American Psychologist

There Is Nothing WEIRD About Basic Research: The Critical Role of Convenience Samples in Psychological Science

Jeffrey W. Sherman

Online First Publication, November 7, 2024. https://dx.doi.org/10.1037/amp0001443

American Psychologist

ISSN: 0003-066X
https://doi.org/10.1037/amp0001443

There Is Nothing WEIRD About Basic Research: The Critical Role of Convenience Samples in Psychological Science

Jeffrey W. Sherman
Department of Psychology, University of California, Davis

Attention to issues of sample diversity and generalizability has increased dramatically in the past 15 years, as psychological scientists have confronted the limitations of relatively homogeneous samples. Though this reckoning was perhaps overdue and has undoubtedly shined a light on some poor research practices, recommendations surrounding sample diversity are sometimes applied to research that does not aim for generalizability across peoples. In this article, I seek to promote discussion about when and why sample diversity and generalizability matter. In doing so, I address problems with language surrounding generalizability, the broader question of generalizability beyond samples, challenges for determining sufficient generalizability, and the inherent question of moderation in psychological science, given the reality of limited time and resources. I then discuss the important roles that basic research plays in understanding group differences, producing generalizable knowledge, and developing applied interventions. Finally, I address issues of equity surrounding sample diversity, emphasizing the distinction between WEIRD samples and convenience samples and the importance of convenience samples for globalizing psychological science.

Public Significance Statement
This article addresses critical questions about the generalizability of psychological research. Whereas generalizability across samples is critical for research designed for practical application, it is not for strictly theoretical work attempting to understand some phenomenon. There are significant implications for the cost and efficiency of research.

Keywords: sample selection, external validity, theory, basic research, generalizability

Since the publication of Henrich et al.'s (2010) landmark "WEIRD" article, in which the authors note that most research participants in empirical psychology come from Western, Educated, Industrialized, Rich, and Democratic (WEIRD) societies, concerns about sample diversity and generalizability have grown significantly in prominence. That work has been heavily cited, receiving over 13,000 citations by the end of 2023, and the rate of citation has grown rapidly, now surpassing 4,000 per year, according to Google Scholar. Thought pieces elaborating on the concern have been prevalent (e.g., Clancy & Davis, 2019; Medin, 2017; Nielsen et al., 2017; Rad et al., 2018; Roberts et al., 2020; Thalmayer et al., 2021), and editorial policies and priorities at top journals and granting agencies have shifted in response to these concerns (e.g., Basnight-Brown et al., 2023; "Towards a Global Psychological Science," 2022). Both the Association for Psychological Science and the American Psychological Association journals now strongly encourage the use of non-WEIRD samples and, in some cases, require "constraints on generality" statements to address homogenous samples (e.g., Kitayama, 2017; Simons et al., 2017). The journal *Psychological Science* explicitly places greater value on submissions that include non-WEIRD samples (Vazire, 2024). Many American Psychological Association journals require that all-White or all-Western samples be justified amid a description of the sample inclusion process (American Psychological Association, 2024). The National Institutes of Health requires grant

Rick Hoyle served as action editor.
Jeffrey W. Sherman https://orcid.org/0000-0002-9020-9814
The author has no competing interests to report. The author thanks Chris Crandall, Keith Maddox, and Andy Todd for their very helpful feedback on earlier drafts of this article and the SoCoLab at the University of California, Davis.

Jeffrey W. Sherman played a lead role in conceptualization, writing–original draft, and writing–review and editing.

Correspondence concerning this article should be addressed to Jeffrey W. Sherman, Department of Psychology, University of California, Davis, One Shields Avenue, Davis, CA 95616, United States. Email: jsherman@ucdavis.edu

Jeffrey W. Sherman

applications to describe and justify the sample diversity of proposed research (National Institutes of Health, 2022).

Henrich et al. (2010) represented a watershed moment in empirical psychology's confrontation with the extent of WEIRD samples and the limits of generalization those samples impose. At the same time, recommendations and requirements for diverse and representative samples are increasingly misapplied to research that does not hold such generalizability as a central goal. Critically, the publication and funding initiatives described above apply to both basic and applied research, even though the former frequently is not concerned with generalizability (more on this below). The point is not that people are all the same or that diverse samples should be avoided; it is far from it. The point is that they are critical when the intent of the research is to establish generalizability across peoples, but not when it is not. Here, it is important to distinguish between WEIRD samples and *convenience* samples, which may or may not be WEIRD. Whereas WEIRD samples, in their own right, confer no particular benefit, convenience samples confer many benefits, whatever types of participants happen to be convenient, even when they are WEIRD.

