Controversy Requires Competence: Comment on Rindermann et al. (2024)

Eric Turkheimer [1],* **and K. Paige Harden** [2],*

[1] University of Virginia
[2] University of Texas at Austin
* Corresponding author: ent3c@virginia.edu (E.T.); harden@utexas.edu (K.P.H.)

Abstract: Rindermann, Klauk, and Thompson (2024) purport to give evidence regarding the determinants of intelligence test scores among refugees who have immigrated to Germany from different countries of origin, and they speculate these intelligence test score differences have negative implications for future economic development in Germany. We describe critical flaws in their measurement, statistical analysis, and interpretation of individual- and country-level differences among the immigrant participants, particularly regarding the authors' specious reference to "evolutionary ancestry." We contrast their pseudoscientific approach with valid scientific methods. Human intelligence and human evolution are controversial areas of scientific inquiry that require the highest levels of scientific rigor and editorial discretion, which are absent here.

Rindermann, Klauk, and Thompson's paper "Intelligence of Refugees in Germany: Levels, Differences and Possible Determinants" in this issue of the *Journal of Controversial Ideas*, professes to be a study of why immigrants to Germany from various countries, about half of them from Syria, perform differently, on average, on a standardized test of matrix reasoning. One predictor of immigrants' test performance is a rating of their "evolution," which was derived from two purported characteristics of their home country – "skin lightness" and "brain size." What is called "skin lightness" was not quantified by measuring the light refracted off people's skin, as has been done, for instance, in carefully conducted genomic studies documenting that Southern African populations with

the oldest genetic lineages have the lightest skin pigmentation within Africa.[1] Instead, a "student", not otherwise described, consulted a map in a mid-century book, *Le razze e i popoli della terra* [Races and peoples of the earth, 1953/1967], by the Italian geographer, Renato Biasutti, who sketched global variation in skin tone based on existing ethnographic reports and his imagination.[2] The reader is left to infer that darker skinned people are less "evolved." Country-level estimates of "brain size" were obtained in similar fashion. Rather than, for example, measuring participants' global brain volume using fMRI (functional magnetic resonance imaging), as is standard in neuroscience, a student simply consulted a map from Beals, Smith, and Dodd (1984),[3] which we reproduce here (Figure 1) – an outdated method, to say the least.
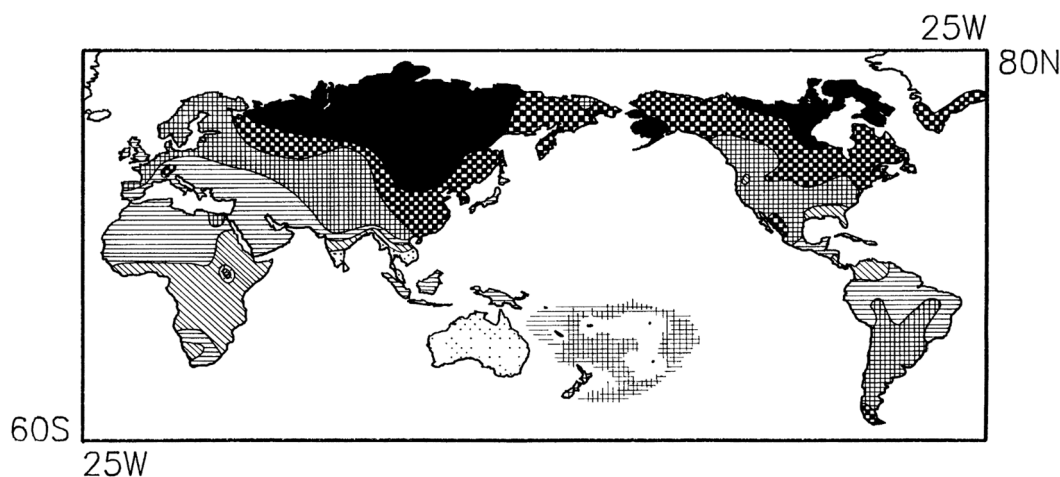

Figure 1: Figure on "clinal depiction of cranial capacity" reproduced from Beals et al. (1984).

