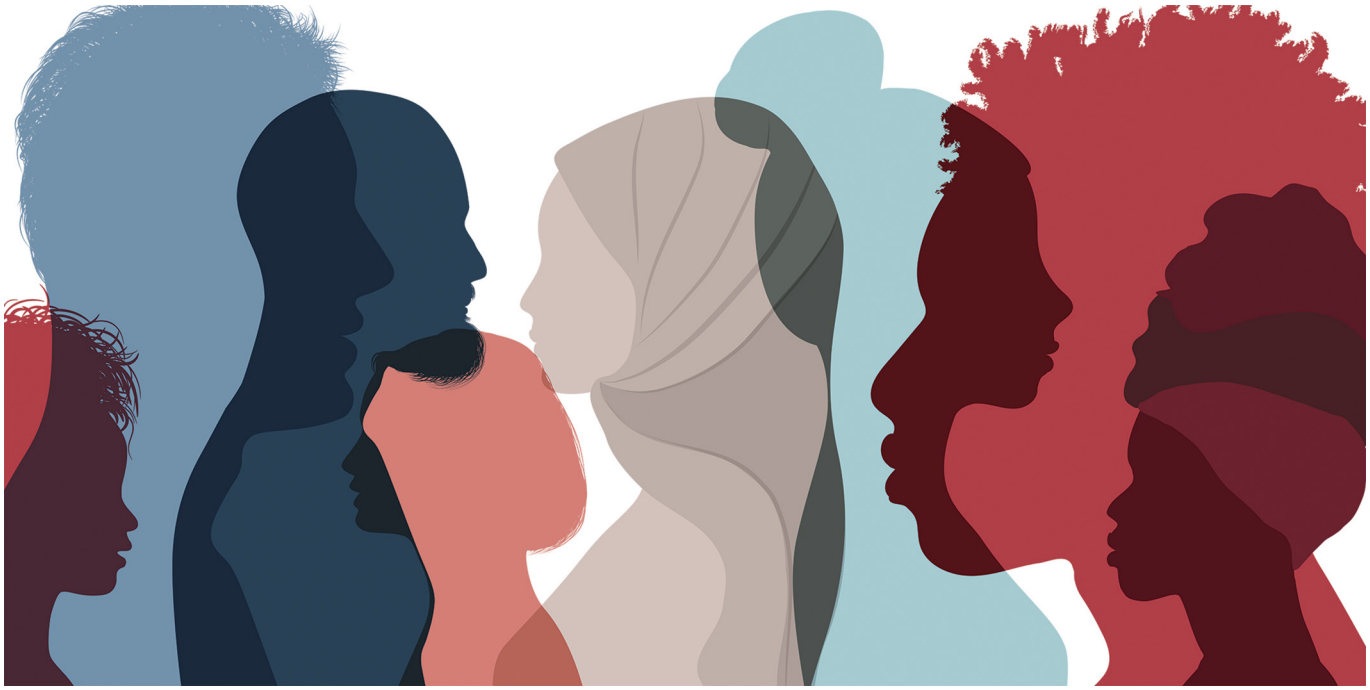
PNAS OPINION

The limits of doing global, cross-cultural behavioral science research

Andrés Gvirtz[a,1] and Anandita Sabherwal[b]



Even for researchers with the best of intentions, it can be challenging to conduct research with diverse, representative sample sets, despite the many tools at scientists' disposal. Image credit: Shutterstock/melitas.

As early-career researchers, we are acutely aware that while behavioral science research typically focuses on Western populations, it aims to draw universal conclusions about human behavior. We were still in high school in 2010 when the Heinrich et al. landmark paper "The weirdest people in the world?" argued that research conducted solely on Western, Educated, Industrialized, Rich, and Democratic (WEIRD) samples might not apply universally (1).

Many studies have highlighted cultural differences in processes that might have reasonably been assumed to be universal, such as visual perception. For example, individuals from Western and non-Western cultures were found, on average, to be susceptible to different optical illusions because of differences in exposure to two-dimensional representations (2). Differences in fundamental psychological phenomena, such as perceptions and beliefs, matter. For example, differences in economic decision making and perceptions of fairness across cultures can scale up to produce societies with vastly different social norms and preferences (3).