In the remainder of this article, I advocate for reliance on convenience samples for basic psychological science. I elaborate on the questions of when and why generalizability across peoples is important and address problematic language that implies overclaiming of such generalizability. I raise broader questions about balancing the many important facets of research generalizability against a backdrop of limited time and resources. I discuss the critical role that basic research with convenience samples plays in both understanding the mechanisms underlying group differences and developing applied interventions. Finally, I address issues of research equity and the positive contributions that basic research with convenience samples can offer to equity goals. The scientific and even moral foundations underlying calls for sample diversity have become nearly ubiquitous articles of faith in psychological science. This article seeks to revive debate about the conditions under which sample diversity is critical and challenge the suggestion that convenience samples (even WEIRD ones) are inherently problematic in a field seeking to expand equity and social justice.

When and Why? Basic Versus Applied Research

In considering the contexts in which sample diversity is critical, the distinction between basic and applied research is of central concern (Mook, 1983). The goal of basic research is to develop and refine theories for the sake of expanding knowledge (e.g., to understand the mental processes through which persuasion succeeds). Sometimes, it takes the form of an existence proof, which seeks to demonstrate only that a particular behavioral outcome is possible and not that it is typical or generalizable. Thus, basic research is not aimed at addressing a specific "real-world" problem (e.g., to develop an effective intervention to stop teens from smoking). Basic science can and often does contribute to solving important problems, which I address in more detail below, but that is not the fundamental aim of the work, and it is not evaluated on its ability to do so.

As such, basic science does not place a premium on representative samples. Consider research conducted to understand the mechanisms underlying cellular activity. This type of research is routinely conducted in samples of genetically identical mice raised in identical settings. Researchers are not concerned that the results may not be representative of all mice in the world. Rather, they are minimizing the variability and random noise among the mice to provide the strongest opportunity to observe theoretically important effects. If a theoretical prediction is correct, it should be observed in *this sample of mice, right here, right now*. No attempt is made to establish the generalizability across mice nor is any claim made to that effect. Of course, this kind of research, which is often conducted to better understand human cellular mechanisms, also does not test the mechanisms in any human sample, never mind a representative human sample. However, should a novel medical treatment for humans result from this basic research, it is absolutely critical that its effectiveness be established in a sample that is representative of the population to which the treatment will be offered.

Analogously, basic psychological science proceeds as such: If Theory X is correct, then the subjects *in this sample, right here, right now* should demonstrate Y. It is not directed toward nor relevant to the question of generalizability across people and does not require a representative sample. Moderators, including those related to sample diversity, can be a vital component of basic research (a point to which I will later

return), but it is not inherent to the enterprise. In contrast, applied research, which aims to solve specific problems, is inherently concerned with generalizability. For example, research designed to build a teen smoking intervention must be concerned that the intervention is effective in any population to which the intervention will be applied and requires a representative sample from that population.

Overgeneralization and the Language Problem

One reason basic psychological science often gets itself into trouble is the language we use to describe our findings. Here, I will build on Mook's (1983) excellent discussion of this matter. Consider the prediction from dual-process models of person perception that people rely on stereotypes to a greater extent when they are under cognitive load (e.g., Brewer, 1988; Fiske & Neuberg, 1990). Let us say I conduct a study with my local convenience sample, in which I ask some participants to count backward by thirteens while they form an impression of a target person. If the dual-process models are correct, then the participants (these participants, right here, right now) doing the counting should report more stereotypical impressions than those who do not do the counting. Let us say that the results support this prediction and show greater stereotyping among participants in the counting condition. I might conclude that people exactly like my participants who count backward by thirteens and form impressions of the target I presented and make judgments on the scales I used stereotype more than people who do not count backward by thirteens. That is certainly accurate and very uninspiring. Alternatively, I might conclude that these results show that *people* in the *real world* stereotype more when cognitively loaded in any way, regardless of the target, and however you measure stereotyping. That would be quite an overstatement, given my data.

