Sub-Saharan Africa has the highest rates of new HIV infections and HIV/AIDS-related mortality in the world. To combat these trends, the World Health Organization (WHO) has worked to introduce new antiretroviral treatments into the region. For a variety of social, economic, scientific, and political reasons, many of the clinical trials that have examined the effectiveness of these drugs were conducted on individuals who were not of African ancestry. Although the WHO has had to confront considerable challenges importing the knowledge obtained from well-controlled medical trials to Africa, we suspect that members rarely worried about the narrowness of the medical data base that tested antiretroviral efficacy: If the treatment was reasonably efficacious when used to raise the immune functioning of European Americans, then it most certainly will be efficacious when used on Africans. Everything we know about biology and HIV so dictates.

But in psychology, we worry about narrowness. Reflecting observations made by Sears (1986), Henry (this issue) notes that in social psychology, prejudice is studied “almost exclusively on student populations with the assumption that findings generalize to a broader population” (p. 55). We agree with Henry’s characterization, but we find both a justification and a critique of this practice in the way that medical practitioners apply medical research. In this article, we distinguish two strategies for drawing inferences from study samples. One approach, adopted in most of the social sciences, uses sampling techniques to establish that a sample is representative of a broader population. The other approach, more common in medical sciences, asks researchers to clarify the dimensions on which the results from a given sample will generalize to a larger population. The two strategies have different implications for research design and the conclusions that are drawn from research, but they often are confused. We seek clarification in light of Henry’s critique.

Representing Populations Versus Generalizing From Samples

Many of the questions that social scientists hope to answer are inextricably bound to specific populations. Investigators who want to know the health consequences of being poor in America, the psychological toll of parental divorce on teens in the United States, or the likelihood that a ballot initiative will pass during the next off-cycle election are focused on questions that draw attention to specific populations. For many American social scientists, for instance, the general population that is typically of interest is the population of adults living in the United States. With more specific research questions, attention will focus on more specific subgroups within the U.S. population.

When a population about which one wants to make statements is known or knowable, concerns for the representativeness of a sample come to the fore. A researcher who has a specific population in mind faces the challenge of locating a sample that can accurately represent this group in an empirical investigation. To this end, the researcher can draw on a class of established population sampling techniques. The nature of these techniques and the way in they are implemented can vary from one study to next, but the spirit underlying each of them is to randomly select cases from the broader population in such a manner that the sample chosen will, on the whole, be every bit as diverse or every bit as homogenous as the population of interest.

The pursuit of a representative sample—common in sociology, political science, and epidemiology—differs from the strategies typically followed in disciplines, including social and experimental psychology. Whereas researchers in the former disciplines often begin with a clear sense of their population, social/experimental psychologists typically start by focusing their empirical scrutiny on a “convenience sample” that they can study easily. They then use this group to make inferences about the psychological dynamics that occur in a larger but often ill-defined group of individuals. More often than not—and certainly in most social psychological studies of prejudice—the sample that is most convenient to researchers is a group of college students, held captive by the shared need for course credit.

Henry (this issue) considered a number of justifications for social psychology’s reliance on college student samples and found them to be lacking. He notes,
for instance, that student samples are rarely as diverse as the general population and that the value of working with restricted groups is often tenuous. We endorse most of his criticisms, but we believe that framing the issue strictly in terms of representativeness is limiting. Because social psychologists rarely have specific populations in mind, it is not clear how they might gauge whether the sample used was “representative” of the “population” being studied. Indeed, individuals who study basic processes often hope to generalize their conclusions to people from generations past as well as future generations (much like medical researchers so hope when they characterize how a heart pumps blood through the body). In this case, the relevant population is hypothetical and potentially infinitely large.

