

The Moderating Role of Culture in the Generalizability of Psychological Phenomena

Robin Schimmelpfennig^{1*}, Rachel Spicer^{2*}, Cindel J. M. White³, Will Gervais⁴, Ara Norenzayan⁵, Steven Heine⁵, Joseph Henrich⁶, and Michael Muthukrishna^{2,7}

¹ Faculty of Business and Economics, University of Lausanne

² Department of Psychological and Behavioural Science, London School of Economics and Political Science

³ Department of Psychology, York University, Canada

⁴ Centre for Culture and Evolution, Psychology, Brunel University London

⁵ Department of Psychology, University of British Columbia

⁶ Department of Human Evolutionary Biology, Harvard University

⁷ Canadian Institute for Advanced Research (CIFAR), Canada

* Shared first authors

Corresponding authors:

Michael Muthukrishna: m.muthukrishna@lse.ac.uk

Robin Schimmelpfennig: robin.schimmelpfennig@unil.ch

Rachel Spicer: r.a.spicer@lse.ac.uk

The Moderating Role of Culture in the Generalizability of Psychological Phenomena

Far more than any other species, humans rely on learning from others, with both genetic and cultural inheritance systems interacting to shape our psychology (Boyd & Richerson, 1985; Henrich, 2016). Large bodies of theory and empirical evidence suggest variability in many aspects of human thinking, perception, and behavior is driven by different ecological niches and social norms (Heine & Norenzayan, 2006; Laland, 2018; Smaldino et al., 2019). Mainstream psychology has historically sampled mostly from a thin slice of humanity (Arnett, 2008; Thalmayer et al., 2020), constraining our knowledge of human diversity and contributing to psychology's limited generalizability and possibly replicability (Adetula et al., 2022; Nosek et al., 2022; Open Science Collaboration, 2015). However, the role of population diversity in replicability and generalizability remains understudied, and robust theories that predict when to expect variation in psychological phenomena remain underused and underdeveloped.

The multi-site project Many Labs 2 (ML2) (Klein et al., 2018) had the important goal of assessing how population, site, and setting variability moderate the replicability of psychological findings. Klein et al. attempted to replicate 28 studies, with 15,305 participants distributed across 125 sample sites in 36 countries. They found that 15 out of 28 studies showed a statistically significant effect in the same direction as the original study. Additionally, there was little detected effect size heterogeneity regarding replicability across different contexts and settings in 11 of the 28. In this commentary, we focus on ML2's exploratory investigation of the moderating role of culture (operationalized as "WEIRDness") and reflect on three important choices related to the selection of samples, measures, and effects. In doing so, we emphasize the central role of theory and connect our arguments with previous recommendations for cross-cultural research (Berry, 2011; Kitayama & Cohen, 2010). Next, we offer suggestions for future research to assess the role of population diversity in generalizability efforts in psychology. We conclude the Commentary by discussing how researchers can thread the needle between searching for cultural variation and confirming replicability and generalizability of psychological phenomena.

Sample Selection: Sampling WEIRD People from Around the World.

Despite the large sample size and number of sites, ML2's samples were culturally relatively homogenous. Multi-site projects have much-improved samples compared to most psychological studies but are still naturally limited by coordination, reachability, and availability constraints. Furthermore, as researchers and the associated participant samples (e.g., university subject pools) self-selected into the project, some level of selection bias is likely. Sampling representatively across the globe is challenging, and random sampling is even more difficult. For instance, the majority of ML2's subject pool were US-based participants (39%) or participants from other WEIRD (Western, Educated, Industrialized, Rich, and Democratic) countries. Indeed, a large fraction of the participants were obtained from a small set of countries (see Figure 1).