The "skin lightness" and "brain size" variables were then combined using "factor analysis." We put "factor analysis" in scare quotes because the researchers did not perform the statistical analysis of variance-covariance matrices that commonly bears that name. One of the most basic rules of factor analysis is that any factor solution requires at least three indicators. Typing "FACTOR" in one's SPSS code is not the same as conducting a meaningful statistical analysis. Their "factor" – essentially the median of two country-level ratings derived from the shadings of fifty-year-old maps – is what they variously call "evolutionary ancestry," "evolution: G factor", or simply "evolution."

We are surprised, to put it mildly, that peer reviewers approved this meaningless analysis and the resulting factor label for publication in a professional scientific journal. There is, to be sure, cutting-edge science being conducted on how human genomes and human cultures have evolved over millennia, some of it even focused on Levant populations.[4] There is also a long-standing body of scientific work regarding "*g*", or general intelligence, which is estimated, as we have done in our own work on the

1 Martin, A. R. et al. An Unexpectedly Complex Architecture for Skin Pigmentation in Africans. *Cell* 171, 1340–1353.e14 (2017).

2 Jablonski, N. G. The Evolution of Human Skin and Skin Color. *Annual Review of Anthropology* 33, 585–623 (2004).

3 Beals, K. L., Smith, C. L. & Dodd, S. M. Brain Size, Cranial Morphology, Climate, and Time Machines. *Current Anthropology* 25(3), 301–330 (1984).

4 Skourtanioti, E. et al. Genomic History of Neolithic to Bronze Age Anatolia, Northern Levant, and Southern Caucasus. *Cell* 181, 1158–1175.e28 (2020).

sensitivity of child cognitive development to environmental contexts,[5] by using data on performance on multiple tests of verbal and visuospatial reasoning. But Rindermann et al., with their dusty maps and specious SPSS code, are not conducting scientific analyses of human evolution or of *g*.

The statistical blundering continues. The authors seem not to understand the role of sampling error in the interpretation of statistical estimates. Using a few general references, they give themselves permission to ignore traditional standards of statistical significance. Not adhering to rigid standards of $p < .05$ is one thing; ignoring precision (or lack thereof) in estimates, or even the units in which the estimates are expressed, is another. The original manuscript we were provided for review did not contain the phrase "standard error"; a month later we were sent the current version that reports them intermittently but makes no effort to use them to interpret results. Without standard errors there is no way to interpret any of the numbers reported, never mind the casual comparisons they indulge in throughout the text.

The path diagrams in Figures 2 and 3 of Rindermann et al.'s paper do not resolve the problem. Neither do they, as the authors claim, describe two separate analyses. They are the same analysis conducted twice, the second time slightly less incorrectly than the first. The only difference between them is that the second analysis correctly specifies the country-level predictors as such, as opposed to the first one in which all variables are treated as though they were measured on the individual level. In any event, the numbers in parentheses following the estimates are not, as would normally be the case, standard errors of the path coefficients; they are the zero-order correlations between the predictors and IQ. Examination of computer output provided to us by the authors gives a clue regarding why the authors did not report the standard errors in the main text: The model is so misspecified that they couldn't compute them correctly. The output on which Figure 3 is based includes a half-page of capitalized error reports, including "THE STANDARD ERRORS OF THE MODEL PARAMETER ESTIMATES MAY NOT BE TRUSTWORTHY."

For a moment, let's set aside, if we can, the substantive point that these pseudo-analyses are meant to support. We ask the reader: In whatever uncontroversial area of science you happen to work, if you were to submit a paper based on a factor analysis of two variables obtained by consulting the map shown in Figure 1, using statistical output that was preceded by extensive error reports, producing results tabled without standard errors, would you expect the paper to be accepted at any legitimate scientific journal? Rindermann et al.'s conclusions about human evolution and its consequences for modern social problems are not empirically supported aruguments. They are speculations accompanied by error-ridden statistical analyses of dubious data.

