International Journal of Primatology
https://doi.org/10.1007/s10764-023-00399-y

Pseudoreplication in Primate Communication Research: 10 Years On

J. Whitehouse[1] · P. R. Clark[2] · J. Micheletta[3] · K. Liebal[4] · K. E. Slocombe[5] · B. M. Waller[1]

Received: 13 July 2023 / Accepted: 11 September 2023

Abstract

Pseudoreplication is the statistical error of collecting numerous datapoints from a single unit (such as an individual), which are not independent, and applying statistical methods that assume independence of data. Importantly, pseudoreplication increases the chances of Type 1 errors (i.e., false positives), bringing findings and conclusions based on pseudoreplicated analyses into question. Ten years ago, Waller et al. (2013) published a paper highlighting the prevalence of statistical pseudoreplication throughout the nonhuman primate communication literature. In this current study, we examined the literature published since the original publication (between 2009 and 2020; 348 papers) to assess whether pseudoreplication is still as widespread as it was, if it has become more problematic, or if the field is beginning to overcome this issue. We find that there has been a significant decrease in pseudoreplication over the past ten years (38.6% then, compared with 23.0% now). This reduction in pseudoreplication appears to be associated with an increase in the use of multilevel models throughout primatology (which allow for nonindependent data to be nested appropriately). Pseudoreplication was historically more prevalent in research using observational (vs. experimental) methods and those working with wild (vs. captive) primates. However, these biases do not seem to exist in more

Badges earned for open practices: Open Data and Open Analytical Code. Experiment materials and data are available in the respository at https://osf.io/SCDFX.

Handling Editor: Joanna (Jo) M. Setchell

Extended author information available on the last page of the article

Published online: 09 October 2023

recent literature with a more comparable likelihood of pseudoreplication seen across the field regardless of methods. Although these current findings relate specifically to primate communication research, we think they will translate broadly across nonhuman communication research, and throughout biology. We continue to emphasise the need to monitor these issues, as although now seen at much lower rates, pseudoreplication is still present and therefore potentially impacting the accuracy of findings.

Keywords Pseudoreplication · Communication · Statistical analysis · Facial expression · Gesture · Vocalization · Multilevel modelling

Introduction

The problems associated with pseudoreplication, also called the pooling fallacy (Machlis *et al*., 1985) or the false treatment of nonindependent data as independent during a statistical analysis, were first highlighted nearly 40 years ago (Hurlbert, 1984). Since then, a variety of scientific fields have highlighted pseudoreplication as an ongoing statistical issue (Freeberg & Lucas, 2009; Heffner *et al*., 1996; Johnson & Freeberg, 2016; Kroodsma *et al*., 2001; Lazic, 2010; Waller *et al*., 2013). Taking repeated samples from a single source (e.g., from an individual or call sequence) without considering that they are not truly independent falsely inflates sample size and statistical power, thus increases the likelihood of incorrect inference (Machlis *et al*., 1985). In an alternate view, Schank and Koehnle (2009) suggested that the term pseudoreplication should be avoided altogether when labelling experimental research and that the use of nonindependent data in statistics is not always a problem. Instead, they suggest we should critique experimental design on a case-by-case basis. Although this debate about whether pseudoreplication remains as problematic continues (Oksanen, 2001; Schank & Koehnle, 2009; Cottenie & De Meester, 2003; Hurlbert, 2004; Freeberg & Lucas, 2009; Oksanen, 2004; David & Gray, 2015; Colegrave & Ruxton, 2018), there is a general consensus among researchers that if pseudoreplication can be avoided, then it should be.

One scientific field where the prevalence of pseudoreplication is relatively high is that of primate communication research (Waller *et al*., 2013). This field may be prone to pseudoreplication due to a variety of reasons. First, repeated observation of a communicative behavior, such as a facial expression, gesture, or call from a single individual, often is necessary to capture the typical variation expected within that behavior and to reduce noise in the data. Therefore, repeated observations are necessary to make an appropriate inference about its function. Second, sample size in primate behavior research is particularly limited by access to subjects, and access to many individual primates for research often is challenging and impractical—especially for more endangered or rare species. Therefore, compromises on sample sizes often are unavoidable, both in terms of subjects for observations/experiments and for the collection and production of experimental stimuli for playback. A combination of these reasons, the repeated observation of behavior on fewer individuals, is likely to inflate the risk of pseudoreplication in this field as individual observations then become a convenient unit for analysis for researchers (Waller *et al.*, 2013). Ten years ago, Waller *et al.* (2013) scrutinised 551 peer-reviewed published papers in the field of primate communication and found that more than one-third contained evidence of pseudoreplicated data, which increased to more than 60% when considering only those articles using naturalistic observational methods. Associated with increased risk of pseudoreplication were those studies using observational methods (vs. experimental) and with a focus on wild subjects (vs. captive).

