COLLOQUIUM PAPER

Should social scientists be distanced from or engaged with the people they study?

Kalonji Nzinga[a], David N. Rapp[b], Christopher Leatherwood[a], Matthew Easterday[a], Leoandra Onnie Rogers[b], Natalie Gallagher[b], and Douglas L. Medin[b,1]

[a]School of Education and Social Policy, Northwestern University, Evanston, IL 60208-2610; and [b]Department of Psychology, Northwestern University, Evanston, IL 60208-2714

Edited by Jennifer A. Richeson, Yale University, New Haven, CT, and approved July 6, 2018 (received for review January 30, 2018)

This commentary focuses on two important contrasts in the behavioral sciences: (*i*) default versus nondefault study populations, where default samples have been used disproportionately (for psychology, the default is undergraduates at major research universities), and (*ii*) the adoption of a distant versus close (engaged) attitude toward study samples. Previous research has shown a strong correlation between these contrasts, where default samples and distant perspectives are the norm. Distancing is sometimes seen as necessary for objectivity, and an engaged orientation is sometimes criticized as biased, advocacy research, especially if the researcher shares a social group membership with the study population (e.g., a black male researcher studying black male students). The lack of diversity in study samples has been paralleled by a lack of diversity in the researchers themselves. The salience of default samples and distancing in prior research creates potential (and presumed) risk factors for engaged research with nondefault samples. However, a distant perspective poses risks as well, and particularly so for research with nondefault populations. We suggest that engaged research can usefully encourage attention to the study context and taking the perspective of study samples, both of which are good research practices. More broadly, we argue that social and educational sciences need skepticism, interestedness, and engagement, not distancing. Fostering an engaged perspective in research may also foster a more diverse population of social scientists.

diversity | generalizability | engagement | insider bias | advocacy research

Our commentary is based on experiences with the following kind of story. A Latina graduate student wants to do research in a Latino/Latina community. She may have a variety of reasons in mind, ranging from wanting to use her Spanish language skills, to having the sense that Latinx perspectives are underrepresented in science, to wanting to "give back" to her community. When she tells her advisor of this interest, the feedback she receives is "You know, it's really important to avoid bias in research" or "Be careful, you don't want to be accused of advocacy research." This scenario, where claims of "insider bias" are levied at the notion of researchers from societally marginalized communities studying their own community, is one that we authors have heard conveyed by numerous colleagues and students. Our starting point for interrogating this scenario was the recurrence of this story in our own academic communities, primarily in the psychological and educational sciences. However, the phenomenon has also been documented across the social sciences; consider, for example, a project by Hendrix (1), a black female scholar from the field of communications, who wrote about her personal experiences, presenting her research and then receiving the question "Did being black introduce bias into your study?" She also surveyed other researchers of color and found that her experience was common rather than unique.

Would a white (or even Latinx) graduate student be cautioned in the same way if he or she wanted to focus on the undergraduates on his or her campus (which is usually majority white) or work in a (heavily) European American community? We seriously doubt that similar cautions would be expressed, given that European American majority communities are the typical default population. Again, this is not a case specific to a particular underrepresented community: A similar scenario could be drawn for a female scholar wanting to study women (especially if she studies what gets stereotyped as "women's issues") or a lesbian gay bisexual transgender or queer (LGBTQ+) scholar wanting to study an LGBTQ+ community. In other words, any departure from the study of "the norm" (white, middle-class heterosexual) can raise questions that do not arise when (Eurocentric) norms are followed. This skepticism seems to be particularly evident when the researcher shares aspects of identity with his or her study participants who have been traditionally stigmatized (2).

The Latina graduate student in our scenario has a potentially strong case to make for doing research within the Latinx community. Her advisors nonetheless caution that she may be "too close" to her Latinx participants. Her advisors also may be worried that her research will be less theoretically motivated and less informed by previous research than it should be. At the core of the advisors' concern is the assumption that good social science research requires a disinterested, distanced perspective that builds systematically on previous research. However, it may reveal the way race and white privilege have shaped and continue to shape social science. As a result, research that centers on an ethnic minority group, or on other nondefault populations, raises concerns that it may be "less than" (less rigorous, less objective, and/or less scientific than) research with the default population. For purposes of our argument, it is time to take a look at that previous research, in part, because it reveals its own form of bias.

Historical Background: Dominance of Default Populations

Henrich et al. (3) noted that social and behavioral research overwhelmingly relies on WEIRD (Western, educated, industrialized, rich, and democratic) samples, samples that they argue are quite literally weird in that they are especially unrepresentative of the world at large. A "heat map" showing the locations of major research universities and who gets studied by them, we hypothesize, would reveal steep mountains and broad, flat plains as the primary contours. The underrepresented groups include communities that have traditionally been minoritized or

This paper results from the Arthur M. Sackler Colloquium of the National Academy of Sciences, "Pressing Questions in the Study of Psychological and Behavioral Diversity," held September 7–9, 2017, at the Arnold and Mabel Beckman Center of the National Academies of Sciences and Engineering in Irvine, CA. The complete program and video recordings of most presentations are available on the NAS website at www.nasonline.org/pressing-questions-in-diversity.