Researchers are well aware of the need to improve sample diversity to generate universal insights. Our experiences illustrate just how challenging it can be to conduct research with diverse, representative sample sets, despite the many tools at researchers' disposal. We found that online data-collection platforms still lack representation from a large proportion of the global population, including Africa, Asia, and South America. Achieving the goal of diverse research will require awareness about platform pluses and minuses, locally sourced alternative sample collection, and collaborations with regional labs.

Seeking Diversity

One of us (A.G.) studied the influence of personality and policy on sheltering-at-home rates during the COVID-19 pandemic (4), while the other (A.S.) is investigating the role of morality in pro-environmental behaviors, such as energy conservation, tree planting, and transport use, around the world (5).

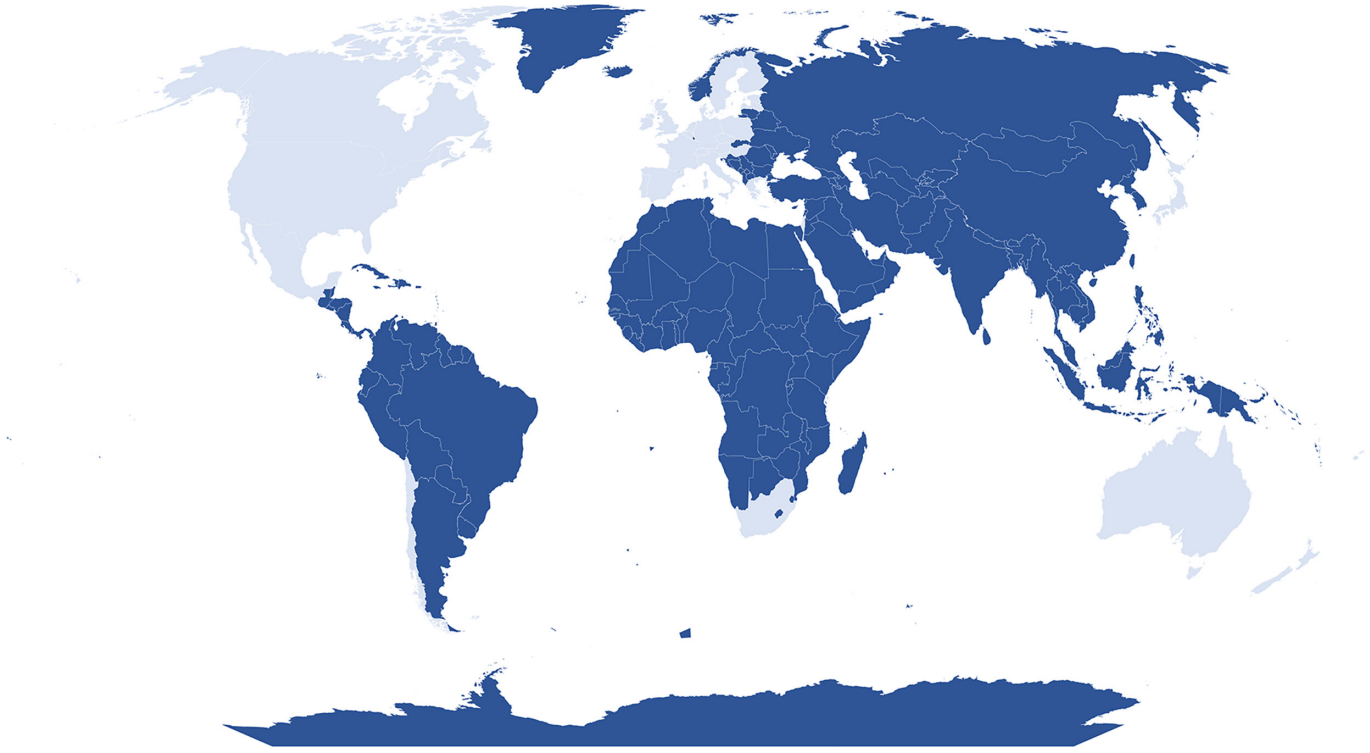
Author affiliations: [a]King's Business School, King's College London, London WC2R 2LS, United Kingdom; and [b]Department of Psychological & Behavioural Science, London School of Economics, London WC2A 2AE, United Kingdom

Author contributions: A.G. and A.S. designed research; performed research; analyzed data; and wrote the paper.

The authors declare no competing interest.

[1]To whom correspondence may be addressed. Email: andres.gvirtz@kcl.ac.uk.

Published August 28, 2024.