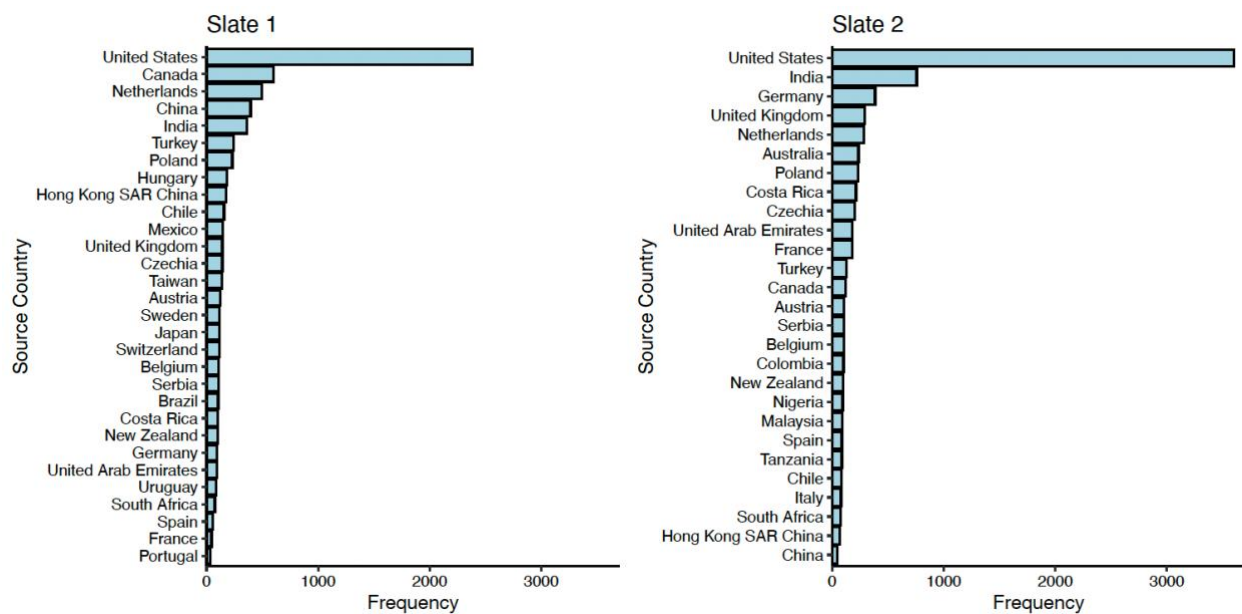


Figure 1. Frequency of participants per source country in which the participants were sampled on Many Labs 2 (Figshare link: <https://doi.org/10.6084/m9.figshare.24038496.v1>).

Furthermore, the participants in ML2 were mostly recruited from university subject pools and Amazon's Mechanical Turk. These samples are likely skewed towards participants with higher education, digital literacy, English proficiency, and socioeconomic status (SES)—especially in less WEIRD countries. Having higher levels of formal education correlates with a greater cultural alignment with the United States (White & Muthukrishna, 2023). As a result, despite cross-country

variation in sampling from 36 countries, the cross-*cultural* variation in many of the samples may have been minimal. The lack of cultural variation is possibly explained by the fact that ML2 had not planned to investigate the moderating role of culture but did so in an ex-post explorative analysis. But future studies that plan to explore how cultural variation affects replicability should opt for a more limited, yet more representative and culturally diverse set of samples. New methodological approaches can help to select samples that show significant cultural variation (e.g., Muthukrishna et al., 2020; Obradovich et al., 2022). Adopting such strategies can offer a more robust test of cultural variability than enlarging the sample size—particularly if the larger sample remains self-selected and culturally uniform. This encompasses comparing populations both between and within countries. Older approaches have already demonstrated that obtaining truly random samples (with participation rates over 95%) from some of the most remote communities on the planet is feasible (Ensminger & Henrich, 2014).

Moving forward, we urge researchers to consider the statistical power of tests that are sensitive to culture's moderating role, rather than solely emphasizing the main effect's power. Simulations can be instrumental in crafting studies with adequate sample sizes that can identify effects, considering anticipated population-level variance. For example, researchers can simulate a multi-site setting to explore how population-level variance influences behavior and can assess whether studies have low power to detect the moderating role despite having large samples. For further exploration of how simulations can inform design choices when investigating the moderating influence of population-level variation on generalizability in psychological studies, see Schimmelpfennig et al. (2023).

Measure Selection: Operationalizing and Measuring Cultural Variation

The selection of adequate measures of culture needs to be underpinned by a theoretical understanding of how to measure cultural variation, and at which level cultural differences are likely to matter. That is, measuring cultural variation necessitates not only an effective tool to gauge

such variation but also careful attention to the unit of analysis, such as differences within or between countries. In ML2, cultural moderation was assessed by decomposing the letters of the WEIRD backronym (Henrich et al., 2010), an approach that has since been adopted by other published articles (e.g., Van Assche et al., 2023). The backronym, however, was originally only intended as a rhetorical device. While the approach to deconstructing the WEIRD backronym may seem pragmatically sound, it lacks grounding in a theoretical understanding of cultural variation. A more effective approach would entail conceptualizing cultural distinctions through empirical and theory-driven methodologies (Muthukrishna & Henrich, 2019). Inductive techniques that use more fine-grained measures for defining cultural differences—such as tightness-looseness (Gelfand et al., 2011), individualistic and collectivistic cultures (Hofstede, 2001), or cultural distance (e.g., CF_{ST} ; (Muthukrishna et al., 2020) are preferable.