In the past decade, statistical approaches have shifted toward the use of multilevel models, such as generalized linear mixed models (GLMMs) for hypothesis testing. These methods initially increased in popularity throughout the field of ecology but have since been widely adopted in the behavioral sciences (Bolker *et al.*, 2009). GLMMs provide a flexible solution to approaching data that have typically been problematic for traditional statistical approaches, such as data that have nonnormal distributions, where observations are missing entirely, and when there is the presence of repeated observation and complex nesting of data (Lee & Nelder, 2001). Originally, without being able to directly account for "random effects" (e.g., additional levels of potential variation in the data) in statistical testing, aggregation of data often was the only solution to avoid problematic pseudoreplication. Aggregation, however, comes with its own problems (Pollet *et al.*, 2015). Not only does aggregation incur a significant loss of information from averaging multiple observations into a single datapoint, but aggregated data are susceptible to other statistical problems, such as the ecological fallacy or Simpson's paradox, where a trend appears across several groups but subsequently disappears when data within each group are pooled. Multilevel modelling is a potential solution to both the issues of pseudoreplication and aggregation (Millar & Anderson, 2004; Pollet *et al.*, 2015; Schank & Koehnle, 2009; Waller *et al.*, 2013), and it is likely that its popularity contributed to increased awareness of these issues in the behavioral sciences in recent years. We should therefore expect that any increase in the popularity of these statistical methods will incur a positive impact on the prevalence of pseudoreplicated data (i.e., a reduction) in more recent research in primate communication.

In this current study, we revisited the question examined by Waller *et al.* (2013) and explored the current state of pseudoreplication in primate communication research post-2009. We hypothesize (and hope) that the increased awareness of both the problem and the possible solutions has reduced the prevalence of pseudoreplication. If this is the case, we predict a greater reduction in pseudoreplication in research using the methods which have historically been particularly prone (e.g., those incorporating observational data collection; Waller *et al.*, 2013) in the post-2009 data compared with the pre-2009 data. We also predict studies which incorporate modern statistical methods, which allow researchers to nest nonindependent data appropriately (e.g., GLMMs) to increase across time, and consequently help to drive a reduction in pseudoreplication. Finally, we revisit the impact of subject species, journal impact, and sample size on the likelihood of pseudoreplication.

Methods

Data

We used two datasets for analysis and comparison in this study. The first dataset is from Waller *et al.* (2013) and consisted of 309 empirical peer-reviewed articles on the topic of nonhuman primate communication, published between the years 1960 to 2008 inclusive, and written in English. We collected a second comparable dataset for articles published between 2009 and 2020 consisting of 347 articles from Liebal *et al.* (2022). We built both datasets by using identical methods via a keyword search of two literature databases (Web of science, Science Direct). More details of this literature search are located in the respective publications and in the supplementary information (SI1). We coded all articles for subject taxa (great ape, or nongreat ape), communicative modality focus (facial, gestural, or vocal), research method (observational, experimental), research environment (wild, or nonwild), the citations per year (as of 2021), and the impact factor/Source Normalized Impact per Paper (SNIP) of the journal published (as of 2021). SNIP scores are a metric quantifying citation potential, or more specifically, the ratio between the article's average citation count and the number of citations expected for the field of that article.

Coding for Pseudoreplication

We read all articles published between 2009 and 2020 and scrutinised data analyses for pseudoreplication following the procedure of Waller *et al.* (2013). We classified each article as 1) reporting pseudoreplicated data, 2) not reporting pseudoreplicated data, 3) presenting no data or statistics, 4) pseudoreplication was undeterminable due to lack of statistical information in the article, and 5) presenting data below the level of the individual (e.g., at the level of the neuron) and considered beyond the scope of this article.