Author contributions: K.N., D.N.R., C.L., M.E., L.O.R., N.G., and D.L.M. wrote the paper; and D.N.R. contributed brainstorming and conceptualizing of ideas.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence should be addressed. Email: medin@northwestern.edu.

Published online November 5, 2018.

absent in WEIRD geocultural spaces, including Native Americans, African American peoples, Latin American peoples, Middle Eastern communities, communities without traditions of Western schooling, the rural poor, and the urban poor. For that matter, the samples drawn from WEIRD populations are not at all representative even of WEIRD people. Instead, they favor select age, educational, income, geographical, and occupational groups. (Note that the "W" in WEIRD fits with white as well as it fits with Western.)

With respect to who gets studied, there is a sense in which all social and educational sciences research necessarily favors some groups over others. If a researcher is studying how to enrich the experience of museum-goers, he or she is necessarily studying people with more education, greater wealth, and greater mobility than people who do not go to museums. Of course, this does not mean that research on museum-goers should be avoided or that any one study is necessarily biased because it fails to recruit a random sample of the world's population. What matters more is the broader profile of a field's research, such as the NIH's portfolio of research covering in aggregate the ecology and severity of diseases. When a disease or community is neglected in NIH research, we see societal pressure for scientific representativeness, as in the initial lack of attention to the AIDS crisis. Social science is largely missing this pressure.

When it comes to this sort of aggregate assessment, psychology's long-standing overreliance on college undergraduates as research participants is particularly striking (4). It means that the field will learn more about how college students think, feel, and act than about how other adults do. Since college students in America are disproportionately middle class, white, and 18–22 y of age, this is the population about which psychology knows the most. Reviews reveal that the situation in developmental research is no better (5) and the negative policy consequences are even more significant (6). Similar concerns can be levied at educational research. Research drawing on middle-class schools close to research campuses may be unrepresentative of schools more broadly and may encourage neglect of issues that are salient elsewhere. Although nothing intrinsic to the social and education sciences favors or supports this sort of "bias" or selectivity in study populations, it is endemic and persists with far too little consideration. This is presumably due, in part, to pragmatic concerns about the feasibility of conducting such work, as well as the strong historical precedent for how research has been done structurally at local and global scales.

These pragmatic and historical considerations, of course, are not intrinsic to the pursuit of rigorous and informative science. However, they appear to reveal either the implicit assumption that the default population is more important than other samples or a surprising lack of curiosity about the generalizability of findings across samples. With this in mind, we pose the following critical question for social scientists to ponder: To what extent is the insider bias of people who have been long overrepresented in the field of psychology (WEIRD researchers studying WEIRD participants) similar or dissimilar to the potential insider bias of nondominant researchers who study their own communities?

Defaults and Selection Biases. If researchers primarily study people who are like themselves, then researcher intuitions may be a good guide to developing stimuli, methods, and appropriate contexts for their studies. Furthermore, the very accessibility of research participants allows for pilot studies and the further development and selection of "what works." It also means that borrowing stimuli and procedures from published studies on WEIRD samples is likely to be most effective for new studies on WEIRD samples. The disproportionate investment in WEIRD samples means researchers have more opportunity to learn from the findings of other researchers and that research methods are likely to converge on effective and efficient measures, measures that are tailored to WEIRD samples and perhaps less apt for non-WEIRD samples. Of course, given these strong selection factors, the stimuli may be anything but a random and representative sample from some domain of interest; consequently, the findings may be less likely to generalize to other stimuli, procedures, and study populations (7).

Is Engaged Research Intrinsically Biased? Are engagement and interest incompatible with good science? One important role for engaged research is in formulating research questions and then translating them into rigorous research programs. It is important to note that the social and educational sciences are not restricted to describing the world as it is but extend to examining what is possible under appropriate implementation.

For example, consider a project in which a researcher seeks to determine the potential benefits of a technological innovation for educational settings. If that researcher is already convinced that the use of technology is the key to better educational outcomes, this need not mean that the work that researcher is doing is necessarily biased or involves questionable research practices. Indeed, if this "interest" disqualified work along these lines, a substantial portion of educational research in computer science, educational psychology, the learning sciences, cognitive psychology, curriculum and instruction, and other related fields would need to be stricken from the research record. Rather, many of these researchers presumably hold the belief that technology can be a valuable tool under certain learning conditions, conditions that the researcher aims to identify.

An analogous interest is related to many researchers', practitioners', and laypersons' beliefs concerning the potential benefits of diversity. Consider research by Gurin and coworkers (8–10) on the educational benefits of campus diversity. While they held the view that distinct benefits could and would be realized, their approach and findings were much more nuanced. Their work showed that it is not that diversity guarantees success but, rather, that positive results of diversity critically depend on (designed) curricular and cocurricular experiences with diverse peers. Their hypothesis concerning the potential benefits of campus diversity led to the careful design of curricula and the application of empirical methods rather than wishful thinking and polemic.

This example demonstrates that an engaged stance is not incompatible with good science. A scholar who wishes to help his or her community is invested in making that support a reality and exploring possibilities for doing so. Given that goal, producing biased research with false-positive outcomes is an exercise in self-deception, and attempts to generalize and apply the findings elsewhere would fail. We now turn to engaged research where the study populations themselves are focal.