Additionally, the sound measurement of cultural variation does not only depend on choosing a measure that accurately measures cultural variation; it also depends on anticipating the levels at which culture varies and should thus be measured. That is, theoretical predictions should precede both the selection of samples representative of the relevant populations and the selection of adequate instruments for measuring cultural variation at that level. Multisite studies often draw on instruments that measure cultural background at the country level based on the country of origin of that site (Gelfand et al., 2011; Hofstede, 2001; Klein et al., 2018; Muthukrishna et al., 2020). While an understandable design choice, clustering by the country of the sample site can conceal potential psychological variation within the country (Greenfield, 2014), like the migration background of participants or their parents. For example, culture is embedded in overlapping distributions of cultural traits within societies, as evidenced in studies focusing on factors such as people's long-term orientation (Harati & Talhelm, 2023), political preferences (Talhelm et al., 2015), religious beliefs (White et al., 2021), SES (Kraus et al., 2017), language (Faessler et al., 2023), and ethnicities (Desmet et al., 2017). Significant cultural divides also sometimes arise when contrasting state- with non-state societies (Henrich et al., 2005), like comparing Tanzanians broadly

with the specific Hadza community of foragers within Tanzania. What dimensions of cultural variation should be measured in a study depends on the research question being asked. When selecting instruments for measuring cultural variation, we thus need to make deliberate choices about what level of cultural variation we are interested in. Country-level measures are generally a poor choice, so psychologists should seek out more finely-grained levels of analysis.

What is more, not all population-level variation is due to cultural transmission. Whereas cultural variation may play a role in explaining psychological and behavioral differences, such differences might also be explained by environmental cues, ecological patterns, or even genetic differences (Uchiyama et al., 2022). Especially in the absence of exogenous variation that could isolate the moderating role of culture (e.g., see Faessler et al., 2023), we should be cautious to ascribe all kinds of population-level variation to be associated with cultural differences.

Effect Selection: Theorizing about Cultural Moderation and Generalizability

Our minds are shaped by millions of years of genetic evolution, thousands of years of cultural evolution, and a short lifetime of individual experiences (Muthukrishna et al., 2021; Schimmelpfennig & Muthukrishna, 2023). Cultural evolutionary theory describes how we evolved as a cultural species, how culture itself evolves, and how this process influences our psychology and behavior (Boyd & Richerson, 1985; Cavalli-Sforza & Feldman, 1981). Cultural evolutionary theory helps us to understand the interplay of culture and psychology and offers a way to make principled predictions about which aspects of human psychology and behavior will vary across populations and across which populations.

Without a sound theory to explain possible aspects that vary, we are running blind (Muthukrishna & Henrich, 2019). Theory matters and helps navigate the much-needed methodological changes in psychological research (Gervais, 2021). Cultural evolutionary theory can serve as an inclusive framework to predict the importance of cultural differences, both cross-societal and cross-temporal (Atari & Henrich, 2023; Muthukrishna et al., 2021). The human psyche

is cumulatively transmitted via social learning, and differences in these social learning dynamics can help explain and predict differences in psychology. The selection of effects for testing the generalizability should therefore be based on theoretical, pre-registered predictions about whether we would expect the effects to vary across populations (Stroebe, 2019). An effective test would align effects with populations known for significant variations in dimensions relevant to the theory (Norenzayan & Heine, 2005). If initial tests validate cultural differences, subsequent examinations can then expand to wider populations. This incremental approach is not only more refined but also minimizes risks before running expansive multi-site projects. It provides a focused yet powerful empirical examination before initiating a broader global generalizability study, which may be challenging to replicate at the same magnitude.

Implications for Future Research

Global research collaborations are crucial for progressing in psychological science and can help address both the replication and generalizability crisis and the narrow sampling from WEIRD populations. Multi-lab projects like ML2 have the potential to facilitate our understanding of cultural variation on cross-societal levels, if theory guides methodological and effect selection choices. Naturally, cross-cultural studies should seek to not just include samples, but also researchers from different cultures. Ultimately, cross-cultural research would also facilitate the empowerment of emancipated research originating from different cultures (Hansen & Heu, 2020).