We coded articles as pseudoreplicated if there was evidence that nested/nonindependent data had not been appropriately accounted for in the statistical procedures. We coded the presence of pseudoreplication if the degrees of freedom of statistical tests exceeded the sample size (with exception of some repeated measures analyses, where degrees of freedom can exceed sample size) or if appropriate random effects were not included in modelling approaches. We coded the presence of pseudoreplication if statistical tests were conducted on the level of the observation (such as individual communicative signal) without aggregation by individual, or without evidence that nonindependence was suitably controlled for. Finally, we coded pseudoreplication where researchers used some exemplars of stimuli repeatedly in experiments/playback studies without accounting for nonindependence (Johnson & Freeberg, 2016). In cases where there was a mixture of pseudoreplicated and nonpseudoreplicated data presented, we coded the article as reporting pseudoreplicated data. When an article presented pseudoreplicated data we further coded this as 1) avoidable or 2) unavoidable. Avoidable pseudoreplication occurred when all information to appropriately nest data was known (e.g., individual or group ID). Unavoidable pseudoreplication occurred when the research did not have all data necessary to infer whether data was independent or not (e.g., when individual ID and true sample size was unknown, and therefore unknowingly collecting numerous datapoints from a single individual was possible). Finally, we noted the statistical methods used; more specifically, we coded whether the article incorporated the use of generalized linear mixed models or not. JW and PC coded the data and both coders were significantly reliable when jointly and independently coding 10.3% (36/347) of the data (88.9% agreement across all codes, Cohens kappa: 0.707). PC blind-coded 9.7% (30/309) of the pre-2009 data (Waller *et al*., 2013) to confirm congruency in coding between the two datasets, this was near perfect (90% agreement across all codes, Cohens kappa: 0.918), and therefore, comparison between the datasets is appropriate.

Statistical Analysis

For all data where we could assess pseudoreplication, we ran a series of generalised linear mixed models with a binomial error structure. First, we assessed the influence of publication year (continuous variable, Model 1) and dataset (Pre/post 2009, Model 2) on the occurrence of pseudoreplication (yes/no). Second, for the post-2009 data only, we built a model that assessed the interaction between publication year and GLMM usage (article contains GLMM, yes/no, Model 3). We included an interaction as we may expect pseudoreplication to further reduce as a product of GLMM usage. Next, we looked at the impact of research methods (observational/experimental, Model 4), research environment (wild/nonwild, Model 5), species taxa (great-ape/nongreat ape, Model 6), and communicative modality studied (gestural/vocal/facial) on the occurrence of pseudoreplication (yes/no, Model 7). Each model included the interaction term "pre-post 2009," where data were labelled as either relating to the older data (Waller *et al*., 2013) or as newly coded data. Finally, for the post-2009 data only, we built a model to assess the influence of sample size, journal SNIP and article citations per year on the occurrence of pseudoreplication (yes/no, Model 8). We used the SNIP rather than the Journal Impact Factor (JIF) as in Waller *et al*. (2013), as SNIP is now more commonly used and is more easily accessible than the JIF. SNIP and JIF are significantly correlated (Oosthuizen & Fenton, 2014). Finally, we built a model to assess the influence of publication year (continuous variable) on GLMM usage (article contains GLMM, yes/no, Model 9).

For all models, we initially included the first and last author as random effects (typically the two key authors: lead and senior author respectively) to control for repeated sampling of work from the same researchers/research groups and to account for pseudoreplication in our own data. Often, inclusion of both these random effects led to singularity and/or model convergence issues, which we think to be a consequence of the complex random error structure. In these cases, we simplified the error structure by dropping "last author" from the random error structure of model, as recommended by Barr *et al*. (2013). To confirm that this was not a problematic exclusion, we compared null models containing the full error structure to a null model containing the reduced error structure using a likelihood ratio test.

We found no significant difference between these different random error structures (all $p > 0.05$). We considered applying random slopes to our models, but this was unfeasible due to the large number of random effects levels (e.g., data from 340 first authors) in relation to the number of observations. We built all models by using function *glmer* package *lme4* in R with Rstudio (Bates *et al*., 2015; R Core Team, 2021; RStudio Team, 2020). We first compared all models with null models containing only the intercept and random effects by using a likelihood ratio test (LRT). If the likelihood ratio test suggested an improved model fit, we then explored the effect of individual predictors by using the *summary* function and present these in the results; details and outputs of LRTs are presented in the SI (SI2). We built all visualisations with ggplot2 for R studio (Wickham, 2016).

Ethical note

This study incorporates exclusively secondary data and has no ethical implications.

Data Availability All data generated or analysed during this study can be found in a data repository at the follow location: https://osf.io/SCDFX (https://doi.org/10.17605/OSF.IO/SCDFX).

Conflict of Interest The authors declare that they have no conflict of interest.

