"Mostly White, Heterosexual Couples": Examining Demographic Diversity and Reporting Practices in Relationship Science Research Samples

Emma L. McGorray[1], Lydia F. Emery[2], Alexandra Garr-Schultz[3], and Eli J. Finkel[1, 2]

[1] Department of Psychology, Northwestern University
[2] Kellogg School of Management, Northwestern University
[3] Department of Psychological Sciences, University of Connecticut

Social and personality psychologists aim to "understand individuals in their social contexts for the benefit of all people" (Society for Personality and Social Psychology, n.d.). Though this mission is admirable, value statements do little, on their own, to create an inclusive, high-quality science that benefits humanity broadly. In this research, we evaluate *relationship science*, a major subfield of social–personality psychology, illustrating both the unique diversity-relevant challenges faced by particular subfields and the barriers to inclusive and diverse research that are shared across research areas. Specifically, we examine the sample diversity and reporting practices of 1,762 studies published in eight mainstream psychology and relationships journals at two time points—(a) 1996–2000 and (b) 2016–2020—and center our analysis around five focal sample characteristics: gender, sexual orientation, regional context, socioeconomic status (SES), and race. We find that reporting practices and representation have not improved for some core demographic characteristics (e.g., socioeconomic status) and that even in domains for which reporting practices have improved (e.g., sexual orientation), reporting remains limited. Further, we find that reporting practices in relationship science frequently center Whiteness (e.g., "participants were mostly White"), obscure or overlook potential sexual orientation diversity (e.g., implying that individuals in man–woman dyads are "heterosexual"), and treat the United States as the contextual default (e.g., participants came from a "large Southeastern university"). In light of these findings, we offer recommendations that we hope will cultivate a more representative and inclusive discipline.

Keywords: diversity, metascience, relationship science, romantic relationships

Supplemental materials: https://doi.org/10.1037/pspi0000417.supp

In 2008, our lab began recruiting couples from the Chicago area for a study about romantic relationships. The sample was limited in diversity along the lines of variables such as race and sexual orientation. In an article reporting results from the study (Finkel et al., 2013), we described the characteristics of the sample in a single sentence, noting that participants were "heterosexual married couples from the Chicago metropolitan area" and reporting only participants' age and relationship length. In 2019, our lab again began recruiting couples from the Chicago area for a new study about romantic relationships. This time, we prioritized diversity and recruited a sample that was more varied with respect to race, socioeconomic status (SES), and sexual orientation. In an article reporting results from the study (Emery & Finkel, 2022), we

dedicated a full paragraph to describing our sample's composition in terms of gender, race, sexual orientation, and SES.

Much work remains to be done, but we, like some other labs, have made changes like these to reckon with the fact that studying college students or convenience samples from university-adjacent communities has produced a scholarly discipline characterized by extreme distortion in our participant samples. Often, this distortion is coupled with insufficient attention to the role of diversity in theorizing and methodology. Like other scholars, we have come to face a series of questions as we grapple with these issues: (a) to what extent are samples in our research area diverse? (the *representation question*); (b) to what extent is researchers' reporting of sample characteristics thorough and inclusive? (the *reporting question*); and (c) to what

extent has the discipline made progress along these lines in recent decades? (the *temporal question*). Questions 2 and 3 have yet to be explored in relationship science, whereas Question 1 has received attention in an analysis that both breaks new ground and leaves much to be uncovered (Williamson et al., 2022). In the present article, we investigate all three questions, focusing on the research area of relationship science to illustrate how insights from area-specific investigations can help to advance improvements for social–personality psychology at large.

Creating an Optimal Scientific Discipline: "Well-Ordered Science"

Philosophers of science suggest that one effective way of evaluating a discipline is to consider whether it is *well ordered*—whether a diverse, informed, and sober-minded group of representatives of the human population from the past, present, and future would endorse the discipline's values, lines of inquiry, research approaches, and research applications (Barker & Kitcher, 2014; Kitcher, 2011). While convening such a group is not practically achievable, comparing the reality of a given science to the ideal of a well-ordered science can be instrumental for scientists hoping to improve their field, as monitoring one's status and progress against a standard is an essential aspect of goal attainment (Harkin et al., 2016).

In psychology generally and in social–personality psychology specifically, scholars aim to use their work to improve the human condition broadly—not just for a subsection of the population but for all of it. The stated mission of the American Psychological Association (2020), for example, is "to promote the advancement, communication, and application of psychological science and knowledge to benefit society and improve lives." The Society for Personality and Social Psychology (n.d.) adopts the mission "to understand individuals in their social contexts for *the benefit of all people*" (emphasis added). Such statements suggest a science that is well ordered with respect to its values, with an orientation toward improving individual and societal conditions for humanity broadly.

But value statements do little, on their own, to create an inclusive, high-quality science. The present research investigates the extent to which personality and social psychology live up to its inclusive rhetoric, focusing on relationship science as an example case. Just as researchers studying different phenomena may require unique approaches to increase power or improve confidence in their research findings (Ledgerwood, 2016), researchers studying different phenomena may confront unique challenges—and require unique approaches to addressing them—related to questions of the diversity and inclusivity of their science. By focusing on relationship science, we simultaneously demonstrate the unique considerations individuals might need to take in a given subfield and discuss issues that will be of broad relevance to scholars from diverse research areas.

In the present research, we contribute to a growing literature examining the diversity of psychological research samples (e.g., Arnett, 2008; Klein et al., 2022; Pollet & Saxton, 2019; Rad et al., 2018; Roberts et al., 2020; Thalmayer et al., 2021) to evaluate whether personality and social psychologists—and relationship scientists in particular—conduct and report on research in ways that enable them to use their work for the benefit of all people. Specifically, through a comprehensive review, we bring extensive data to bear regarding the three broad questions introduced above—the representation question (to what extent are research samples in our research area diverse?); the

reporting question (to what extent is researchers' reporting of sample characteristics thorough and inclusive?); and the temporal question (to what extent has the discipline made progress along these lines in recent decades?).

Why Sample Diversity?

Without examining a diverse array of people and the diverse environments in which they live, the explanatory scope of psychological theories and the applicability of research findings remain unclear. One major criterion in theory evaluation is generality, or the extent to which a theory can explain a broad rather than a narrow set of phenomena or experiences (Gawronski & Bodenhausen, 2015). A theory that explains domain-general processes of prejudice, for example, would be higher in generality than (and thus preferable to) a theory that explains only processes related to gender-based prejudice. Similarly, a theory that explains the relationship maintenance processes of individuals across contexts and identities would be higher in generality than (and thus preferable to) a theory that explains only the relationship maintenance processes of heterosexual people in the United States. This is not to say that the unique experiences of individual subgroups or the unique dynamics of individual subdomains are irrelevant or should be overlooked in the name of focusing on psychological "universals" experienced across time and place, but instead to say that theories that integrate these unique features into a single comprehensive framework offer an advantage over separate, disjointed theories that each speak to just one kind of experience.

Diversity in research samples may play a key role in the development and refinement of theories high in generality. As researchers develop new theoretical models, drawing on evidence from diverse samples or conducting exploratory studies in diverse contexts can help them to discover processes that may not be present in typical samples, ensure that the tenets they advance in their models apply broadly, and identify moderators and appropriate boundary conditions from the outset. Existing evidence suggests that many psychological phenomena vary across cultures, SES groups, social identities, and contexts (e.g., Karney, 2021; Levine et al., 1995; Markus & Kitayama, 1991; Mays et al., 2007; Van Bavel et al., 2016). Testing ideas in diverse populations and settings can help researchers uncover this variation (Medin et al., 2017), identify important moderators (Brewer, 2000), and incorporate these discoveries to broaden and strengthen their theories. Like efforts to explore multiple operationalizations of a variable or test an idea in a new domain, efforts to increase sample diversity are key to developing high-quality theories with a broad explanatory scope.

Sample diversity is also essential to establishing research findings' *external validity*, a construct that encompasses factors such as an effect's robustness (does it generalize across places, historical moments, and people?) and relevance (is it applicable to real-world issues?; Brewer, 2000). The psychological research literature's overemphasis on studying people from Western, educated, industrialized, rich, and democratic (WEIRD) societies (Henrich et al., 2010) poses a barrier to addressing these questions, as researchers have a large amount of knowledge about a subset of the world population that is arguably not just nonrepresentative but is in fact an outlier. Conversely, by using diverse samples, researchers can (a) test whether an effect holds across places, time, and people; (b) ascertain

whether the phenomenon of interest occurs in the day-to-day lives of a diverse group of participants; and (c) identify a broader set of real-world issues in need of researcher attention, not just the subset of issues affecting a single homogenous population. Establishing external validity in this way is important not only for building the basic scientific knowledge base and developing the kinds of high-quality theories discussed above but also for providing crucial information to researchers who are planning applied interventions, forming policy recommendations, or engaging in other efforts to use their science for the benefit of human society. Without testing ideas in diverse samples and contexts, researchers are left in the dark about to whom their findings apply, whether their efforts to use social psychological science for good are beneficial to more than a select few, and to what extent they are overlooking important problems and experiences. If psychologists plan interventions based on research conducted on a single or highly nonrepresentative sector of the human population, their intervention may benefit members of that one sector while doing little to improve conditions for everyone else. For example, couple therapy interventions validated on higher SES couples may help higher SES couples without offering any benefits to lower SES couples (Karney, 2021).

More broadly, because of psychology's focus on participant samples from the United States—one of the wealthiest nations on Earth—the portions of humanity that do benefit are unlikely to be the populations that are arguably most disadvantaged by the problems that psychologists may be able to ameliorate (Arnett, 2008). A focus on problems affecting more advantaged (compared to disadvantaged) populations marks a stark departure from a well-ordered science (Kitcher, 2011) and is incompatible with social psychologists' goals of using their work to benefit all people. To create a more well-ordered science, social psychologists must use diverse samples to identify a broad array of pressing social problems and, through an appreciation of the boundary conditions of their theories and potential moderators of their research findings, find effective and broadly beneficial ways of addressing them. Evaluating sample diversity, then, is essential to gauge whether we are approaching the ideal of a well-ordered science that broadly benefits human society.

A recent investigation from Williamson et al. (2022) provides a major step toward evaluating the nature of sample diversity in relationship science, yet much also remains unknown. In their analysis of race, for example, the researchers categorized the samples in their analysis as "Primarily White," "Racially and Ethnically Diverse," or "Primarily Non-White." Based on their analysis, we know that primarily White samples are more common than the other types of samples, but not how frequently individual racial groups are represented. Similarly, we know that only a small number of studies include any same-sex couples, but not the rates of representation for individual sexual orientation and gender identity groups. Without such specific information, the picture remains unclear—samples are generally not very diverse, but what is the specific nature of this problem? Who is in our studies, and who is missing from them? The present investigation adds depth and breadth to address these and other issues.

Why Reporting Practices?

Upon reflection, it is easy to see why biased research samples undermine generalizability and impair the processes of theory development and testing. In comparison, reporting practices may seem like mere technical issues in scientific writing. The reality, however, is that they add up to an essential component of a well-ordered science. Thorough reporting practices can facilitate meta-analytic or integrative data analytic approaches that enable researchers to investigate how phenomena might play out for demographic groups that are not well-represented in any individual study but that may be well-represented enough across studies that they can be examined with relatively highly powered analyses. If researchers report the full gender breakdown of their sample, for example, other researchers seeking out studies including nonbinary participants can better identify which articles might be eligible for inclusion in an integrative data analysis they are conducting. More thorough reporting practices will also enable researchers to make use of the sometimes limited data that already exist as they work as a field to collect samples that are more diverse.

Additionally, the language researchers use to describe their samples can marginalize or obscure diversity that is important—and that a science invested in benefitting all of society should take seriously. The language researchers use can also reinforce certain populations as the baseline, perpetuating the idea that homogenous samples made up of participants from societally dominant groups in WEIRD societies are the default sample against which all other kinds of samples must be compared.

Noninclusive reporting practices can manifest in many ways, including how authors describe commonly reported sample characteristics such as race, gender, and regional context (i.e., where a study took place). One example is authors' tendency to center Whiteness in their discussion of the race of their samples. For example, they might report only the percentage of participants in the sample who are White, reference Whiteness even when referring to people of color (e.g., by using the term "non-White"), or solely investigate the experiences of people of color in comparison to White people's (McLoyd & Randolph, 1985). Moreover, researchers might not think to report on race at all if their sample is predominantly White since that population is often treated as the default (Cundiff, 2012; Remedios, 2022). Such approaches can (perhaps unintentionally) reinforce Whiteness as a default, perpetuate the notion that White perspectives are neutral (Roberts & Mortenson, 2022), imply that people of color are a single homogenous group, and obscure variability within and between racial groups.

Another example is the practice of reporting only the proportion of a single gender group in one's sample (e.g., writing "51% female" and providing no further information). This approach leaves the gender representation of the sample unclear—are readers meant to assume that the remaining 49% of the sample are men (an assumption that relies on the implication that the only possible gender categories are "men" and "women")? Though concise, the approach of reporting only the proportion of a single gender group (without also stating, if relevant, that only men and women were present in the sample) reinforces the assumption that gender is binary, a notion that can perpetuate the limited acknowledgment of nonbinary identities in psychological research.

Reporting practices can reinforce existing hierarchies related not just to individual identities but also to world regions by treating especially dominant and wealthy nations as standard. In the social sciences broadly and in psychology specifically, samples from the United States and other WEIRD societies are overrepresented (Arnett, 2008; Henrich et al., 2010) and seemingly considered

default. For example, in their article titles, authors mentioned the geographic region in which their study took place more frequently when their samples came from the global South or "non-WEIRD" countries than when they came from WEIRD countries, particularly the United States (Castro Torres & Alburez-Gutierrez, 2022; Cheon et al., 2020; Kahalon et al., 2021). Researchers sometimes also use statements like "this study took place in the northeast," relying on the idea that readers will assume the study took place in the northeastern United States rather than, say, northeastern Kazakhstan. This practice can hinder the diversification of our field by reinforcing the notion that the United States is the default context that need not be named (in comparison to other world regions, which contrastingly tend to be acknowledged and discussed).

Such reporting practices are unlikely to result from any sort of malicious intent on the part of individual researchers. Indeed, they are driven in part by externally imposed constraints (e.g., article word limits). Yet regardless of researchers' intentions or motivations, the accumulation of these practices creates a field that overlooks important identities and experiences and perpetuates the notion that dominant groups are default, reinforcing the overemphasis on dominant groups in researchers' participant samples. This is yet another reason why evaluating reporting practices is central to efforts to assess the state of the field.

Why Relationship Science?

In the present analysis, we focus on the research area of relationship science. In general, analyses of a broader scope help to illuminate the nature of the field at a macro level, whereas analyses with a narrower scope can examine issues broadly relevant across social–personality psychology research areas (e.g., race) and issues that have implications specific to particularly research topics, which may be just as important to address but are more difficult to examine in broad-level analyses. For example, while participants' sexual orientation may not be viewed as sufficiently relevant to warrant inclusion in a fieldwide analysis of sample characteristics, researchers studying certain topics (e.g., stigma, identity, relationships) may find knowing the sexual orientation diversity of their samples key to understanding the implications and generalizability of their findings. Indeed, while previous assessments have demonstrated that psychology research tends to rely on samples from WEIRD societies (Arnett, 2008; Henrich et al., 2010; Pollet & Saxton, 2019; Rad et al., 2018; Sears, 1986) and focuses primarily on dominant groups within those societies (Roberts et al., 2020), these existing critiques often exclude assessments of the representation of lesbian, gay, bisexual, transgender, and queer people in research samples.

The nature and implications of limited demographic diversity may also differ from subfield to subfield in ways that have not yet been illuminated. Different subfields within social psychology may have particularly low representation of certain sociodemographic groups, may have different reasons for that limited representation, and may require different strategies for addressing their low sample diversity and appropriately leveraging their findings for societal benefit. For example, in research areas that primarily conduct research on individuals' perceptions of others, diversification may require not only increasing the diversity of participants but also increasing the diversity of study stimuli (Cook & Over, 2021), concerns less relevant for research areas that do not rely on the same kind of visual stimuli.

Researchers in different subfields may also engage in noninclusive reporting practices that are specific to their research area. In relationship science, for example, researchers often use the phrase "heterosexual couples" to describe their participants. While researchers may use this phrase to indicate the couples are made up of men and women, the phrase can imply that individual couple members are heterosexual, an implication that risks obscuring or miscategorizing participants' identities. For example, a bisexual woman in a relationship with a man might be described as a member of a "heterosexual couple," whereas a bisexual woman in a relationship with a woman might be described as a member of a "lesbian couple." Both labels fail to recognize that couple members may not identify as heterosexual or as gay, which can conceal sexual orientation diversity (Fingerhut & Peplau, 2012) and invisibilize the existence of sexual orientations such as bisexuality, an identity associated with a unique constellation of experiences within relationships that are important for researchers to understand (e.g., Feinstein & Dyar, 2018). By focusing specifically on relationship science, we can examine issues like these, providing examples of subfield-specific issues that limit the inclusivity of our science while also examining issues of representation and reporting practices that relationship science shares with researchers from different areas (e.g., the tendency to use language that centers Whiteness when reporting on participants' racial identities).

Relationship science is also an area in which it is particularly clear that sample diversity and inclusivity matter; as the International Association for Relationship Research (2020) diversity statement notes, "understanding and addressing the lived experiences of underrepresented groups are intrinsic to understanding interpersonal relationships." Relationship processes and outcomes vary across cultures and sociodemographic groups (e.g., Eastwick et al., 2009; Emery & Finkel, 2022; Gottman et al., 2003; Kurdek, 2001; Ross et al., 2019; Timmer et al., 1996; Wu et al., 2021), making limited sample diversity particularly consequential and affording illustrative examples of the consequences of such limited diversity.

Finally, relationship science is a relatively new subfield within social–personality psychology. While topics such as intergroup relations, social perception, and attitudes and attitude change have been of continued interest to researchers since the field's early days (Ross et al., 2010), relationship science only began to coalesce into a major research area in the late 20th century (Berscheid, 1999; Reis et al., 2013). Since the "greening" of relationship science in the late 1990s (Berscheid, 1999), relationship researchers have outlined a path toward a "ripened" (Reis, 2007) and "blossoming" science (Campbell & Simpson, 2013). During this time, the field has become more robust and influential, gaining new prominence in journals, textbooks, and professional organizations within psychology and other social sciences (Reis et al., 2013). Relationship science has no doubt developed substantially since its emergence, but it remains to be seen whether this development has included a notable increase in attention to issues of diversity, which are key to developing a robust science. The last few decades of rapid growth and evolution in relationship science provide a unique opportunity to examine a field as it develops—and make analyses of differences between time periods particularly insightful.