The appropriate conclusion has an entirely different flavor. It only refers back to the theory that guided the research: These data are consistent with *the theory that people use stereotypes more when they are under cognitive load*. That is it. That is all I can say. There is no universalist goal or claim (Kroupin et al., 2024). The conclusion does not refer at all to the specificity/generality of the sample, the judgment target, the operationalization of cognitive load, or the operationalization of stereotyping. The research was not designed or intended to establish any such generalities. If I wish to address those questions of external validity, I will need to do more research. As is, I can only note how the findings relate to the theory and its constructs.

Unfortunately, as is noted by many concerned with WEIRD samples, we often state conclusions in the form of the following: This research shows that people stereotype more when they are under cognitive load. The problem, of course, is that this conclusion implies that the result applies to all people. By the same logic, it also implies that the results apply to all kinds of judgment targets, cognitive loads, and indicators of stereotyping. In all cases, those conclusions are not warranted. The findings are simply moot with regard to these questions. I have certainly been guilty many times of drawing just such overgeneral conclusions. It is difficult not to do; it feels like the natural and clearest description in plain language. However, all that is required to avoid this error is to refer to the theory and its predicted relationships among relevant constructs and not to the subjects, targets, or operations of the study in any way. The data are consistent or not with the theory that x, y, z.

In my reading of the calls for greater sample diversity, this language problem has played an oversized role. Yet, in my experience, when people say: This research shows that people stereotype more under cognitive load, they do not believe or intend to claim that the finding establishes the result for all people and all relevant operations. Certainly, when I have made such overgeneralizations, I did not intend to make such claims. Whatever the case, one possible solution is for researchers to actually demonstrate that the finding holds with all kinds of people (operations, etc.) to make sure that their language is accurate. This is often the implied solution among critics of WEIRD samples. A different solution to making the language match the data is to simply be more careful with our language and make sure that the conclusions refer only to the theory and its constructs, without making any claims implicating the extent of external validity. That is, rather than change the nature of our research to fit our sloppy language, we can change our sloppy language to more accurately reflect our research (see also Yarkoni, 2022).

How Much Generalization?

Even when generalization across people is an important research goal, it is difficult to say how much generalization is enough. Though the situation seems to be changing, the overwhelming majority of research on sample diversity in empirical psychology has compared North American/Western European cultures with East Asian cultures. This research has been highly impactful and has contributed substantially to the recognition that psychology may differ across cultures. At the same time, this does not reflect a representative sample of humanity or a tremendous amount of cultural diversity in a planetary context. There are hundreds of human cultures, and we know very little about how they may differ from one another psychologically. To be sure, knowing about potential differences among two, three, or four cultures provides greater generalizability than knowing about one. However, if the goal is to be justified in drawing broad conclusions about human behavior, this type of research does not achieve it.

I would argue that generalization to all peoples is simply not an attainable goal in the behavioral sciences. In the most extreme example, one could take a single finding and spend the better part of a career simply testing the worldwide

generalizability of that single result and do little else. That would be extremely difficult and expensive research, and not many people would sign up for such an endeavor, foregoing other important goals in building a research program and understanding human behavior. This raises inevitable questions about how much generalizability is enough and which peoples may be excluded. In addition, it reflects that, even when attempting to establish greater generality, the broader samples we access most frequently share features of convenience samples. Simply, some cultures are easier to access than others, and that is where we tend to focus our generalizability efforts. Is that sufficient?

I certainly do not wish to imply that no amount of generalizability is worthwhile if we cannot examine total generalizability. However, we would do well to recognize that the goal of conducting research that permits complete or even broad generalization is not attainable and should not be invoked as a standard or leveled as a critique of research findings. We also should recognize that our attempts at establishing sample diversity are bounded by the same pragmatic concerns that have led to the widespread use of Intro Psych subject pools. Research is always a trade-off in resources. Acknowledging this truth places the discussion of sample diversity into a broader discussion about the many different potential moderators one might investigate and the many reasons for choosing among them. The choice to use a convenience sample often is a choice to focus on different important moderators of our effects.

What Kinds of Generalization?

Indeed, beyond the question of how much generalizability across peoples is sufficient lies an even more fraught question: Which kinds of generalizability should we prioritize? There are many important aspects of external validity beyond those pertaining to generalization across human groups. Questions pertaining to the generalizability of manipulations and measures can be particularly important. In the study about stereotyping and cognitive load described above, does the counting backward manipulation of cognitive load generalize to other kinds of cognitive loads? How representative is the judgment target of all possible judgment targets? Does the chosen measure of stereotyping generalize to other indices of stereotyping? In a world of limited time and resources, investigating these aspects of generalizability necessarily competes with the choice to examine generalization across peoples.