If pressed to identify the population they are trying to target, many of the social psychologists who rely on college student samples would probably respond that they are interested in studying basic processes that are true of human beings in general. But if a social psychologist was confronted with anthropological data suggesting that the type of prejudice her or she has been studying is not known to a small tribe living in the Nile River Valley, we doubt that this would generate any concern. This researcher could simply argue that the effect being studied will generalize to “many humans, but maybe not humans like those living in that small tribe in the Nile River Valley.” Problem solved.

As this example illustrates, social psychologists rarely concern themselves with the task of trying to represent a known population. Borrowing from Mook (1983), they more typically are interested in testing the link between their findings and a broader theory than between their findings and a broader population. It would be wrong, however, to say that social psychologists have no interest in populations. The fact that they use inferential statistics reveals their often tacit assumption that the individuals they study are a random sample from a population. The question then becomes “Who is this population?” Stated another way, social psychologists turn the traditional approach of specifying factors one should consider when generalizing results to broader populations. In the case of college students, the population that students are construed as representing might vary dramatically—depending on the theoretical or applied questions being asked—a point that we now elaborate.

Suppose there is a population that is half males and half females and where 18% of the population is prejudiced. Suppose further that 18% of the males are prejudiced and 18% of the females are prejudiced. A researcher selects a sample of 500 individuals to estimate the amount of prejudice, but for practical reasons adopts a sampling strategy that ends up yielding only males in his or her sample. She estimates the percentage of prejudiced individuals in the population from this “biased” sample of all males. It turns out, the researcher will be just as accurate in her estimates, on average, as if she had sampled 250 males and 250 females from this population and based her estimates on this half-males and half-females sample. In this case, the sample bias on gender is inconsequential because gender is unrelated to the phenomenon we are making inferences about. For purposes of estimating prejudice in this population, a sample consisting of all males is “representative” and effectively functions as a random sample from the population. If a variable of bias is unrelated to the phenomena of interest, bias on it will be inconsequential.

As another example, consider Figure 1. This presents a hypothetical scatterplot between two variables, education and prejudice, in a small population of individuals. Suppose six different researchers want to characterize the relationship between these variables and each decides to use the unstandardized regression coefficient for regressing prejudice onto education for doing so. Each researcher selects a biased sample of students, and each estimates the unstandardized regression coefficient. In one case, R1 has a sample of only college students, in another case R2 has a sample of only white students, and so on. A researcher looking at R1 might be surprised to see that the amount of prejudice is not related to education, whereas a researcher looking at R5 might be surprised to see a stronger relationship. This doesn’t mean that education has nothing to do with prejudice overall. The researchers are simply sampling from different populations, and the samples are not generalizable to each other.

### Figuring Out Who the Sample Represents: Why Bias May Not Matter

Although we share Henry’s (this issue) sense that it is questionable to argue that college student samples are representative of what he often termed the general adult population, we take his analysis a step further by specifying factors one should consider when generalizing results to broader populations. In the case of college students, the population that students are construed as representing might vary dramatically—depending on the theoretical or applied questions being asked—a point that we now elaborate.

Suppose there is a population that is half males and half females and where 18% of the population is prejudiced. Suppose further that 18% of the males are prejudiced and 18% of the females are prejudiced. A researcher selects a sample of 500 individuals to estimate the amount of prejudice, but for practical reasons adopts a sampling strategy that ends up yielding only males in his or her sample. She estimates the percentage of prejudiced individuals in the population from this “biased” sample of all males. It turns out, the researcher will be just as accurate in her estimates, on average, as if she had sampled 250 males and 250 females from this population and based her estimates on this half-males and half-females sample. In this case, the sample bias on gender is inconsequential because gender is unrelated to the phenomenon we are making inferences about. For purposes of estimating prejudice in this population, a sample consisting of all males is “representative” and effectively functions as a random sample from the population. If a variable of bias is unrelated to the phenomena of interest, bias on it will be inconsequential.