Why Should We Examine These Issues Over Time?

As societal and research attention has increasingly been directed toward the importance of diversity, equity, and inclusion, knowing

whether the inclusivity of relationship science has changed over time has become particularly pressing. This growing attention is evident in social developments and advocacy efforts (e.g., the election of the United States' first Black president, the advent of a global movement in support of trans rights, advances in marriage equality and antidiscrimination protections, the Black Lives Matter movement), as well as in scholarly critiques of longstanding issues with the diversity of research samples, including overrepresentation of societally dominant groups (e.g., Roberts et al., 2020) from WEIRD societies (Henrich et al., 2010). Focusing on relationship science in particular allows us to track how the field's evolution, along with society's evolving understanding of diversity and psychologists' evolving awareness of the homogeneity of their samples, is related to changes in sample diversity and reporting inclusivity.

Has relationship science increased the diversity of its samples in response to the increased awareness of diversity ushered in by a changing society, pivotal articles such as Henrich et al. (2010) WEIRD critique, and the maturation of the research area? Or, has the field mostly carried on as it did 20 years ago, at a time when diversity in research sample was no less important but was less widely considered? Relatedly, have relationship scientists made their reporting practices more inclusive over time—have they increased the extent to which they report on their samples in ways that (a) capture the nuances of diverse identities and (b) avoid centering dominant groups? The answers to these questions are invaluable to understanding what strategies might be effective as researchers attempt to increase diversity and improve the field, but as of yet, they are unknown. In the present article, we address all these unanswered questions, using relationship science as an example case to illustrate problems—and recommendations for addressing them—facing social–personality psychologists across research areas.

Research Overview

In the present research, we examine demographic diversity and reporting practices in social psychological research over time, highlighting the field of relationship science as a case study. Focusing on relationship science enables us to demonstrate the importance of critically evaluating subfield-specific barriers to diverse samples and inclusive reporting practices while also enabling us to speak to broad issues that affect researchers across subfields. We compare samples from studies published between 1996 and 2000 (the "earlier era") to those published between 2016 and 2020 (the "later era") in eight major social psychology and relationship science journals: *Journal of Experimental Psychology: General*, *Journal of Experimental Social Psychology*, *Journal of Personality and Social Psychology* (all three sections), *Journal of Social and Personal Relationships*, *Personal Relationships*, *Personality and Social Psychology Bulletin*, *Psychological Science*, and *Social Psychological and Personality Science*. Adding to existing assessments of sample diversity in relationship science that have examined broad characteristics at a single window of time (e.g., Williamson et al., 2022), we evaluate if and how sample diversity and specific reporting practices have changed over time. In comparing time periods separated by the substantial length of 20 years, we create a liberal test of whether change has occurred. If we observe little to no change over time, that stagnancy would be

particularly striking, especially in light of pivotal societal changes (e.g., growing social justice efforts such as lesbian, gay, bisexual, transgender, and queer advocacy and the Black Lives Matter movement) and scholarly developments (e.g., growing awareness of the low representativeness of research samples and the maturation of relationship science; Arnett, 2008; Henrich et al., 2010; Reis et al., 2013) that occurred in the time between the two eras.

Our analysis focuses on five focal sample characteristics: gender, sexual orientation, regional context, SES, and race. While a wide range of sample characteristics contribute to sample diversity, we focus on these characteristics because we believe that (a) they are likely to be important across many research areas and (b) they afford an illustration of the ways that the goals and practices of a given subfield can generate additional considerations for researchers as they attempt to make their samples more diverse and their reporting more inclusive. Based on the findings of our analysis, we conclude by offering recommendations to scholars hoping to improve their sample diversity and reporting practices, while also considering some challenges that researchers may face as they try to make these improvements.

Method

Transparency and Openness

Data, analysis code, and a codebook with information about variables analyzed in this study are available on the Open Science Framework (https://osf.io/u8czb/?view_only=ac0737b01ee54d84a bcd46481367bdbd). We conducted our analyses using RStudio (RStudio Team, 2022) using R Version 4.2.1 (R Core Team, 2022). This study was not preregistered.

Identification of Eligible Studies

Our analysis focuses on studies relevant to romantic relationships that were published in eight outlets for social psychology and/or relationship science research during two 5-year periods: (a) 1996–2000 and (b) 2016–2020. We selected six high-impact journals publishing social psychology research (*Journal of Experimental Psychology: General*, *Journal of Experimental Social Psychology*, *Journal of Personality and Social Psychology*, *Psychological Science*, *Personality and Social Psychology Bulletin*, *Social Psychological and Personality Science*) and the two official journals of the International Association for Relationship Research (*Journal of Social and Personal Relationships* and *Personal Relationships*). While relationship science often integrates research from psychology, communications, sociology, and other disciplines, we selected primarily journals publishing social–personality psychology research given our intention to evaluate and make recommendations that are especially relevant to social and personality psychology.

To examine the current state of demographic diversity in relationship science, we narrowed our focus to two eras: (a) 2016–2020, which was the most recent 5-year period at the time of the project's undertaking, and which enabled us to examine the state of our science a decade after the publication of pivotal critiques of psychological research samples (e.g., Henrich et al., 2010); and (b) 1996–2000, the 5-year window 20 years before, as the field of relationship science was coalescing (Berscheid, 1999) and before the world had

encountered the societal and scholarly changes that characterize the current era. In our initial identification of articles with studies eligible for our analysis, we conceived of romantic relationship studies broadly, retaining for further review any study that in some way examined variables directly relevant to romantic or sexual relationships. While relationship scientists also examine an array of relationship types such as friendships or family bonds, relationship science research in social psychology tends to predominantly focus on romantic relationships (Chopik, 2019). This meant that (a) a focus on romantic relationships would capture a large, characteristic proportion of the relationship science literature, and (b) it was feasible to develop workable inclusion criteria for what constituted relevant articles for our analysis, given enhanced theorizing within social psychology on romantic relationships. The Preferred Reporting Items for Systematic Reviews and Meta-Analyses diagram depicting the process of study selection is depicted in Figure 1.

To identify eligible studies, we examined all volumes of the eight journals published between 1996 and 2000 and between 2016 and 2020. For each volume, we examined the list of articles published in each of its issues. For any articles whose titles suggested they would be relevant to relationships (e.g., if they specifically mentioned romantic relationships) or were sufficiently ambiguous that their topic was unclear, we next read the article abstract. For any articles whose abstracts suggested they would include a relationship-relevant study (e.g., by mentioning the article focuses on a longitudinal analysis of married couples) or were ambiguous with regards to whether they would include a romantic relationship-relevant study (e.g., by mentioning "close relationships" or "interpersonal relationships" but not specifying the nature of those relationships), we then proceeded to scan the article to determine if it included a study relevant to romantic relationships. Our analysis focused on samples of human participants reported on in the context of research reports. We omitted commentaries/responses, meta-analyses, samples reported informally in theoretical or review articles, or samples that were included in articles primarily focused on illustrating a methodological or statistical technique.

In total, we counted 7,356 articles (including issue introductions and corrigenda) published in these eight journals between 1996 and 2000 or between 2016 and 2020. Based on our review of all article titles, we examined 1,646 abstracts from these journals, which led us to scan 1,345 articles to further evaluate their eligibility. We ultimately identified 1,164 articles that on this initial pass seemed to include at least one relationship-relevant study. These articles included 2,051 studies that we screened to determine their eligibility.

We considered each sample in alignment with how the authors discussed it. For example, if several samples were treated as a single study and combined for analysis and sample-by-sample analyses were not reported, we considered it a single study. If several samples were reported separately, with information about each sample and primary analyses separately reported for each, we considered each individual sample independently (i.e., as its own study). We included pilot or preliminary studies in this total count of 2,051 studies if they were reported on separately from main studies (e.g., if they were set apart under their own headings) but not if they were described briefly in passing or as part of the methods section of another study (e.g., mentioned briefly as justification for the suitability of stimuli). We included only studies that were presented within a primary article; we did not include studies reported in full only in the Supplemental Material.

At this stage of the screening process, we counted quantitative studies as eligible if they involved at least one analyzed variable or manipulation related to existing, potential, or dissolved romantic or sexual relationships (e.g., variables related to relationship status, relationship processes, relationship outcomes, relationship beliefs, relationship tendencies) or included analyzed variables or manipulations that were focused on phenomena contextualized within a romantic/sexual relationship or the romantic/sexual relationship domain. We counted qualitative studies, including content analyses, as eligible if they included focused analysis or discussion of one of the broadly defined aforementioned relationship phenomena.
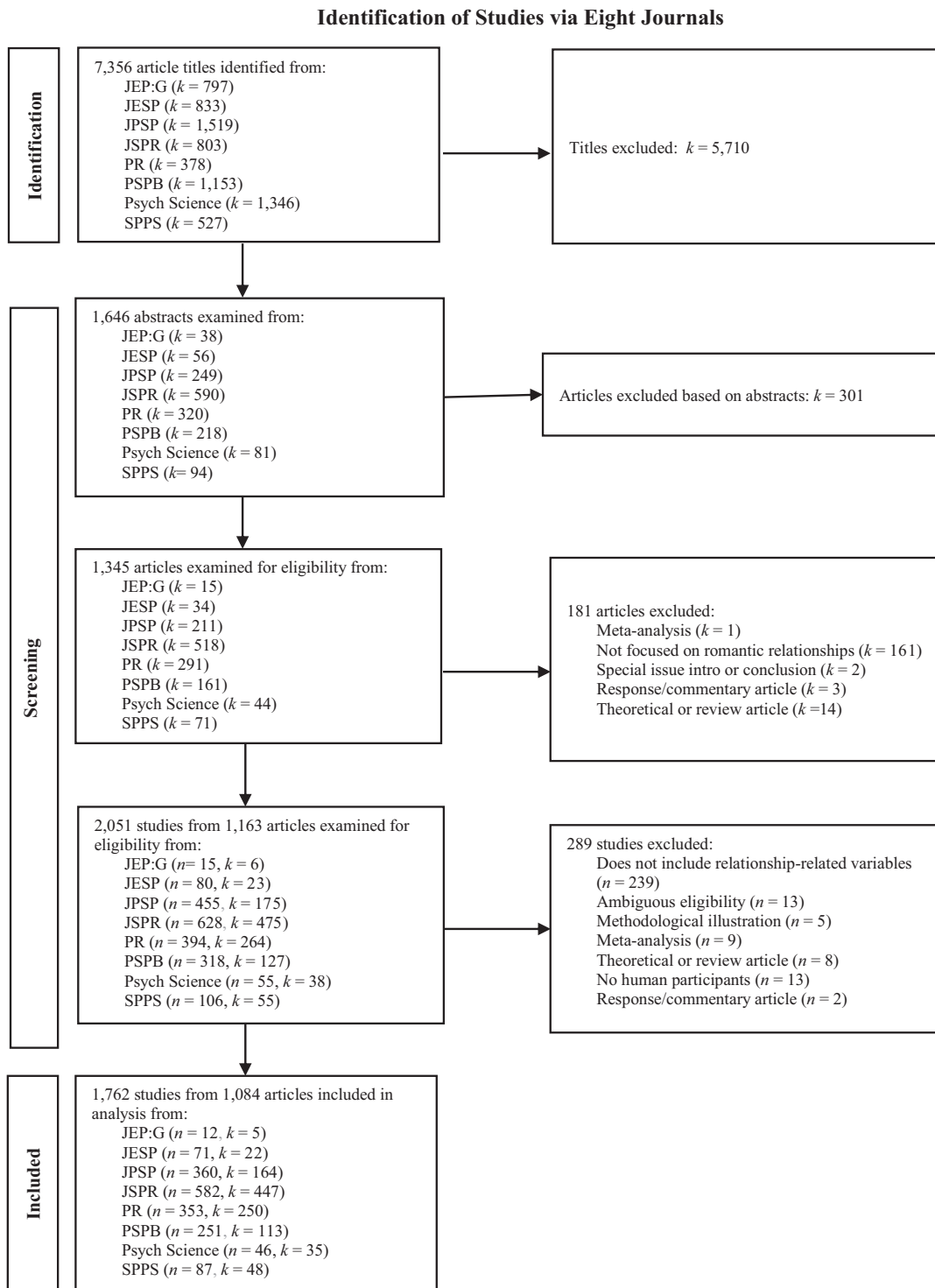
Some studies that might appear on their face to be relationship-related were ultimately excluded from our analysis because we determined they did not meet the criteria described above (i.e., were outside the scope of romantic and sexual relationships). Examples included studies that (a) examined the role of relationship variables as merely a robustness check (e.g., testing for the effect of relationship type as a robustness check; Pederson & Mclaren, 2017); (b) focused on sexual health outcomes, such as condom use or sexually transmitted infection prevention (e.g., a study examining the relationship between alcohol consumption and condom use; MacDonald et al., 1996, Study 1); (c) examined nonconsensual behaviors, such as harassment and sexual coercion (e.g., the association between power and the likelihood of engaging in sexual harassment; Williams et al., 2017, Study 3); (d) examined family relationships without any focused discussion of specific romantic or sexual relationships within the family context (e.g., the disclosure of family secrets to outsiders; Vangelisti & Caughlin, 1997); (e) examined basic physiological responses to sexual stimuli (e.g., a study examining testosterone reactivity in response to erotica; Zilioli et al., 2016); (f) had implications for relationship phenomena but did not include directly relationship-relevant variables (e.g., studies examining attitudes toward same-sex marriage; van der Toorn et al., 2017); or (g) examined relationship-related variables such as attachment without a specific focus on romantic relationships (e.g., Popa-Velea et al., 2019).

In cases where the first author determined that the eligibility of a study was unclear ($n = 76$), the first and second authors separately evaluated the study. In cases of disagreement or continued lack of clarity about the study's eligibility ($n = 32$), the study was included if at least one author could nominate an unambiguous justification for why the study met our inclusion criteria. The study was excluded if neither author could generate such a justification or at least one author could provide unambiguous justification for why the article should not be included ($n = 19$). This process ultimately resulted in a total of 1,762 studies from 1,084 articles included in our analysis.

Data Extraction

We trained a team of 30 research assistants to extract information related to the sample characteristics of eligible studies. They then coded several practice articles, and the authors then reviewed and provided feedback on their coding before allowing research assistants to begin their primary coding. At least two research assistants independently coded each study. After two coders completed the coding for each study, a research assistant flagged any discrepancies

Figure 1

Process of Identification of Studies Eligible for Analysis

Identification of Studies via Eight Journals

Identification

7,356 article titles identified from:
JEP:G (k = 797)
JESP (k = 833)
JPSP (k = 1,519)
JSPR (k = 803)
PR (k = 378)
PSPB (k = 1,153)
Psych Science (k = 1,346)
SPPS (k = 527)

Titles excluded: k = 5,710

Screening

1,646 abstracts examined from:
JEP:G (k = 38)
JESP (k = 56)
JPSP (k = 249)
JSPR (k = 590)
PR (k = 320)
PSPB (k = 218)
Psych Science (k = 81)
SPPS (k = 94)

Articles excluded based on abstracts: k = 301

1,345 articles examined for eligibility from:
JEP:G (k = 15)
JESP (k = 34)
JPSP (k = 211)
JSPR (k = 518)
PR (k = 291)
PSPB (k = 161)
Psych Science (k = 44)
SPPS (k = 71)

181 articles excluded:
Meta-analysis (k = 1)
Not focused on romantic relationships (k = 161)
Special issue intro or conclusion (k = 2)
Response/commentary article (k = 3)
Theoretical or review article (k = 14)

2,051 studies from 1,163 articles examined for eligibility from:
JEP:G (n = 15, k = 6)
JESP (n = 80, k = 23)
JPSP (n = 455, k = 175)
JSPR (n = 628, k = 475)
PR (n = 394, k = 264)
PSPB (n = 318, k = 127)
Psych Science (n = 55, k = 38)
SPPS (n = 106, k = 55)

289 studies excluded:
Does not include relationship-related variables (n = 239)
Ambiguous eligibility (n = 13)
Methodological illustration (n = 5)
Meta-analysis (n = 9)
Theoretical or review article (n = 8)
No human participants (n = 13)
Response/commentary article (n = 2)

Included

1,762 studies from 1,084 articles included in analysis from:
JEP:G (n = 12, k = 5)
JESP (n = 71, k = 22)
JPSP (n = 360, k = 164)
JSPR (n = 582, k = 447)
PR (n = 353, k = 250)
PSPB (n = 251, k = 113)
Psych Science (n = 46, k = 35)
SPPS (n = 87, k = 48)

Note. Here, k = number of articles; n = number of studies; JEP:G = *Journal of Experimental Psychology: General*; JESP = *Journal of Experimental Social Psychology*; JPSP = *Journal of Personality and Social Psychology*; JSPR = *Journal of Social and Personal Relationships*; PR = *Personal Relationships*; PSPB = *Personality and Social Psychology Bulletin*; Psych Science = *Psychological Science*; SPPS = *Social Psychological and Personality Science*.

between the two. The first author then resolved these discrepancies by comparing the coding against the original study and recording the accurate information.[1]

The coding guide containing the instructions research assistants followed when coding is available on Open Science Framework (https://osf.io/u8czb/?view_only=ac0737b01ee54d84abcd4648136 7bdbd), as is a codebook describing the cleaned variables used in our analyses. Our analyses centered around five focal umbrella categories: gender, sexual orientation, regional context, SES, and race/ethnicity. We highlight these focal characteristics because of their broad relevance to social psychology and/or unique implications for understanding the state of relationship science research. See Table 1 for the possible coded values for each of the variables related to these characteristics, along with examples that would receive each possible code.

Gender

Coders extracted three key pieces of information about the gender representation of participants. First, they determined whether the authors reported the gender breakdown of the sample. They recorded whether the breakdown was (a) reported, (b) not reported, or (c) implied but not reported (e.g., if the authors referenced "heterosexual couples" but did not specify that the sample was composed of men and women). We did not initially include an additional category ("partially reported") that we later deemed relevant. The first author reviewed coders' work to verify their coding and to extract the "partially reported" category from their data; this category represents samples for which the percentage of just a single gender category was reported (e.g., if authors reported only the percentage of women).

Second, coders entered the percentage of men, women, and nonbinary participants reported for the sample, as well as the percentage of participants with another gender identity. Because authors largely seemed to be using the terms "males" and "females" as interchangeable with "men" and "women," we did not distinguish between the terms in our coding scheme. For example, a sample reported to be 50% men and a sample reported to be 50% male would be coded equivalently. The final gender-relevant piece of information research assistants coded was whether the authors reported the inclusion of any transgender participants in their sample, entering "not reported" if not and entering the exact information reported if they did. When coding gender (as well as when coding other categories), research assistants entered 0% for a category (even if it was not explicitly reported as 0%) if it was reasonably clear from the numbers presented that that category was likely not represented (e.g., a research assistant would record that 0% of participants were nonbinary if researchers reported that 50% of the participants were men and 50% were women,[2] or as 0% gay/lesbian if 100% of the participants were heterosexual).