Results

Frequency of Pseudoreplication

Of the 348 articles coded on the topic of primate communication published between 2009 and 2020, 17 (5%) articles contained no inferential statistics; in 22 (7%) articles, presence/absence of pseudoreplication could not be determined, and 26 articles (8%) contained data at the level below the individual (e.g., neuron) and were therefore not scrutinised further. Of the remaining articles, 76 (23%) had evidence of pseudoreplication, and 207 (62.5%) did not. Compared with articles published

Table I Occurrence of pseudoreplicated data in in primate communication articles: a comparison of articles published pre- and post-2009. Only those articles containing inferential statistics are included

All papers	1960–2008*	2009–2020
	Number of articles (%)	
Total (of articles containing inferential statistics)	420	331
No pseudoreplication	174 (41.4)	207 (62.5)
Pseudoreplication	162 (38.6)	76 (23.0)
Undeterminable	64 (15.2)	22 (6.6)
Level other than individual	20 (4.8)	26 (7.9)

*Waller *et al*., 2013

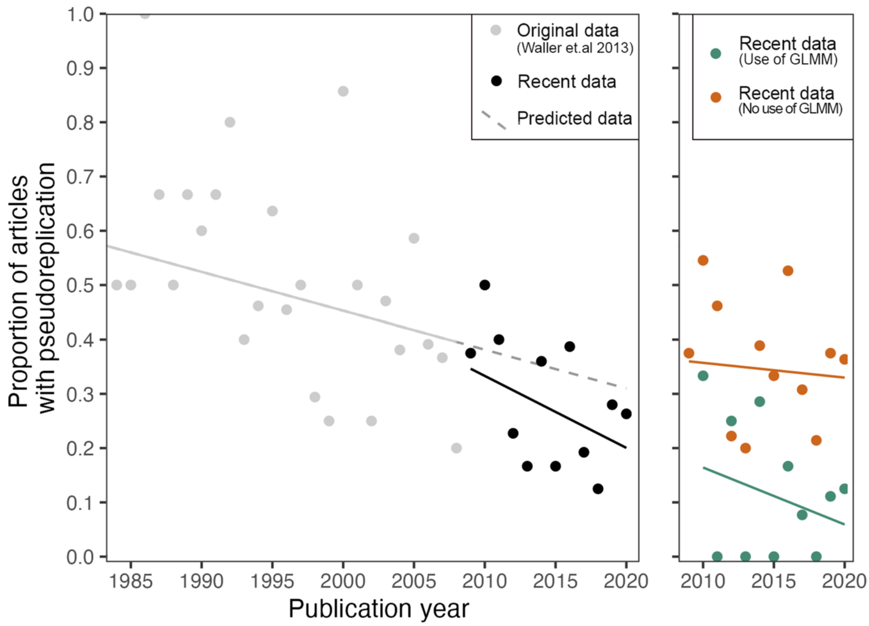

Fig. 1 Occurrence of pseudoreplicated data in primate communication articles across time and according to GLMM usage. For visualization purposes, datapoints before 1980 have been omitted as for many years only a single article was published. Each data point represents the proportion for a single year. (**Left**) All data regardless of statistical approach. (**Right**) A breakdown of recent data, split into those who incorporated GLMMs (green) and those who did not (red)

before 2009, the proportion of pseudoreplicated articled decreased (Table I). Of the 76 articles with evidence of pseudoreplication, 58 (76%) included avoidable pseudoreplication.

Pseudoreplication over Time and Multilevel Model Usage

The likelihood of pseudoreplication was significantly lower in those articles published after 2009, compared with those published before 2009 (Model 1, $ß = 0.948$, $SE = 0.220$, $z = 4.313$, $p < 0.001$). Similarly, when considering all data, there has been a significant linear decrease in pseudoreplication across time (Model 2, year 1960–2020, Fig. 1, $ß = -0.050$, $SE = 0.011$, $z = -4.687$, $p < 0.001$). Whether or not the authors incorporated a multilevel modelling (with the inclusion of random error structures) approach seems to be a key driver of this effect in the post-2009 data, as we found a significant decrease in those articles incorporating GLMM (Model 3, Fig. 1, $ß = -1.467$, $SE = 0.407$, $z = -3.602$, $p < 0.001$) (down to ~ 10% pseudoreplication rate in papers implementing GLMMs compared with a ~ 30% in non-GLMM papers). Similarly, GLMM usage across time increased significantly (Model 9, $ß = 1.094$, $SE = 0.361$, $z = 3.029$, $p = 0.002$), rising from ~ 10% prevalence in 2009, to ~ 50% in 2020. Full model outputs can be found in the supplementary materials.