Theory is also the key to understanding how culture influences the relationship between generalizability and replicability in psychology. Replicating findings in the same population as the original study is critical to a rigorous scientific protocol. First and foremost, they are a tool to ensure that a given scientific effect is not an idiosyncratic finding of a specific research approach. Notably, many psychological phenomena may represent universal human states while others may not and to test these replications across samples and settings (i.e., generalizability studies) are needed. A better theoretical understanding of how a cultural context matters thus defines the

differences between searching for generalizability and seeking to identify cultural differences. Practically speaking, if theory predicts the universality of a given effect, one could show this by testing it in a few very different cultural contexts. If theory predicts a moderating role by culture, we should selectively target confirmation/and non-confirmation in different contexts (House et al., 2013, 2020). This would help us to understand when and why to expect deviations from theoretical predictions (such as when they are driven by cultural factors), and when these deviations may not be so surprising after all.

Conclusion

Many Labs 2 was a milestone in testing variation in the generalizability of psychological findings across 36 countries. We discuss the degree to which their data can inform us about how cultural variation moderates the generalizability of psychological phenomena. By emphasizing targeted sample selection, appropriate cultural measures, and theoretically informed effect choices, we hope to encourage greater attention to population diversity in replicability efforts, and a nuanced approach to understanding the complex role of culture in shaping psychological phenomena.

Author contributions: *Conceptualization:* Robin Schimmelpfennig, Rachel Spicer, Cindel White, Will Gervais, Ara Norenzayan, Steven Heine, Joseph Henrich, Michael Muthukrishna. *Writing – Original Draft Preparation:* Robin Schimmelpfennig, Rachel Spicer, Michael Muthukrishna; *Writing – Review & Editing:* Robin Schimmelpfennig, Rachel Spicer, Cindel White, Will Gervais, Ara Norenzayan, Steven Heine, Joseph Henrich, Michael Muthukrishna.

Conflict of Interest: “The author(s) declare that there were no conflicts of interest with respect to the authorship or the publication of this article.”

Acknowledgments: None.

Funding: None.

Pre-registration: None.

Supplemental Material: None.

Bibliography

- Adetula, A., Forscher, P. S., Basnight-Brown, D., Azouaghe, S., & IJzerman, H. (2022). Psychology should generalize from—Not just to—Africa. *Nature Reviews Psychology*, 1(7), 370–371. <https://doi.org/10.1038/s44159-022-00070-y>
- Arnett, J. J. (2008). The neglected 95%: Why American psychology needs to become less American. *American Psychologist*, 63(7), 602–614. <https://doi.org/10/c54kcm>
- Atari, M., & Henrich, J. (2023). Historical Psychology. *Current Directions in Psychological Science*, 32(2), 176–183. <https://doi.org/10.1177/09637214221149737>
- Berry, J. W. (Ed.). (2011). *Cross-cultural psychology: Research and applications* (3rd ed). Cambridge University Press.
- Boyd, R., & Richerson, P. J. (1985). *Culture and the Evolutionary Process*. University of Chicago Press.
- Cavalli-Sforza, L. L., & Feldman, M. W. (1981). *Cultural Transmission and Evolution: A Quantitative Approach*. Princeton University Press.
- Desmet, K., Ortuño-Ortín, I., & Wacziarg, R. (2017). Culture, Ethnicity, and Diversity. *American Economic Review*, 107(9), 2479–2513. <https://doi.org/10.1257/aer.20150243>
- Ensminger, J., & Henrich, J. (2014). *Experimenting with Social Norms: Fairness and Punishment in Cross-Cultural Perspective*. Russell Sage Foundation.
- Faessler, L., Lalive, R., & Efferson, C. (2023). How Culture Shapes Choices Related to Fertility and Mortality: Causal Evidence at the Swiss Language Border. *CESifo Working Paper*, No. 10515. <https://doi.org/10.2139/ssrn.4492290>
- Gelfand, M. J., Raver, J. L., Nishii, L., Leslie, L. M., Lun, J., Lim, B. C., Duan, L., Almaliach, A., Ang, S., Arnadottir, J., Aycan, Z., Boehnke, K., Boski, P., Cabecinhas, R., Chan, D., Chhokar, J., D'Amato, A., Ferrer, M., Fischlmayr, I. C., ... Yamaguchi, S. (2011). Differences between tight and loose cultures: A 33-nation study. *Science*, 332(6033), 1100–1104. <https://doi.org/10/dzt3n2>