Sexual Orientation

Coders first recorded whether (a) information about participants' sexual orientation was not reported, (b) information about sexual orientation was reported, or (c) information about individual participants' sexual orientation was not reported, but authors referred to couples or relationships using sexual orientation labels

(e.g., "heterosexual relationships," "lesbian relationships"). For instances where sexual orientation was reported, research assistants then recorded the percentage of participants who were heterosexual, the percentage who were bisexual, the percentage who were gay/lesbian, and the percentage who were of another sexual orientation. To encompass a greater number of sexual minorities in our analysis, participants with other identities involving attraction to multiple gender groups, such as pansexual, were combined with the percentage of bisexual participants to form a "bisexual+" category when we cleaned the data. We named this category "bisexual+" to reflect that it encompasses multiple labels that describe attraction to multiple gender groups, consistent with the use of the term by other scholars (e.g., Davila et al., 2019).

Regional Context

Coders recorded any information about the exact location of each study sample and/or the country or countries participants were from. If the authors explicitly reported the country associated with the sample or reported a more specific location that would enable a reader to identify the country (e.g., "Peking University," "Houston, Texas"), the regional context was considered "reported" and the appropriate country was recorded. If any inexact information about the location was reported (e.g., "a Southern University"), research assistants also recorded that information. The first author then reviewed this information and coded the location as "implied but not reported" if the study authors used phrases like "a Southwestern University" or "from multiple states" (i.e., if they gave location-based information in a manner that seemed to rely on the implication that the study took place in a particular country, usually the United States). Otherwise, the country was marked as "not reported." In cases where authors reported in-person and online recruitment but did not specify the regional context of participants recruited online, we classified the study as taking place in whichever context the in-person recruitment took place, if specified.

Before conducting analyses, we organized single-country studies (i.e., studies that appeared to have been conducted in a single country) into nine regions: Sub-Saharan Africa, Northern Africa and Western Asia, Central and Southern Asia, Eastern and South-Eastern Asia, Latin America and the Caribbean, Oceania, Europe, the United States, and Canada. The first six of these categories are taken directly from regional categories used by the United Nations Sustainable Development Goals initiative (United Nations, n.d.; see Castro Torres & Alburez-Gutierrez, 2022, for a similar

---

[1] Ideally, the coded information would be the features of the sample used for data analysis (i.e., the final sample after any exclusions). However, in some instances, information was only presented for the sample before exclusions, or it was unclear whether demographic information was describing the full sample or the analytic sample. The extracted data represent the coders' and authors' best efforts to extract accurate numbers in light of these complexities.

[2] While individuals sometimes identify as both nonbinary and as men or women (e.g., Beischel et al., 2022), our impression was that researchers were generally not measuring gender in a way that captured these nuances, which was why we treated men, women, and nonbinary as mutually exclusive categories. However, it is possible that some of the participants who described themselves as men and women are also nonbinary, even if that was not captured in the gender measures they completed.

Table 1

Coded Values for Demographic Variables

Attribute	Codes	Examples
Gender	Gender identity breakdown	
	Not reported	
	Implied but not reported	"Participants were 50 heterosexual couples"; "The average age was 20 for women and 20.5 for men"
	Partially reported	"100 participants (79 females)"
	Reported	"The sample included 300 participants (200 men, 100 women)"
	Inclusion of transgender participants	
	Not mentioned	
	Mentioned	"1% identified as transgender"; "50% male, 48% female, 2% transgender"
Sexual orientation	Not reported	
	Description of couples (e.g., "heterosexual couples")	"The sample included 40 heterosexual couples and one gay couple"
	Reported	"78% heterosexual, 20% bisexual, 1% gay/lesbian, 1% other"
Regional context	Not reported	
	Reported	"Participants were students at the National University of Singapore"; "Participants were MTurk workers from Canada"
Race	Race/ethnicity breakdown	
	Not reported	
	Implied but not reported	"The sample was predominantly White"; "The population from which we sampled was mostly Black"
	Reported	"10% of participants were White, 20% were Black, and 70% were Latinx"
	Reported only the percentage of White participants	
	Yes	"80% of the participants were White"
	No	"80% were White, 10% were Asian, 5% were Multiracial, and 5% were Black"
	Reported only the percentage of participants from one minoritized racial group	
	Yes	"20% of the participants were Black"
	No	"80% were White, and 20% were Black"
Socioeconomic status	Not reported	
	Implied but not reported	"The sample was mostly middle class"
	Reported	"Median education level was a bachelor's degree"; "The mean household income was $80,000"

Note. The examples listed here are not all in line with inclusive reporting practices, as sample characteristics were considered reported even if they were not reported in inclusive ways. The quotes given here are presented as a demonstration of what types of things were assigned to each code, not as examples of inclusive reporting. See Tables 6 and 7, for examples of inclusive reporting practices and the shortcomings of common reporting practices, including some represented in the examples in this table.

approach). In the United Nations' categorization, the last three categories are classified together in the category "Europe and Northern America." In order to examine the United States separately, which was a research goal considering evidence of an overemphasis on the United States in psychological research (e.g., Arnett, 2008), we separated the "Europe and Northern America" category into Europe, the United States, and Canada. A list of the countries we classified under each region can be seen in the Supplemental Material. To enable examination of the dominance of the United States in research samples, we categorized studies with samples drawing from multiple countries as either "multiple countries (including the United States)" or "multiple countries (not including the United States)."

In order to gain a sense of the regional contexts in which lead authors were situated, coders also extracted information about the institutional affiliation of the first author of each article, as well as the country in which that institution was located. We classified these countries into the same set of world regions described above. This provided us with information about whether *researchers* from some

regions published at higher rates than others and enabled analyses related to authors' regional context (e.g., whether authors from certain regions more frequently reported the regional context in which their study took place).

Socioeconomic Status

Coders recorded any information about the SES of participants in each sample, including participants' educational attainment, individual or household income, or subjective SES. Reports that a sample was composed of undergraduate students were not considered to be instances of authors reporting SES information; we return to this point in the Discussion section. If no information about SES was given, coders entered that no information was reported. After coders entered the SES information, we verified the coding and separated out what SES metric was used (e.g., education, income, occupation, subjective SES). If the authors gave a description of the sample but did not report specific numbers (e.g., "the sample was

socioeconomically diverse"; "middle-class couples"), the study was categorized as implying but not reporting SES.

Race/Ethnicity

Race and ethnicity are socially constructed, such that what race means and who is racialized are shaped by sociohistorical, legal, and political context (e.g., Atkin et al., 2022; Haney Lopez, 2004; Smedley & Smedley, 2005). This reality makes analyses of racial/ethnic representation that collapse across disparate countries or regions often inadequate, since categories that are meaningful within one nation may have little meaning in another. For example, describing a U.S. participant as "Chinese" in an American context gives us some insight into that individual's experience of race in the United States. On the other hand, describing participants as "Chinese" (with no further elaboration) when referring to a study that took place in China only tells us the nationality of the individual, not necessarily the aspects of their race/ethnicity that are important for understanding their experiences of ethnicity or racialization within China.

The descriptor "Chinese participants" with no further elaboration does not tell us (a) whether the person has an Asian racial/ethnic background or is a non-Asian citizen of China or (b) how well individual racial/ethnic groups within China are represented within the sample. Additionally, in some countries, race and ethnicity information is not commonly collected and use of the term "race" is avoided, in some cases as a "colorblind" effort to address historical issues of discrimination (Juang et al., 2022). In contexts like those, individuals' migration background and heritage can be one major way of defining who is minoritized (Juang et al., 2022; Simon, 2017). Because immigration background—and when, why, and from where one's family has immigrated—plays a large role in some countries in determining who is minoritized and which groups are affected by prejudice and racism (e.g., Gyberg et al., 2018; Leinonen, 2012; Moffitt & Juang, 2019) that constellation of information may sometimes be the most informative when attempting to depict the composition of a sample.

Because of this cross-regional variability in how race and ethnicity are defined, understood, acknowledged, and assessed, it was not sensible for us to assess reporting of race/ethnicity and representation of racial and ethnic minorities in the same way for all the countries included in our analysis. At the same time, we felt that separately analyzing race and ethnicity reporting and representation for each country in our analysis would limit us from drawing useful conclusions about the entire corpus of studies. As a result, we took a two-phase approach to coding race/ethnicity information. In the first phase, given the expected preponderance of studies from the United States and the U.S.-based author team's firsthand insights into U.S. race/ethnicity categories, we coded information about race/ethnicity focusing on categories that were meaningful in the U.S. context, similar to recent analyses conducted by other U.S. researchers (e.g., Roberts et al., 2020). In the second phase, drawing on insights from our coding of race in the first phase, we coded information about the reporting of race/ethnicity in studies that took place outside of the United States, albeit with full recognition that future research will be required to fine-tune this analysis for any given cultural context.

Phase 1: U.S. Racial/Ethnic Categories.   Coders recorded how the authors reported the race/ethnicity breakdown for each sample; if no information was given, coders recorded that no information was reported. Research assistants then entered the percentage of

participants who were African American, Black, African, or Caribbean; who were Asian American, Asian, or Pacific Islander; who were European American, White, Anglo, or Caucasian; who were Hispanic American, Latinx, or Chicanx; who were Native American or American Indian; who were Middle Eastern; and who identified as another race. In cases where separate percentages were reported for subsamples within the sample and the proportion of each subsample in the entire sample was not made explicitly clear (e.g., race was reported for husbands and wives), we recorded the values for the respective subsamples but did not compute an average (except in cases where the values for each of the subsamples were the same, in which case the average was computable even without information about the proportional representation of each subsample). This information was extracted for all samples, but the subset of samples of participants outside of the United States were excluded in later analyses using these data, as the racial categories used in this phase of analysis have greater meaning for U.S. samples than for samples from other countries, as discussed above.

Based on the information that research assistants extracted, we then evaluated whether race was reported, not reported, or implied but not reported (e.g., the authors stated that the sample was "mostly White" but did not provide specific numbers); whether the study only reported the percentage of White participants specifically (e.g., "90% of participants were White"); and whether the study only reported the percentage of any other racial group of participants, and if so, what the group was (e.g., "90% of participants were Black").

Phase 2: Race/Ethnicity Reporting for Countries Other Than the United States.   In Phase 2, coders reviewed studies that took place in countries other than the United States. Coders recorded any information about race/ethnicity that was reported in the study, including representation of specific racial/ethnic groups; background (e.g., "European background"); immigration status/background; or any other information that appeared to be included for the purpose of painting a picture of the race/ethnicity of the sample. We defined race/ethnicity in this broad way due to the variety of ways that race is defined and assessed internationally, with the goal of providing a general sense of how frequently race/ethnicity was reported in the literature. Due to the variability described above in how race is described and understood across contexts, we focused our analysis on the extent to which information about race/ethnicity was reported rather than on computations of individual racial/ethnic group representation in these studies. Based on the information that research assistants extracted, we evaluated whether race was reported, not reported, or implied but not reported (e.g., the authors stated that the sample was "mostly White" but did not provide specific numbers).

Results

For each of the focal characteristics in our analysis (gender, sexual orientation, regional context, SES, and race/ethnicity), we computed the frequency of the reporting of that characteristic in our sample and, where possible, the rates of representation of specific demographic groups (median and mean values). Because the distribution for variables containing information about the representation of individual groups was frequently skewed or contained outliers (e.g., the vast majority of studies conducted between 1996 and 2000 included zero gay participants, with the exception of a small number of studies containing *only* gay participants), we focus the presentation and interpretation of our results on the median values we computed,

though we also present means and standard deviations for the sake of completeness. In Supplemental Table S2, we also present values representing the approximate percentage of the research population that is a member of each demographic group; these values are generally similar to the average percentages we report in the main text.

Demographic characteristics were frequently incompletely reported or unreported, so our mean and median values for each characteristic were computed based on only a subset of the entire set of studies (i.e., those that reported the characteristic in question and did so in a way that allowed us to extract a single numerical estimate for the representation of individual groups). As a result, we present our results with the caveat that the subsamples that we used to compute the rates of representation for individual groups may not be representative of the entire corpus of studies. In some cases, the subsamples we used to compute our estimates may in fact be *more* diverse than the body of research at large, given that (a) researchers may be more inclined to report demographic characteristics fully when their samples contain notable diversity and (b) researchers who are more attuned to thorough reporting practices may be more inclined to recruit more diverse samples.

We foreground descriptive statistics in the presentation of our results, though we also report the results of chi-squared hypothesis tests.[3] For analyses conducted using the entire sample of 1,762 studies, we had .80 power to detect an effect size of Cohen's $w = .08$ for tests with three degrees of freedom and $w = .07$ for tests with one or two degrees of freedom. For chi-squared analyses conducted using a subset of 1,380 studies with two degrees of freedom (see the results related to race/ethnicity), we had .80 power to detect an effect size of Cohen's $w = .08$. For chi-squared analyses conducted using a subset of 965 studies with one degree of freedom (see the results related to studies from the United States vs. other countries), we had .80 power to detect an effect size of Cohen's $w = .09$. For chi-squared analyses conducted using a subset of 382 studies with two degrees of freedom (see the results related to reporting of race in studies conducted outside of the United States), we had .80 power to detect an effect size of Cohen's $w = .16$. In other words, across the hypothesis tests we present here, we had .80 power to detect small (or, in the case of our analysis of race reporting outside of the United States, small-to-medium) effect sizes. We conduct follow-up tests for our chi-square analyses using the chisq.posthoc.test package in R (Ebbert, 2019) and report 95% confidence intervals, computed using the prop.test function in the R stats package (R Core Team, 2022), for the difference in proportions for pairwise comparisons that emerged as significant according to those post hoc tests. Additional analyses examining the role of journal type (relationship science journals vs. mainstream psychology journals) are reported in the supplement.[4]

Gender

Across the entire sample of studies, 52.2% reported participants' gender; 21.2% partially reported participants' gender; 21.4% implied but did not report participants' gender; and 5.2% did not report participants' gender. To determine if reporting status changed over time,[5] we conducted a chi-square test of independence examining the relationship between reporting status and time period. Reporting status and time period were related, $\chi^2(3, N = 1,762) = 97.66, p < .001$. As can be seen in Table 2, the proportion of studies for which gender was partially reported was greater in the later time period (27.3%) than in the earlier time period (6.7%), 95%

CI $[-.24, -.17], p < .001$. Additionally, the proportion of samples for which gender was reported was significantly greater in the earlier time period (63.8%) than in the later time period (47.3%), 95% CI [.11, .22], $p < .001$. The inclusion of transgender participants was mentioned in only 1.9% of all studies ($n = 33$). All 33 of these samples came from the 2016 to 2020 time period. When authors did mention transgender participants, they often treated transgender status and identification as a man or woman as mutually exclusive (e.g., by implying that "men," "women," and "transgender" are three distinct and nonoverlapping groups).

The median proportion of men per sample was 43.5% ($M = 40\%$, $SD = 20\%$, $n = 1,002$); the median proportion of women per sample was 57.2% ($M = 60\%$, $SD = 19\%$, $n = 1,189$); and the median proportion of nonbinary people per sample was 0% ($M = 0.02\%$, $SD = 0.4\%$, $n = 745$).[6] As can be seen in Table 3, the percentage of men, women, and nonbinary people remained relatively consistent across time periods.

Sexual Orientation

Across the entire corpus of articles, 63.8% did not report participants' sexual orientation, 19.1% used terms like "heterosexual relationship" or "lesbian couple" rather than reporting individual participants' orientation, and the remaining 17% reported participants' sexual orientation. A chi-squared test of independence revealed that reporting status and time period were related, $\chi^2(2, N = 1,762) = 55.61, p < .001$. Post hoc testing revealed that the proportion of samples for which authors reported participants' sexual orientation was significantly greater (95% CI $[-.16, -.10], p < .001$) in the 2016–2020 time period (20.9%) than it was in the 1996–2000 time period (7.9%), and the proportion for which sexual orientation was *not* reported was significantly lower (95% CI [.12, .22], $p < .001$) in the later time period (58.9%) than in the earlier time period (75.7%).

---

[3] Because some of our data points may not be statistically independent (e.g., because two studies might come from the same article), we repeated our hypothesis tests on a subset of our dataset to verify that our conclusions remained substantively similar when addressing the potential issue of nonindependence. This subset included all studies from single-study articles and one randomly selected article from each multistudy article. The results of those analyses are presented in the supplement. The overall patterns we observed were overwhelmingly similar, with the exception of minor differences that we detail in the supplement and that do not alter the major conclusions we draw and elaborate on in the discussion.

[4] The most notable differences in reporting rates across journal types were that regional context, SES, and race were reported at greater rates in relationship science journals than in psychology journals. Additional details can be seen in the supplement.

[5] We wanted to ensure that any potential changes over time were not driven by the fact that studies from Social Psychological and Personality Science were impossible to include in the first time period, because Social Psychological and Personality Science was founded in 2010. We repeated our hypothesis tests on the subsample of studies from the other seven journals. Tables of descriptive statistics and the results of hypothesis tests are available in the supplement. Again, the overall patterns we observed were overwhelmingly similar, with the exception of minor differences that we detail in the supplement and that do not alter the major conclusions we draw and elaborate on in the discussion.

[6] Due to substantial heterogeneity in the reporting practices associated with and categories falling under the "other" gender category, computing numerical results for that category was infeasible, and we do not report results for that category. For similar reasons, we do not present results associated with the "other" race and sexual orientation categories in later sections.