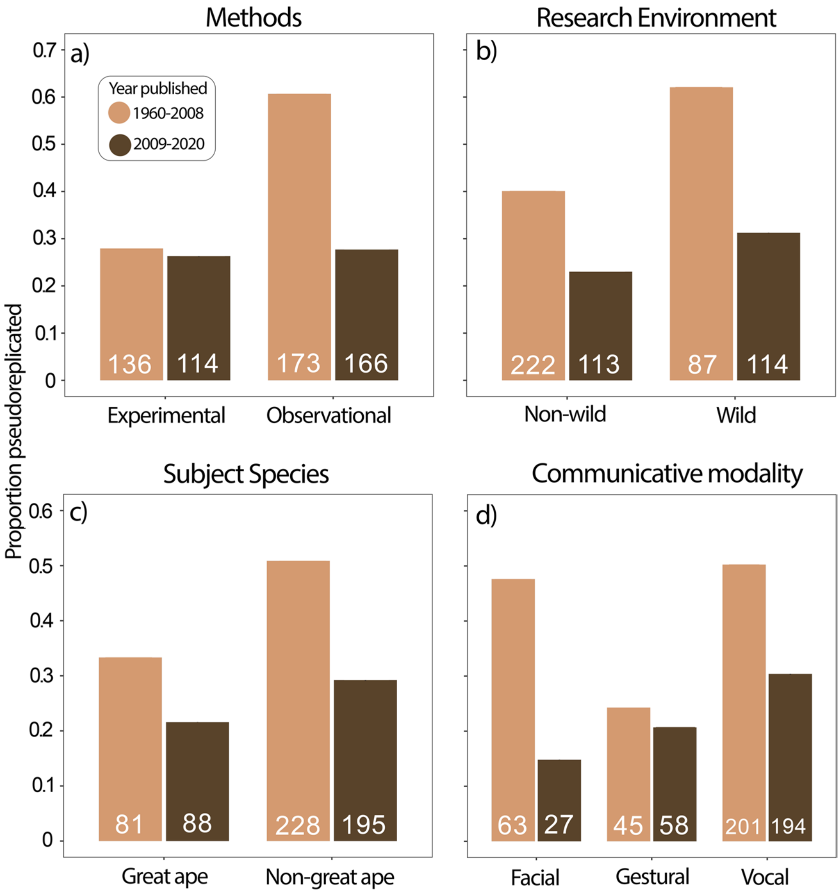

Fig. 2 Impact of (**a**) methods, (**b**) research environment, (**c**) subject species, and (**d**) communicative modality on pseudoreplication in primate communication research

Impact of Methods, Research Environment, Subject Species, and Communicative Modality on Pseudoreplication

In the pre-2009 data, the probability of pseudoreplication occurring was higher in those studies using observational methods, compared with those using experimental methods (Waller *et al.*, 2013). The proportion of pseudoreplication among those studies incorporating observational methods was significantly lower in the post-2009 data, compared with the pre-2009 data (Model 4, Fig. 2a, $ß = 1.572$, $SE = 0.299$, $z = 5.244$, $p < 0.001$). The proportion of pseudoreplication among experimental research, however, was not significantly different across the two datasets ($ß = 0.113$, $SE = 0.400$, $z = 0.331$, $p = 0.740$). The proportion of pseudoreplication was lower in studies using wild animals in the post-2009 compared with the pre-2009 data (Model 5, Fig. 2b, $ß = 1.420$, $SE = 0.352$, $z = 4.029$, $p < 0.001$) and lower for studies involving nonwild animals ($ß = 0.892$, $SE = 0.303$, $z = 2.945$, $p = 0.003$). We found a lower proportion of pseudoreplication within studies using nongreat ape subjects in post-2009 data (Model 6, Fig. 2c, ß = 0.997, SE = 0.249, z = 3.997, $p < 0.001$), as well as a greater decrease in articles studying facial communication (Fig. 2d, Model 7, ß = 1.783, SE = 0.650, z = 2.742, $p = 0.006$) and vocal communication (ß = 0.883, SE = 0.249, z = 3.545, $p < 0.001$). We found no change in the proportion of pseudoreplication for papers studying gestural communication (ß = 0.381, SE = 0.546, z = 0.697, $p = 0.486$). Full model outputs can be found in the supplementary materials.

Impact of Sample Size, Citations per Year, and Source Normalized Impact per Paper on Pseudoreplication

We found no impact of sample size, mean citation count per year, or Source Normalized Impact per Paper (SNIP), on proportion of pseudoreplication (Model 8, LRT: $\chi^2 = 6.64$, $p = 0.084$).

Discussion

In all primate communication research published after 2009, approximately one quarter (23%) contained evidence of pseudoreplication. Although this appears quite high, this was significantly less than the 38.3% reported in Waller *et al.* (2013) for an earlier period of study in the same field. This also is lower than that reported in other fields sampled 10 or more years ago (40% in zoo research; Kuhar, 2006, up to 36% in neuroscience; Lazic, 2010, and 48% in ecology; Hurlbert, 1984). The observed decrease in pseudoreplication appeared to be linear across time, suggesting it is continuing to fall. Our analyses show that there has been a rise in popularity of GLMMs in the past 10 years in primate communication research, with up to 50% of all articles published containing a mixed-modelling approach in 2020 (from ~ 10% in 2009). These articles containing GLMMs were significantly less likely to present pseudoreplicated data compared with those without, and their incorporation seems to be a primary driver for the observed reduction in pseudoreplication overall.

