani L, Giordano M, Day LB, Schlinger BA (2007b): High-speed video analysis reveals individual variability in the courtship displays of male golden-collared manakins. *Ethology* 113:964–972.
- Fuxjager MJ, Barske J, Du S, Day LB, Schlinger BA (2012a): Androgens regulate gene expression in avian skeletal muscles. *PLoS One* 7:e51482.
- Fuxjager MJ, Eaton J, Lindsay WR, Salwiczek LH, Rensel M, Barske J, Schlinger B (2015): Evolutionary patterns of adaptive acrobatics and physical performance predict expression profiles of androgen receptor – but not oestrogen receptor – in the forelimb musculature. *Funct Ecol* 29:1197–1208.
- Fuxjager MJ, Schultz JD, Barske J, Feng NY, Fusani L, Mirzamani A, Schlinger BA (2012b): Spinal motor and sensory neurons are androgen targets in an acrobatic bird. *Endocrinology* 153:3780–3791.
- Garamszegi LZ (2014): Uncertainties due to within-species variation in comparative studies: measurement errors and statistical weights; in: *Modern Phylogenetic Comparative Methods and Their Application in Evolutionary Biology*. Berlin, Springer, pp 157–199.
- García-Berthou E (2001): On the misuse of residuals in ecology: testing regression residual vs the analysis of covariance. *J Anim Ecol* 70: 708–711.
- García-Peña GE, Sol D, Iwaniuk AN, Székely T (2013): Sexual selection on brain size in shorebirds (Charadriiformes). *J Evol Biol* 26:878–888.
- Goodnight JH, Harvey WR (1978): Least-Squares Means in the Fixed Effect General Linear Model. Cary, SAS Institute.
- Grace-Martin K (2012): When assumptions of ANCOVA are irrelevant. *StatsBlogs*. <http://www.statsblogs.com/2012/10/15/when-assumptions-of-ancova-are-irrelevant/> (accessed February 1, 2016).
- Gutierrez-Ibanez C, Iwaniuk AN, Wylie DR (2016): Relative brain size is not correlated with display complexity in manakins: a reanalysis of Lindsay et al. (2015). *Brain Behav Evol*, DOI 10.1159/000446312, in press.
- Herculano-Houzel S (2012): The remarkable, yet not extraordinary, human brain as a scaled-up primate brain and its associated cost. *Proc Natl Acad Sci* 109:10661–10668.
- Hoverman JT, Relyea RA (2008): Temporal environmental variation and phenotypic plasticity: a mechanism underlying priority effects. *Oikos* 117:23–32.
- Jerison H (1973): *Evolution of the Brain and Intelligence*. New York, Academic Press.
- Lindsay WR, Houck JT, Giuliano CE, Day LB (2015): Acrobatic courtship display coevolves with brain size in manakins (Pipridae). *Brain Behav Evol* 85:29–36.
- McCoy MW, Bolker BM, Osenberg CW, Miner BG, Vonesh JR (2006): Size correction: comparing morphological traits among populations and environments. *Oecologia* 148:547–554.
- Münkemüller T, Lavergne S, Bzeznik B, Dray S, Jombart T, Schiffrers K, Thuiller W (2012): How to measure and test phylogenetic signal. *Methods Ecol Evol* 3:743–756.
- Newman MC (1993): Regression analysis of log-transformed data: statistical bias and its correction. *Environ Toxicol Chem* 12:1129–1133.
- Orme D (2013): The caper package: comparative analysis of phylogenetics and evolution in R, package version 5.
- Pano G, Lindsay WR, Day LB (2015): Do the sexiest dancers have the largest little brain? Associations between display complexity and both brain volume and cerebellar granular layer volume in manakins (Pipridae). *Society for Neuroscience. Program No. 756.09*. 2015 Neuroscience Meeting Planner. Chicago.
- Pérez-Barbería FJ, Shultz S, Dunbar RIM (2007): Evidence for coevolution of sociality and relative brain size in three orders of mammals. *Evolution* (NY) 61:2811–2821.
- Revell LJ (2009): Size-correction and principal components for interspecific comparative studies. *Evolution* 63:3258–3268.
- Revell LJ (2010): Phylogenetic signal and linear regression on species data. *Meth Ecol Evol* 1:319–329.
- Revell LJ, Reynolds RG (2012): A new Bayesian method for fitting evolutionary models to comparative data with intraspecific variation. *Evolution* 66:2697–2707.
- Sayol F, Lefebvre L, Sol D (2016): Relative brain size and its relation with the associative pallidum in birds. *Brain Behav Evol*, Epub ahead of print.
- Schlinger BA, Barske J, Day L, Fusani L, Fuxjager MJ (2013): Hormones and the neuromuscular control of courtship in the golden-collared manakin (*Manacus vitellinus*). *Front Neuroendocrinol* 34:143–156.
- Schlinger B, Day LB, Fusani L (2008a): Behavior, natural history and neuroendocrinology of a tropical bird. *Gen Comp Endocrinol* 157: 254–258.
- Schlinger B, Fusani L, Day L (2008b): Hormonal control of courtship in male golden-collared manakins (*Manacus vitellinus*). *Ornitol Neotrop* 19:229–239. http://docente.unife.it/leonida.fusani/copy_of_corso-modelli-animali/il-corteggiamento-del-manachino-col-lodorato/Schlinger_etal2008.pdf.
- Searle SR, Speed FM, Milliken GA (1980): Population marginal means in the linear model: an alternative to least squares means. *Am Stat* 34: 216–221.
- Shultz S, Dunbar RIM (2010): Species differences in executive function correlate with hippocampus volume and neocortex ratio across nonhuman primates. *J Comp Psychol* 124: 252–260.
- Siegfried T (2015): P value ban: small step for a journal, giant leap for science. *Science News*. <https://www.sciencenews.org/blog/context/p-value-ban-small-step-journal-giant-leap-science> (accessed February 3, 2016).
- Smaers JB, Soligo C (2013): Brain reorganization, not relative brain size, primarily characterizes anthropoid brain evolution. *Proc Biol Sci* 280: 20130269.
- Smith RJ (1999) Statistics of sexual size dimorphism. *J Hum Evol* 36:423–458.
- Swanson EM, Holekamp KE, Lundrigan BL, Arsznov BM, Sakai ST (2012): Multiple determinants of whole and regional brain volume among terrestrial carnivores. *PLoS One* 7:e38447.
- Symonds MRE, Blomberg SP (2014): A primer on phylogenetic generalised least squares; in: *Garamszegi LZ (ed): Modern Phylogenetic Comparative Methods and Their Application in Evolutionary Biology*. Berlin, Springer, pp 451–464.
- Symonds MRE, Weston MA, Robinson RW, Guay PJ (2014): Comparative analysis of classic brain component sizes in relation to flightiness in birds. *PLoS One* 9:e91960.
- Vincze O, Vágási CI, Pap PL, Osvath G, Möller AP (2015): Brain regions associated with visual cues are important for bird migration. *Biol Lett* 11:1–4.
- Weisbecker V, Goswami A (2010): Brain size, life history, and metabolism at the marsupial/placental dichotomy. *PNAS* 107:16216–16221.