Dark blue countries are unavailable on Prolific, applying a threshold of 100 active workers; mTurk declined any request for information, but previous research has suggested that Amazon mTurk's sample diversity is limited. See ref. 6.

This past year, we designed a study to probe differences in prosocial behavior—actions intended to benefit others—and, more specifically, the social and personality-related factors that have an impact on people's willingness to donate their time for charitable causes. With cultural and national nuances likely to influence such behaviors, our goal was to explore how these variations manifest across countries and across the online platforms that researchers use to recruit participants for studies.

We failed to conduct a truly global study—not because of a lack of effort, but because current research platforms are much less suitable for research done between countries than is commonly assumed.

When starting the project, we adopted a systematic approach to choose which countries to include. We wanted to recruit a sample that was representative of key demographic characteristics (such as age and ethnicity) from the most populous country of every continent. If a representative sample was not available for the most populous country, we wanted to settle for a convenience sample. These samples consist of participants who are easily reachable—for example, because they are students in a behavioral science department—but they are unlikely to be representative of the country's wider demographics. If a convenience sample was also not available, we considered the next largest country, looking again for a representative sample. Our primary tools were online data-collection platforms Prolific and Amazon Mechanical Turk (mTurk). Both are market-leading companies used for research, with Prolific being specifically designed for academic research, while mTurk is a broader platform for various tasks, including surveys. We had used both platforms in the past successfully to recruit mainly Western samples.

We started with Prolific, where the system allows users to set recruitment criteria and returns the available sample sizes. We went one by one through every single country. The continents of Africa, Asia, and South America had only one country each represented in Prolific's participant pool: South Africa, Israel, and Chile, respectively. This held true even when we applied minimal inclusion criteria, which required only having at least 100 active workers in that country within the last 90 days. In other words, 99.81% of African, 95.69% of Asian, and 95.53% of South American populations are inaccessible when using this popular research tool.

The makeup of Amazon mTurk's participant pool appeared much less transparent. While the mTurk system also allows us to set recruitment criteria, such as countries, the system does not indicate whether there are actually any workers available that fulfill the criteria provided. Reaching out to their email support team, we requested the demographic information, but were told that while they have a global sample, they do not have any specific information they can disclose. In addition to previous studies raising concerns about the data quality and validity when using mTurk (7), there is evidence to suggest that we cannot assume global sample diversity.

Until 2019, Amazon did not make cash payments to participants outside the United States and India. It instead relied on Amazon gift cards, making participation less viable in countries lacking an Amazon presence or in those where participants were interested in cash payments rather than gift cards (8). Over time, workarounds have emerged to allow participants in other countries to cash out their payments. But these remain cumbersome and create an additional barrier to broadening participation. Indeed, the largest study on the demographics of Amazon mTurk

workers (n > 40,000) found that although 30 countries were represented, over 90% of participants were based either in the United States (75%) or India (16%), highlighting limited sample diversity (6). Furthermore, it is unclear how representative this study was of Amazon mTurk at the time, and there does not appear to be a way of finding out whether the situation has evolved.

The lack of sample diversity on Prolific and mTurk may not come as a surprise to some, given that these platforms were developed by Western companies. And our concerns may sound especially familiar to researchers in the Global South and in other underrepresented countries such as Chile, Lebanon, and South Africa, where researchers have consistently underscored the limited access to large samples in their local context, waging an uphill battle against Western publishing standards that emphasize multistudy papers with large, and increasingly also representative, samples (9).

Conducting research that is informed by, and benefits, diverse groups requires that we collaborate with scholars local to the research context.

Seeking better-quality data from overlooked countries, we turned to market research companies. That option proved shockingly expensive. A leading company told us that a 10-minute study with 500 participants in Sri Lanka, without translation, would cost £4,498 or $5,712. This is more than the departmental research budget either of us had during our PhD programs.

Better Tools

So how can and should cross-cultural research be conducted? We believe there are several lessons for the field. First, researchers at all stages of their careers must be made more aware of the limitations of popular research platforms. Data-collection platforms need to clearly communicate the limitations of their participant pools. While Prolific did not feature the sample diversity we needed, at least this platform is fully transparent about the nationalities of its participants. We believe that other platforms must also pursue such transparency to enable informed research decisions about data-collection platforms.

