

Commentary

Olena Shcherbakova, Volker Gast, Simon J. Greenhill, Damián Blasi, Russell D. Gray and Hedvig Skirgård*

Different models, different assumptions, different findings: commentary on “Replication and methodological robustness in quantitative typology” by Becker and Guzmán Naranjo

<https://doi.org/10.1515/lingty-2025-0022>

Received March 7, 2025; accepted May 17, 2025; published online July 28, 2025

We are the authors of Shcherbakova et al. (2022). We welcome Becker and Guzmán Naranjo (2025) (henceforth B&GN)’s initiative to re-visit our findings. However, we find the framing of their paper misleading. While replication has been defined in many different ways (Clemens 2017: 333), the most common interpretation is that it involves reassessing a specific hypothesis with new data and the same methods (e.g. Minocher et al. 2021; National Academies of Sciences, Engineering, and Medicine 2019). Using the same data and different methods is commonly referred to as

***Corresponding author: Hedvig Skirgård** [**ˈhɛdː.vɪɡ ˈʃiː.ɡoːd**], Department of Linguistic and Cultural Evolution, Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany, E-mail: hedvig_skirgard@eva.mpg.de. <https://orcid.org/0000-0002-7748-2381>

Olena Shcherbakova [**ɔ.ˈlɛ.na ʃtʃɛr.bɛ.ˈko.və**], Department of Linguistic and Cultural Evolution, Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany, E-mail: olena_shcherbakova@eva.mpg.de

Volker Gast [**ˈfɔl.ke ɡast**], Department of English and American Studies, Friedrich Schiller University Jena, Jena, Germany, E-mail: volker.gast@uni-jena.de

Simon J. Greenhill [**ˈsaɪ.mən ɡiːn.hɪl**], Department of Linguistic and Cultural Evolution, Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany; and School of Biological Sciences, University of Auckland, Auckland, New Zealand, E-mail: simon.greenhill@auckland.ac.nz

Damián Blasi [**ˈdamːi.an ˈblas.i**], Catalan Institution for Research and Advanced Studies, Barcelona, Spain; and Center for Brain & Cognition, Pompeu Fabra University, Barcelona, Spain, E-mail: dblasi@fas.harvard.edu

Russell D. Gray [**ˈɹʌs.əl ɡreɪ**], Department of Linguistic and Cultural Evolution, Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany; and School of Psychology, University of Auckland, Auckland, New Zealand, E-mail: russell_gray@eva.mpg.de

“robustness analysis” (see Gollwitzer 2020). This distinction is important: differences in methods carry with them differences in the assumptions and hypotheses that are being tested. In this case, B&GN’s addition of a spatial term brings a different theoretical understanding of the processes involved, and switches from co-evolutionary analysis to correlational analysis. This has implications for our understanding of the causation. We argue that the differences in methods between our study and that of B&GN are significant and that these differences are not adequately addressed in their paper.

1 Crucial differences between our study and the study of B&GN

Here, we point out three assumptions that substantially differentiate the authors’ statistical modelling from ours, making the term “replication” invalid. Firstly, the difference between our study and B&GN’s arises from their focus on modelling correlation between the variables that ignores the interdependence between the changes in these variables throughout time (co-evolutionary modelling). Secondly, including a spatial relationship predictor in their model further widens the divide between the two studies’ theoretical assumptions. We note that B&GN’s use of Gaussian Processes using simple continuous longitude and latitude values is also not without drawbacks. It assumes the planet is a two-dimensional rectangle, which leads to incorrect and biased distances between some language pairs, especially towards the poles. It also does not account for the anti-meridian correctly, such that the distance from for example Fiji to Samoa becomes almost the circumference of the earth (>38,000 km), instead of the shortest path across the antimeridian (~1,206 km). This is a known problem in statistical modelling of the earth and it is why researchers are recommended to model the earth as a 3D sphere or ellipsoid, see for example Banerjee et al. (2004:17). The prevalence and severity of these distortion problems increase as the distances get larger. The precise effects of this are unclear and need to be carefully investigated in each particular case to assess the impact of these methodological shortcomings. However, given that most models of language features find that spatial covariance loses its strength to explain the variation in the response at distances over 20–500 km, and that these distortions are much less of a problem at these smaller scales, it is likely that the detrimental effect is small and needs to be assessed on a case-by-case basis. Thirdly, B&GN model 13 variables simultaneously in a single multivariate Beta regression model in contrast to our original pairwise analyses. The authors justify using this model over the pairwise model by claiming a better fit of the former on the synthetic data. However, when it comes to real (typological) data, potential multicollinearity between predictors could seriously affect parameter estimates and thereby invalidate them. This means that combined

analyses are not a better option *a priori* (McElreath 2020: 161–191). While the authors' contribution aimed to compare results between original studies and their analyses, empirical studies devoted to specific research questions should thoroughly account for the causal relationships between modelled variables, considering that some typological features are expected to be interdependent.