As another example, consider Figure 1. This presents a hypothetical scatterplot between two variables, education and prejudice, in a small population of individuals. Suppose six different researchers want to characterize the relationship between these variables and each decides to use the unstandardized regression coefficient for regressing prejudice onto education for doing so. Each researcher selects a biased sample
relative to the variables involved. Researcher 1 selects only individuals in the bottom third of attained education (see panel R1 in Figure 1), Researcher 2 selects only individuals in the middle third of attained education (see panel R2 in Figure 1), and Researcher 3 selects only individuals in the top third of attained education (see panel R3 in Figure 1). Another set of researchers selects based on degree of prejudice. Researcher 4 selects only individuals in the bottom third of prejudice, Researcher 5 selects only individuals in the middle third of prejudice, and Researcher 6 selects only individuals in the top third of prejudice. Even though each investigator has a distinctly different biased sample, all will end up with the same estimates of the unstandardized regression coefficient (except for sampling error) and all will end up with unbiased estimators of the population coefficient. To be sure, the statistical power of their tests of these coefficients and the associated confidence intervals surrounding them may differ, but the parameter estimates themselves should be comparable. As in the previous case, the sample bias is unrelated to the parameter being estimated, hence the bias can be ignored.1

When social psychologists endeavor to generalize their results to broader populations, they need to consider (a) the characteristics of their sample, (b) the characteristics of the broader group of individuals to whom they want to generalize their results, (c) the disparities between their sample and this broader “population” on these characteristics, and (d) whether these disparities will create bias or limit generalizability for the particular parameter(s) being estimated. It is rare that this is done in a systematic and informed fashion in prejudice research.

Identifying Boundary Conditions

In practice, a sample and the broader population that one wants to estimate can differ on an unknown range of characteristics (e.g., height, weight, shoe size, education, ethnicity, personality), and so the choice of which sample–population disparities to focus on for purposes of generalizability analysis needs to be theoretically guided, vis-à-vis the questions being asked. This way of conceptualizing external validity in terms of generalizability rather than representativeness is compatible with McGuire’s (1989) perspectivist theory of knowledge. McGuire argued that social psychologists should be less concerned with asserting that their effects occur in some general or absolute sense than with identifying the conditions under which their effects do and do not occur. We thus see in McGuire’s analyses—and in other seminal works on social psychological methods (e.g., Aronson & Carlsmith, 1968; Mook, 1983)—a rationale for focusing empirical attention on strategically selected (and possibly biased) convenience samples that explicate boundary conditions of a phenomena. Perhaps instead of seeking representativeness to address generalizability, some thoughtfully selected convenience samples that will allow one to make confident statements about boundary conditions will do the trick.

In our opinion, the primary challenge facing prejudice researchers is not so much to “seek diversity” for the sake of doing so. Rather, the challenge facing them is to articulate the nature of their effects in ways that promote more pointed testing of the dimensions around which their effects generalize, so that these hypotheses might be tested. This means clarifying not only the hypothesized mechanisms underlying effects but also the variables that might moderate them, as well as any boundary conditions that constrain the researcher’s ability to make firm predictions. By giving more explicit attention to the dimensions of generalizability, researchers can move beyond abstracts calls for more diverse samples and towards more focused attention on the populations that might expand theory.

Consider an example from medical research. Suppose that a medical researcher has developed a new technique for souchering medical incisions. Although we know little about the latest souchering techniques, we suspect that the medical community would not expect this researcher to test this new technique on a sample that is representative of the entire human population, the entire U.S. population, or even the entire population of individuals who are likely to have surgery. Doctors know that every person when cut will bleed. Moreover, variability in the way we bleed probably is not so great that there would be calls to test if medical trials on narrow samples generalize to the broader public. In contrast, a medical researcher interested in testing the efficacy of various hypertension medications would have good reason for being concerned. Suppose the initial clinical trials on a new drug were run on samples comprising White men. Research clearly indicates that the factors leading to high blood pressure in men differ some from those leading to high blood pressure in women, just as the factors typically leading to high blood pressure in European Americans differ some from those typically leading to high blood pressure in African Americans. An investigator aware of such differences should thus point out the need for running clinical trials that include samples drawn from new populations.