Departments, research supervisors, and young researchers must do more to find and use platforms local to countries that remain underrepresented in research. Often, these platforms are in nascent stages and do not have resources for marketing at international conferences. When planning our study, we came across newer and local online platforms. However, lacking sufficient knowledge about these national contexts, we were unable to confidently determine their sample quality, credibility, and participant payment practices.

Partnerships and connections with research labs in the countries we sample are the way forward. Collaborating with colleagues who are local and familiar with a context can enhance the cultural sensitivity of our hypotheses and modes of data collection. Not only will such colleagues be more proficient in assessing the credibility and representativeness of local data-collection platforms, but they may also facilitate in-person data collection. A prime example of this approach is many labs' international collaborations, in which researchers across many countries come together to co-create a study protocol and then collect data using this protocol in their respective local contexts. As such, many labs' studies ensure sample quality by entrusting local researchers with data collection and are equipped to comment on the generalizability of psychological phenomena across cultural contexts. These large-scale investigations are time- and cost-intensive endeavors and, therefore, might be especially suited for research questions surrounding fundamental psychological phenomena that require generalizability (such as perception and values), as well as issues with global immediacy and relevance, such as those pertaining to climate mitigation action (10) and public health support (11).

For our small study, we have resigned ourselves to using Prolific, a platform that was transparent and provided us all the information required to make an informed decision. This has meant that our research, which we are currently preparing for publication, relies on the few countries that are available on the platform.

We drew a map of the research with borders, drawing all countries unavailable using prominent online data-collection methods. We hope to work toward a future in which research encapsulates insights beyond the few countries currently represented. To make progress, the academic community, including researchers and participant recruitment platforms, must be transparent about the inadequacy of the current system. Conducting research that is informed by, and benefits, diverse groups requires that we collaborate with scholars local to the research context. Only then will we be able to make these studies a little less "WEIRD."

1. J. Henrich, S. J. Heine, A. Norenzayan, The weirdest people in the world? *Behav. Brain Sci.* **33**, 61–83 (2010), 10.1017/S0140525X0999152X.
2. M. H. Segall, D. T. Campbell, M. J. Herskovits, "The influence of culture on visual perception" in *Social Perception*, H. Toch, C. Smith, Eds. (Bobbs-Merrill, Indianapolis, IN, 1966).
3. M. Muthukrishna *et al.*, Beyond Western, Educated, Industrial, Rich, and Democratic (WEIRD) psychology: Measuring and mapping scales of cultural and psychological distance. *Psychol. Sci.* **31**, 678–701 (2020), 10.1177/0956797620916782.
4. F. M. Götz, A. Gvirtz, A. D. Galinsky, J. M. Jachimowicz, How personality and policy predict pandemic behavior: Understanding sheltering-in-place in 54 countries at the onset of COVID-19. *Am. Psychol.* **76**, 39–49 (2021), 10.1037/amp0000740.
5. Moral Balancing Project, The moral balancing project. https://moralbalancing.weebly.com. Accessed 21 August 2024.
6. D. Difallah, E. Filatova, P. Ipeirotis, Demographics and dynamics of Mechanical Turk workers. *Proc. ACM Int. Conf. Web Search Data Min.* **11**, 135–143 (2018), 10.1145/3159652.3159661.
7. M. A. Webb, J. P. Tangney, Too good to be true: Bots and bad data from Mechanical Turk. *Perspect. Psychol. Sci.*, 10.1177/17456916221120027 (2022). Accessed 21 August 2024.
8. Amazon Mechanical Turk, "Amazon Mechanical Turk workers in 25 countries outside of the US can now transfer their earnings" (2019). https://blog.mturk.com/amazon-mechanical-turk-workers-in-23-countries-outside-of-the-us-can-now-transfer-their-earnings-98ec29ef7f7f
9. F. Bou Zeineddine, R. Saab, B. Lášticová, A. H. Ayanian, A. Kende, "Unavailable, insecure, and very poorly paid": Global difficulties and inequalities in conducting social psychological research. *J. Soc. Polit. Psychol.* **10**, 723–742 (2022), 10.5964/jspp.8311.
10. M. Vlasceanu *et al.*, Addressing climate change with behavioral science: A global intervention tournament in 63 countries. *Sci. Adv.* **10**, eadj5778 (2024), 10.1126/sciadv.adj5778.
11. J. J. Van Bavel *et al.*, National identity predicts public health support during a global pandemic. *Nat. Commun.* **13**, 517 (2022), 10.1038/s41467-021-27668-9.