One concerning finding of the pre-2009 data was that studies incorporating an observational (or, nonexperimental) approach were significantly more likely to be pseudoreplicated than other methods. The reasons as to why data collected through observational methods were more prone to these issues have not been fully explored, but we expect it is because of the nature of observing spontaneous communicative behavior (often leading to varying amounts of repeated measures of individuals) compared with collecting data through a more controlled experimental design. This previous disparity between observational and experimental data appears to have disappeared in more recent data. In the post-2009 data we see a ~ 50% decrease in pseudoreplication in observational studies but almost no difference in experimental studies, which has ultimately led to more equal proportions of pseudoreplication between these two approaches. Although this means the prevalence of pseudoreplication in experimental research has remained unchanged, the original bias toward pseudoreplication in observational research has now disappeared. In a somewhat similar pattern, we find that all methods that were originally more prone to pseudoreplication showed greater reductions; research on wild subjects saw a greater decrease in pseudoreplication compared with nonwild subjects, research on non-great apes a greater decrease than great apes, and vocalization and facial expressions research saw a greater decrease than gestural research. These patterns suggest that we are now less statistically constrained by our methodological approach, with more comparable likelihood of pseudoreplication across the field regardless of data collection method, choice of subject species, or environment in which we choose to conduct research.

If our results were due to increasing awareness and intentional action regarding these issues as we and others may hope (i.e., as a response to papers specifically highlighting them; Eisner, 2021; Kroodsma *et al*., 2001; Lazic, 2010; Millar & Anderson, 2004; Waller *et al*., 2013), we would expect a uniform decrease throughout papers incorporating all types of statistical approaches. But we do not see this. This raises the question of the whether researchers are modifying their statistical approach to purposefully avoid pseudoreplication or instead the decrease in pseudoreplication is a fortunate by-product of something else. The flexibility of GLMMs allow other added benefits for communication researchers, for example, their suitability when data do not conform to assumptions required by many parametric statistics (e.g., violation of a normal distribution, another common occurrence in research data with primates or when there are missing data from some individuals). There also are other reasons why GLMMs are likely to be now more commonplace, such as a push toward open science and reproduceable data handling (Foster & Deardorff, 2017; Hampton *et al*., 2015) and an increase in the use of open-source statistical tools, such as R (Lai *et al*., 2019) both in research and in taught university curriculum (Rode & Ringel, 2021). It could be that this popularity of GLMMs is being driven primarily by these other factors, and pseudoreplication also has been reduced in parallel, or perhaps more simply, it could be that those that are aware of these issues are choosing to opt for a mixed-modelling approach. Although our results are overall positive, these data do not provide evidence for an increase in awareness or knowledge, but it could be argued that the cause of the decrease is not as important as the decrease itself. Regardless, we feel that continued emphasis and further communication of the issues surrounding pseudoreplication are still necessary, at least within primate communication research.

We found a decrease in both avoidable pseudoreplication (88.0% down to 76.0%) and a decrease in the number of articles where the statistical approaches were undeterminable (15.7% decreased to 6.7%). In some research approaches, individual identities of subjects cannot be realistically detected with any certainty (e.g., during research incorporatingn passive recording of vocalizations in free-ranging primates; Pérez-Granados & Schuchmann, 2021), and in these cases, pseudoreplication often is unavoidable. Research with these constraints should not be overly criticized as the benefits of obtaining any data from these hard-to-reach populations could outweigh the problems. The fact that these unavoidable cases now make up an increased proportion of the reported pseudoreplication articles is a positive. This means that not only is overall pseudoreplication lower, but a higher proportion of these cases could be considered understandable. However, any generalizations based on these articles should still be approached with great caution. These studies may be case-specific, where results are specific to a single location (such as an impact assessment of a specific environment) and where wider generalizability is not as important. Thus, the impact of pseudoreplication on wider inference is less troublesome (Davis & Gray, 2015; Cottenie & De Meester, 2003). We also find that the reporting of statistical methods was more transparent in post-2009 articles. A higher proportion of articles contained sufficiently detailed descriptions of how the data were managed and, therefore, could be scrutinized in this study. This is additionally promising and further evidence that we are heading toward more open and accessible science (Nosek *et al*., 2015).