Among researchers conducting basic, theory-driven research, establishing generalization across operations is often of much greater concern than generalization across peoples. If the primary goal of the work is to establish theoretically derived relationships among constructs, then ensuring that we are accurately manipulating and measuring those constructs is of critical importance (Flake et al., 2017). This is precisely what generalization across manipulations and measures provides. The point is not just to show generalization for the sake of generalization. The point is to establish construct validity by demonstrating that multiple operationalizations of the same constructs produce the same outcomes. This kind of work is absolutely critical for testing and improving theories and advancing science (Crandall & Sherman, 2016; Yarkoni, 2022). The suggestion that these questions should be secondary to establishing generalizability across peoples undermines basic research, and one may reasonably ask how cross-group generalizability matters if we do not know what we are manipulating or measuring. Moreover, establishing such construct validity may be necessary for any novel group to be tested, as the same operations may not induce or measure the same constructs among different populations. Indeed, demonstrating the invariance of manipulations and measures across groups can also increase confidence in the generalizability of the primary research findings.

The Broader Issue of Moderators in Generalization

Other conflicting goals pertain to research priorities that have nothing to do with establishing generalizability per se. You may be more interested in extending your theoretical work than in establishing its generality. Such extensions often involve testing theoretically meaningful moderators to an observed outcome. Cross-cultural moderators play exactly this role in basic research aiming to understand the psychological mechanisms associated with specific cultural differences. However, there is practically an infinite variety of meaningful moderators that may inform theory and research. Researchers may be interested in how different motivations, cognitive states (e.g., cognitive load), social contexts (e.g., alone or with others), or personalities (e.g., Person × Situation interactions) moderate a result. Those moderators may be the ones that shed the most light on a theory and its implications. For example, in the stereotyping project described above, a second key prediction of dual-process models is that stereotyping should increase when motivation to process carefully is low. As such, the next step after testing cognitive load as a moderator may be to test accuracy motivation as a moderator.

Limited time and resources are unavoidable constraints. One simply cannot test every kind of generalizability or examine every potentially interesting moderator. When researchers are required to discuss constraints on generality, the correct response is always that there is an infinite variety of ways in which the research may not generalize. It is presumptuous and counterproductive to insist that researchers concern themselves first and foremost with different peoples as a potential moderator or limitation on generalizability. Researchers can judge for themselves which moderators

matter the most for advancing their research and what kinds of generalizability are most critical.

Basic Research Is Critical for Understanding Group Differences

In and of themselves, observed differences among human groups may tell us very little about human psychology. Culture, ethnicity, race, gender/sex, and other *subject variables* are not psychological mechanisms that can explain variation in human behavior. Rather, observed differences in behavior among these groups presumably reflect some important psychological process(es) that differs across groups. Empirical psychologists are not demographers. Our job is not to simply identify and catalog group differences. Our job is to explain the psychology underlying those differences (e.g., Galinsky et al., 2024). Thus, having observed group differences, further research is required to understand the meanings of those differences via theoretically derived predictions about and tests of psychological mechanisms. This kind of work can be characterized as basic research aimed at identifying mechanisms underlying group differences and does not aim to generalize beyond the specific groups being compared.

However, even basic research with convenience samples contributes to understanding human diversity in powerful ways. In an earlier call to diversify research samples, Sears (1986) discussed the drawbacks of primarily using college students as participants. Specifically, he argued that self-perception processes in attitude expression (Bem, 1972) are most likely to unfold among people who have relatively uncrystallized attitudes. He further suggested that because college students are more likely to have such weak attitudes, their overuse in research has exaggerated the role of self-perception in attitude expression. How does Sears know that attitude certainty is a critical moderator of self-perception? He cites basic research conducted with convenience samples by Chaiken and Baldwin (1981), Wood (1982), and Taylor (1975).

As an example, Chaiken and Baldwin (1981) identified University of Toronto students who had earlier expressed attitudes about environmentalism that were either crystallized or not and then induced them to behave in either a pro- or antienvironmental fashion. Subsequently, when the participants' attitudes were again measured, those with initially weaker attitudes were influenced by their induced behavior: Those who were induced to behave in a proenvironmental manner expressed more positive attitudes about environmentalism, and those induced to counter-environmental behavior expressed more negative attitudes. This kind of attitudinal inference from one's behavior is the hallmark of self-perception theory. In contrast, those with strong attitudes were not affected by their pro- or antienvironmental behavior. In this way, Chaiken and Baldwin (1981) were able to demonstrate the important role of attitude certainty in self-perception using a convenience sample of college students. In turn, this finding provided a foundation for Sears' prediction that younger participants would engage in self-perception processes to a greater extent than older participants.

