Response to Comment on "Phonemic Diversity Supports a Serial Founder Effect Model of Language Expansion from Africa"

Quentin D. Atkinson[1,2]*

Jaeger *et al.* use statistical simulations to show that the serial founder effect analysis I reported has an inflated type 1 error rate. Crucially, however, their simulations also reveal that the strength of the observed relationship between phonemic diversity and distance from Africa is unlikely to be due to chance, even accounting for multiple comparisons and geographic clustering of phonemic diversity.

I hypothesized that phoneme inventory size is subject to a serial founder effect like that observed in population genetics (*1*). I show that global variation in phonemic diversity is clinal and, like our genetic diversity, fits a serial founder effect model of expansion from Africa. Although there is reason for caution when interpreting any such correlational finding, I show in the paper that this result is robust in the face of alternative explanations, including the impact of geographic variation in modern demography (speaker population size, area, and density), language density, postglacial expansion, and statistical nonindependence due to relatedness between languages.

In their commentary, Jaeger *et al.* (*2*) raise a further concern. They use a range of statistical simulations to evaluate the type 1 error rate of a serial founder effect analysis like the one I report. Although they find an inflated type 1 error rate, Jaeger *et al.*'s simulations also elegantly demonstrate that a relationship between phonemic diversity and distance from Africa as strong as that observed in the real data is unlikely to be due to chance.

Jaeger *et al.* report the results of three simulations using different methods to randomly or

semirandomly reassign languages to locations around the globe. The logic behind this approach is that if language locations are shuffled around the globe, a serial founder effect analysis should only infer a significant geographic cline in ~5% of samples (the standard type 1 error rate), because any global geographic patterning in the data should have been removed by the location reassignment.

Simulation 1 randomly assigns language locations within language families. This is an unusual approach to testing for type 1 error, because the test being conducted is on a global sample of languages and the simulated data preserves the global pattern of diversity between families. Detecting a global effect of distance from the origin in this case is not a type I error, and all such simulations still show the significant effect of distance from Africa preserved across language families.

Simulations 2 and 3 are more useful for assessing type 1 error rates. In simulation 3, languages are simply randomly shuffled around the globe, whereas simulation 2 shuffles language locations with the constraint that sets of related languages (the recognized language families) should cluster together in a manner similar to the clustering we observe in the real data (*2*). Using a random sample of 1000 of these simulated data sets (*3*), a serial founder effect analysis finds a significant global decline in phonemic diversity with distance from origin in between 15% (simulation 3) and 20% (simulation 2) of samples, suggesting an inflated type 1 error rate

when searching for a clinal pattern from anywhere on the globe. Many of these cases would not, however, have constituted clear support for a serial founder effect, because for the simulated data the strongest relationship with distance from all putative origins often shows a positive slope, contrary to the predictions of the model. Removing these cases leaves 18% (simulation 2) and 12% (simulation 3) of the samples. Between 1.5% (simulation 3) and 2.5% (simulation 2) meet these criteria and show the strongest relationship with distance from a putative origin in Africa—that is, the simulations imply that between 1.5% and 2.5% of the time, we would find support for a significant cline from Africa by chance.

What we really want to know, however, is the probability of finding an effect of distance from any origin by chance that is at least as large as the effect we observe in the real data. The Dunn-Šidák correction for significance offered by Jaeger *et al.* is inappropriate because the tests being conducted are not independent. However, Jaeger *et al.*'s simulations allow us to answer the question. Among the sample data sets run under both simulation 2 and simulation 3 (and a further simulation with increased geographic clustering reported in their supporting online material), between 0.2% and 0.5% show a relationship with distance from any of the 2667 possible origins that is as strong as the relationship we observe between distance from Africa and the actual global distribution of phonemic diversity. Jaeger *et al.*'s simulations therefore demonstrate that despite inflated type 1 error rates, the pattern we observe in the data is unlikely to have occurred by chance ($P < 0.005$), even accounting for the multiple comparisons being made and the geographic clustering of families with similar phonemic inventories.

Jaeger *et al.*'s work shows the utility of simulation-based approaches to significance testing and suggests that caution is needed when interpreting the results of serial founder analyses. The type 1 error rates inferred here are specific to this data set and do not necessarily apply to other applications of a serial founder analysis, but it would be interesting to test their generality.

References and Notes

1. Q. D. Atkinson, *Science* **332**, 346 (2011).
2. T. F. Jaeger, D. Pontillo, P. Graff, *Science* **335**, 1042 (2012); www.sciencemag.org/cgi/content/full/335/6072/1042-a.
3. Data provided by Jaeger *et al.*

[1]Department of Psychology, University of Auckland, Private Bag 92019, Auckland 1142, New Zealand. [2]Institute of Cognitive and Evolutionary Anthropology, University of Oxford, 64 Banbury Road, Oxford OX2 6PN, UK.

*E-mail: q.atkinson@auckland.ac.nz

Science

Response to Comment on "Phonemic Diversity Supports a Serial Founder Effect Model of Language Expansion from Africa"

Quentin D. Atkinson

ARTICLE TOOLS	http://science.sciencemag.org/content/335/6072/1042.2
RELATED CONTENT	http://science.sciencemag.org/content/sci/332/6027/346.full http://science.sciencemag.org/content/sci/335/6072/1042.1.full
REFERENCES	This article cites 2 articles, 2 of which you can access for free http://science.sciencemag.org/content/335/6072/1042.2#BIBL
PERMISSIONS	http://www.sciencemag.org/help/reprints-and-permissions

Use of this article is subject to the Terms of Service