Engaged Research and Nondefault Samples. Engaged research with non-WEIRD study populations draws on samples that are less conveniently accessible. It employs methods and measures that have had fewer opportunities for selection effects and less opportunity for convergence on the most efficient and effective stimuli, methods, and contexts. Engaged research is therefore more likely to be in exploratory stages with respect to these factors. This difference from research with default samples has mixed consequences. As we have seen, selection may lead to unrepresentative stimuli and settings as well as procedures tailored to default samples. If the Latina in our scenario has concerns about the advisability of directly "building" on prior research by adopting methods that have not been selected for her study population, her worries may be well justified.

Perhaps more significant is the fact that research with nondefault populations is outside the mainstream. Because of the American academy's history of excluding the research and the participation of scholars of color throughout the 19th and 20th centuries, engaged researchers (specifically researchers of color studying non-WEIRD populations) are often likely to draw on

research theories, epistemologies, and methodologies that were developed by previous generations of non-WEIRD scholars who operated on the margins of the academic mainstream. For example, the Association of Black Psychologists (ABPsi) is a professional association of African American psychologists and researchers founded in 1968 during a time when many American universities were quite hostile to black students and professors, and were just beginning to racially desegregate their campuses (11). The founders of the ABPsi made an intentional decision to remain independent from the American Psychological Association, and many of the scholars, although trained in some of the most prestigious predominantly white American universities, decided to join faculties at distinguished historically black colleges and universities like Howard University. They were wary of joining institutions with a tradition of supporting research theories with racist logics (e.g., strong deficit models, eugenics).

The ABPsi thus represents a community of engaged scholars who have been developing empirically tested methods, theories, and instruments for researching African American communities for at least the past 50 y. Many researchers at predominantly white institutions still view these theories as less credible since they were published in journals with low-impact factors, including outlets like ABPsi's *Journal of Black Psychology*.

The question of credibility invoked here is not merely about the research process itself but about the power structures in which that research takes place, in which scientists decide who and what counts as high-quality research. We suggest that research that challenges the status quo or questions the existing "truths" (theories and approaches) of the field faces disproportional critique with respect to its quality. Those scholars who are intentionally choosing research theories and methodologies that have been separate (and, in some cases, blatantly excluded) from the intellectual communities and rely on language, concepts, and theories that are "foreign" to mainstream scholarly communities often get inappropriately flagged as biased.

This tendency for scientists to scrutinize and disqualify research that defies the norms of the mainstream paradigm is related to social processes that Kuhn (12) identified when he studied stasis and change in scientific disciplines. We suggest that the current research paradigm involves primarily WEIRD researchers using WEIRD participants (and protocols developed for WEIRD samples) as the basis for (presumably universal) generalizations. Consequently, approaches that start elsewhere are often viewed with suspicion.

Summary. The comparison of distant versus engaged perspectives is inevitably confounded with WEIRD versus non-WEIRD samples, vastly unequal previous investment in them, and (often implicit) default assumptions that derive from fields dominated by WEIRD researchers. The diversity of researchers and the researched is intertwined, and calls for increased diversity of scholars and scholarship (13, 14) face research environments dominated by default (WEIRD) perspectives and values, including the perspective that objectivity requires distance. It is important to bear these factors in mind in considering how engagement may interact with risk factors and virtues in research. In what follows, we provide our consensual and preliminary assessment of specific risks associated with engaged versus disinterested research.

Engagement, Disinterest, and Risk Factors in Research

Generalized accusations of real or potential bias do not take into account the fact that a number of specific kinds of research biases and questionable research practices have been identified and discussed in the literature, going back at least to Campbell and Stanley's classic analysis of various threats to the validity and generalizability of research (15), which remains relevant today (16). In addition, the considerable literature associated with what has been called "replication crises" has called attention to a number of questionable research practices, including selective reporting of results, unreported researcher degrees of freedom, and hypothesizing after the results are known (17). Finally, there is a robust literature on biases associated with comparisons of two or more study populations (e.g., refs. 7, 18) when one of them is the default (WEIRD) sample.

With respect to these biases, we raise the issue of whether they are more likely to appear from a disinterested perspective or an engaged perspective, or, equally likely, are independent of perspective. To foreshadow, our analysis suggests limitations of each perspective, and it is well to bear those in mind rather than to adopt a blanket dismissal or endorsement of either. Furthermore, we identify some virtues of engaged research that may be useful in all social science research.

The social and educational sciences are extremely diverse with respect to research methodologies employed both within and across disciplines. As a consequence, the aptness of the factors we describe will also vary (e.g., concerns about how statistical tests are done obviously are not relevant if no statistics are done) within and across disciplines. Nonetheless, we believe that most of the factors discussed have fairly broad applicability.

For simplicity, we write as if engaged versus distant stances were a categorical distinction, even while acknowledging that degree of engagement is a continuous variable. Note also that researchers inevitably rely on their own intuitions to some degree such that even the most distanced perspective is likely to favor research participants who are like the researcher and sometimes disfavor those who are not.