If we had to guess, we suspect that the psychology of prejudice is probably more like hypertension than

---

1The described sample bias in the selection of individuals in this example will lead to inaccurate statements about the correlation coefficient between the variables in the population as well as inaccurate statements about the population standardized regression coefficient, which is why many social scientists prefer unstandardized coefficients to correlations or standardized regression coefficients when making statements about relationships between variables.
bleeding. This calls for research on broader groups than just college students. However, it may not be realistic for a researcher to become concerned with all of the dimensions highlighted by Henry’s (this issue) analysis, nor is it even clear why these specific dimensions might threaten any given research inquiry rather than other ones (although Henry offers many informative observations). Researchers must approach their selection of samples in the same judicious, theory-driven manner that a medical researcher might decide to expand the medical data base on a new medical procedure. Even when validity threats are identified, this does not necessarily mean that a researcher should pursue samples that are diverse. Just as a medical researcher might follow a study on a narrowly defined sample of European Americans with a study on a narrowly defined sample of African Americans, a prejudice researcher might move from one narrow group to another.

**Promoting Representativeness Versus Promoting Generalizability**

Although we believe that thoughtfully selected convenience samples are justifiable, our statements should not be taken as support for prejudice researchers going about their business as usual. To the contrary, we believe that more attention needs to be given to the “generalizability question,” and that doing so might shake the field out of some conceptual ruts. Henry (this issue) proposes his own set of strategies for shaking things up. His attention is focused mostly on ways to provide incentives that encourage researchers to work with more diverse, nonstudent samples. Although we see value in his suggestions, they are premised on the need to produce more representative samples in social psychological research. If representation truly is the goal that our field should be pursuing, then we would propose a somewhat more radical agenda than Henry suggests. We would argue that prejudice researchers should begin their investigations by (a) identifying the specific populations of interest to their study and then (b) employing sampling techniques to locate representative samples of those populations that they can study. But such research could be costly and perhaps only feasible if considerable groundwork has been laid through pointed investigations of smaller, more practical samples. An alternative strategy, one that might be more realistic, is to seek out theoretically meaningful convenience samples that will elucidate the boundary conditions of effects.

**The Holier-Than-Thou Mentality of Theoretical Innovation**

A justification for studying just college students not mentioned by Henry (this issue) is the argument that such research is a necessary first step in theoretical innovation. Before moving into costly research with representative samples, one needs to do more focused research to firmly establish a phenomenon and/or the possible mechanisms underlying it. If a phenomenon or mechanism can’t be established under optimal conditions (e.g., in carefully controlled studies with college students), then there is no sense in pursuing more costly representative samples in pursuit of it. Indeed, we know social psychologists who use college students in their research and who argue that they focus on the “hard part” of science, namely theoretical innovation, and they leave it to others to do the more mundane work of exploring generalizability and application.

Although it is reasonable to work with convenience samples at the early stages of establishing phenomena or until specific threats to generalizability are identified, we agree with Henry (this issue) that broad adherence to this practice by a discipline can lead to collective blind spots. Henry questions, for instance, if our reliance on college student samples is leading to our interest in the subtle forms of prejudice that might occur commonly in colleges but that have little relevance to many of the more common and more damaging forms of prejudice found outside university settings. We share his concerns (Blanton & Jaccard, in press). We doubt, however, that researchers will spontaneously pursue more diverse samples because they appreciate the possibility that they might have to grapple with the factors that can prevent them from observing the effects they originally had predicted. No, we see basic theory on the social psychology of prejudice continuing to move forward mostly through its use of convenience samples. At specific junctures, researchers will pursue “convenience samples” to address cogent validity threats, but such efforts likely will be vastly outnumbered by studies using student samples.