Table 2
Reporting Status of Demographic Groups by Era

Attribute	Reporting status	1996–2000 (n = 522) %	2016–2020 (n = 1,240) %	Total (N = 1,762) %
Gender	Gender identity			
	Not reported	7.1	4.4	5.2
	Implied but not reported	22.4	21.0	21.4
	Partially reported	6.7	27.3	21.2
	Reported	63.8	47.3	52.2
	Inclusion of trans participants			
	Not mentioned	100	97.3	98.1
	Mentioned	0	2.7	1.9
Sexual orientation	Not reported	75.7	58.9	63.8
	Description of couples (e.g., "heterosexual couples")	16.5	20.2	19.1
	Reported	7.9	20.9	17.0
Regional context	Not reported	28.2	37.6	34.8
	Implied but not reported	12.5	5.9	7.8
	Reported	59.4	56.5	57.4
Socioeconomic status	Not reported	72.4	73.4	73.1
	Implied but not reported	2.9	0.6	1.3
	Reported	24.7	26.0	25.6
Race	Race/ethnicity breakdown (U.S. context)			
	Not reported	42.4	31.0	34.4
	Implied but not reported	6.5	0.8	2.5
	Reported	51.1	68.2	63.0
	Race/ethnicity breakdown (international context, excluding United States)			
	Not reported	89.7	72.0	77.0
	Implied but not reported	1.9	1.8	1.8
	Reported	8.4	26.2	21.2
	Reported only the percentage of White participants (among U.S. studies reporting race)			
	Yes	27.4	20.3	21.9
	No	72.6	79.8	78.0
	Reported only the percentage of participants from one minoritized racial group (among U.S. studies reporting race)			
	Yes	0.9	0.2	0.3
	No	99.1	99.8	99.7

Note. The information presented here for race (U.S. context) is based on the set of studies for which the regional context was either not reported or was the United States (for race/ethnicity breakdown, n = 415 for 1996–2000, n = 965 for 2016–2020; for reporting only White or minoritized racial group participants, n = 212 for 1996–2000, n = 658 for 2016–2020). The information presented here for race (international context, excluding United States) is based on the set of studies for which the regional context was reported to be something other than the United States, including multicountry studies (n = 107 for 1996–2000, n = 275 for 2016–2020).

The median proportion of heterosexual people per sample was 96.2% (M = 89%, SD = 22%, n = 292), whereas the median proportion of gay or lesbian people was 0% (M = 8%, SD = 22%, n = 226), and the median proportion of bisexual+ people was 0% (M = 3%, SD = 7%, n = 221). As can be seen in Table 3, the median proportion of gay, lesbian, and bisexual+ people did not meaningfully change over time (0% in both time periods).

Regional Context

Across the entire sample of studies, 34.8% did not report the regional context in which the study took place, 7.8% implied but did not report the context, and 57.4% of the studies did report regional context. Time period and reporting status were significantly related, $\chi^2(2, N = 1,762) = 30.11$, $p < .001$, with a significantly greater proportion (95% CI [−.14, −.05], $p < .001$) not reporting regional context in the later time period (37.6%) than in the earlier time period (28.2%) and a significantly lower proportion (95% CI [.03, .10], $p < .001$) implying but not reporting regional context in the later time period (5.9%) than in the earlier time period (12.5%). Reporting status differed by the affiliation of the first author, $\chi^2(2, N = 1,762) = 93.03$, $p < .001$. Studies led by first authors from the United States reported information about the regional context less often compared to those led by first authors from other regions (51.3% vs. 69.0%; 95% CI [−.23, −.13], $p < .001$). Studies led by first authors from the United States also included descriptions that implied but did not report regional context at greater rates (11.8% vs. 0.3%; 95% CI [.09, .13], $p < .001$) and failed to report regional context at greater rates (37.0% vs. 30.6%, 95% CI [.02, .11], $p = .009$).

Of the studies reporting regional context (n = 1,011), the largest proportion of studies came from the United States (62.2%), followed

Table 3

Representation of Demographic Groups by Era

Demographic category	1996–2000			2016–2020			Total		
	n	Mdn	M (SD)	n	Mdn	M (SD)	n	Mdn	M (SD)
Gender									
Men (%)	336	44.7	42 (19)	666	43.0	39 (21)	1,002	43.5	40 (20)
Women (%)	355	56.0	59 (18)	834	58.0	61 (19)	1,189	57.2	60 (19)
Nonbinary (%)	312	0	0 (0)	433	0	0.04 (.5)	745	0	0.02 (0.4)
Sexual orientation									
Heterosexual (%)	40	100	83 (36)	252	95.1	90 (18)	292	96.2	89 (22)
Gay or lesbian (%)	36	0	18 (38)	190	0	6 (17)	226	0	8 (22)
Bisexual+ (%)	35	0	0.2 (1)	186	0	4 (7)	221	0	3 (7)
Race (U.S. context)									
African American, Black, African, or Caribbean (%)	138	5.9	10 (12)	464	7.1	10 (14)	602	7.0	10 (14)
Asian American, Asian, or Pacific Islander (%)	115	3.0	7 (11)	437	7.0	12 (13)	552	6.3	11 (13)
European American, White, Anglo, or Caucasian (%)	193	84.0	79 (19)	608	74.9	69 (20)	801	76.2	71 (20)
Hispanic American, Latinx, or Chicanx (%)	109	2.7	7 (10)	425	6.4	10 (11)	534	6.0	9 (11)
Middle Eastern (%)	42	0	0.2 (1)	94	0	0.8 (2)	136	0	0.6 (2)
Native American or American Indian (%)	53	0	3 (14)	203	1	1 (2)	256	0.8	2 (6)

Note. The race information presented here is based on the set of studies for which the regional context was either not reported or was the United States.

by Europe (12.5%) and Canada (8.5%). In other words, the United States was represented more than four times more than the second most-represented region and was more represented than every other region combined. With regard to studies reporting national context, 3.4% of the total sample used multiple countries (including the United States), whereas 1.2% used multiple countries (not including the United States). Using the sample of single-country studies reporting regional context ($n = 965$), we tested whether the proportion of U.S. versus non-U.S. studies significantly changed between the earlier time period (67.2% U.S. studies) compared to the later time period (64.3% U.S. studies). It did not, $\chi^2(1, N = 965) = 0.68$, 95% CI [−.04, .10], $p = .410$. See Table 4 for full details on the representation of each of the world regions in our analysis, including by time period.

A similar overall pattern emerged regarding the regional context of lead authors. The largest proportion of studies was reported on in the context of articles authored by researchers with institutional affiliations in the United States (65.0%), followed by Canada (12.3%) and Europe (11.5%). Era and the predominance of first authors at institutions outside of the United States were related, $\chi^2(1, N = 1,762) = 38.25$, $p < .001$, such that the proportion of lead authors from the United States versus other countries was lower in the later era compared to the earlier era (61.0% versus 76.4%, 95% CI [.11, .20]. See Table 5 for full details on the proportion of studies authored by authors with U.S. affiliations versus affiliations in other countries.

Socioeconomic Status

Across the studies, 73.1% did not report, 1.3% implied but did not report (e.g., "the sample was largely middle class"), and 25.6% reported participants' SES. A chi-square test revealed that reporting status and time period were related, $\chi^2(2, N = 1,762) = 14.25$, $p < .001$. Post hoc testing revealed that the proportion of studies that implied but did not report SES was lower (95% CI [.006, .04], $p < .001$) in the later time period (0.6%) than in the earlier time period (2.9%), but the proportion of studies that reported SES did not significantly increase over time, and

the proportion of studies that did not report SES did not decrease over time. Of the studies that did report SES, current personal educational attainment was the most common metric (80.9%), followed by income (36.6%), subjective social class (17.5%), or some other metric (5.8%).[7]

We did not calculate the percentage of participants from specific SES groups, because the reporting of SES across the studies was highly inconsistent. For example, even within a single metric of SES, such as education, some authors reported median education level, whereas others reported mean education level or the percentage of participants with a 4-year degree. This made it difficult to calculate meaningful estimates of SES. We return to this issue in the Discussion section.

Race/Ethnicity

Race/Ethnicity in the United States

For this analysis, we excluded any studies that we could definitively determine were not conducted in the United States (i.e., any studies that were explicitly reported as having been conducted in another country).[8] This left 1,380 single-country studies. Of these, 63.0% reported participants' race, 34.4% did not report race, and 2.5% implied but did not report race. There were differences in reporting between the two time periods, $\chi^2(2, N = 1,380) = 61.35$, $p < .001$. As indicated in Table 2, fewer studies implied but did not report race in the later time period (0.8%) than in the earlier time period (6.5%; 95% CI [.03, .08], $p < .001$). Additionally, more studies reported race in the later time period (68.2%) than during the earlier time period (51.1%; 95% CI [−.23, −.11], $p < .001$), and fewer studies failed to report

---

[7] We counted each metric separately, so if an article reported multiple indices of SES, each index counted toward the total for that metric; as a result, percentages do not add up to 100%.

[8] Results are similar if we restrict to studies that are definitively from the United States (rather than studies from the United States and studies for which the study's regional context was not reported). See the Supplemental Material for these results.

Table 4

Representation of World Regions Studied by Time Period

Region	1996–2000		2016–2020		Total	
	n	*%*	*n*	*%*	*n*	*%*
Sub-Saharan Africa	0	0	0	0	0	0
Northern Africa and Western Asia	11	3.5	50	7.1	61	6.0
Central and Southern Asia	0	0	4	0.6	4	0.4
Eastern and South-Eastern Asia	3	1.0	21	3.0	24	2.4
Latin America and the Caribbean	0	0	6	0.9	6	0.6
Oceania	18	5.8	11	1.6	29	2.9
Europe	34	11.0	92	13.1	126	12.5
United States of America	203	65.5	426	60.8	629	62.2
Canada	33	10.6	53	7.6	86	8.5
Multiple countries (including the United States)	4	1.3	30	4.3	34	3.4
Multiple countries (not including the United States)	4	1.3	8	1.1	12	1.2

Note. The information presented here is based on the set of 1,011 studies for which the regional context was reported.

race in the later time period (31.0%) than in the earlier time period (42.4%; 95% CI [.06, .17], $p < .001$). However, among those articles that did report race, 21.9% only reported numbers of White participants (e.g., "the sample was 85% White"), compared to 0.3% reporting only numbers of any other group ($n = 3$ studies, which reported only the number of Black participants). Patterns of reporting only the proportion of White participants broken down by era can be seen in Table 2.

Next, we calculated the median percentage of each race/ethnicity reported in the articles that did report race. As indicated in Table 3, the median sample had limited participant diversity in both time periods. Overall, across the two time periods, the median racial representation was 6.3% Asian ($M = 11\%$, $SD = 13\%$, $n = 552$), 7.0% Black ($M = 10\%$, $SD = 14\%$, $n = 602$), 6.0% Latinx ($M = 9\%$, $SD = 11\%$, $n = 534$), 0% Middle Eastern ($M = 0.6\%$, $SD = 2\%$, $n = 136$), 0.8% Native American ($M = 2\%$, $SD = 6\%$, $n = 256$), and 76.2% White ($M = 71\%$, $SD = 20\%$, $n = 801$). Descriptively, the predominance of White participants decreased somewhat between the two eras—the median proportion of White participants in the earlier era was 84%, but was 75% in the later era. However, increases in the representation of individual minoritized racial groups were quite modest (see Figure 2 and Table 3, for the representation of each racial group across the two eras).

Race/Ethnicity Outside the United States

For this analysis, we included all studies that were coded as taking place outside of the United States, including multicountry studies. This left 382 studies. Of these, 21.2% reported participants' race, 77.0% did not report race, and 1.8% implied but did not report race. There were differences in reporting between the two time periods, $\chi^2(2, N = 382) = 14.62$, $p < .001$. As indicated in Table 2, more studies reported race in the later time period (26.2%) than during the earlier time period (8.4%; 95% CI [−.26, −.10], $p < .001$), and fewer studies failed to report race in the later time period (72.0%) than in the earlier time period (89.7%; 95% CI [.09, .26], $p < .001$).

Discussion

In the current research, we examined the demographic diversity and reporting practices associated with relationship science research samples, comparing studies published between 1996 and 2000 to studies published between 2016 and 2020. Overall, we found that improvements have occurred on some fronts but not others, and much remains to be done to increase the inclusivity of our science. We begin by reviewing our findings, followed by discussion of the implications of our findings for psychological theory and methodological practice.

Table 5

Representation of Studies Authored by Researchers With Affiliations in Each World Region

Region	1996–2000		2016–2020		Total	
	n	*%*	*n*	*%*	*n*	*%*
Sub-Saharan Africa	0	0	0	0	0	0
Northern Africa and Western Asia	13	2.5	52	4.2	65	3.7
Central and Southern Asia	0	0	1	0.1	1	0.1
Eastern and South-Eastern Asia	2	0.4	34	2.7	36	2.0
Latin America and the Caribbean	0	0	3	0.2	3	0.2
Oceania	19	3.6	58	4.7	77	4.4
Europe	54	10.3	148	11.9	202	11.5
United States of America	399	76.4	747	60.2	1,146	65.0
Canada	35	6.7	182	14.7	217	12.3
Multiple countries (including the United States)	0	0	9	0.7	9	0.5
Multiple countries (not including the United States)	0	0	6	0.5	6	0.3

Note. The information presented here is based on the set of studies for which the regional context was reported.

Figure 2

Proportion Representation Per Racial Group by Era for U.S. Studies Reporting Race
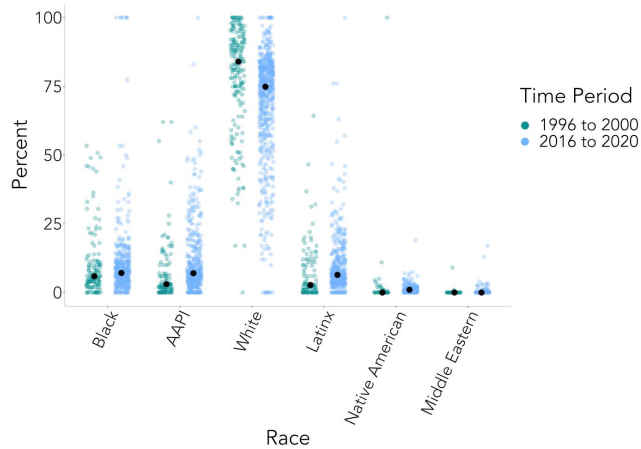
Note. AAPI = Asian American or Pacific Islander. Each dot corresponds to a data point. Black dots represent the median value. See the online article for the color version of this figure.

Is the Field in a Better Place Now Than It Was 20 Years Ago?

Overall, with regard to the diversity of its samples and the inclusivity and comprehensiveness of its reporting practices, the field has made strides in some areas (e.g., increased reporting of sexual orientation over time) while remaining stagnant in others (e.g., predominance of research in the United States), with much room left to grow even in areas of improvement. Gender was fully reported somewhat commonly across eras (52% of the time), but it was fully reported at lower rates in the later era than in the earlier era, perhaps in part because of greater rates of partially reporting gender in the later era (e.g., writing "The sample was 54% men" rather than "The sample was 54% men and 46% women"). As noted previously, this practice of partially reporting gender may serve to reinforce the gender binary and overlook the possibility of participants who are nonbinary, an issue that is important to attend to given that the median percentage of nonbinary participants across both eras was 0%. Mentioning the inclusion of transgender participants, too, was extremely rare in the studies in our analysis (mentioned in 0% of studies from the earlier era and 2.7% of studies from the later era). That all studies mentioning transgender participants occurred in the later era may be the result of evolving terminology and growing awareness of transgender identity, and time will tell if this growing awareness corresponds to greater understanding of trans people's experiences.

With regard to sexual orientation, in the later time period, sexual orientation was more frequently reported (and less frequently "not reported"). This encouraging trend suggests greater awareness of the importance of sexual orientation; it may also reflect reduced stigma toward sexual minorities over time, which may have empowered more researchers to ask about and report on sexual orientation without worrying as much about adverse consequences for participants. Less encouraging is that the median percentage of

gay, bisexual+, and lesbian participants remained zero in both eras, suggesting that typical samples continue to be composed primarily of heterosexual people.

We observed a similar pattern with the reporting and representation of race. For studies occurring in a U.S. context, race was frequently reported (63.0% of the time), and it was reported more often in the later era than in the earlier era (and less frequently "not reported"). This increasing rate of reporting of race across time also emerged for studies taking place outside of the United States, though levels of reporting were lower. The median percentage of many racial groups among studies taking place within the United States context remained relatively low across eras. For example, the median representation of African American, Black, African, or Caribbean people in the most recent era was 7.1%, whereas 13.6% of the United States identifies as monoracial Black or African American (U.S. Census Bureau, 2021). If researchers wish to understand the experiences of marginalized populations, including members of those populations at representative rates at minimum will be necessary; arguably, statistical *over*representation may be required. We hope that greater awareness of the importance of reporting and attending to identities such as gender, sexual orientation, and race will also translate to greater representation of marginalized populations and greater understanding of their experiences.

Reporting of SES did not follow the same encouraging trends that sexual orientation and race did. Instead, the percentage of articles reporting information about participants' SES remained relatively stagnant between the two eras, as did the percentage who failed to report SES. Only around a quarter of studies included information about participants' SES, which is especially concerning given recent evidence about the role of SES in relationship phenomena (e.g., Emery & Finkel, 2022; Karney, 2021; Ross et al., 2019). We were unable to compare rates of socioeconomic diversity across time periods due to variability in how SES was reported, but as understanding of the role of SES in relationship processes (and psychological processes more generally) grows, we hope to see both greater reporting of SES and greater SES diversity in the coming years.

We also hope to see increasing diversification of the regions represented in relationship science and social psychology literature, as we found that the United States was overwhelmingly the most frequent context in which studies in our analysis took place across both time periods. Fewer studies included authors "implying but not reporting" context in the later time period (e.g., writing "this study took place in the Northeast"), but this was not accompanied by an increase in fully reporting context in the later time period. Additionally, although the proportion of studies authored by researchers with U.S. institutional affiliations declined over time, the United States remained the most frequent regional context of first authors' institutions in the later time period. The continued dominance of the United States (in terms of both study context and first author institutional affiliation) is especially striking given that the overrepresentation of the United States and other regions in the Global North were so widely discussed between our earlier and later periods (e.g., Arnett, 2008; Henrich et al., 2010).

All in all, the field has made important strides forward—but there is much left to do, some of which involves addressing issues that psychologists have been raising for years. Pushing the field toward realizing its goal of using science for societal benefit will require changes to theory and methodology. We discuss each in turn.

Implications for Theory

Psychological theory and research reciprocally influence one another. Theories are necessary for researchers to make sense of their findings (Muthukrishna & Henrich, 2019), and they shape and motivate the research questions that scholars pursue and the methods they use to investigate them. Research findings, in turn, are key to developing and refining theories. Given this reality, both increasing sample diversity and promoting greater theoretical consideration of diversity are promising routes to improving the field.

Limited sample diversity (or limited attention to sample diversity) can prevent researchers from fully understanding the constraints of existing theories and from developing innovative theoretical ideas that could introduce or further advance fruitful programs of research. For example, only by using SES-diverse samples have researchers been able to understand the boundary conditions of dynamics related to existing theories of risk regulation and relationship conflict and SES differences in these processes (Emery & Finkel, 2022; Ross et al., 2019). Additionally, when studying diverse samples, researchers can uncover the nuances of theoretical constructs such as familism (i.e., the cultural value placed on family relationships; Campos et al., 2014), novel insights that can help to generate theories that more comprehensively account for how people across cultures and contexts can most easily attain the well-being benefits of close relationships. As is evident in these examples, increasing sample diversity has the potential to strengthen and refine psychological theories and generate new theories that focus on understudied and overlooked phenomena.