- Gervais, W. M. (2021). Practical Methodological Reform Needs Good Theory. *Perspectives on Psychological Science*, 16(4), 827–843. <https://doi.org/10/ghxc54>
- Greenfield, P. M. (2014). Sociodemographic Differences Within Countries Produce Variable Cultural Values. *Journal of Cross-Cultural Psychology*, 45(1), 37–41. <https://doi.org/10.1177/0022022113513402>
- Hansen, N., & Heu, L. (2020). All Human, yet Different. *Social Psychology*, 51(6), 361–369. <https://doi.org/10.1027/1864-9335/a000436>
- Harati, H., & Talhelm, T. (2023). Cultures in Water-Scarce Environments Are More Long-Term Oriented. *Psychological Science*, 34(7), 754–770. <https://doi.org/10.1177/09567976231172500>
- Heine, S. J., & Norenzayan, A. (2006). Toward a Psychological Science for a Cultural Species. *Perspectives on Psychological Science*, 1(3), 251–269. <https://doi.org/10/cwtzfw>
- Henrich, J. (2016). *The secret of our success: How culture is driving human evolution, domesticating our species, and making us smarter*. Princeton University Press.
- Henrich, J., Boyd, R., Bowles, S., Camerer, C., Fehr, E., Gintis, H., McElreath, R., Alvard, M., Barr, A., Ensminger, J., Henrich, N. S., Hill, K., Gil-White, F., Gurven, M., Marlowe, F. W., Patton, J. Q., & Tracer, D. (2005). “Economic man” in cross-cultural perspective: Behavioral experiments in 15 small-scale societies. *Behavioral and Brain Sciences*, 28(6), 795–815. <https://doi.org/10/cp87gg>
- Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences*, 33(2–3), 61–83. <https://doi.org/10.1017/S0140525X0999152X>
- Hofstede, G. (2001). *Culture's Consequences: Comparing Values, Behaviors, Institutions and Organizations Across Nations*. SAGE.
- House, B. R., Kanngiesser, P., Barrett, H. C., Broesch, T., Cebiglu, S., Crittenden, A. N., Erut, A., Lew-Levy, S., Sebastian-Enesco, C., Smith, A. M., Yilmaz, S., & Silk, J. B. (2020).

- Universal norm psychology leads to societal diversity in prosocial behaviour and development. *Nature Human Behaviour*, 4(1), 36–44. <https://doi.org/10/gf9hmp>
- House, B. R., Silk, J. B., Henrich, J., Barrett, H. C., Scelza, B. a, Boyette, A. H., Hewlett, B. S., McElreath, R., & Laurence, S. (2013). Ontogeny of prosocial behavior across diverse societies. *Proceedings of the National Academy of Sciences of the United States of America*, 110(36), 14586–14591. <https://doi.org/10.1073/pnas.1221217110>
- Kitayama, S., & Cohen, D. (2010). *Handbook of Cultural Psychology, First Edition*. Guilford Press.
- Klein, R. A., Vianello, M., Hasselman, F., Adams, B. G., Adams, R. B., Alper, S., Aveyard, M., Axt, J. R., Babalola, M. T., Bahník, Š., Batra, R., Berkics, M., Bernstein, M. J., Berry, D. R., Bialobrzeska, O., Binan, E. D., Bocian, K., Brandt, M. J., Busching, R., ... Nosek, B. A. (2018). Many Labs 2: Investigating Variation in Replicability Across Samples and Settings. *Advances in Methods and Practices in Psychological Science*, 1(4), 443–490. <https://doi.org/10.1177/2515245918810225>
- Kraus, M. W., Park, J. W., & Tan, J. J. X. (2017). Signs of Social Class: The Experience of Economic Inequality in Everyday Life. *Perspectives on Psychological Science*, 12(3), 422–435. <https://doi.org/10.1177/1745691616673192>
- Laland, K. N. (2018). *Darwin's Unfinished Symphony: How Culture Made the Human Mind*. Princeton University Press.
- Muthukrishna, M., Bell, A. V., Henrich, J., Curtin, C. M., Gedranovich, A., McInerney, J., & Thue, B. (2020). Beyond Western, Educated, Industrial, Rich, and Democratic (WEIRD) Psychology: Measuring and Mapping Scales of Cultural and Psychological Distance. *Psychological Science*, 31(6), 678–701. <https://doi.org/10/ggxkjb>
- Muthukrishna, M., & Henrich, J. (2019). A problem in theory. *Nature Human Behaviour*, 3(3), 221–229. <https://doi.org/10/gfvdx8>
- Muthukrishna, M., Henrich, J., & Slingerland, E. (2021). Psychology as a Historical Science. *Annual Review of Psychology*, 72(1), 717–749. <https://doi.org/10/ghrnb6>