One trade-off that we must consider in our attempts to account for nonindependent data through modelling is the complexity of the random error structure we decide upon. To take this paper as an example, there are conceivably many ways in which our data could be nested and arguably considered to be nonindependent. The first author, final author, any other positioned author, the journal each paper was published in, the affiliations of authors, research groups, etc., could potentially all lead to additional variation in the data. However, building such a complex error structure whilst also maintaining stable and meaningful models often can be an unrealistic goal given the typical sample sizes that tend to prevail in the behavioral sciences (Schank & Koehnle, 2009). A pragmatic approach to this should be taken to avoid unhelpful accusations of pseudoreplication. Barr *et al*. (2013) suggests that a goal should be to find a compromise between accuracy (how much variance is explained by the random effects) and complexity (how many random effects to include in a model). By statistically comparing the fits of models with and without certain error terms that have been identified, we can find a resulting error structure which hopefully includes the maximal number of terms and whilst allowing the exclusion of terms that add little to no explanatory value and that add unnecessary complexity (Barr *et al*., 2013). Ultimately, our goal is to reduce the risk of compromising the generalizability of our models through overfitting and avoid common issues, such as failed convergence.

We hope that this lower occurrence of pseudoreplication ultimately means more reliable inference throughout the field and, thus, higher-quality science. Although this finding is derived only from primate communication research, we have no reason to think this field is an anomaly and would hope these findings translate broadly within communication research, primatology, and biology. We emphasize the need to monitor these issues continually and to advocate awareness and solutions to pseudoreplication to ensure robust progression of the field.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s10764-023-00399-y.

Acknowledgements The authors, JW, PRC, and BMW, were supported by European Research Council Consolidator Grant FACEDIFF (864694). Thanks to Federica Amici, Blanca Striegler, and Damilola Olaoba for their help in the coding of the literature.

References

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3). https://doi.org/10.1016/j.jml.2012.11.001

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*, 1–48. https://doi.org/10.18637/jss.v067.i01

Bolker, B. M., Brooks, M. E., Clark, C. J., Geange, S. W., Poulsen, J. R., Stevens, M. H. H., & White, J.-S.S. (2009). Generalized linear mixed models: A practical guide for ecology and evolution. *Trends in Ecology & Evolution, 24*(3), 127–135. https://doi.org/10.1016/j.tree.2008.10.008

Colegrave, N., & Ruxton, G. D. (2018). Using biological insight and pragmatism when thinking about pseudoreplication. *Trends in Ecology & Evolution, 33*(1), 28–35.

Cottenie, K., & De Meester, L. (2003). Comment to Oksanen (2001): Reconciling Oksanen (2001) and Hurlbert (1984). *Oikos, 100*(2), 394–396.

Eisner, D. A. (2021). Pseudoreplication in physiology: More means less. *Journal of General Physiology*, *153*(2), e202012826. https://doi.org/10.1085/jgp.202012826

Foster, E. D., & Deardorff, A. (2017). Open science framework (OSF). *Journal of the Medical Library Association: JMLA, 105*(2), 203.

Freeberg, T. M., & Lucas, J. R. (2009). Pseudoreplication is (still) a problem. *Journal of Comparative Psychology, 123*(4), 450–451. https://doi.org/10.1037/a0017031

Hampton, Stephanie E., et al. (2015). The Tao of open science for ecology. *Ecosphere*, *6.7*, 1–13.

Heffner, R. A., Butler, M. J., & Reilly, C. K. (1996). Pseudoreplication Revisited. *Ecology, 77*(8), 2558–2562. https://doi.org/10.2307/2265754

Hurlbert, S. H. (1984). Pseudoreplication and the Design of Ecological Field Experiments. *Ecological Monographs, 54*(2), 187–211. https://doi.org/10.2307/1942661

Hurlbert, S. H. (2004). On misinterpretations of pseudoreplication and related matters: A reply to Oksanen. *Oikos, 104*(3), 591–597.

Johnson, W. T., & Freeberg, T. M. (2016). Pseudoreplication in use of predator stimuli in experiments on antipredator responses. *Animal Behaviour, 119*, 161–164. https://doi.org/10.1016/j.anbehav.2016.07.006

Kroodsma, D. E., Byers, B. E., Goodale, E., Johnson, S., & Liu, W.-C. (2001). Pseudoreplication in playback experiments, revisited a decade later. *Animal Behaviour, 61*(5), 1029–1033. https://doi.org/10.1006/anbe.2000.1676

Kuhar, C. W. (2006). In the deep end: Pooling data and other statistical challenges of zoo and aquarium research. *Zoo Biology, 25*(4), 339–352. https://doi.org/10.1002/zoo.20089

Lai, Jiangshan, et al. (2019). Evaluating the popularity of R in ecology. *Ecosphere*, *10.1*, e02567.