2 The reasons for our methodological choices

The primary goal of our study was to test the diachronic implications of the trade-off hypothesis. To do this, we opted for an explicit diachronic analysis to detect the co-evolution between the features (i.e. Bayestraits, Pagel et al. 2004; Levinson et al. 2011). This is in contrast to the approach used by B&GN, which measures the correlation of nominal and verbal complexity given how much they are jointly predicted by phylogenetic and spatial covariance (multivariate Beta regression modelling). Our methodological choices enabled us to explore the processes of simplification and complexification in nominal and verbal domains over time by analysing if one feature led to changes in the states of another feature. Our inferences about the extent to which feature pairs co-evolved not only incorporates the information about relatedness between languages from phylogenetic trees but crucially accounts for how features developed over time.

When comparing results of our original study and the authors' analyses, it is evident that some findings diverge. However, we reject their characterization of the findings as “very different results”. In our study, we do not find substantial support for our nominal and verbal complexity metrics having co-evolved on the global tree using modelling in BayesTrait (Bayes Factor = 1.81; $r = 0.09$, Shcherbakova et al. 2022: 160). B&GN find weak-to-moderate positive support for the nominal and verbal complexity metrics correlating in their multivariate Beta regression (mean correlational estimate of 0.32 in the phylogeny-only model and 0.45 in the model with both phylogeny and spatial covariance). Note, a trade-off relationship would manifest as a negative correlation. Given the differences in the methodological approaches between the two studies, we find their characterization of a null result versus a weak positive correlation as “very different results” to be an overstatement.

In conclusion, we welcome the careful scrutiny of published results, but oppose the watering down of the label “replication”. We would classify all of the analyses in their paper as “robustness analyses”. The study by B&GN applies a significantly different model that entails different assumptions and hypotheses. We are not convinced that it is an appropriate approach to investigate the diachronic relationship between nominal and verbal complexity. It is a synchronic correlational analysis. Nevertheless, despite these key differences in modelling approaches, we note

that there are only minor differences between the two studies, suggesting that our findings are robust.

References

- Becker, Laura & Matías Guzmán Naranjo. 2025. Replication and methodological robustness in quantitative typology. *Linguistic Typology* 29(3). 463–505.
- Banerjee, S., B. P. Carlin & A. E. Gelfand. 2004. *Hierarchical modeling and analysis for spatial data*. Chapman and Hall/CRC.
- Clemens, Michael A. 2017. The meaning of failed replications: A review and proposal. *Journal of Economic Surveys* 31(1). 326–342.
- Gollwitzer, Mario. 2020. DFG Priority program SPP 2317 Proposal: A meta-scientific program to analyze and optimize replicability in the behavioral, social, and cognitive sciences (META-REP). PsychArchives. <https://doi.org/10.23668/PSYCHARCHIVES.3010>.
- Levinson, Stephen C., Simon J. Greenhill, Russell D. Gray & Michael Dunn. 2011. Universal typological dependencies should be detectable in the history of language families. *Linguistic Typology* 15. 509–534.
- McElreath, Richard. 2020. *Statistical rethinking: A Bayesian course with examples in R and Stan*, 2nd edn. CRC Press.
- Minocher, Riana, Silke Atmaca, Claudia Bavero, Richard McElreath & Bret Beheim. 2021. Estimating the reproducibility of social learning research published between 1955 and 2018. *Royal Society Open Science* 8(9). 210450.
- National Academies of Sciences, Engineering, and Medicine. 2019. *Reproducibility and replicability in science*. Washington, DC: The National Academies Press.
- Pagel, Mark, Andrew Meade & Daniel Barker. 2004. Bayesian estimation of ancestral character states on phylogenies. *Systematic Biology* 53(5). 673–684.
- Shcherbakova, Olena, Volker Gast, Damián E. Blasi, Hedvig Skirgård, Russell D. Gray & S. J. Greenhill. 2022. A quantitative global test of the complexity trade-off hypothesis: The case of nominal and verbal grammatical marking. *Linguistics Vanguard* 9(s1). 155–167.