Challenges and Risk Factors Linked to Engaged Research. Some limitations associated with engaged research derive from the historical and ongoing underinvestment in nondefault study populations. For example, engaged research is often in a more exploratory phase, which can limit the robustness of statistically significant results (19, 20). In that sense, the risk factors only become questionable research practices when exploratory studies are judged by the standards of more established research. Other limitations are associated with the need to preserve a meaningful researcher and participant relationship. Here, we provide a list of our considerations along these lines:

i) Data mining: One questionable research practice involves conducting statistical tests and analyses that have not been preplanned and failing to adjust for multiple tests that go unreported. It typically takes more effort to do research with advocacy populations, and it is possible researchers will probe further into their data and consider a wider range of measures. Arguably, this is what they should do. In exploratory research, it may be difficult or impossible to specify all potentially relevant dependent variables in advance. For example, one notable result reported by Unsworth et al. (21) in their examination of children's ecological knowledge was cultural differences that emerged in their spontaneous imitations of sounds of animals included in probes. This observation is intriguing but was not preplanned. What may be a virtue for exploratory research is nonetheless a serious risk factor (given, for example, that it might enable so-called "fishing expeditions"), so it is important to fully report analyses; apply extra caution concerning generalizability; and, whenever feasible, include systematic replication.

ii) Conducting underpowered studies: The issue of limited numbers and accessibility of samples for engaged research means that studies often will be underpowered. Low power is associated with difficulty with replication. At major research universities or with online MTurk studies, samples of 100 or more participants per condition may be routine and feasible to collect in a few days or weeks compared with

many months or even years for a population in which there is an engaged interest. This challenge is common in cultural research or when the sample of interest is small (e.g., children with Williams' syndrome). Low power is often an inevitable reality, but engaged researchers need to bear this problem in mind in discussing the scope of their findings and the claims to be derived from them (22).

iii) Selective reporting of studies and file drawer problems: Given null results, engaged researchers may be more likely to reflect on how their methods and measures could be improved and to treat early studies as pilot studies. In addition, the question of when and why piloting is stopped, and when "real" data collection has started, can be fuzzy. On the other hand, when sample size is limited and data collection is difficult, researchers may be more likely to report their data, hopefully noting their limitations, rather than to discard the data in favor of a more pristine design.

iv) Hawthorne effect and reactive measures: It is important to separate effects of independent variables on performance from results driven simply by receiving attention from researchers (the latter has been dubbed the "Hawthorne effect," named after a study of worker performance in a Western Electric factory in the Hawthorne suburb of Chicago). Given the dominance of WEIRD samples in research, non-WEIRD samples may find being in a study to be a novel experience and exert more effort, pay more attention, and be more concerned about looking good (This presumes that researchers have taken steps to build rapport and trust; without these steps, participants may react with indifference). In the same way, researcher enthusiasm may elicit more effort and attention from non-WEIRD participants. In contrast, default research participants (i.e., undergraduates completing studies for course credit, extra credit, or pay; typical MTurk workers have been participants in hundreds of studies) may be only engaged enough to complete the study requirements. In that sense, the latter populations may be evidencing a reverse Hawthorne effect, so the issue of effort and attention cuts both ways.

v) Protecting community by selective use of dependent variables: Close engagement with a study population may bias researchers to avoid anything that might tend to make the population look bad. For example, the Nisbett and Cohen (23) *Culture of Honor* monograph examined the question of why US murder rates are higher in the South than in the North. Had the authors deeply identified with the South, they might have instead focused on the issue of why military service to country is higher in the South than in the North and ignored murder rates. This can also relate to concerns about developing and testing for reasonable alternative hypotheses, which should necessarily be part of any research project (24). Some engaged projects might fail to consider such reasonable alternatives, or even to generate them, if the focus is on supporting the invested groups and engagements.

Risk Factors Linked to Disinterested Research. Some limitations of disinterested research derive from the unquestioned assumption that what one is studying is a human universal that should emerge in any population. Others arise from a reliance on prior research primarily conducted in WEIRD populations, and failure to recognize how this qualifies theories and methodology. We outline some of these issues below:

i) Inattention to description of study populations: It is important for descriptions of participant samples to include factors that may be relevant to the pattern of results. However, disinterested research may tend to neglect this aspect of research, sometimes even failing to describe the participant sample at all (25). Even statements such as "participants were college students at a large Midwestern university" ignores research suggesting that social science majors may reason differently from natural science or humanities majors (e.g., ref. 26) or that decision-making processes may differ between psychology and economics majors (27). The very fact that nondefault samples are departures from the norm means that descriptions of them are more elaborate and understood as potentially relevant to understanding variability, while descriptions of default populations are less likely to include details that highlight the nuances of social context.

ii) Overgeneralization of research findings (both explicitly and implicitly): One stance on generalizability is to assume that any effects observed are universal until proven otherwise. Henrich et al. (3) suggest that this stance is unrealistic, and others have noted that when the universalist assumption is translated into policy, it is deeply problematic [e.g., Keller (28)]. Although the replication crisis in the social sciences has led to a focus on "exact replications," potential limitations on generalizability are related to study samples, measures, contexts, and procedures (29). Interestingly, the OpenScience project focuses on exact replications (30–34), but implicitly assumes that "exact" is not affected by nationality, language, or other demographic variability in study samples (35). One corrective action for any tendency to overgeneralize is to describe results using past tense rather than present tense, as in "mothers spoke" rather than "mothers speak" (36).