Reciprocally, increasing theoretical considerations of diversity has the potential to increase the diversity of research samples. For example, in a discussion of the suffocation model of marriage in America, Finkel et al. (2014) speculate that their model's key tenets apply across U.S. sociodemographic groups, even as there may be variation in the extent to which those tenets apply. While the model engages with the possibilities of sociodemographic variation along some dimensions, initial formulations of the model do not account for whether historical differences in the meaning of marriage across sexual orientation groups might limit the model's explanatory scope to relationships between men and women. Incorporating greater attention to diversity with regard to sexual orientation might lead researchers to test the model using not the typical population of heterosexual people in male/female relationships but instead a sample of people with diverse sexual orientations and relationship types.

Considering diversity in the process of theory development and refinement does not mean that all theories will or should generate predictions that phenomena will differ across demographic groups or cultural contexts, as there well may be phenomena that researchers have no reason to expect identity-related or contextual factors to influence. But the claim that a theory is universal should be made explicitly and with justification reflecting engagement with—rather than avoidance of—questions surrounding diversity, and it should be tested (Sue, 1999), just as other claims and predictions of a theory might be. A theory's purported universality should not be justification for the appropriateness of a homogenous sample of White undergraduate students. Instead, it should motivate researchers to collect samples that vary along multiple dimensions of diversity, so that they can test whether the theory is, in fact, universal. Regardless of whether researchers expect contextual or identity-based variation, then, greater theoretical attention to diversity should facilitate greater sample diversity.

In sum, the ways research and theory mutually inform one another mean that efforts to build an inclusive and well-ordered science would benefit from attention to both researchers' theoretical formulations and their methodological practices. The relationship between research and theory can perpetuate the shortcomings of psychological science. More optimistically, however, the reciprocal relationship between research and theory presents an opportunity to build positive momentum, as increasing both sample diversity and theoretical considerations of diversity can trigger the development of a cycle wherein increasingly diverse samples lead to theories that increasingly take into consideration the role of diversity, which leads to more diverse samples to test relevant theoretical tenets, which leads to greater theory refinement, and so forth.

Implications and Recommendations for Methodological Practice

Improving the diversity of psychological research samples requires many changes, including at the structural level. At the same time, the individual choices of researchers influence those structures and cumulatively contribute to the low rates of marginalized populations in our research samples, making individual-level changes important as well. We turn to structural recommendations toward the end of this section, but we primarily focus here on individual-level changes, strategies that individual researchers can implement starting today.

Though we acknowledge the financial and logistical constraints that may limit researchers' efforts to recruit more diverse samples, we urge researchers to consider sample diversity seriously at every stage of the research process. Regardless of what samples they are able to collect or analyze, all researchers can contribute to improving the field by adopting thorough and inclusive practices when reporting the participant demographic information that is available to them. Such practices will help to paint a clearer picture of who is represented in the field's research samples and make the sample diversity (or lack thereof) explicit rather than implicit, a necessary first step on the road to a well-ordered science. Additionally, when combined with open science research practices such as sharing de-identified data, clear and inclusive reporting can facilitate integrative data analyses that allow researchers to draw together data from members of groups that may be sparsely represented within datasets but more well-represented across them. Below, we offer recommendations that researchers can consider regarding how they reflect on and write about their samples. Our recommendations include both suggestions that apply across research areas and suggestions that are specific to relationship science, which we offer as demonstrations of the ways research area-specific concerns can inform how researchers discuss their samples.

In offering our suggestions, we also acknowledge the practical constraints that may prevent researchers from reporting on all of the characteristics that would paint the clearest picture of their sample. Just as pursuing one aspect of a high-quality science (e.g., replicability) can involve trade-offs related to other desirable scientific features (e.g., external validity; Finkel et al., 2017), increasing the diversity of the collective participant population along one dimension may mean researchers are unable to increase diversity along another dimension or report all potentially relevant sample characteristics. For example, in some countries, same-sex sexual behavior remains illegal and is sometimes punished with measures as

extreme as death (International Lesbian, Gay, Bisexual, Trans and Intersex Association [ILGA] World, 2020). These countries are largely outside of the North American and European regions that our analysis revealed are overrepresented in the literature (ILGA World, 2020), meaning studies conducted in these countries are needed if researchers wish to increase the field's geographical diversity. Yet conducting studies in these regions may mean not collecting or reporting information about participants' sexual orientation or same-sex relationship history, as even participants who are promised anonymity may feel fear and mistrust when asked questions about highly stigmatized identities or behaviors that are associated with negative legal and personal consequences. Due to the legal context, researchers who do venture to ask these sensitive questions may receive data that is incomplete (e.g., participants abstaining from answering the question) or inaccurate (e.g., participants saying they have never had a same-sex relationship when they have in order to avoid potentially negative consequences). While these researchers may be unable to comprehensively assess or report on their participants' sexual orientation and thus may not visibly contribute to sample diversity along that dimension, they will have contributed to the literature a much-needed study that examines participants in an underexamined global region.

We anticipate trade-offs like these to be common, and we hasten to emphasize that improving participant diversity and reporting inclusivity relies not on individual studies that meet every goal perfectly, but instead on a series of complementary studies that each make their own contributions, much in the way researchers achieve research programs high in validity through a set of studies with complementary strengths (Brewer, 2000; Finkel et al., 2017). Our recommendations should be considered not as stringent standards each individual study must meet, as each research effort comes with its own unique set of constraints and challenges, but instead as suggestions offered to help guide the field forward at the collective level as we work to improve diversity and inclusivity.

Which Sample Characteristics Should Be Reported?

In principle, there is an infinite set of potential sample characteristics that researchers could report on, from sociodemographic characteristics to psychological individual difference variables. We hesitate to offer any one-size-fits-all recommendations, as the dimensions that are important to report on and feasible to collect information on may vary based on the research question at hand, the researchers' goals, and the geographical context of the study. However, when deciding which sample characteristics to report on, researchers might ask themselves whether the reader has a clear sense of how representative the sample is of the broader population, particularly with respect to characteristics that have theoretical implications for the research question. Making determinations about which characteristics are theoretically relevant will become increasingly feasible as theoretical considerations of diversity increase and researchers increase the diversity of their samples. Below, we offer recommendations for collecting and reporting the focal characteristics included in our analysis, which we believe psychologists across research areas will often find important to report on. For additional suggestions on how to describe research participants in inclusive and respectful ways, we also recommend that researchers consult the American Psychological Association's (2021) inclusive language guidelines. We also point readers toward Call et al. (2022)'s ethical,

social-justice-oriented approach to working with demographic data, which offers questions to consider as researchers collect and report on participant demographics. In Table 6, we offer an example of how authors might report on sample characteristics in an inclusive way, given one particular example set of research goals. In Table 7, we offer a summary of some reporting practices with shortcomings, the issues associated with them, and recommended alternatives.

Gender

When formulating and discussing questions about participants' demographic characteristics and identities, researchers would benefit from evaluating *why* they are asking these questions. Is it for the purpose of painting a picture of who the respondents are, or are these variables going to be used to test specific hypotheses? In the latter case, researchers should ensure they are asking questions in ways that provide them with the hypothesis-relevant information they need. In the case of gender/sex,[9] researchers might have hypotheses relevant to a range of facets, from gender identity and expression to hormonal or anatomical characteristics, and they should ensure the questions they ask to tap into the variables of interest (Lindqvist et al., 2021; Price, 2018). Asking only participants' gender or sex assigned at birth, for example, tells researchers little information about how individuals identify or express themselves, what gendered experiences they encounter, or what hormonal profile they have. If researchers are interested in the role of gender expression, gendered experiences, or specific hormones, they will best capture the information they need by asking questions that directly assess those variables (Lindqvist et al., 2021; Price, 2018), rather than merely inferring that information from questions about participants' identity labels or assigned gender/sex at birth. Being clear about how they are defining the terms they use (i.e., when they use terms like "sex," do they mean sex assigned at birth or something more like gender identity?) will also help them to ensure participants are providing the relevant information and that they are accurately capturing their experiences. Lowik et al. (2022) offer guidelines for considering how to inclusively communicate and determine eligibility in ways that are attentive to gender's many facets, and we recommend that researchers consult this guide as they develop their research plans.

Regardless of the role of gender in their studies, in their discussions of participants' gender and in the survey questions they use to gather this information, researchers should take care not to reinforce the gender binary or otherwise overlook or erase gender diversity. For example, researchers could report the complete gender breakdown of their samples rather than reporting just the percentage of participants who are in one gender group, which was common among studies in our analysis and which can reinforce the conceptualization of gender as binary (e.g., "Our sample comprised 300 participants, 54% men"). Clearer and more comprehensive reporting practices would make the gender diversity of research samples more transparent.

When asking about gender identity, providing a variety of options, including both binary and nonbinary identities and the opportunity

---

[9] We follow other researchers in using the term "gender/sex" to acknowledge that "gender and sex are, to some degree, mutually constitutive categories that must be considered in tandem" and to "refer to phenomena that cannot be easily separated into sociocultural or biological/bodily components, like whole identities" (Schudson et al., 2019, p. 449).

Table 6
Sample Research Goals and Corresponding Measurement Items and Write-Ups

Focal characteristic	Sample research goal(s) related to variable	Potential measurement item(s)	Example of reporting in "participants" section
Gender	Provide a thorough description of how participants describe their gender	How do you describe your gender? • Man • Woman • Nonbinary • Another gender identity or multiple identities (please specify): Do you describe yourself as "transgender"? • Yes • No • Other (please specify):	Most participants were women (70% women, 24% men, 4% nonbinary, 2% agender). Three percent of the total sample was transgender (4.2% of the men, 1.4% of the women, and 25% of the nonbinary participants).
Sexual orientation	Provide a thorough description of how participants describe their sexual orientation	What is the primary label you use to describe your sexual orientation? • Heterosexual • Gay • Lesbian • Bisexual • Pansexual • Queer • Fluid • Asexual • I include more than one of these terms in my primary sexual orientation label (please specify) • Another identity label (please specify):	The sample was fairly diverse with respect to sexual orientation: 40% were heterosexual, 20% were gay or lesbian, 20% were bisexual, 10% were pansexual, 5% were queer, and 1% each were asexual, fluid, and demisexual.
Regional context	Give appropriate context for where the study took place	Not applicable in this example (regional context is reported here based on researchers' recruitment strategy)	Participants were recruited from a university in the Northeastern United States.
Socioeconomic status (U.S. context)	Test for differences between lower SES versus higher SES participants; give SES context for the sample	What is the highest level of education you have completed? • No or some high school • High school completion • Some college, no degree • Associate degree • Bachelor's degree • Master's degree • Professional degree (e.g., JD, MD) • Doctorate (PhD) What is your annual household income before taxes?	The sample was fairly socioeconomically diverse; 52% of participants had at least a 4-year college degree, whereas 48% did not. The median household income was $50,000.
Socioeconomic status (U.K. context)		What is the highest level of education you have completed? • Primary education • Lower secondary degree (e.g., GCSE, O-levels, or standard grades) • Upper secondary degree (e.g., A-levels or highers/advanced highers) • Vocational degree (e.g., SVQ, HNC, HND) • Undergraduate degree (e.g., BSc, BA, LLB)	

(table continues)

Table 6 (*continued*)

Focal characteristic	Sample research goal(s) related to variable	Potential measurement item(s)	Example of reporting in "participants" section
		• Master's degree (e.g., MSc, MPhil, MBA) • Doctorate or professional degree (e.g., PhD, PsyD, DPhil, MD, JD) • Other educational degree (please specify in the space below): _____	
Race (U.S. context, briefer measure)	Provide a relatively thorough description of how participants describe their race	How do you describe your race? (Check all that apply) • African American, Black, African, or Caribbean • Asian American, Asian, or Pacific Islander • European American, White, Anglo, or Caucasian • Hispanic American, Latinx, or Chicanx • Native American or American Indian • Middle Eastern or North African • Multiracial (i.e., having biological parents from multiple racial backgrounds; feel free to further specify if desired) • Another identity (please specify)	The sample was 24% Black, 20% Asian or Pacific Islander, 50% White, 10% Latinx, 1% Native American, and 5% Middle Eastern or North African. Percentages sum to more than 100 because participants could select multiple racial groups; 24% endorsed "Multiracial" as one of the labels describing their race.
Race (U.S. context, more detailed)	Test for differences in an outcome variable between people of color (POC) and non-POC; provide a thorough description of how participants describe their race; gather information needed to examine variability across different Asian subgroups	How do you describe your race? (Check all that apply) • African American, Black, African, or Caribbean • East Asian or East Asian American • Southeast Asian or Southeast Asian American • South Asian or South Asian American • Pacific Islander • European American, White, Anglo, or Caucasian • Hispanic American, Latinx, or Chicanx • Native American or American Indian • Middle Eastern or North African • Multiracial (i.e., having biological parents from multiple racial backgrounds; feel free to further specify if desired) • Another identity (please specify) Do you identify as a person of color? • Yes • No	The sample was 24% Black, 5% East Asian or East Asian American, 10% Southeast Asian or Southeast Asian American, 3% South Asian or South Asian American, 2% Pacific Islander, 50% White, 10% Latinx, 1% Native American, and 5% Middle Eastern or North African. Percentages sum to more than 100 because participants could select multiple racial groups; 24% endorsed "Multiracial" as one of the labels describing their race. Fifty-four percent of the sample identified as a person of color.

Note. Unless otherwise specified, the questions presented here have a "select one of the choices below" format. The examples represented here are not intended as one-size-fits-all suggestions; they are simply one reasonable (and, in our view, inclusive) way of presenting sample characteristics. The goals presented here are offered as an illustration of how to generate demographic items based on researcher's goals. Many types of goals beyond those represented here are possible for each category. SES = socioeconomic status.

Table 7

Common Reporting Practices, Their Shortcomings, and Potential Alternatives

Example	Shortcomings	Potential alternatives
"The sample was predominantly made up of women (61%)"	Relies on readers to infer that the remaining percentage of the sample is male (an assumption that, without actively naming the gender binary, nonetheless reinforces it)	"The sample was predominantly made up of women (61% women, 39% men)"
"70% men, 29% women, 1% transgender"	Seems to treat men, women, and transgender as mutually exclusive categories, when in fact people often simultaneously identify with categories such as "man" or "woman" *and* transgender	"The sample was 70% men and 30% women, and 1.4% of the men self-described as transgender. The remainder of the sample (99%) self-described as cisgender"
"The sample was made up of 67 heterosexual couples and three lesbian couples."	Uses language that can be misinterpreted to mean that the sample contains 67 couples with heterosexual participants and three couples with lesbian participants; does not provide information about individual participants' sexual orientation	"The sample included 67 man-woman couples and three woman-woman couples. Among participants in the man-woman couples, 85% were heterosexual and 15% were bisexual; among participants in the same-gender couples, 67% were gay/lesbian and 33% were bisexual."
"Participants came from a university in the South."	Relies on the assumption that the default context in which a study took place is in the United States	"Participants came from a university in the Southern United States."
"The sample was predominantly White."	Reports only the percentage of a societally dominant demographic group (White people) without discussing the existence of any other groups; is vague (what is meant by "predominantly"?)	"The sample was 85% White, 6% Black, 4% Asian, 3% Latinx, 2% Middle Eastern, and 1% Native American."

Note. The potential alternatives represented here are not intended as one-size-fits-all suggestions; they are simply one reasonable (and, in our view, inclusive) way of addressing the shortcomings of common reporting practices.

for participants to self-describe their identities, can improve measure inclusivity (Price, 2018). When deciding which options to provide, researchers would benefit from attending to the context in which their study takes place, as some cultures and communities use unique gender labels that may be less common in other contexts (e.g., two-spirit and other nonbinary gender identities within Native American communities; Indian Health Service, n.d.).

Researchers should also carefully consider how transgender and nonbinary participants will encounter their survey questions about gender, as the appropriateness and clarity of the questions can influence the accuracy of the data researchers will receive and can subsequently report. Researchers should not treat "man" (or male), "woman" (or female), and "transgender" as mutually exclusive gender categories when they report or ask questions about participants' gender, and they should not provide only those three categories as options, as seemed to be common among studies examined in our analysis that did mention the inclusion of transgender participants (98.1% of studies did not). Providing only those three options overlooks a diverse array of nonbinary identities and can carry the implication that transgender people cannot be women or men (Ansara & Hegarty, 2014). Moreover, questions that provide these as mutually exclusive response options make it difficult to accurately discern the proportion of cisgender and transgender respondents (Tate et al., 2013). Transgender and gender-diverse participants who were asked to provide suggestions on writing gender questions recommended another approach: ask participants' gender identity and then ask a separate question about whether they identify as transgender and/or gender diverse (Puckett et al., 2020). These approaches, combined with clarity about the reasons for asking participants about their gender, can help to produce the most accurate reflection of participants' gender identities and experiences (Puckett et al., 2020).

Sexual Orientation

The median proportion of gay or lesbian and bisexual+ people in the samples in our analysis was 0%. One contributor to these low rates of representation may be the exclusion of sexual minorities from relationship research (Andersen & Zou, 2015), either by virtue of not being recruited or by being dropped from analyses (Junkins et al., 2022). We noted this trend informally in the coding process and observed that authors provided an array of justifications for their exclusions. Some researchers defaulted to studying heterosexual people without an explanation for why, whereas others excluded sexual minorities because doing so enabled more straightforward statistical analyses or because they lacked a sufficiently large number to allow tests of differences between sexual minority and heterosexual participants. We urge researchers to think carefully about why they are excluding sexual minorities from their analyses (or from their samples to begin with) and to offer clear justifications for their exclusions, justifications that go beyond untested assumptions that sexual minorities must fundamentally differ from heterosexual people and would therefore contribute error to one's analyses. Resources related to conducting dyadic data with indistinguishable dyads (e.g., same-gender couples) have grown in recent years (e.g., Kenny et al., 2020; Sakaluk et al., 2021), a pattern we hope continues and that we hope researchers take advantage of in order to prevent the unnecessary exclusion of sexual minority participants.

In general, for researchers with a small number of sexual minorities in their sample, we recommend including rather than excluding sexual minority participants if they have no compelling theoretical or ethical reason not to. If researchers exclude sexual minorities and *are* motivated by a theoretical rationale for doing so, we recommend that they use those exclusions as a reflection point. Researchers

might ask themselves whether the reason the exclusion is theoretically appropriate is because the theory focuses only on heterosexual people (without acknowledging sexual minorities) or because the theory has engaged with issues of diversity and generated predictions about variability on the basis of sexual orientation. If the former is true, researchers might ask themselves whether and how the underlying theory might be refined to better acknowledge the range of human sexual orientation. If the latter is true, researchers should consider explicitly and intentionally testing for the hypothesized variation using samples with the representation of people from different sexual orientation groups.