This example highlights the value that basic research on psychological processes among convenience samples can bring to identifying important group differences, accounting for them, once observed, and contributing to generalizability. The findings of this research apply not only to participants of different ages but to all groups that vary in terms of attitude certainty. Any group observed to have particularly high or low attitude certainty can now be expected to exhibit self-perception processes consistent with the observed attitudes. This represents an extremely powerful form of generalizability, in which a mechanism, once identified, can be applied to all participants for whom the mechanism is relevant. This example also illustrates how any psychological mechanism proposed to account for group differences can be directly manipulated or measured in the lab in a convenience sample to probe the viability of the mechanism as an explanation (e.g., Galinsky et al., 2024). This is critical work for not just documenting group differences but explaining them.

Equity in Research: Convenience Samples Versus WEIRD Samples

Beyond concerns about generalizability, some writers have expressed criticism that WEIRD samples are unequitable. These authors have pointed out that WEIRD samples have tended to include mostly White participants and that, as a result, members of other racial groups do not benefit equitably from the research (e.g., Clancy & Davis, 2019; Roberts et al., 2020). Obviously, this is a vital concern that must be part of any calculus in designing a research program.

However, on this matter, I would also argue that the extent of the concern varies tremendously, depending on the purpose of the research. For any kind of applied research that aims to directly improve the lives of people, equitable representation is critical. If an intervention is being developed to help people at large, then it must ensure that all members of society benefit equally from the work. For basic research aimed at theory testing, where generalizability is not part of the calculus, this is much less of a concern. Certainly, there are exceptions. If a group is being excluded specifically because researchers expect its members to respond differently, that is clearly a problem. This has been an ongoing issue for biomedical fields that, overwhelmingly, rely on male models (e.g., in mice) for basic research. The primary reason for the exclusion of female models is that they are *expected* to operate differently as a result of having menstrual cycles. Thus, this work intentionally narrows the relevance of the basic research. It would be similarly troubling for a psychological scientist to intentionally exclude women or

Black people from basic research because they were expected to operate differently than men or White people.

Some have indeed argued that the choice of a convenience sample reflects a desire to only study and understand some groups (historically, White, male, and WEIRD) to the exclusion of others (e.g., Clancy & Davis, 2019; Roberts et al., 2020). However, there is a big difference between intentionally excluding groups that are expected to differ and using convenience samples due to the lack of specific reasons to expect important differences across groups. Others have charged that the use of convenience samples reflects an implicit assumption that the sample represents and generalizes to all humans (e.g., Cheon et al., 2020). However, again, there is a big difference between assuming that one's work will apply to all groups because people are all the same and lacking clear reasons to expect specific differences. As described above, basic research makes no claims pro- or contra-generalizability. It is moot with respect to the question. Conclusions support or fail to support a theoretical prediction, which implies nothing about generalizability. The use of convenience samples in these cases is driven not by ideological concerns but, rather, by the desire to maximize resources with the cheapest and easiest samples available.

Let me use my own lab's history as an example. Whenever possible, we use the Intro Psych subject pool because it maximizes the amount of research that we can do with limited time and resources. Prior to arriving at UC Davis, I was a professor at Northwestern University, where 20 years ago, our convenience sample was approximately 67% White, 15% Asian, 5% Black, and 5% Latino. It was approximately 50% female and 50% male. When I arrived at UC Davis in 2005, our convenience sample was approximately 50% White, 20% Asian, 5% Black, and 20% Latino. It was approximately 60% female and 40% male. Currently, the pool at UC Davis is approximately 15% White, 60% Asian, 3% Black, and 20% Latino. It is approximately 70% female and 30% male. At no point in the 30 years across these settings have we concerned ourselves with these numbers. As the convenience pool has skewed increasingly Asian and female, we have never worried that we were not getting enough White or male participants. Simply, whatever hypothesis we are testing, we predict that, if it is correct, it should be true *right here, right now, with these participants*. Have we failed to detect group differences in some of this work? Perhaps. But nothing where we had clear reasons for predicting them. Might others have predicted some differences? Perhaps. Do we claim that our results hold universally? Certainly not. That would require considerable additional work. Occasionally, we have had specific predictions about group differences, and, in those cases, we have powered our samples to detect them. I suspect that our approach to research is representative of how most psychological scientists conduct basic research.