But, for all the incompetence in the conduct of the analyses, the reported results – when stripped of unsubstantiated references to "evolution" – are not so much wrong as obvious. What do the results, such as they are, show? Some refugee immigrants, fleeing from countries beset by poverty and political strife, now torn apart by unrelenting war, obtain upon arrival in Germany lower scores on IQ tests than the resident German population. The best predictors of the immigrants' IQ scores are the economic and educational levels in their countries of origin, and, when the authors deign to treat their

---

5   Engelhardt, L. E., Church, J. A., Harden, K. P. & Tucker-Drob, E. M. Accounting for the Shared Environment in Cognitive Abilities and Academic Achievement with Measured Socioecological Contexts. *Developmental Science* 0, e12699.

research participants as individuals, differences in education levels among the immigrants. Is this controversial? Although we claim no expertise in the economics or politics of immigration, we find it unsurprising that the worst-off immigrants, those from the poorest countries with the least education and wealth, also have the lowest IQ scores and present the greatest social and economic difficulties to their host countries. Given the current sociopolitical state of the world, it is even entirely plausible that the most desperate immigrants would have darker skin than resident Germans, although referring to this state of affairs with the word "evolution" is as gross an affront to the legitimate science of human evolution as it is to the refugees it is intended to demean.

What to make of immigrants' IQ test scores? The authors speak of cognitive ability as though it were an essential feature baked into immigrants, an ultimate cause of the chaos in their home countries and fraught migration to new homes in Europe. The skills tested by a matrix reasoning task, we are told, are the ultimate cause of wealth and democracy, rather than their result. There is, however, no evidence this is the case. These people are impoverished economically, educationally, and medically, in the process of uprooting themselves from a shattered home and finding their way to a radically different culture. As has been well-known for more than a century, IQ scores are correlated with social outcomes, but that was just as true of the Italian, Irish, and yes, German immigrants to the US in 1924 as it is for Syrian and Somali immigrants to Europe in 2024. IQ scores have always provided a convenient essentialist proxy for racist anti-immigrant sentiment. The authors do not have a scintilla of data supporting their contention that the test scores of immigrants have anything whatsoever to do with their "evolution." They have only centuries-old prejudices about strangers with dark skin.

If the *Journal of Controversial Ideas* wished to publish ideas related to the biology and psychology of racial and ethnic differences, there is no shortage of legitimate controversies. The National Academies of Science, Engineering, and Medicine, for example, recently released a report titled, "Using Population Descriptors in Genetics and Genomics Research"[6] that makes some controversial recommendations about how population categories should (and mostly should not) be used in human research. To give another example, we participated in a collaborative effort of geneticists, philosophers, and social scientists that examined the problems and possibilities of including genetics in scientific investigations of social outcomes, and some of our conclusions provoked sharp disputes within and beyond the academy.[7] What united these efforts was the belief that controversial ideas, precisely because they provoke the most passionate disagreements, must be addressed using the most rigorous and dispassionate scholarship, using the highest standards of evidence and argumentation. To do less than that makes a mockery of the ideals of reasoned inquiry.

In the end, Rindermann et al.'s paper did make us contemplate a controversial idea: Is the academic journal dead? After all, publishing today has nothing to do with printing presses and nothing to do with free speech. Anyone can disseminate anything in five minutes. By calling a collection of papers posted online an "academic journal," the editorial board is implicitly making a claim about the benefits of expertise for discerning whether an idea is worthy of discussion, and moreover, whether the idea has been developed in

---

[6] *Using Population Descriptors in Genetics and Genomics Research: A New Framework for an Evolving Field*. (National Academies Press, Washington, DC, 2023). link to this article.

[7] Meyer, M. N. et al. Wrestling with Social and Behavioral Genomics: Risks, Potential Benefits, and Ethical Responsibility. *Hastings Center Report* 53, S2–S49 (2023).

a way that merits the attention of the scholarly community. If the paper is representative of the discernment of some of the world's most eminent philosophers and scientists, we fear for the future of academic publishing.