iii) Inattention to the social context of research and positionality: Despite classic research on experimenter expectancy effects by Rosenthal and coworkers (37–39), the social context of research is often ignored as if the experimenter were invisible. It is now uncommon for methods sections of research papers to describe the experimenters who collected the data or to mention whether they were blinded to the hypotheses guiding the research. One obvious counterexample to this trend is the recent upsurge of interest in the role of stereotype threat in participant performance (40). Notably, this work has been driven by an engaged perspective (41).

iv) Insensitivity to selection effects from intuitions; evolution of stimuli, methods, and materials; and focus on efficiency: Given the dominance of nonadvocacy research and the tendency to adopt or reuse materials "that work," it is likely that effect sizes will be misleadingly large since researchers are incentivized to design studies to produce effects. Competitive pressures may encourage convenience sampling and efficiency in data collection, particularly for pretenure scholars, and this is a serious risk for selection effects in research.

v) Insensitivity to default assumptions in cross-group comparisons of WEIRD and non-WEIRD samples: The hazards associated with having a default population as one of the groups in cross-group comparisons are considerable, and we provide a partial list here: (*i*) use of measures selected and normed for WEIRD samples as the standard for comparison (7); (*ii*) use of outdated, inappropriate norms [e.g., until only the past few years, reliance on the word frequency norms of Kučera and Francis (42)] that have been developed using restricted groups of participants or that are too old to accommodate contemporary considerations of linguistic diversity and variability; and (*iii*) insensitivity to alternative measures and intuitions that may be appropriate for non-WEIRD samples. These biases will tend to favor WEIRD over non-WEIRD samples and encourage a deficit orientation toward non-WEIRD samples.

vi) Default samples leading to asymmetrical attention to groups: When researchers describe findings with nondefault samples (e.g., African American), it is common to ask how these results compare with those from default samples (e.g.,

European American), but the reverse does not hold when results from default samples are described.

vii) Focus on the average/assumption of homogeneity: Distanced research tends to be characterized by a focus on central tendencies. This can lead to a blind spot as to whether default populations can be meaningfully subdivided. A good practice is to explicitly test for within-sample variation rather than just assuming that samples are homogeneous (e.g., as in ref. 43; also ref. 44).

Summary of Risk Factors. Our analysis suggests that both "disinterested" and "engaged" perspectives on study populations have associated risk factors that constitute questionable research practices in some cases. Many of the limitations associated with a disinterested perspective would be addressed by employing more diverse study samples. In the same way, the primary risk factors for an engaged perspective derive from the strong correlation between the use of non-WEIRD samples and more exploratory stages of research. Again, a more balanced investment in study samples would address many of the potential limitations of an engaged perspective.

So far, our focus has been on risk factors and potential limitations of an engaged research perspective with respect to study samples. We have shifted from generalized accusations of bias associated with advocacy research to specific risk factors associated with engaged versus disinterested perspectives. This, however, neglects potential virtues associated with engagement, and we turn to them now.

Virtues of Engagement

Consider the following contrasting scenarios:

> In one case, a researcher reviews the literature, identifies a "gap" in the knowledge, and designs a study to test a novel hypothesis. A sample is recruited from the psychology subject pool or via MTurk, and standard measures are used based on prior literature. The researcher collects the data following appropriate design procedures, analyzes the data, and writes up the results for publication in a peer-review journal.
>
> In an alternative scenario, our Latina graduate student researcher identifies a community in which to conduct a new study. She does background research on the community and location, and meets with members of the community to identify key partners and contacts to carry out the work and discuss effective strategies for recruiting community members to participate in the study. She conducts small focus groups and pilot tests the materials and survey instruments to ensure they are valid and appropriate for the population. To build trust and rapport, she also volunteers once per week at a local community center. The research is carried out rigorously, with attention to systematic protocols, control groups, and design. The results are analyzed and presented to community members for discussion. Based on this work, she submits a manuscript for publication in scholarly outlets.

The former description sounds more similar to traditional notions of "basic science," whereas the latter calls forth what we have referred to as engaged research. Both researchers carried out an empirical study of human behavior following the standards of social science, but from different starting points: The first is what we refer to in this paper as "distanced" research, and the second is an example of engaged research. We suggest that rather than a subtype of research, the qualities that often delineate engaged research from distanced research may be reframed as useful and effective methodological approaches for a wide array of social science inquiries.

Why might research, for example, that considers the culture and context of the sample, builds genuine relationships and rapport with participants, and ensures the ecological validity of measures and research activities be viewed with a skeptical eye? Might not all behavioral science research benefit from such engaged practices? Why has research by minority scholars with minority populations, particularly in minority-serving spaces (e.g., communities, churches, schools), been delineated as both more engaged and less scientific than research with dominant populations conducted by majority-group scholars? Why is it that research with and by "majority"-group populations is not equally held to the standards of humane, equitable, engaged research practices? Much in the way that Ladson-Billings (45) reframed "culturally relevant teaching" for communities of color as "just good teaching," we propose that systematic, empirical science that is responsive to communities, policies, cultures, and contexts is "just good research." Rather than a unidirectional comparison whereby the "deficiencies" of engaged research are compared with the "rigors" of distanced research, we consider how a more engaged paradigm in the social sciences can strengthen the quality, rigor, and impact of all research. We believe that there are compelling reasons for researchers to take an engaged stance on their study populations (even when the study samples are WEIRD and even while bearing risk factors in mind).