Within the realm of relationship science, we encourage researchers reporting on couples to use clear terminology that allows readers insight into participants' sexual orientation and the type of relationship they are in. We discourage researchers from using sexual orientation labels (which describe individuals) to characterize couples. Phrases like "heterosexual couple" and "gay couple" that were common in the studies we analyzed (19% of all studies) are unclear—do "heterosexual" and "gay" refer to the gender make-up of the couple or to the sexual orientation of its members? As discussed earlier, both labels fall short of acknowledging that couple members may not identify as heterosexual or as gay just because they are in different-gender or same-gender relationships, and bisexual people may feel that neither label aligns with their own self-identification. Practices that overlook bisexuality may contribute to feelings of identity denial and invisibility among bisexual people, which are linked to poor well-being (e.g., Garr-Schultz & Gardner, 2021).

Rather than referring to couples composed of men and women as "heterosexual couples" or couples composed of two men or two women as "gay couples" or "lesbian couples," we recommend using language that more directly describes the gender make-up of the couple and separately describes the sexual orientation of its members. For example, if researchers want to convey something about the gender make-up of couples in their sample, they might consider reporting the number of relationships between men and women, between two women, between a nonbinary person and a woman, and so forth. When attempting to capture the relationship experiences of participants, even terms like same-gender/sex and different-gender/sex can be limiting in that they do not necessarily capture the experiences of nonbinary people, who may hold a different gender identity from their partners but whose relationship experiences may more closely resemble those in "same-gender" relationships. If researchers use terms such as "same-gender" or "different-gender," we suggest that they be clear about how nonbinary people fit into these definitions. In some cases, these terms might be entirely unsuitable, as in cases where participants do not identify with a gender at all (e.g., are agender) or reject the notion of classifying their relationship in gendered terms altogether. Where possible, we suggest that researchers ask participants how they categorize their relationships rather than making a classification for them. If researchers are particularly interested in whether participants experience stigma or bias on the basis of their relationships, they also might consider asking that directly rather than attempting to infer that information based on the gender of participants and their partners or their sexual orientation.

We also recommend that, when possible, relationship researchers report the individual sexual orientation identities of their participants in addition to the gender make-up of the relationships participants are in. Sexual orientation cannot be straightforwardly inferred from the gender composition of a relationship, and relationship phenomena can vary based on sexual orientation. For example, some research suggests that while being in a relationship is associated with reduced psychological distress for gay and lesbian individuals, it is associated with increased distress for some bisexual individuals (Whitton et al., 2018). Other research suggests that one's *partner's* sexual orientation can be influential to participants' psychological experiences (e.g., Xavier Hall et al., 2021), further demonstrating the importance of presenting information about the sexual orientation of couple members.

For researchers across research areas reporting on sexual orientation, we recommend considering which aspect of sexual orientation researchers are most interested in. Sexual orientation is multidimensional, encompassing sexual identity, behavior, and attraction (Wolff et al., 2017), and different research questions or motivations may make different aspects of sexual orientation particularly relevant (Salomaa & Matsick, 2019). A researcher interested in studying sexual health behaviors might find it especially important to ask about sexual behavior (e.g., the gender of one's sexual partners and their specific sexual experiences with those partners), whereas a researcher interested in studying identity might find it essential to ask about the sexual identity label participants use. Asking about all three dimensions may help to paint the clearest picture of participants' sexual orientation (e.g., Fu et al., 2019), though researchers hoping to provide a description of how participants describe themselves may find it suitable to ask only about participants' identity labels. When doing so, we recommend that researchers provide an array of potential labels and allow participants to provide their own labels if their identity is not represented among the answer choices. Attending to how the labels they present will be received by participants (e.g., being aware of evidence that gay and lesbian participants anticipate that institutions using the term "homosexual" vs. "lesbian/gay" in demographic questionnaires will be less welcoming and understanding; Matsick et al., 2022) will aid researchers in creating an inclusive survey experience. As mentioned earlier in the discussion, recruiting sexual minority participants and reporting on participants' sexual orientation are not always feasible. When sexual orientation information might be relevant but is unavailable or impractical to collect, we recommend researchers note this aspect of their context.

Regional Context

In our discussion of regional context, we focus on three separate but related issues: geographical location, historical context, and linguistic diversity.

Geographical Location. Particularly given robust evidence that many psychological phenomena are shaped by and can vary across cultures and regions (e.g., Heine & Hamamura, 2007; Kagitcibasi & Berry, 1989; Lehman et al., 2004; Markus & Kitayama, 1991; Triandis & Suh, 2002; Wang, 2021), researchers should be explicit about the country or region where their research took place. While many studies did report the national context of their samples in our analysis (57.4%), many that did *not* report the national context left the reader to infer where a sample was collected (e.g., by reporting that a sample was collected at a "Southern university" without specifying that the university is in the United States; 7.8% of all studies). This notion of the United States as default, which pervades researchers'

discussion of their work (e.g., Castro Torres & Alburez-Gutierrez, 2022; Cheon et al., 2020; Kahalon et al., 2021), is one researchers must actively work against in the interest of greater sample diversity. Given our findings that studies led by first authors from the United States were less likely to report the regional context of their studies, we especially encourage U.S. authors to keep this issue top of mind when reporting on their samples.

Reporting on the country where a research study took place is, as noted previously, also key in relationship to the other demographic features of a sample. For example, perceptions of who is considered prototypically Asian differ by country (Goh & McCue, 2021), demonstrating the importance of considering national context when thinking and writing about racial and ethnic categories. Additionally, indicators of SES such as annual income may carry very different consequences in some countries than others—an income equivalent to $10,000 U.S. dollars might put somebody lower on the SES hierarchy in the United States than they might be in other nations, for example.

Historical Context. Historical context can give rise to or help to explain geographic or cross-cultural variation (Muthukrishna et al., 2021), and national and historical context can work hand in hand to shape the meaning of identity labels and which groups are marginalized and in what ways. For example, after the 9/11 terrorist attacks in the United States, Muslim Americans were racialized to a greater extent, repeatedly questioned by others about their citizenship and loyalties to the United States (Selod, 2015). Knowing that a study of Muslim Americans took place before or after September 11, 2001, would be essential to fully understanding the context those participants were living in, just as the historical moment in which a study of South Africans took place would be essential to understanding the nature of apartheid at the time and its effect on participants' experiences of social relationships, discrimination, and belonging. When studying same-sex couples, knowing the historical context (in combination with the geographical context) may also be key to understanding those couples' experiences of stress—are they living in a place and time when their relationships are socially or legally forbidden? For a recent and widespread example of the importance of attending to context, consider the COVID-19 pandemic. A study of stress conducted in April 2020 (during a relatively early wave of the pandemic) would have different implications than a study of stress conducted in April 2019 (prepandemic) or April 2022 (2 years into the pandemic). Additionally, attending to historical context can help researchers to identify appropriate interventions that address issues such as racism—when considering historical context, researchers can look beyond the individual moment and gain a deeper understanding of systemic issues at play (Trawalter et al., 2022). These examples illustrate the value of reporting when data were collected in addition to where a given study took place.

Linguistic Diversity. Conducting research across a diverse number of regions, including regions with different official languages, does not necessarily mean that speakers of a diverse group of languages are represented. For example, a researcher might recruit participants from a variety of countries but only recruit participants from those countries who are fluent in English, perhaps because they do not have the expertise or resources required to translate their measurement items into different languages. This approach has the obvious limitation of generating a group of participants who may differ in key ways from the typical citizens

of the sampled regions (and of the world), preventing the researchers from fully understanding the experiences of people in those regions. We encourage researchers to transparently report linguistic criteria they may have used if they take such an approach, but we also encourage them to combat the limitations of this tactic by collaborating with scholars from the region of interest, who have both linguistic and experiential expertise that may be of value to the researcher. One useful resource for developing such a collaboration might be the Psychological Science Accelerator, a formalized research network of researchers from around the world that enables scholars to collect large and diverse samples, including samples with participants from different cultures and who speak different languages (Moshontz et al., 2018).

Socioeconomic Status

Researchers frequently did not report SES, and we found no change over time in the percentage of studies reporting SES metrics. Given that SES is robustly linked to differences in cultural values, norms, and social experiences (Stephens et al., 2014; Stephens et al., in press), knowing the SES of participants is crucial for understanding the extent to which constructs in relationship science vary across SES contexts (e.g., Emery & Finkel, 2022), or whether findings from participants in higher SES contexts even generalize to those in lower social class contexts (Ross et al., 2019). Moreover, the ways that SES shapes psychology depend on the broader cultural context in which participants are embedded, with evidence of differences between culturally independent and interdependent cultures (e.g., Miyamoto et al., 2018). Given the overwhelming focus on the United States as the site of research, even work that does report SES may largely do so in the context of independent cultures, leaving much left unknown about SES in other contexts.

In our analysis, we found that there was considerable variation in the aspects of SES reported, as well as variation in how interpretable these metrics were. Consider educational attainment as an example. Reporting the percentages of participants with or without a 4-year degree is especially informative because having a 4-year degree (or not) is a key difference in the material elements, cultural context, and structural constraints of people's lives (Stephens et al., 2014). However, even when samples reported percentages of participants with or without a 4-year degree, this information was not always interpretable if the samples included undergraduate students or individuals of traditional undergraduate age. A sample reporting that 70% of participants did not have a 4-year degree seems socioeconomically diverse, but if 65% are currently undergraduates, this number is misleading. Many studies also reported much less interpretable metrics of education; for example, articles might report the median education level of the sample (e.g., "the median level of education was college") without specifying what their scale points were.

Difficulties in interpretation also arose when researchers reported subjective SES. Although researchers commonly use the MacArthur Ladder of subjective social class (Adler et al., 1994), there were also metrics of subjective social class presented without guidance regarding how to interpret them (e.g., reporting a mean of 4.3 on a given measure without indicating the anchor points, response options, or comparative data for interpretation). When researchers present measures of subjective social class, they should provide a clear elaboration of what their metric means and how to understand its implications for the sample.

Household income is generally more straightforward to interpret, although there are some complexities in making sense of it. Couples often do not know precisely what their household income is and may give substantially different estimates than their partner. In a recent dyadic sample we collected (see Emery & Finkel, 2022), partners' estimates of their household income differed from each other, on average, by $28,652, and only 9% of the sample gave exactly the same estimate as their partner. As a result, it may be more useful to report income in a few different ways—the median, the range, as well as percentages that fall above or below a meaningful number (e.g., the median household income where the data were collected).

Overall, we urge researchers both to report participant social class on multiple indices and in a way that is meaningful for readers to interpret (see, e.g., Diemer et al., 2013). For example, researchers reporting education should only do so for the portion of the sample that is older than, say, 25 years old so that a lack of a 4-year degree does not simply reflect someone's status as a traditional undergraduate. Researchers should also consider which measures of social class are most relevant to their research questions. Different measures of social class are appropriate for different questions—education is optimal for questions about cultural differences or models of self and identity, whereas income is more relevant to questions tied to possessing resources (Stephens et al., in press). Furthermore, the measure of social class that is most relevant may depend on the national and historical context. Educational attainment, for example, is more predictive of outcomes in the modern United States than is occupation (e.g., Adler & Rehkopf, 2008). In other cultural contexts, occupation may be more relevant than education. Reporting and interpreting meaningful, relevant metrics of social class will enable researchers to identify who they are studying and whether their work represents people across the socioeconomic spectrum.

Race/Ethnicity

In our analysis, we found that race was relatively frequently reported for studies taking place in the United States (63.1%). However, for studies taking place in the United States, it was somewhat common for researchers to report only the percentage of White participants in their sample while neglecting to mention other races (22.0%); reporting in this way while focusing on other racial groups (e.g., "The sample was 20% Black") was vanishingly uncommon (0.9%). This tendency to focus on Whiteness was evident in the studies included in our analysis, and echoing others (e.g., Garay & Remedios, 2021; Roberts & Mortenson, 2022), we encourage researchers to decenter Whiteness from their thinking about and reporting on race. Rather than reporting only the percentage of White and "non-White" participants, for example, researchers could instead report the racial breakdown of their samples comprehensively to better acknowledge participant diversity. Our point is not that grouping participants based on broader shared identities (e.g., as people of color) is never useful or theoretically appropriate, but instead that researchers should recognize the nuances within their samples and take care not to inherently other people of color in their discussion of race.

Race was less frequently reported in studies taking place outside of the United States compared to studies in a U.S. context, with around 21% of those studies reporting race. There are several potential explanations for this pattern, which we encourage future researchers to explore. One is that the specific sociohistorical context of the United States makes attending to race more central when researchers who have conducted studies in the United States are reporting on their samples, and that other constructs figure more prominently in other national contexts and are accordingly reported more commonly. Another possibility is that measurement of race/ethnicity is limited in some regions, limiting reporting rates. One set of authors, for example, noted that the country in which their study took place (France) had legislation prohibiting asking questions about ethnicity (IJzerman et al., 2018). We encourage researchers to provide context like this when discussing race, as highlighting contextual variation surrounding race will help to broaden the field's understanding of the topic beyond the U.S. context that characterizes much of the literature.

When considering what specific racial categories to present as options to participants and subsequently report on, researchers should think carefully about how race might influence the phenomena they are investigating. In some cases, the categories researchers often default to may not be specific or inclusive enough to enable the most useful and illuminating analyses or capturing how people really identify. For example, in the United States, when not presented with a Middle Eastern or North African (MENA) response option, MENA participants tended to report that they were White (80%); when a MENA option was available, only 10% identified as only White (Maghbouleh et al., 2022). The questions researchers ask can shape what they're able to investigate and answer—a researcher who offers MENA options would be better able to identify disparities between MENA and White participants that a researcher who offers no such option would never uncover. Additionally, examining Asian Americans as a univalent category rather than disaggregating the category into subgroups can obscure important differences within the Asian American category and between individual Asian American subgroups and other racial categories (e.g., Adia et al., 2020; Holland & Palaniappan, 2012; Schwartz & Jahn, 2022). In some European contexts, where reporting on participants' "cultural identification" (including things like migration background) may have more meaning than reporting on race/ethnicity (Juang et al., 2022), reporting only on participants' "migrant background" without also reporting on other aspects of their experience, such as their heritage and generation status, can also conceal important differences in individuals' experiences (Vietze et al., 2022). Racial, ethnic, and immigrant groups are not homogenous, and the way researchers ask about participant ethnicity and background should enable researchers to capture heterogeneity (Buchanan et al., 2021) and avoid reinforcing exclusionary ideals (Moffitt & Juang, 2019). Researchers would benefit from considering what questions related to race they are interested in asking to ensure that their measures accurately capture important variation.

We also suggest that researchers be inclusive of multiracial identities, especially because failing to present questions about race that capture multiracial identity has been shown to undermine multiracial participants' self-esteem and motivation (Townsend et al., 2009). Researchers should allow participants to select multiple race/ethnicity categories ("check all that apply") so that participants are able to report on all parts of their identity rather than being forced to choose between options that fail to capture their identities. They might also consider including Multiracial as an option, an option suggested by multiracial survey takers themselves (Minniear & Atkin, 2022). If offering "multiracial" as a response option,

researchers should allow participants to provide information about their specific multiracial identity, as multiracial people's experiences can differ on the basis of their specific racial backgrounds (Garay & Remedios, 2021; Parker et al., 2015). When presenting race options, researchers should be clear about how they are defining the terms they present to participants, particularly when providing multiracial options, as terms like "multiracial" and "biracial" are complex and nuanced and are often used in different ways by different people (Atkin et al., 2022).

Finally, we encourage researchers to be mindful of the social, political, historical, and cultural context in which they are attempting to measure race/ethnicity. Even countries within the same region or continent (e.g., Europe) have different ways of understanding and discussing race, meaning that different contexts will require different forms of measurement and elicit unique issues to attend to when reporting on one's sample. Attending to the context in which the research took place will sensitize researchers to issues that may arise when describing their samples. For example, researchers in a German context might be sensitive to the fact that national identity labels in countries like Germany are sometimes used in exclusionary ways to denote membership in the ethnic majority rather than to merely describe one's citizenship (Juang et al., 2022). As a result, to avoid this exclusionary labeling pattern and maintain clarity about which aspect of participants' background they are referring to, researchers might instead choose to report more specific information, such as participants' heritage, migration background status, and generational status (Vietze et al., 2022). For an insightful exploration of measuring constructs related to race/ethnic identity across different regional contexts, we recommend researchers consult Juang et al. (2022)'s discussion of adapting a U.S.-based race/ethnicity intervention to five different countries in Europe.

Fieldwide Recommendations

We join others in advocating for journals to make serious, transparent commitments to diversity in terms of who participates in the research they publish (Roberts et al., 2020). Many journals, spurred in part by the replicability crisis, added "badges" to articles to reward researchers' engagement in open science practices (e.g., preregistration), whereas other practices contributing to a high-quality science (e.g., using non-WEIRD samples) have largely gone unrewarded (Finkel et al., 2017). These open science badges speak to the priority the field places on building a more replicable science but also raise questions about why the field has not enacted similar policy changes to respond to the issue of limited sample diversity, which has been discussed for decades as a threat to the quality and inclusivity of psychological research (e.g., Arnett, 2008; Cundiff, 2012; Graham, 1992; Henrich et al., 2010; Nagayama Hall & Maramba, 2001; Rad et al., 2018; Sears, 1986; Sue, 1999). If the field wishes to increase the diversity of its samples, it must take the issue seriously and enact policy changes that reward and place value on the collection of diverse or understudied samples. Such changes can take any number of forms, including offering badges for publications with samples from underrepresented regions; developing journal diversity task forces; training editorial teams about the value of research focused on marginalized populations; and introducing journal policies requiring the reporting of demographic information, constraints on generality statements, and a certain level of sample diversity if authors wish to generalize their

conclusions to the human population (Buchanan et al., 2021; Cundiff, 2012; Roberts et al., 2020; Simons et al., 2017). Changes like these have the potential to incentivize researchers to collect more diverse samples and to place greater value on the types of work that have been devalued and excluded from the esteemed journals that researchers arguably reap the greatest career rewards from publishing (Nagayama Hall & Maramba, 2001; Settles et al., 2021).