But there is a potential check in the system. This check has not exerted much influence on the way prejudice research has been conducted in the past 20 years, but we think it has an important role to play that could move researchers away from student samples, if it is taken more seriously. This check is the need to apply the social psychological literature on prejudice to real-world social problems. We believe that prejudice researchers would become more proactive in their use of diverse samples if they became serious about exploring the implications of their research for social problems. Although a researcher may argue that their research using convenience samples will generalize to many of the populations that might be interest, a skeptic who must invest resources into a program or policy may question that effects shown with college students generalize to a specific group, acting in a particular manner, in a specific applied context.

We see some evidence that when researchers try to study applied problems rather than the broad theories
surrounding them, they become more focused on specific “real-world” populations to which they apply. Consider, for instance, research examining race effects on weapons identification (Payne, 2001). This work focused attention on the influence of race on shooter decisions (Correll, Park, Judd, & Wittenbrink, 2002), which quickly led to studies examining the influence that race has on the shooter decisions of trained law enforcement professionals (Correll et al., 2007). Although a sample of police officers was theoretically interesting, in that they could be used to test the ability of training to override automatic processes, we suspect that the desire to generalize findings to this particular group was driven by an interest in applying theory to law enforcement.

Unfortunately, concern for application does not always move researchers toward more relevant samples, and this can lead to undesirable effects if researchers still try to make pointed conclusions. For example, strong claims have been made, suggesting that the social psychological literature on implicit biases indicates that antidiscrimination employment law should be altered to accommodate unconscious forms of racism (Blasi & Jost, 2006; Jolls & Sunstein, 2006). Included in this is Kang and Banaji’s (2006) argument that affirmative action procedures should stay in place until national surveys indicate that implicit biases have been reduced to zero. We find such claims provocative because not a single study that has examined the influence of implicit race biases on the employment-related decisions of real employers, working in actual organizational settings. Given this, and given the broader publishing trends noted by Henry (this issue), such aggressive applied claims strike us as dubious.

In summary, whereas interest in basic process may give researchers license to work with student samples until and unless there is a need for tests of generality, interest in research applications should move researchers toward samples that are representative of specific groups, often those that are nonstudent populations. With the increasing emphasis on securing grants and contracts by research universities, this practice can be expected to increase. Social psychology’s reliance on convenience samples may not be so much a threat to its ability to test important theoretical propositions about the nature of prejudice (but see our previous discussion of blind spots). Rather, it may more represent a threat to its ability to test for the specific causes and consequences of prejudice in specific applied contexts. Researchers can thus restrict their samples, if they also restrict their conclusions.

Comparing Diverse Groups: Metric Considerations

As an alternative or complementary strategy to studying representative samples, we have argued for the study of two or more groups that are strategically selected to test the boundary conditions of phenomena. Such research designs engender methodological issues that too often are glossed over when research is focused on a single group of relatively homogenous college students. We discuss one such issue here, namely, that surrounding metric equivalence across groups. This issue speaks to the empirical data presented by Henry (this issue) in his comparisons between student and community samples.

In the realm of psychometrics, most social psychologists are well versed in the concepts of reliability and validity, but unfortunately they tend to pay scant attention to issues surrounding metrics and metric equivalence (Blanton & Jaccard, 2006; Borsboom, 2005, 2006; Kline, 1998; Michell, 2000; see McGrath, 2006). Proper consideration of metrics should lead analysts away from the usual uncritical reliance on standardized coefficients (such as correlations, multiple correlations, and standardized regression coefficients) and brings to the forefront the idea that metric properties can differ from one diverse group to the next, thereby clouding our usual interpretations of group differences. Consider a simple bivariate regression, one in which we regress a measure of income onto the number of years of education to determine the “value” of a year of education. The analysis is conducted in two different ethnic groups, African Americans and European Americans. Suppose that the analysis yields identical standardized regression coefficients in the two groups, indicating that for every 1.0 standard deviation that education changes, income is predicted to change 0.50 standard deviations. One might conclude from this that the “value” of education is the same in the two groups. Suppose, however, that the standard deviation for education is 3.0 in both groups but that for income it is 15,000 for European Americans and 6,000 for African Americans. Such a state of affairs yields unstandardized regression coefficients of 2,500 for European Americans and 1,000 for African Americans. Whereas for European Americans an additional year of education is predicted to be worth $2,500, for African Americans it is only worth $1,000. There is a clear disparity between the groups that was not evident when one focused on standardized coefficients.

