Reply to Biersteker: When methods matter

We appreciate Biersteker's comments (1) on our research (2). Moreover, we agree with many of her points so wholeheartedly that our paper addresses them in detail: We devote whole sections in the main text and supporting information to the incompleteness of the Index Translationum, the imperfect quality of the language detector, and the limitations of the Wikipedia dataset, among others.

However, besides restating previously acknowledged limitations, Biersteker does not describe how these limitations change our findings. We intuit that her comments stem from a misunderstanding of our methods and would like to provide a few clarifications.

First, we never claimed that our networks are globally representative. We explicitly state on page 3 that "the resulting networks represent patterns of linguistic coexpression not among the entire human population but among the kinds of speakers and texts that contributed to the respective datasets." The very reason we trace three different networks is to characterize the linguistic patterns of the distinct elites captured by each network.

Second, our method controls for the presence of a language in each medium by considering only links that are statistically significant; that is, links that are not explained merely by the number of articles, tweets, or books recorded in the data. Thus, languages with a larger presence in a medium are not necessarily more central. Biersteker asks whether "Winaray [is] more important than Spanish because it has 100,000 more Wikipedia articles?," when in our article, we make the opposite claim: that Spanish is more central than Winaray. Although Winaray has more Wikipedia articles than Spanish (1.25 vs. 1.15 million),* Spanish is more central because its editors are statistically more likely than Winaray editors to edit Wikipedias in other languages at a rate that goes beyond what we would expect from the number of articles in Spanish.

Finally, Biersteker questions our methodology of assigning globally famous individuals to each language by suggesting that "The language Wikipedias with the most articles produced, in this case, the most 'famous people.'" This is not true. For instance, the Dutch Wikipedia ranks fourth by number of articles, but Dutch ranks only ninth by number of famous people produced. The Vietnamese Wikipedia is the 11th largest Wikipedia, but Vietnamese ranks only 50th by number of famous people produced. Biersteker deems our methodology wrong because Switzerland produced more globally famous people than China, but this is not surprising considering that our definition of global fame involves having Wikipedia articles in at least 25 languages. Large populations, such as the Chinese, can produce individuals known to many people but whose fame is contained within one region or language.

In sum, we appreciate Biersteker's comments but feel that our methods and conclusions stand strong.

Shahar Ronen[a], Bruno Gonçalves[b,c,d], Kevin Z. Hu[a], Alessandro Vespignani[b], Steven Pinker[e], and César A. Hidalgo[a,1]

[a]Macro Connections, The MIT Media Lab, Massachusetts Institute of Technology, Cambridge, MA 02139; [b]Department of Physics, Northeastern University, Boston, MA 02115; [c]Aix-Marseille Université, CNRS, Centre de Physique Théorique, UMR 7332, 13288 Marseille, France; [d]Université de Toulon, CNRS, Centre de Physique Théorique, UMR 7332, 83957 La Garde, France; and [e]Department of Psychology, Harvard University, Cambridge, MA 02138

1 Biersteker AJ (2015) Links that speak only some languages. *Proc Natl Acad Sci USA* 112:E1814.

2 Ronen S, et al. (2014) Links that speak: The global language network and its association with global fame. *Proc Natl Acad Sci USA* 111(52):E5616–E5622.

Author contributions: S.R., B.G., K.Z.H., A.V., S.P., and C.A.H. wrote the paper.

The authors declare no conflict of interest.

[1]To whom correspondence should be addressed. Email: hidalgo@mit.edu.

*As a side note, the Winaray Wikipedia has very few human editors and most of its articles were created by bots, which we do not consider for the purpose of identifying connections between languages.

Downloaded from https://www.pnas.org by 37.19.197.239 on October 16, 2023 from IP address 37.19.197.239.