Building Relationships and Rapport, Diminished Stereotype Threat. As stated earlier, a dispassionate researcher often relies on the fact that participants will be conveniently available. Mandatory participation of students enrolled in "Introduction to Psychology" courses in research studies means that participants will be involved despite the fact that the researchers may have made no personal connection with them beforehand. Engaged researchers, on the other hand, usually have to venture off campus, out of the WEIRD campus community, spending a significant amount of time building relationships with a local community, school, or organization to gain access to participants. During this process of relationship building, an impassioned stance is often a criterion for entry into the community. Institutional gatekeepers to the community (and potential participants themselves) are often quite skeptical of researchers who do not show an explicit advocacy for the population in question, given that many non-WEIRD communities are vulnerable or have been exploited by researchers in the past (46, 47). This tendency toward relationship building with participants can be an asset when engaging in certain methodologies, for example, building rapport during a clinical interview. While a dispassionate interviewer may be better at avoiding acquiescence bias or social desirability bias, the engaged interviewer is often read by informants as more authentic and trustworthy, which encourages them to open up, thereby providing more complete and accurate data.

Awareness of Positionality. As illustrated by the earlier example of the Latina graduate student cautioned by her advisor, engaged researchers are often asked to think about how their identity might affect their perception of the research question and data. Since the Latina student is likely to be a member of a student cohort in which very few of her classmates are asked this question (especially with regard to race), it will feel like unfair scrutiny. This scrutiny, although it is certainly doled out inequitably, may lead engaged researchers to a deep examination of the epistemology of their work. They cannot help but ask "How do my life experiences bias my epistemology in a way that makes me see a research question in a certain way?"

There are some areas of social science, anthropology in particular, that include this kind of self-questioning in research training. Perhaps this question should be a prerequisite for other students as well: "How does the fact that I am a college-educated white male psychology researcher in an American college town

affect the way that I would approach answering this research question?" Asking this question would be helpful for a field dominated by this demographic, even if the only effect were to encourage them to consistently question their tendency to default to WEIRD undergraduate participants and to restrict themselves to research methodologies and theories created in the majority-culture West.

Taking Laboratory-to-Field Steps Seriously. One of the most enduring problems in social science research is the fact that experimental results achieved in a controlled laboratory setting often do not translate to the messiness of the environments that humans actually inhabit (16). However, engaged researchers can be very integrated in the community that they are studying in a way that encourages attention to laboratory-to-field application. If they are collecting data from a community center, they are necessarily spending time building relationships there, volunteering their time as an instructor, or conducting extended ethnographic observations. They know some of the potential beneficiaries of the research by name. Therefore, the process of thinking about how research results can be deployed in the field is ongoing. It is never an abstract mental exercise, and it is seen as a core component of the work rather than a possible extension. This concern of applicability to real-life settings happens during the design and implementation of the study. In fact, researchers are often thinking about the research study itself doubling as an activity that is useful for the participants involved. If you are deeply impassioned to produce research that can be directly applied within your population of interest, the design of the research is executed with that goal in mind. All researchers can also learn from this by developing field studies as part of the initial stages of research rather than follow-ups once in-laboratory phenomena have been demonstrated. In general, laboratory-based research would benefit not only from attention to internal validity but also from greater attention to external validity (35).

Awareness of Audience and Politics/Policy. Engaged researchers are often informed by a deep interest in the political implications of their research on social policy. For example, how will research on the environmental practices of Native Americans inform US policy on native land rights? As another example, how will research on African American language practices influence political discourse about addressing achievement gaps or opportunity gaps? While a disinterested stance is often associated with political neutrality and abstaining from policy discussions, such research becomes implicated in policy nonetheless, if only in the form of preserving the status quo.

Engaged researchers are deeply informed by the political discourse related to their subject matter. They are likely more cognizant about how a particular research finding relates to the current political arguments about that phenomenon (48). They may better anticipate how their findings will be used by various political agendas, and thus know how to couch their arguments to prevent misuse and distortion of findings. All researchers can benefit from paying attention to the sociopolitical context of their research, especially as part of disseminating research to the public or to policy audiences.

The science of science communication needs to include attention to potential policy implications (and, in some cases, potential misunderstandings). With that said, we also acknowledge the benefits of showing open-mindedness and intellectual independence from any one political party or agenda. This allows for exposure to and consideration of effective policy solutions that do not coincide with theories of social change that are currently in vogue in one party or the other. Sociopolitical independence is also useful to effectively and credibly communicate research to an audience across the political spectrum. This tactfulness should not be confused with being distant from the sociopolitical issues that affect the communities we study and making explicit value commitments, such as human rights and social justice.

Discussion

The "OpenScience" initiative has led to a healthy reexamination of social science research practices. It has illuminated a number of questionable research practices and fostered changes in practices aimed at establishing robust empirical findings. To date, it has primarily focused on replicability and avoiding practices that lead to unreliable results. The present commentary can be seen as an extension of this movement, one that takes up issues of generalizability and diversity, especially across study samples, as well as biases associated with cross-group comparisons.