- Norenzayan, A., & Heine, S. J. (2005). Psychological universals: What are they and how can we know. *Psychological Bulletin*, 131(5), 763–784. <https://doi.org/10/bgtcvr>
- Nosek, B. A., Hardwicke, T. E., Moshontz, H., Allard, A., Corker, K. S., Dreber, A., Fidler, F., Hilgard, J., Kline Struhl, M., Nuijten, M. B., Rohrer, J. M., Romero, F., Scheel, A. M., Scherer, L. D., Schönbrodt, F. D., & Vazire, S. (2022). Replicability, Robustness, and Reproducibility in Psychological Science. *Annual Review of Psychology*, 73(1), 719–748. <https://doi.org/10.1146/annurev-psych-020821-114157>
- Obradovich, N., Özak, Ö., Martín, I., Ortuño-Ortín, I., Awad, E., Cebrián, M., Cuevas, R., Desmet, K., Rahwan, I., & Cuevas, Á. (2022). Expanding the measurement of culture with a sample of two billion humans. *Journal of The Royal Society Interface*, 19(190), 20220085. <https://doi.org/10.1098/rsif.2022.0085>
- Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, 349(6251), aac4716–aac4716. <https://doi.org/10/68c>
- Schimmelpfennig, R., & Muthukrishna, M. (2023). Cultural evolutionary behavioural science in public policy. *Behavioural Public Policy*, 1–31. <https://doi.org/10.1017/bpp.2022.40>
- Schimmelpfennig, R., Spicer, R., White, C., Gervais, W. M., Norenzayan, A., Heine, S., Henrich, J., & Muthukrishna, M. (2023). *A Problem in Theory and More: Measuring the Moderating Role of Culture in Many Labs 2*. PsyArXiv. <https://doi.org/10.31234/osf.io/hmnrx>
- Smaldino, P. E., Lukaszewski, A., von Rueden, C., & Gurven, M. (2019). Niche diversity can explain cross-cultural differences in personality structure. *Nature Human Behaviour*, 3(12), Article 12. <https://doi.org/10.1038/s41562-019-0730-3>
- Stroebe, W. (2019). What Can We Learn from Many Labs Replications? *Basic and Applied Social Psychology*, 41(2), 91–103. <https://doi.org/10.1080/01973533.2019.1577736>
- Talhelm, T., Haidt, J., Oishi, S., Zhang, X., Miao, F. F., & Chen, S. (2015). Liberals Think More Analytically (More “WEIRD”) Than Conservatives. *Personality and Social Psychology Bulletin*, 41(2), 250–267. <https://doi.org/10.1177/0146167214563672>

- Thalmayer, A. G., Toscanelli, C., & Arnett, J. J. (2020). The neglected 95% revisited: Is American psychology becoming less American? *American Psychologist*, 76(1), 116–129. <https://doi.org/10/gg4vbg>
- Uchiyama, R., Spicer, R., & Muthukrishna, M. (2022). Cultural Evolution of Genetic Heritability. *Behavioral and Brain Sciences*, 2022, e152. <https://doi.org/10/gkct6d>
- Van Assche, J., Swart, H., Schmid, K., Dhont, K., Al Ramiah, A., Christ, O., Kauff, M., Rothmann, S., Savelkoul, M., Tausch, N., Wölfer, R., Zahreddine, S., Saleem, M., & Hewstone, M. (2023). Intergroup contact is reliably associated with reduced prejudice, even in the face of group threat and discrimination. *American Psychologist*, 78(6), 761–774. <https://doi.org/10.1037/amp0001144>
- White, C., & Muthukrishna, M. (2023). *More educated, more WEIRD: Higher education predicts worldwide cultural similarity to the United States and other WEIRD countries*. PsyArXiv. <https://doi.org/10.31234/osf.io/8wr5d>
- White, C., Muthukrishna, M., & Norenzayan, A. (2021). Cultural similarity among coreligionists within and between countries. *Proceedings of the National Academy of Sciences*, 118(37), e2109650118. <https://doi.org/10/gmv9pd>