Equity in Research: The Scientific Importance of Diverse Scientists

This again raises the important point that a convenient sample is not the same thing as a WEIRD sample. For basic research, I argue for the use of convenience samples, not for the use of White or WEIRD samples. This means that basic researchers all over the world should use whatever samples are most convenient. This is a question of not just science but of equity. Researchers at small colleges or in countries with fewer research resources should not additionally face the demand of demonstrating the generalizability of their basic research across peoples. No researcher conducting basic research should face that demand, but the burden on those with fewer resources is particularly heavy. Demands that researchers outside of North America or Western Europe provide a WEIRD sample as some sort of "control" condition are entirely inappropriate and should be rejected in the strongest possible terms. The tendency in our field to require the specification of the use of non-WEIRD samples in the titles of research articles is equally odious. The solution to this problem is not to require that the title of every article specifies the nature of the sample. Quite the contrary, the solution, particularly for basic research, is to require no articles to specify the nature of the sample if it is not relevant to the theoretical questions tested within. If a convenience sample is used, it is fully and equally valid, regardless of the nature of the sample.

This approach is crucial for facilitating the globalization of scientific psychology. Specifically, I would argue that encouraging a more diverse population of scientists is much more critical for scientific progress than using diverse samples, certainly where basic research is concerned. Whereas the lack of diverse samples constrains conclusions about the generalizability of research findings, the lack of global diversity among psychological scientists constrains the very questions that are considered and investigated. Scientists from different parts of the world, with different cultures and different experiences, will inevitably produce different ideas about the nature of human behavior. If we are serious about advancing the breadth and diversity of our science, promoting the development of global researchers is the single most important step we can take. Recognizing the importance of convenience samples for all researchers in all locations is vital for these efforts.

Equity in Research: Basic Research Promotes Application

Finally, in terms of equity, there is a strong argument that using the same resources to conduct more basic research with convenience samples versus less research with varied samples will ultimately produce better understanding of group differences and increased applicability. Lewin's (1943) famous maxim that "there is nothing as practical as a good

theory" applies. First, understanding mechanisms helps us identify which groups might respond similarly or differently from one another and why. As described above in the example of self-perception theory, basic research contributed significantly to the recognition and understanding of potential differences across age groups in attitude change mechanisms. Second, strong theories about the operation of basic psychological mechanisms are critical for developing effective practical interventions, including understanding when different interventions are needed for different constituencies. Again, using self-perception theory as an example, basic research suggests that successful interventions aimed at people with less certain attitudes could benefit from a component that targets self-perception processes (e.g., inducing subtle behavior change to affect attitudes). Other basic research suggests that influencing the attitudes and behaviors of people with strongly held attitudes requires an approach that recognizes the full complexity of an issue and presents both sides rather than one side of an argument (Hovland et al., 1953). Lacking theory, interventions are based on intuition or worse, with little understanding of how or why they are effective or not. That increases the likelihood of wasting vital resources on poorly designed interventions and makes it very hard to improve interventions and tailor them for groups known to differ in important ways. Thus, basic research with convenience samples helps to understand people beyond those sampled in the research.

Summary

In this article, I have tried to delineate the circumstances under which sample diversity is and is not critical and why. Sample choice should be based on the goals of the research. When the goal is to develop or test an application, the sample must be representative of the target population of the intervention. By contrast, when the goal is to test a theoretical prediction, convenience samples are sufficient, keeping in mind that no claims should be made as to the generalizability of the findings across peoples. Convenience does not mean WEIRD or White or male. Convenience means convenient, wherever the research is being conducted and in whatever context. The major advantages of convenience samples are that they are cheap and efficient and help maximize knowledge creation, given limited time and resources. In turn, the use of such samples promotes the internationalization of psychological science.

Of course, there are always moderators in psychological research, and the type of group or culture is sometimes a very important one. If a researcher deems generalizability across peoples to be of greatest importance, then they should absolutely test that. However, there are many other important moderators, as well, having to do with motivations, mental states, context, and so forth. Those conducting basic research may be particularly interested in the extent to which key findings generalize across or are moderated by different operationalizations of manipulations and measures. These kinds of generalizations support construct validation, a critical component of theory development. Research is not one size fits all, and researchers know best how to spend their resources and which potential moderators are of greatest importance in their work.