In the interest of building a field that has an inclusive climate for its members and a body of work that is broadly beneficial to the public, psychologists would also benefit from working to implement changes that support diversity (with regards to race, gender, sexual orientation, first language, SES, etc.) in who writes and edits psychological research. The diversity of the field can be influential in efforts to build a well-ordered science, as incorporating the perspectives of diverse researchers can broaden the field's understanding of which people and concepts are important to study, what methodological approaches can be used to study them, and the extent to which the topics scientists study are broadly relevant to society (Kozlowski et al., 2022; Medin et al., 2017). Crucially, the diversity of the field's decisionmakers can shape what research topics and methods are valued and highlighted (Avery et al., 2022)—for example, fewer articles highlighting race were published in psychology journals when editors were White (Roberts et al., 2020), and authors from editor-in-chiefs' home countries are published more often in the journals under those editors' direction (Lin & Li, 2022). The decisions and values of field leaders can have downstream implications for what studies are conducted using what samples, which scholars are given credence and esteem, and who remains in versus leaves the field. Decisions about whom to study and who should study them can reflect whose perspectives are valued in our science (Lewis, 2021), and building a well-ordered science will require not only changes to the individual practices researchers engage in but to the entire research ecosystem that influences how individuals' work is received and rewarded.

Limitations and Future Directions

Although our analysis contributes a number of new insights to our understanding of the samples used in social psychological research, it has a number of limitations. One limitation is that we focused our analysis on just five focal sample characteristics in studies related to romantic relationships published in just eight journals. A number of other characteristics (such as participants' disability status, immigration status, and religion) may also be of interest to researchers, and we encourage further investigation into how these and other qualities are represented and discussed in the literature. We especially encourage investigations from researchers of diverse backgrounds and research areas who may have unique insights into the characteristics relevant in their own region or research domains. While we selected the journals in our analysis because they are high-impact outlets for relationship research within social psychology, the samples included in our analysis are not necessarily representative of all the research samples used to investigate questions related to romantic relationships. The results of our analysis may reflect not what researchers are studying but rather what research these high-impact journals are willing to publish (Diaz & Bergman, 2013), as studies focusing on marginalized, understudied populations are often published in lower impact "specialty" journals (Atherton, 2021; Cortina et al., 2012), a practice exemplifying epistemic

exclusion (Buchanan et al., 2021). Future research could examine differences in the representativeness of samples in lower impact versus higher impact journals to determine how sample diversity is linked to perceived journal prestige. In our own supplementary analyses, field-wide, higher impact journals seemed to report certain characteristics (regional context, race) at lower rates than did "specialty" relationship science journals. We would be interested in seeing this finding further explored in other domains.

Another limitation is that we were unable to compute estimates of the representation of certain demographic groups, such as multiracial people, people of varying SES groups, and some specific racial and sexual orientation groups. This was in part because of the messy ways in which data were reported (i.e., that there were substantial inconsistencies in how authors of different articles acknowledged or reported on these groups). Even the data for the groups we were able to conduct analyses on were often reported imprecisely or inconsistently across articles (e.g., different sets of authors combining different racial groups in distinct ways from one another, sometimes in ways that did not allow us to extract all of the provided information in a way that fit into our coding scheme, which prevented us from generating the most accurate possible estimates). Future researchers should take on the challenge of computing meaningful estimates of the representation of these unexamined groups to give us a deeper understanding of sample diversity in the field, an effort that will be aided by other researchers increasing the thoroughness and inclusivity of their reporting practices.

One promising future direction for this line of work is examination of the diversity of relationship science (and social–personality psychology) researchers themselves. As others have discussed (e.g., Medin et al., 2017), the diversity of researchers can be highly influential to what work is valued, and examining the diversity of researchers—along with the patterns of inequity that affect marginalized scholars (Ledgerwood et al., 2022)—could help us to better understand the larger context that has given rise to the state of the science as exemplified in our analysis.

Conclusion

Although they endeavor to use their science to benefit all people, social and personality psychologists frequently study only a small, nonrepresentative sector of the population. In our analysis of 1,762 relationship-relevant studies published in psychological and relationship science outlets between 1996 and 2000 and between 2016 and 2020, we find that improvement on this issue has been modest at best. While some improvement has occurred in the extent to which researchers even report or acknowledge demographic characteristics such as sexual orientation and race, samples continue to include only small proportions of people of color, gender and sexual minorities, and participants from countries outside the United States. Moreover, scholars continue to use language in their reporting of demographic characteristics that centers on dominant groups (e.g., White people) and overlooks marginalized populations (e.g., nonbinary people). In other words, researchers tend to report on "mostly White, heterosexual couples," rather than reporting on diverse samples in inclusive ways. If scientists wish to develop an inclusive science that broadly benefits humanity, they must work to improve both their reporting practices and the diversity of their samples. By drawing greater attention to these issues, we are hopeful that researchers will make greater progress in the next 20 years than has occurred in the last 20.

References

Adia, A. C., Nazareno, J., Operario, D., & Ponce, N. A. (2020). Health conditions, outcomes, and service access among Filipino, Vietnamese, Chinese, Japanese, and Korean adults in California, 2011–2017. *American Journal of Public Health*, *110*(4), 520–526. https://doi.org/10.2105/AJPH.2019.305523

Adler, N. E., Boyce, T., Chesney, M. A., Cohen, S., Folkman, S., Kahn, R. L., & Syme, S. L. (1994). Socioeconomic status and health. The challenge of the gradient. *American Psychologist*, *49*(1), 15–24. https://doi.org/10.1037/0003-066X.49.1.15

Adler, N. E., & Rehkopf, D. H. (2008). U.S. disparities in health: Descriptions, causes, and mechanisms. *Annual Review of Public Health*, *29*(1), 235–252. https://doi.org/10.1146/annurev.publhealth.29.020907.090852

American Psychological Association. (2020, May). *About APA*. https://www.apa.org/about

American Psychological Association. (2021). *Inclusive language guidelines*. https://www.apa.org/about/apa/equity-diversity-inclusion/language-guidelines.pdf

Andersen, J. P., & Zou, C. (2015). Exclusion of sexual minority couples from research. *Health Science Journal*, *9*(6), 1–9.

Ansara, Y. G., & Hegarty, P. (2014). Methodologies of misgendering: Recommendations for reducing cisgenderism in psychological research. *Feminism & Psychology*, *24*(2), 259–270. https://doi.org/10.1177/0959353514526217

Arnett, J. J. (2008). The neglected 95%: Why American psychology needs to become less American. *American Psychologist*, *63*(7), 602–614. https://doi.org/10.1037/0003-066X.63.7.602

Atherton, O. E. (2021, July). Deconstructing problematic peer reviews in personality psychology and some calls to action. In L. A. King & J. A. Booker (Chairs), *"How open is personality psychology to researchers from marginalized communities?"* Paper presented at the biennial meeting of the Association for Research in Personality (virtual).

Atkin, A. L., Christophe, N. K., Stein, G. L., Gabriel, A. K., & Lee, R. M. (2022). Race terminology in the field of psychology: Acknowledging the growing multiracial population in the U.S. *American Psychologist*, *77*(3), 381–393. https://doi.org/10.1037/amp0000975

Avery, D. R., Dumas, T. L., George, E., Joshi, A., Loyd, D. L., van Knippenberg, D., Wang, M., & Xu, H. H. (2022). Racial biases in the publication process: Exploring expressions and solutions. *Journal of Management*, *48*(1), 7–16. https://doi.org/10.1177/01492063211030561

Barker, G., & Kitcher, P. (2014). *Philosophy of science: A new introduction*. Oxford University Press.

Beischel, W. J., Schudson, Z. C., Hoskin, R. A., & van Anders, S. M. (2022). The gender/sex 3× 3: Measuring and categorizing gender/sex beyond binaries. *Psychology of Sexual Orientation and Gender Diversity*. Advance online publication. https://doi.org/10.1037/sgd0000558

Berscheid, E. (1999). The greening of relationship science. *American Psychologist*, *54*(4), 260–266. https://doi.org/10.1037/0003-066X.54.4.260

Brewer, M. B. (2000). Research design and issues of validity. In H. T. Reis & C. M. Judd (Eds.), *The handbook of research methods in personality and social psychology* (pp. 3–16). Cambridge University Press.

Buchanan, N. T., Perez, M., Prinstein, M. J., & Thurston, I. B. (2021). Upending racism in psychological science: Strategies to change how science is conducted, reported, reviewed, and disseminated. *American Psychologist*, *76*(7), 1097–1112. https://doi.org/10.1037/amp0000905

Call, C., Eckstrand, K., Kasparek, S. W., Boness, C. L., Blatt, L. R., Jamal-Orozco, N. P., & Foti, D. (2022). An ethics and social justice approach to collecting and using demographic data for psychological researchers. *Perspsectives on Psychological Science*. Advance online publication. https://doi.org/10.1177/17456916221137350

Campbell, L., & Simpson, J. A. (2013). The blossoming of relationship science. In J. A. Simpson & L. Campbell (Eds.), *The Oxford handbook of close relationships* (pp. 3–12). Oxford University Press.

Campos, B., Ullman, J. B., Aguilera, A., & Dunkel Schetter, C. (2014). Familism and psychological health: The intervening role of closeness and social support. *Cultural Diversity & Ethnic Minority Psychology*, *20*(2), 191–201. https://doi.org/10.1037/a0034094

Castro Torres, A. F., & Alburez-Gutierrez, D. (2022). North and South: Naming practices and the hidden dimension of global disparities in knowledge production. *Proceedings of the National Academy of Sciences of the United States of America*, *119*(10), Article e2119373119. https://doi.org/10.1073/pnas.2119373119

Cheon, B. K., Melani, I., & Hong, Y. 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 & Personality Science*, *11*(7), 928–937. https://doi.org/10.1177/1948550620927269

Chopik, W. J. (2019, February). *Friendship relationships across the lifespan and around the world: Mechanisms linking relationship quality to well-being, health, and mortality* [Paper presentation]. The Close Relationships Preconference at the annual meeting of the Society for Personality and Social Psychology, Portland, OR.

Cook, R., & Over, H. (2021). Why is the literature on first impressions so focused on White faces? *Royal Society Open Science*, *8*(9), Article 211146. https://doi.org/10.1098/rsos.211146

Cortina, L. M., Curtin, N., & Stewart, A. J. (2012). Where is social structure in personality research?: A feminist analysis of publication trends. *Psychology of Women Quarterly*, *36*(3), 259–273. https://doi.org/10.1177/0361684312448056

Cundiff, J. L. (2012). Is mainstream psychological research "womanless" and "raceless"? An updated analysis. *Sex Roles*, *67*(3), 158–173. https://doi.org/10.1007/s11199-012-0141-7

Davila, J., Jabbour, J., Dyar, C., & Feinstein, B. A. (2019). Bi+ visibility: Characteristics of those who attempt to make their bisexual+ identity visible and the strategies they use. *Archives of Sexual Behavior*, *48*(1), 199–211. https://doi.org/10.1007/s10508-018-1284-6

Diaz, I., & Bergman, M. E. (2013). It's not us, it's you: Why isn't research on minority workers appearing in our "top-tier" journals? *Industrial and Organizational Psychology: Perspectives on Science and Practice*, *6*(1), 70–75. https://doi.org/10.1111/iops.12010

Diemer, M. A., Mistry, R. S., Wadsworth, M. E., López, I., & Reimers, F. (2013). Best practices in conceptualizing and measuring social class in psychological research. *Analyses of Social Issues and Public Policy*, *13*(1), 77–113. https://doi.org/10.1111/asap.12001

Eastwick, P. W., Richeson, J. A., Son, D., & Finkel, E. J. (2009). Is love colorblind? Political orientation and interracial romantic desire. *Personality and Social Psychology Bulletin*, *35*(9), 1258–1268. https://doi.org/10.1177/0146167209338524

Ebbert, D. (2019). *chisq.posthoc.test: A post hoc analysis for Pearson's Chi-squared test for count data*. R package Version 0.1.2. https://CRAN.R-project.org/package=chisq.posthoc.test

Emery, L. F., & Finkel, E. J. (2022). Connect or protect? Social class and self-protection in romantic relationships. *Journal of Personality and Social Psychology*, *122*(4), 683–699. https://doi.org/10.1037/pspi0000368

Feinstein, B. A., & Dyar, C. (2018). Romantic and sexual relationship experiences among bisexual individuals. In D. J. Swan & S. Habibi (Eds.), *Bisexuality* (pp. 145–163). Springer. https://doi.org/10.1007/978-3-319-71535-3_9

Fingerhut, A. W., & Peplau, L. A. (2012). Same-sex romantic relationships. In C. J. Patterson & A. R. D'Augelli (Eds.), *Handbook of psychology and sexual orientation* (pp. 165–178). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199765218.003.0012

Finkel, E. J., Eastwick, P. W., & Reis, H. T. (2017). Replicability and other features of a high-quality science: Toward a balanced and empirical

approach. *Journal of Personality and Social Psychology*, *113*(2), 244–253. https://doi.org/10.1037/pspi0000075

Finkel, E. J., Hui, C. M., Carswell, K. L., & Larson, G. M. (2014). The suffocation of marriage: Climbing Mount Maslow without enough oxygen. *Psychological Inquiry*, *25*(1), 1–41. https://doi.org/10.1080/1047840X.2014.863723

Finkel, E. J., Slotter, E. B., Luchies, L. B., Walton, G. M., & Gross, J. J. (2013). A brief intervention to promote conflict reappraisal preserves marital quality over time. *Psychological Science*, *24*(8), 1595–1601. https://doi.org/10.1177/0956797612474938

Fu, T. C., Herbenick, D., Dodge, B., Owens, C., Sanders, S. A., Reece, M., & Fortenberry, J. D. (2019). Relationships among sexual identity, sexual attraction, and sexual behavior: Results from a nationally representative probability sample of adults in the United States. *Archives of Sexual Behavior*, *48*(5), 1483–1493. https://doi.org/10.1007/s10508-018-1319-z

Garay, M. M., & Remedios, J. D. (2021). A review of White-centering practices in multiracial research in social psychology. *Social and Personality Psychology Compass*, *15*(10), Article e12642. https://doi.org/10.1111/spc3.12642

Garr-Schultz, A., & Gardner, W. L. (2021). "It's just a phase": Identity denial experiences, self-concept clarity, and emotional well-being in bisexual individuals. *Self and Identity*, *20*(4), 528–544. https://doi.org/10.1080/15298868.2019.1625435

Gawronski, B., & Bodenhausen, G. V. (2015). Theory evaluation. In B. Gawrosnki & G. V. Bodenhausen (Eds.), *Theory and explanation in social psychology* (pp. 3–23). Guilford Press.

Goh, J. X., & McCue, J. (2021). Perceived prototypicality of Asian subgroups in the United States and the United Kingdom. *Journal of Experimental Social Psychology*, *97*, Article 104201. https://doi.org/10.1016/j.jesp.2021.104201

Gottman, J. M., Levenson, R. W., Swanson, C., Swanson, K., Tyson, R., & Yoshimoto, D. (2003). Observing gay, lesbian and heterosexual couples' relationships: Mathematical modeling of conflict interaction. *Journal of Homosexuality*, *45*(1), 65–91. https://doi.org/10.1300/J082v45n01_04

Graham, S. (1992). Most of the subjects were White and middle class": Trends in published research on African Americans in selected APA journals, 1970–1989. *American Psychologist*, *47*(5), 629–639. https://doi.org/10.1037/0003-066X.47.5.629

Gyberg, F., Frisén, A., Syed, M., Wängqvist, M., & Svensson, Y. (2018). "Another kind of Swede" Swedish youth's ethnic identity narratives. *Emerging Adulthood*, *6*(1), 17–31. https://doi.org/10.1177/2167696817690087

Haney Lopez, I. F. (2004). The social construction of race. In J. Rivkin & M. Ryan (Eds.), *Literary theory: An anthology* (2nd ed., pp. 964–974). Blackwell Publishing.

Harkin, B., Webb, T. L., Chang, B. P. I., Prestwich, A., Conner, M., Kellar, I., Benn, Y., & Sheeran, P. (2016). Does monitoring goal progress promote goal attainment? A meta-analysis of the experimental evidence. *Psychological Bulletin*, *142*(2), 198–229. https://doi.org/10.1037/bul0000025

Heine, S. J., & Hamamura, T. (2007). In search of East Asian self-enhancement. *Personality and Social Psychology Review*, *11*(1), 4–27. https://doi.org/10.1177/1088868306294587

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

Holland, A. T., & Palaniappan, L. P. (2012). Problems with the collection and interpretation of Asian-American health data: Omission, aggregation, and extrapolation. *Annals of Epidemiology*, *22*(6), 397–405. https://doi.org/10.1016/j.annepidem.2012.04.001

IJzerman, H., Neyroud, L., Courset, R., Schrama, M., Post, J., & Pronk, T. M. (2018). Socially thermoregulated thinking: How past experiences matter in thinking about our loved ones. *Journal of Experimental Social Psychology*, *79*, 349–355. https://doi.org/10.1016/j.jesp.2018.08.008

Indian Health Service. (n.d.). *Two-spirit: Health resources*.

International Association for Relationship Research. (2020). *International association for relationship research diversity, equity, & inclusion statement.* https://iarr.org/diversitystatement.html

International Lesbian, Gay, Bisexual, Trans and Intersex Association World. (2020). *Lucas Ramon Mendos, Kellyn Botha, Rafael Carrano Lelis, Enrique López de la Peña, Ilia Savelev and Daron Tan, State-Sponsored Homophobia 2020: Global Legislation Overview Update.*

Juang, L. P., Umaña-Taylor, A. J., Schachner, M. K., Frisén, A., Hwang, C. P., Moscardino, U., Motti-Stefanidi, F., Oppedal, B., Pavlopoulos, V., Abdullahi, A. K., Barahona, R., Berne, S., Ceccon, C., Gharaei, N., Moffitt, U., Ntalachanis, A., Pevec, S., Sandberg, D. J., Zacharia, A., & Syed, M. (2022). Ethnic-racial identity in Europe: Adapting the identity project intervention in five countries. *European Journal of Developmental Psychology.* Advance online publication. https://doi.org/10.1080/17405629.2022.2131520

Junkins, E., Dugan, K., Chehab, S., Han, S., Lu, T., Yuan, J., & Derringer, J. (2022, February 16–19). *Exclusion of LGBTQ+ and nonbinary folks in romantic relationships research* [Poster presentation]. Society for Personality and Social Psychology, San Francisco, CA.

Kagitcibasi, C., & Berry, J. W. (1989). Cross-cultural psychology: Current research and trends. *Annual Review of Psychology*, 40(1), 493–531. https://doi.org/10.1146/annurev.ps.40.020189.002425

Kahalon, R., Klein, V., Ksenofontov, I., Ullrich, J., & Wright, S. C. (2021). Mentioning the sample's country in the article's title leads to bias in research evaluation. *Social Psychological & Personality Science.* Advance online publication. https://doi.org/10.1177/19485506211024036

Karney, B. R. (2021). Socioeconomic status and intimate relationships. *Annual Review of Psychology*, 72(1), 391–414. https://doi.org/10.1146/annurev-psych-051920-013658

Kenny, D. A., Kashy, D. A., & Cook, W. L. (2020). *Dyadic data analysis.* Guilford Press.