With respect to generality over study samples, there are distinct challenges growing out of the historical (disproportionate) focus on WEIRD samples, which has led to such samples being the "default" rather than one of a wide range of samples on which to build a science. Departure from a default captures attention, and in the case of scholars of color who wish to study people from their own ethnic and racial groups, this attention has often taken the form of the belief that the research may be compromised by the researchers' closeness to those being researched. A parallel argument exists for researchers who are members of other marginalized groups, for instance, LGBTQ+ individuals, women, or people from low-income backgrounds.

As American society, in general, and social science, in particular, continue the process of addressing human diversity, it is important that we understand we are not just socializing black and brown students and faculty into "the mainstream," but that we will have to reconcile multiple distinct traditions of intellectual inquiry [e.g., Western science, Native science (49)] that have remained segregated due to political and social processes. Both the history of segregation and racism in the academy and the correlated disproportionate investment of research in default study populations have important consequences for research practices, research environments, and perceptions of bias.

Our goal has been to move beyond vague accusations of (possible) bias and the idea that engaged research is necessarily biased by examining specific research practices and asking whether or not they are associated with an engaged perspective. With respect to potential risk factors, our analysis suggests a mixed picture, one that tends to undermine favoring a distanced perspective over engagement. Some risk factors may be linked to an engaged perspective, but others are associated with a distanced one. Furthermore, we have identified distinct virtues associated with engagement, virtues that might also benefit research with WEIRD samples.

In retrospect, it is something of a puzzle that psychological research, in particular, has so often ignored the very people it studies. A former colleague of one of the authors stated, "I only want to study what is universal." Indeed, if some phenomenon is universal, study samples do not matter. Equally obvious, however, one cannot establish that some finding is universal without evidence. Arguably, science consists of a search for patterned variability that it seeks to understand and explain. A willful neglect of study sample variability represents an abdication of a fundamental moral and scientific principle.

Merton (50) argued that effective science is organized around norms that include (*i*) universalism (applying the same standards of evidence for claims, regardless of who is making them), (*ii*) disinterestedness (avoiding ideological or political temptations to distort the truth), and (*iii*) skepticism (creation of accountability systems dedicated to evenhanded norm enforcement). His first point should provide encouragement for non-WEIRD researchers, but where he suggests we need distancing and skepticism, we suggest that the social and educational sciences need skepticism, interestedness, and engagement, not distancing.

ACKNOWLEDGMENTS. We thank Monnica Williams for her cogent comments and suggestions. This work was supported, in part, by National Science Foundation Grants DRL 1713368, BCS 1623555, and BCS 1647219 (to D.L.M.); Air Force Office of Scientific Research Grant FA9550-14-1-0030 (to D.L.M.); and National Science Foundation Graduate Research Fellowship 1000234799 (to N.G.).