The problem with the standardized analysis is that it created different metrics for the two groups. The standardized metric was in units of $15,000 for the European Americans but it was in units of $6,000 for African Americans. Comparing groups on these different metrics is like measuring income in American dollars for one group but British pounds for another and then comparing groups, without acknowledging the group metric differences. Unless one is confident of comparable standard deviations in the groups, comparing standardized coefficients between groups compares groups on measures with different metrics, which can
mask important group differences or similarities. This also is true when comparing groups on correlations, because in the bivariate case, for example, correlations are standardized regression coefficients.

Even if one focuses on unstandardized coefficients and raw means, issues of metric equivalence in the different groups remain important. For example, in the case of comparing means in two diverse groups, a measure can be completely reliable and valid in both groups yet still yield artificial mean differences based on the scales having different “working” metrics. Consider a thermometer scale ranging from 1 to 10 that measures favorability toward Black people. It is well known that ratings of stimuli are relativistic and are made in the context of some frame of reference or comparison standard (Parducci, 1984). If a researcher does not provide individuals with a comparison standard, they invent their own. Individual may make the exact same absolute cognitive judgment but translate that judgment onto a rating scale differently depending on the comparison standards they invoke when assigning a number on the rating scale (Wyer, 1974). Perhaps college students invoke a different comparison standard than a more general “adult” sample does when completing a thermometer rating scale, leading to mean differences that are measurement artifacts rather than real. To be sure, the scale might be reliable and valid for discriminating different levels of prejudice within each group considered separately, as members of a given group invoke roughly the same comparison standard. But across-group comparisons run the risk of characterizing group differences that derive from differences in the use of the rating scale rather than true differences on the underlying dimension being measured.

There is a growing literature on metric equivalence and methods for establishing it when comparing groups (e.g., Millsap, 2005). The diverse-group strategy demands consideration of this literature. Metric nonequivalence can undermine the interpretation of group differences in means, standard deviations, percentages, standardized regression coefficients, and unstandardized regression coefficients. College students almost certainly interact with different types of people than non-college students do, and the comparison standards they might invoke when confronted with an unfamiliar rating scale that they now must assign a number to could be quite different than non-college students. Good research practice with diverse groups requires that the investigator provide individuals with explicit comparison standards so as to remove measurement artifacts because of differential anchoring and differential strategies for defining scale categories.

Conclusions

The tendency to rely on student samples dominates the social psychological analysis of prejudice. Comfort with student samples probably does limit the literature on prejudice in ways we may not appreciate, but one also needs to be careful when encouraging the use of “representative” samples. We suspect that attempts to promote the pursuit more sample diversity probably will be unsuccessful. This is, in part, because it often is harder to study more diverse samples and, in some cases, they may not be any better equipped than college samples in terms of their ability to advance generalizable principles and the theoretical specification of boundary conditions. We see interest in sample diversity being stimulated by researchers who develop pointed interests in applied problems, grant funding, and the identification of boundary conditions for purposes of theory building. For better or worse, prejudice researchers rarely concern themselves with these matters and indeed typically maintain that they are at the forefront of theory innovation, hence their focus on samples and settings that are most favorable to their theories is justified. We agree with Henry (this issue) that this is a narrow perspective and that it underestimates the value of immersion in real-world phenomena with diverse populations. It is a shame that it so difficult removing the blinders.

Note

Address correspondence to Hart Blanton, Department of Psychology, Texas A&M University, 4235 TAMU, College Station, TX 77843. E-mail: hblanton@gmail.com

References