Kitcher, P. (2011). Science in a democratic society. In W. J. Gonzalez (Ed.), *Scientific realism and democratic society: The philosophy of Philip Kitcher* (pp. 95–112). Brill.

Klein, V., Savaş, Ö., & Conley, T. D. (2022). How *WEIRD* and androcentric is sex research? global inequities in study populations. *Journal of Sex Research*, 59(7), 810–817. https://doi.org/10.1080/00224499.2021.1918050

Kozlowski, D., Larivière, V., Sugimoto, C. R., & Monroe-White, T. (2022). Intersectional inequalities in science. *Proceedings of the National Academy of Sciences of the United States of America*, 119(2), e2113067119. https://doi.org/10.1073/pnas.2113067119

Kurdek, L. A. (2001). Differences between heterosexual-nonparent couples and gay, lesbian, and heterosexual-parent couples. *Journal of Family Issues*, 22(6), 727–754. https://doi.org/10.1177/019251301022006004

Ledgerwood, A. (2016). *The start-local approach* [Conference session]. Talk presented at the Navigating the New Era of Social and Personality Psychology Preconference of the 2016 Annual Convention of the Society for Personality and Social Psychology.

Ledgerwood, A., Hudson, S. T. J., Lewis, N. A., Jr., Maddox, K. B., Pickett, C. L., Remedios, J. D., Cheryan, S., Diekman, A. B., Dutra, N. B., Goh, J. X., Goodwin, S. A., Munakata, Y., Navarro, D. J., Onyeador, I. N., Srivastava, S., & Wilkins, C. L. (2022). The pandemic as a portal: Reimagining psychological science as truly open and inclusive. *Perspectives on Psychological Science*, 17(4), 937–959. https://doi.org/10.1177/17456916211036654

Lehman, D. R., Chiu, C. Y., & Schaller, M. (2004). Psychology and culture. *Annual Review of Psychology*, 55(1), 689–714. https://doi.org/10.1146/annurev.psych.55.090902.141927

Leinonen, J. (2012). Invisible immigrants, visible expats? Americans in Finnish discourses on immigration and internationalization. *Nordic Journal of Migration Research*, 2(3), Article 213. https://doi.org/10.2478/v10202-011-0043-8

Levine, R., Sato, S., Hashimoto, T., & Verma, J. (1995). Love and marriage in eleven cultures. *Journal of Cross-Cultural Psychology*, 26(5), 554–571. https://doi.org/10.1177/0022022195265007

Lewis, N. A., Jr. (2021). What counts as good science? How the battle for methodological legitimacy affects public psychology. *American Psychologist*, 76(8), 1323–1333. https://doi.org/10.1037/amp0000870

Lin, Z., & Li, N. (2022). Global diversity of authors, editors, and journal ownership across subdisciplines of psychology: Current state and policy implications. *Perspectives on Psychological Science.* Advance online publication. https://doi.org/10.1177/17456916221091831

Lindqvist, A., Sendén, M. G., & Renström, E. A. (2021). What is gender, anyway: A review of the options for operationalising gender. *Psychology and Sexuality*, 12(4), 332–344. https://doi.org/10.1080/19419899.2020.1729844

Lowik, A., Cameron, J. J., Dame, J., Ford, J., Pulice-Farrow, L., Salway, T., van Anders, S., & Shannon, K. (2022). *Gender and sex in methods & measurement—Research equity toolkit. "Tool #1: Determining & communicating eligibility.* Centre for Gender and Sexual Health Equity, University of British Columbia.

MacDonald, T. K., Zanna, M. P., & Fong, G. T. (1996). Why common sense goes out the window: Effects of alcohol on intentions to use condoms. *Personality and Social Psychology Bulletin*, 22(8), 763–775. https://doi.org/10.1177/0146167296228001

Maghbouleh, N., Schachter, A., & Flores, R. D. (2022). Middle Eastern and North African Americans may not be perceived, nor perceive themselves, to be White. *Proceedings of the National Academy of Sciences of the United States of America*, 119(7), Article e2117940119. https://doi.org/10.1073/pnas.2117940119

Markus, H. R., & Kitayama, S. (1991). Culture and the self: Implications for cognition, emotion, and motivation. *Psychological Review*, 98(2), 224–253. https://doi.org/10.1037/0033-295X.98.2.224

Matsick, J. L., Kruk, M., Palmer, L., Layland, E. K., & Salomaa, A. C. (2022). Extending the social category label effect to stigmatized groups: Lesbian and gay people's reactions to "homosexual" as a label. *Journal of Social and Political Psychology*, 10(1), 369–390. https://doi.org/10.5964/jspp.6823

Mays, V. M., Cochran, S. D., & Barnes, N. W. (2007). Race, race-based discrimination, and health outcomes among African Americans. *Annual Review of Psychology*, 58(1), 201–225. https://doi.org/10.1146/annurev.psych.57.102904.190212

McLoyd, V. C., & Randolph, S. M. (1985). Secular trends in the study of Afro-American children: A review of child development, 1936–1980. *Monographs of the Society for Research in Child Development*, 50(4–5), 78–92. https://doi.org/10.2307/3333865

Medin, D., Ojalehto, B., Marin, A., & Bang, M. (2017). Systems of (non-) diversity. *Nature Human Behaviour*, 1(5), Article 0088. https://doi.org/10.1038/s41562-017-0088

Minniear, M., & Atkin, A. (2022). Exploring Multiracial identity, demographics, and the first period identity crisis: The role of the 2020 United States Census in promoting monocentric norms. *Journal of Applied Communication Research.* Advance online publication. https://doi.org/10.1080/00909882.2022.2107401

Miyamoto, Y., Yoo, J., Levine, C. S., Park, J., Boylan, J. M., Sims, T., Markus, H. R., Kitayama, S., Kawakami, N., Karasawa, M., Coe, C. L., Love, G. D., & Ryff, C. D. (2018). Culture and social hierarchy: Self- and other-oriented correlates of socioeconomic status across cultures. *Journal of Personality and Social Psychology*, 115(3), 427–445. https://doi.org/10.1037/pspi0000133

Moffitt, U., & Juang, L. P. (2019). Who is "German" and who is a "migrant?" Constructing Otherness in education and psychology research. *European Educational Research Journal*, 18(6), 656–674. https://doi.org/10.1177/1474904119827459

Moshontz, H., Campbell, L., Ebersole, C. R., IJzerman, H., Urry, H. L., Forscher, P. S., Grahe, J. E., McCarthy, R. J., Musser, E. D., Antfolk, J.,

Castille, C. M., Evans, T. R., Fiedler, S., Flake, J. K., Forero, D. A., Janssen, S. M. J., Keene, J. R., Protzko, J., Aczel, B., … Chartier, C. R. (2018). The Psychological Science Accelerator: Advancing psychology through a distributed collaborative network. *Advances in Methods and Practices in Psychological Science*, *1*(4), 501–515. https://doi.org/10.1177/2515245918797607

Muthukrishna, M., & Henrich, J. (2019). A problem in theory. *Nature Human Behaviour*, *3*(3), 221–229. https://doi.org/10.1038/s41562-018-0522-1

Muthukrishna, M., Henrich, J., & Slingerland, E. (2021). Psychology as a historical science. *Annual Review of Psychology*, *72*(1), 717–749. https://doi.org/10.1146/annurev-psych-082820-111436

Nagayama Hall, G. C., & Maramba, G. G. (2001). In search of cultural diversity: Recent literature in cross-cultural and ethnic minority psychology. *Cultural Diversity & Ethnic Minority Psychology*, *7*(1), 12–26. https://doi.org/10.1037/1099-9809.7.1.12

Parker, K., Morin, R., Horowitz, J. M., Lopez, M. H., & Rohal, M. (2015, June 11). *Proud, diverse and growing in numbers*. Pew Research Center. https://www.pewresearch.org/social-trends/2015/06/11/multiracial-in-america/

Pederson, J. R., & Mclaren, R. M. (2017). Indirect effects of supportive communication during conversations about coping with relational transgressions. *Personal Relationships*, *24*(4), 804–819. https://doi.org/10.1111/pere.12214

Pollet, T. V., & Saxton, T. K. (2019). How diverse are the samples used in the journals 'Evolution & Human Behavior' and 'Evolutionary Psychology'? *Evolutionary Psychological Science*, *5*(3), 357–368. https://doi.org/10.1007/s40806-019-00192-2

Popa-Velea, O., Diaconescu, L. V., Ion, F. A., Ştefancu, D., & Truţescu, C. (2019). Attachment style and perceived discrimination: Associations with depression among young lesbian, gay, and bisexual Romanian adults. *Journal of Social and Personal Relationships*, *36*(10), 3259–3272. https://doi.org/10.1177/0265407518815775

Price, D. (2018). *Bad gender measures and how to avoid them*. Medium. https://devonprice.medium.com/bad-gender-measures-how-to-avoid-them-23b8f3a503a6

Puckett, J. A., Brown, N. C., Dunn, T., Mustanski, B., & Newcomb, M. E. (2020). Perspectives from transgender and gender diverse people on how to ask about gender. *LGBT Health*, *7*(6), 305–311. https://doi.org/10.1089/lgbt.2019.0295

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

R Core Team. (2022). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Reis, H. T. (2007). Steps toward the ripening of relationship science. *Personal Relationships*, *14*(1), 1–23. https://doi.org/10.1111/j.1475-6811.2006.00139.x

Reis, H. T., Aron, A., Clark, M. S., & Finkel, E. J. (2013). Ellen Berscheid, Elaine Hatfield, and the emergence of relationship science. *Perspectives on Psychological Science*, *8*(5), 558–572. https://doi.org/10.1177/1745691613497966

Remedios, J. D. (2022). Psychology must grapple with whiteness. *Nature Reviews Psychology*, *1*, 125–126. https://doi.org/10.1038/s44159-022-00024-4

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

Roberts, S. O., & Mortenson, E. (2022). Challenging the White = Neutral framework in psychology. *Perspectives on Psychological Science*. Advance online publication. https://doi.org/10.1177/17456916221077117

Ross, J. M., Karney, B. R., Nguyen, T. P., & Bradbury, T. N. (2019). Communication that is maladaptive for middle-class couples is adaptive for socioeconomically disadvantaged couples. *Journal of Personality and Social Psychology*, *116*(4), 582–597. https://doi.org/10.1037/pspi0000158

Ross, L., Lepper, M., & Ward, A. (2010). History of social psychology: Insights, challenges, and contributions to theory and application. In S. T. Fiske, D. T. Gilbert, & G. Lindzey (Eds.), *Handbook of social psychology* (pp. 3–50). Wiley.

RStudio Team. (2022). *RStudio: Integrated development environment for R*. RStudio, PBC. http://www.rstudio.com/

Sakaluk, J. K., Fisher, A. N., & Kilshaw, R. E. (2021). Dyadic measurement invariance and its importance for replicability in romantic relationship science. *Personal Relationships*, *28*(1), 190–226. https://doi.org/10.1111/pere.12341

Salomaa, A. C., & Matsick, J. L. (2019). Carving sexuality at its joints: Defining sexual orientation in research and clinical practice. *Psychological Assessment*, *31*(2), 167–180. https://doi.org/10.1037/pas0000656

Schudson, Z. C., Beischel, W. J., & van Anders, S. M. (2019). Individual variation in gender/sex category definitions. *Psychology of Sexual Orientation and Gender Diversity*, *6*(4), 448–460. https://doi.org/10.1037/sgd0000346

Schwartz, G. L., & Jahn, J. L. (2022). Disaggregating Asian American and Pacific Islander risk of fatal police violence. *PLOS ONE*, *17*(10), Article e0274745. https://doi.org/10.1371/journal.pone.0274745

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), 515–531. https://doi.org/10.1037/0022-3514.51.3.515

Selod, S. (2015). Citizenship denied: The racialization of muslim american men and women post-9/11. *Critical Sociology*, *41*(1), 77–95. https://doi.org/10.1177/0896920513516022

Settles, I. H., Jones, M. K., Buchanan, N. T., & Dotson, K. (2021). Epistemic exclusion: Scholar(ly) devaluation that marginalizes faculty of color. *Journal of Diversity in Higher Education*, *14*(4), 493–507. https://doi.org/10.1037/dhe0000174

Simon, P. (2017). The failure of the importation of ethno-racial statistics in Europe: debates and controversies. *Ethnic and Racial Studies*, *40*(13), 2326–2332. https://doi.org/10.1080/01419870.2017.1344278

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

Smedley, A., & Smedley, B. D. (2005). Race as biology is fiction, racism as a social problem is real: Anthropological and historical perspectives on the social construction of race. *American Psychologist*, *60*(1), 16–26. https://doi.org/10.1037/0003-066X.60.1.16

Society for Personality and Social Psychology. (n.d.). *About SPSP*. https://www.spsp.org/about

Stephens, N. M., Emery, L. F., & Townsend, S. S. M. (in press). Social class. In D. T. Gilbert, S. T. Fiske, E. J. Finkel, & W. B. Mendes (Eds.), *The Handbook of Social Psychology* (6th ed.). HSP Foundation.

Stephens, N. M., Markus, H. R., & Phillips, L. T. (2014). Social class culture cycles: How three gateway contexts shape selves and fuel inequality. *Annual Review of Psychology*, *65*(1), 611–634. https://doi.org/10.1146/annurev-psych-010213-115143

Sue, S. (1999). Science, ethnicity, and bias: Where have we gone wrong? *American Psychologist*, *54*(12), 1070–1077. https://doi.org/10.1037/0003-066X.54.12.1070

Tate, C. C., Ledbetter, J. N., & Youssef, C. P. (2013). A two-question method for assessing gender categories in the social and medical sciences. *Journal*

of Sex Research, *50*(8), 767–776. https://doi.org/10.1080/00224499.2012.690110

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

Timmer, S. G., Veroff, J., & Hatchett, S. (1996). Family ties and marital happiness: The different marital experiences of Black and White newlywed couples. *Journal of Social and Personal Relationships*, *13*(3), 335–359. https://doi.org/10.1177/0265407596133003

Townsend, S. S., Markus, H. R., & Bergsieker, H. B. (2009). My choice, your categories: The denial of multiracial identities. *Journal of Social Issues*, *65*(1), 185–204. https://doi.org/10.1111/j.1540-4560.2008.01594.x

Trawalter, S., Higginbotham, G. D., & Henderson, K. (2022). Social Psychological research on racism and the importance of historical context: Implications for policy. *Current Directions in Psychological Science*. Advance online publication. https://doi.org/10.1177/09637214221114092

Triandis, H. C., & Suh, E. M. (2002). Cultural influences on personality. *Annual Review of Psychology*, *53*(1), 133–160. https://doi.org/10.1146/annurev.psych.53.100901.135200

United Nations. (n.d.). *SDG indicators—SDG indicators*. Retrieved September 29, 2022, from https://unstats.un.org/sdgs/indicators/regional-groups/

U.S. Census Bureau. (2021). *Quick facts*. https://www.census.gov/quickfacts/fact/table/US/PST045221

Van Bavel, J. J., Mende-Siedlecki, P., Brady, W. J., & Reinero, D. A. (2016). Contextual sensitivity in scientific reproducibility. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(23), 6454–6459. https://doi.org/10.1073/pnas.1521897113

van der Toorn, J., Jost, J. T., Packer, D. J., Noorbaloochi, S., & Van Bavel, J. J. (2017). In defense of tradition: Religiosity, conservatism, and opposition to same-sex marriage in North America. *Personality and Social Psychology Bulletin*, *43*(10), 1455–1468. https://doi.org/10.1177/0146167217718523

Vangelisti, A. L., & Caughlin, J. P. (1997). Revealing family secrets: The influence of topic, function, and relationships. *Journal of Social and Personal Relationships*, *14*(5), 679–705. https://doi.org/10.1177/0265407597145006

Vietze, J., Schwarzenthal, M., Moffitt, U., & Civitillo, S. (2022). Beyond 'migrant background': How to select relevant, social justice oriented, and feasible social categories in educational research. *European Journal of Psychology of Education*. Advance online publication. https://doi.org/10.1007/s10212-022-00611-2

Wang, Q. (2021). The cultural foundation of human memory. *Annual Review of Psychology*, *72*(1), 151–179. https://doi.org/10.1146/annurev-psych-070920-023638

Whitton, S. W., Dyar, C., Newcomb, M. E., & Mustanski, B. (2018). Romantic involvement: A protective factor for psychological health in racially-diverse young sexual minorities. *Journal of Abnormal Psychology*, *127*(3), 265–275. https://doi.org/10.1037/abn0000332

Williams, M. J., Gruenfeld, D. H., & Guillory, L. E. (2017). Sexual aggression when power is new: Effects of acute high power on chronically low-power individuals. *Journal of Personality and Social Psychology*, *112*(2), 201–223. https://doi.org/10.1037/pspi0000068

Williamson, H. C., Bornstein, J. X., Cantu, V., Ciftci, O., Farnish, K. A., & Schouweiler, M. T. (2022). How diverse are the samples used to study intimate relationships? A systematic review. *Journal of Social and Personal Relationships*, *39*(4), 1087–1109. https://doi.org/10.1177/02654075211053849

Wolff, M., Wells, B., Ventura-DiPersia, C., Renson, A., & Grov, C. (2017). Measuring sexual orientation: A review and critique of US data collection efforts and implications for health policy. *Journal of Sex Research*, *54*(4–5), 507–531. https://doi.org/10.1080/00224499.2016.1255872

Wu, D. C., Kim, H. S., & Collins, N. L. (2021). Perceived responsiveness across cultures: The role of cultural fit in social support use. *Social and Personality Psychology Compass*, *15*(9). Article e12634. https://doi.org/10.1111/spc3.12634

Xavier Hall, C. D. X., Feinstein, B. A., Sales, J. M., Girod, C., & Yount, K. M. (2021). Outness, discrimination, and depressive symptoms among bi+ women: The roles of partner gender and sexual identity. *Journal of Bisexuality*, *21*(1), 24–41. https://doi.org/10.1080/15299716.2021.1886219

Zilioli, S., Ponzi, D., Henry, A., Kubicki, K., Nickels, N., Wilson, M. C., & Maestripieri, D. (2016). Interest in babies negatively predicts testosterone responses to sexual visual stimuli among heterosexual young men. *Psychological Science*, *27*(1), 114–118. https://doi.org/10.1177/0956797615615868