I also have argued here that the mere identification of group differences is not sufficient for empirical psychology, which aims to understand the mechanisms underlying human behavior. Having identified group differences, it is critical to then conduct basic research examining how and why group type moderates some outcome. Basic research using convenience samples can play a powerful role in understanding group differences by delineating the critical mechanisms that produce some outcome. A good grasp of the mechanisms helps identify which groups might be expected to respond similarly or differently and why. This promotes generalization to all groups and people for whom the mechanism applies.

Finally, I argue that the use of convenience samples to conduct basic research is equitable in important ways. First, the use of convenience samples opens up research and publication for scientists with fewer resources, such as those working in a small college setting and those in less wealthy parts of the world. Second, in delineating basic mechanisms, basic research contributes to the development and refinement of effective applied interventions and helps identify when different interventions are needed for different groups of people. The research enterprise works best when basic and applied research work in concert, each informing the other. Indeed, the second part of Lewin's point about the practicality of good theory is that those theories, when relevant, should subsequently be tested in applied research, which must then feed back into theory improvement. Each individual scientist need not divide their attention and resources between basic and applied work. Rather, a healthy division of labor with different researchers doing different kinds of work (including investigating different aspects of generalizability) can be the most efficient state of affairs (Kitcher, 1993), without compelling anyone to use their resources on research that they do not view as a priority.

Conclusion

WEIRD samples limit the generalizability of psychological research. This is particularly concerning when that research is aimed at developing practical interventions to change behavior or help people. Concerns surrounding this issue have come to the fore in empirical psychology, where there is increasing demand to deliver generalizable results via broad or representative sampling. Though the impetus behind this movement is constructive, it is being overapplied to basic research that does not claim generalizability as a goal or outcome. This unnecessarily burdens basic research and slows

the advancement of knowledge, the development of theories that contribute to the development of effective interventions, and, ultimately, the international growth of scientific psychology. In our zeal to make our science more applicable and equitable, we should not lose sight of the critical role of basic science in achieving those goals.

References

American Psychological Association. (2024, August). *Equity, diversity, and inclusion in APA journals*. https://www.apa.org/pubs/authors/equity-diversity-inclusion

Basnight-Brown, D., Janssen, S. M. J., & Thomas, A. K. (2023). Exploration of human cognitive universals and human cognitive diversity. *Memory & Cognition*, *51*(3), 505–508. https://doi.org/10.3758/s13421-023-01410-w

Bem, D. J. (1972). Self-perception theory. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 6, pp. 1–62). Academic Press.

Brewer, M. B. (1988). A dual process model of impression formation. In T. K. Srull & R. S. Wyer (Eds.), *Advances in social cognition* (Vol. 1, pp. 1–36). Lawrence Erlbaum.

Chaiken, S., & Baldwin, M. W. (1981). Affective-cognitive consistency and the effect of salient behavioral information on the self-perception of attitudes. *Journal of Personality and Social Psychology*, *41*(1), 1–12. https://doi.org/10.1037/0022-3514.41.1.1

Cheon, B. K., Melani, I., & Hong, Y. (2020). How USA-Centric is psychology? An archival study of implicit assumptions of generalizability of findings to human nature based on origins of study samples. *Social Psychological and Personality Science*, *11*(7), 928–937. https://doi.org/10.1177/1948550620927269

Clancy, K. B. H., & Davis, J. L. (2019). Soylent is people, and WEIRD is White: Biological anthropology, whiteness, and the limits of the WEIRD. *Annual Review of Anthropology*, *48*(1), 169–186. https://doi.org/10.1146/annurev-anthro-102218-011133

Crandall, C. S., & Sherman, J. W. (2016). On the scientific superiority of conceptual replications for scientific progress. *Journal of Experimental Social Psychology*, *66*, 93–99. https://doi.org/10.1016/j.jesp.2015.10.002

Fiske, S. T., & Neuberg, S. L. (1990). A continuum of impression formation, from category-based to individuating processes. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (pp. 1–74). Academic Press.