1. Hendrix KG (2002) "Did being black introduce bias into your study?": Attempting to mute the race-related research of black scholars. *Howard J Commun* 13:153–171.
2. Banks JA (1998) The lives and values of researchers: Implications for educating citizens in a multicultural society. *Educ Res* 27:4–17.
3. Henrich J, Heine SJ, Norenzayan A (2010) Most people are not WEIRD. *Nature* 466:29.
4. Arnett JJ (2008) The neglected 95%, a challenge to psychology's philosophy of science. *Am Psychol* 64:571–574.
5. Nielsen M, Haun D, Kärtner J, Legare CH (2017) The persistent sampling bias in developmental psychology: A call to action. *J Exp Child Psychol* 162:31–38.
6. Keller EF (2017) Climate science, truth, and democracy. *Stud Hist Philos Biol Biomed Sci* 64:106–122.
7. Medin D, Bennis W, Chandler M (2010) Culture and the home-field disadvantage. *Perspect Psychol Sci* 5:708–713.
8. Gurin P (1999) Selections from the compelling need for diversity in higher education, expert reports in defense of the University of Michigan. *Equity Excel Educ* 32:36–62.
9. Gurin P, Dey E, Hurtado S, Gurin G (2002) Diversity and higher education: Theory and impact on educational outcomes. *Harv Educ Rev* 72:330–367.
10. Gurin P, Nagda BRA, Lopez GE (2004) The benefits of diversity in education for democratic citizenship. *J Soc Issues* 60:17–34.
11. Williams RL (2008) A 40-year history of the Association of Black Psychologists (ABPsi). *J Black Psychol* 34:249–260.
12. Kuhn TS (1962) *The Structure of Scientific Revolutions* (Univ of Chicago Press, Chicago).
13. Stewart D-L (2017) Remembering alma mater. *Black Collegians' Experiences in US Northern Private Colleges* (Springer Nature, New York), pp 183–192.
14. Medin D, Ojalehto B, Marin A, Bang M (2017) Systems of (non-) diversity. *Nat Hum Behav* 1:0088.
15. Campbell DT, Stanley JC (1966) Experimental and quasi-experimental designs for research. *Handbook of Research on Teaching*, ed Gage NL (Rand McNally, Chicago), pp 171–246.
16. Campbell DT, Stanley JC (2015) *Experimental and Quasi-Experimental Designs for Research* (Ravenio Books, United Kingdom).
17. Ioannidis JP (2005) Why most published research findings are false. *PLoS Med* 2:e124.
18. Cole M, Scribner S (1974) *Culture and Thought: A Psychological Introduction* (Wiley, New York).
19. Baribault B, et al. (2018) Metastudies for robust tests of theory. *Proc Natl Acad Sci USA* 15:2607–2612.
20. Shiffrin RM, Börner K, Stigler SM (2018) Scientific progress despite irreproducibility: A seeming paradox. *Proc Natl Acad Sci USA* 115:2632–2639.
21. Unsworth SJ, et al. (2012) Cultural differences in childrens' ecological reasoning and psychological closeness to nature: Evidence from Menominee and European American Children. *J Cogn Cult* 12:17–29.
22. Geuter S, Qi G, Welsh RC, Wager TD, Lindquist MA (2018) Effect size and power in fMRI group analysis. *bioRxiv*:10.1101/295048.
23. Nisbett RE, Cohen D (1996) *Culture of Honor: The Psychology of Violence in the South* (Westview Press, Boulder, CO).
24. Warren B, Ballenger C, Ogonowski M, Rosebery AS, Hudicourt-Barnes J (2001) Rethinking diversity in learning science: The logic of everyday sense-making. *J Res Sci Teach* 38:529–552.
25. Salari Rad M, Martingano A, Ginges J (2018) Towards a psychology of *Homo sapiens*: Making psychological science more representative of the human population. *Proc Natl Acad Sci USA* 115:11401–11405.
26. Lehman DR, Nisbett RE (1990) A longitudinal study of the effects of undergraduate training on reasoning. *Dev Psychol* 26:952–960.
27. Frank RH, Gilovich T, Regan DT (1993) Does studying economics inhibit cooperation? *J Econ Perspect* 7:159–171.
28. Keller H (2018) The universality claim of attachment theory: Fact or fancy? Children's socioemotional development across cultures. *Proc Natl Acad Sci USA* 115: 11414–11419.
29. Simons DJ, Shoda Y, Lindsay DS (2017) Constraints on generality (COG): A proposed addition to all empirical papers. *Perspect Psychol Sci* 12:1123–1128.
30. Eerland A, et al. (2016) Registered replication report: Hart & Albarracín (2011). *Perspect Psychol Sci* 11:158–171.
31. Alogna VK, et al. (2014) Registered replication report: Schooler and Engstler-Schooler (1990). *Perspect Psychol Sci* 9:556–578.
32. Hagger MS, et al. (2016) A multilab preregistered replication of the ego-depletion effect. *Perspect Psychol Sci* 11:546–573.
33. Cheung I, et al. (2016) Registered replication report: Study 1 from Finkel, Rusbult, Kumashiro, & Hannon (2002). *Perspect Psychol Sci* 11:750–764.
34. Acosta A, et al. (2016) Registered replication report: Strack, Martin, & Stepper (1988). *Perspect Psychol Sci* 11:917–928.
35. Sue S (1999) Science, ethnicity, and bias: Where have we gone wrong? *Am Psychol* 54: 1070–1077.
36. Gutiérrez KD, Rogoff B (2003) Cultural ways of learning: Individual traits or repertoires of practice. *Educ Res* 32:19–25.
37. Rosenthal R (1966) *Experimenter Effects in Behavioral Research* (Appleton-Century-Crofts, East Norwalk, CT).
38. Rosenthal R, Jacobson L (1968) Pygmalion in the classroom. *Urban Rev* 3:16–20.
39. Rosenthal R, Rosnow RL (2009) *Artifacts in Behavioral Research: Robert Rosenthal and Ralph L. Rosnow's Classic Books* (Oxford Univ Press, New York).
40. Steele CM, Aronson J (1995) Stereotype threat and the intellectual test performance of African Americans. *J Pers Soc Psychol* 69:797–811.
41. Steele CM (2010) *Whistling Vivaldi: How Stereotypes Affect Us and What We Can Do* (W. W. Norton & Co, New York).
42. Kučera H, Francis WN (1967) *Computational Analysis of Present-Day American English* (Dartmouth Publishing Group, Hanover, NH).
43. Romney AK, Weller SC, Batchelder WH (1986) Culture as consensus: A theory of culture and informant accuracy. *Am Anthropol* 88:313–338.
44. Oravecz Z, Vandekerckhove J, Batchelder WH (2014) Bayesian cultural consensus theory. *Field Methods* 26:207–222.
45. Ladson-Billings G (1995) But that's just good teaching! The case for culturally relevant pedagogy. *Theory Pract* 34:159–165.
46. Drabiak-Syed K (2010) Lessons from Havasupai tribe v. Arizona State University Board of Regents: Recognizing group, cultural, and dignity harms as legitimate risks warranting integration into research practice. *J Health Biomed Law* 6:175–225.
47. Suite DH, La Bril R, Primm A, Harrison-Ross P (2007) Beyond misdiagnosis, misunderstanding and mistrust: Relevance of the historical perspective in the medical and mental health treatment of people of color. *J Natl Med Assoc* 99:879–885.
48. Hatzenbuehler ML (2016) Structural stigma: Research evidence and implications for psychological science. *Am Psychol* 71:742–751.
49. Cajete G (1999) *Igniting the Sparkle: An Indigenous Science Education Model* (Kivaki Press, Skyland, NC).
50. Merton RK (1973) *The Sociology of Science: Theoretical and Empirical Investigations* (Univ of Chicago Press, Chicago).