Lazic, S. E. (2010). The problem of pseudoreplication in neuroscientific studies: Is it affecting your analysis? *BMC Neuroscience, 11*(1), 5. https://doi.org/10.1186/1471-2202-11-5

Lee, Y., & Nelder, J. A. (2001). Hierarchical Generalised Linear Models: A Synthesis of Generalised Linear Models, Random-Effect Models and Structured Dispersions. *Biometrika, 88*(4), 987–1006.

Liebal, K., Slocombe, K. E., & Waller, B. M. (2022). The language void 10 years on: Multimodal primate communication research is still uncommon. *Ethology Ecology & Evolution*, 1–14. https://doi.org/10.1080/03949370.2021.2015453

Machlis, L., Dodd, P. W. D., & Fentress, J. C. (1985). The Pooling Fallacy: Problems Arising when Individuals Contribute More than One Observation to the Data Set. *Zeitschrift Für Tierpsychologie, 68*(3), 201–214. https://doi.org/10.1111/j.1439-0310.1985.tb00124.x

Millar, R. B., & Anderson, M. J. (2004). Remedies for pseudoreplication. *Fisheries Research, 70*(2–3), 397–407. https://doi.org/10.1016/j.fishres.2004.08.016

Nosek, B. A., Alter, G., Banks, G. C., Borsboom, D., Bowman, S. D., Breckler, S. J., Buck, S., Chambers, C. D., Chin, G., Christensen, G., Contestabile, M., Dafoe, A., Eich, E., Freese, J., Glennerster, R., Goroff, D., Green, D. P., Hesse, B., Humphreys, M., … Yarkoni, T. (2015). Promoting an open research culture. *Science*, *348*(6242), 1422–1425. https://doi.org/10.1126/science.aab2374

Oosthuizen, J. C., & Fenton, J. E. (2014). Alternatives to the impact factor. *The Surgeon, 12*(5), 239–243.

Oksanen, L. (2001). Logic of experiments in ecology: Is pseudoreplication a pseudoissue? *Oikos, 94*(1), 27–38. https://doi.org/10.1034/j.1600-0706.2001.11311.x

Oksanen, L. (2004). The devil lies in details: Reply to Stuart Hurlbert. *Oikos, 104*(3), 598–605.

Pérez-Granados, C., & Schuchmann, K. (2021). Passive acoustic monitoring of the diel and annual vocal behavior of the Black and Gold Howler Monkey. *American Journal of Primatology*, *83*(3). https://doi.org/10.1002/ajp.23241

Pollet, T. V., Stulp, G., Henzi, S. P., & Barrett, L. (2015). Taking the aggravation out of data aggregation: A conceptual guide to dealing with statistical issues related to the pooling of individual-level observational data. *American Journal of Primatology, 77*(7), 727–740. https://doi.org/10.1002/ajp.22405

Rode, J. B., & Ringel, M. M. (2021). Undergraduate student perceptions of R and SPSS: An experimental comparison from a one-time lab activity. *Scholarship of Teaching and Learning in Psychology, 7*(2), 93.

RStudio Team. (2020). *RStudio: Integrated Development Environment for R.* (1.3.1056) [Mac OS]. RStudio. http://www.rstudio.com/. Accessed 07/06/2023

R Core Team. (2021). R: A language and environment for statistical computing. R Foundation Statistical Computing, Vienna, Austria. https://www.R-project.org/. Accessed 07/06/2023

Schank, J. C., & Koehnle, T. J. (2009). Pseudoreplication is a pseudoproblem. *Journal of Comparative Psychology, 123*(4), 421–433. https://doi.org/10.1037/a0013579

Waller, B. M., Warmelink, L., Liebal, K., Micheletta, J., & Slocombe, K. E. (2013). Pseudoreplication: A widespread problem in primate communication research. *Animal Behaviour, 86*(2), 483–488. https://doi.org/10.1016/j.anbehav.2013.05.038

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag. https://ggplot2.tidyverse.org. Accessed 07/06/2023

Authors and Affiliations

J. Whitehouse[1] · P. R. Clark[2] · J. Micheletta[3] · K. Liebal[4] · K. E. Slocombe[5] · B. M. Waller[1]

✉ J. Whitehouse
Jamie.whitehouse@ntu.ac.uk

1 Department of Psychology, Nottingham Trent University, Nottingham, UK

2 School of Psychology, University of Lincoln, Lincoln, UK

3 Centre for Comparative and Evolutionary Psychology, Department of Psychology, University of Portsmouth, Portsmouth, UK

4 Institute of Biology, Faculty of Life Science, Leipzig University, Leipzig, Germany

5 Department of Psychology, University of York, York, UK