Flake, J. K., Pek, J., & Hehman, E. (2017). Construct validation in social and personality research: Current practice and recommendations. *Social Psychological and Personality Science*, *8*(4), 370–378. https://doi.org/10.1177/1948550617693063

Galinsky, A. D., Turek, A., Agarwal, G., Anicich, E. M., Rucker, D. D., Bowles, H. R., Liberman, N., Levin, C., & Magee, J. C. (2024). Are many sex/gender differences really power differences? *PNAS Nexus*, *3*(2), Article pgae025. https://doi.org/10.1093/pnasnexus/pgae025

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

Hovland, C. I., Janis, I. L., & Kelley, H. H. (1953). *Communication and persuasion; psychological studies of opinion change*. Yale University Press.

Kitayama, S. (2017). Journal of Personality and Social Psychology: Attitudes and social cognition [Editorial]. *Journal of Personality and Social Psychology*, *112*(3), 357–360. https://doi.org/10.1037/pspa0000077

Kitcher, P. (1993). *The advancement of science: Science without legend, objectivity without illusions*. Oxford University Press.

Kroupin, I., Davis, H. E., & Henrich, J. (2024). Beyond Newton: Why assumptions of universality are critical to cognitive science, and how to finally move past them. *Psychological Review*. Advance online publication. https://doi.org/10.1037/rev0000480

Lewin, K. (1943). Psychology and the process of group living. *The Journal of Social Psychology*, *17*(1), 113–131. https://doi.org/10.1080/00224545.1943.9712269

Medin, D. L. (2017). Psychological science as a complex system: Report card. *Perspectives on Psychological Science*, *12*(4), 669–674. https://doi.org/10.1177/1745691616687746

Mook, D. G. (1983). In defense of external invalidity. *American Psychologist*, *38*(4), 379–387. https://doi.org/10.1037/0003-066X.38.4.379

National Institutes of Health. (2022, October). *Inclusion of women and minorities as participants in research involving human subjects*. https://grants.nih.gov/policy/inclusion/women-and-minorities.htm

Nielsen, M., Haun, D., Kärtner, J., & Legare, C. H. (2017). The persistent sampling bias in developmental psychology: A call to action. *Journal of Experimental Child Psychology*, *162*, 31–38. https://doi.org/10.1016/j.jecp.2017.04.017

Rad, M. S., Martingano, A. J., & Ginges, J. (2018). Toward a psychology of Homo sapiens: Making psychological science more representative of the human population. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(45), 11401–11405. https://doi.org/10.1073/pnas.1721165115

Roberts, S. O., Bareket-Shavit, C., Dollins, F. A., Goldie, P. D., & Mortenson, E. (2020). Racial inequality in psychological research: Trends of the past and recommendations for the future. *Perspectives on Psychological Science*, *15*(6), 1295–1309. https://doi.org/10.1177/1745691620927709

Sears, D. O. (1986). College sophomores in the laboratory: Influences of a narrow data base on social psychology's view of human nature. *Journal of Personality and Social Psychology*, *51*(3), 515–530. https://doi.org/10.1037/0022-3514.51.3.515

Simons, D. J., Shoda, Y., & Lindsay, D. S. (2017). Constraints on Generality (COG): A proposed addition to all empirical papers. *Perspectives on Psychological Science*, *12*(6), 1123–1128. https://doi.org/10.1177/1745691617708630

Taylor, S. E. (1975). On inferring one's attitudes from one's behavior: Some delimiting conditions. *Journal of Personality and Social Psychology*, *31*(1), 126–131. https://doi.org/10.1037/h0076246

Thalmayer, A. G., Toscanelli, C., & Arnett, J. J. (2021). The neglected 95% revisited: Is American psychology becoming less American? *American Psychologist*, *76*(1), 116–129. https://doi.org/10.1037/amp0000622

Towards a global psychological science. (2022). *Nature Reviews Psychology*, *1*, Article 369. https://doi.org/10.1038/s44159-022-00087-3

Vazire, S. (2024). The next chapter for Psychological Science. *Psychological Science*, *35*(7), 703–707. https://doi.org/10.1177/09567976231221558

Wood, W. (1982). Retrieval of attitude-relevant information from memory: Effects on susceptibility to persuasion and on intrinsic motivation. *Journal of Personality and Social Psychology*, *42*(5), 798–810. https://doi.org/10.1037/0022-3514.42.5.798

Yarkoni, T. (2022). The generalizability crisis. *Behavioral and Brain Sciences*, *45*, Article e1. https://doi.org/10.1017/S0140525X20001685

Received June 20, 2024
Revision received August 29, 2024
Accepted September 4, 2024 ■