Bringing an Intersectional Lens to “Open” Science: An Analysis of Representation in the Reproducibility Project

Natalie J. Sabik, Jes L. Matsick, Kaitlin McCormick-Huhn, and Elizabeth R. Cole

Abstract
Feminist psychologists have called for researchers to consider the social and historical context and the multidimensionality of participants in research studies. The Reproducibility Project documents the degree to which the findings from mainstream psychological studies are reproduced. Drawing on intersectionality theory, we question the value of reproducing findings while ignoring who is represented, intersecting social and group identities, sociohistorical context, and the power and privilege that likely influence participants’ responses in psychology experiments. To critically examine the Reproducibility Project in psychology, we analyzed the 100 replication reports produced between 2011 and 2014 (Open Science Collaboration, 2015). We developed an intersectional analytic framework to investigate (a) representation, (b) whether demographic and identity factors were considered through a multidimensional or intersectional lens, (c) explanations of non-replication, and (d) whether socio-cultural context was considered. Results show that reports predominantly include WEIRD samples (people from Western, educated, industrialized, rich, and democratic countries). Context and identity were rarely considered, even when study design relied on these factors, and intersectional identities and structures (considering power, structural issues, discrimination, and historical context) were absent from nearly all reports.

Keywords
open science, feminism, reproducibility, intersectionality, WEIRD, representation

Beginning in November 2011, the Open Science Collaboration, a group of scholars aiming to “increase openness, integrity, and reproducibility of scholarly research (Nosek, 2017, p. 6),” began the Reproducibility Project to conduct replications of 100 experimental and correlational studies selected from three leading psychology journals (Open Science Collaboration, 2015). Noting that scientific debates are meaningless when the evidence being debated is not reproducible (Open Science Collaboration, 2015), this large-scale effort relied on collaboration among researchers to repeat, as closely as possible, the studies published in 2008 in “three important psychology journals: Psychological Science (PSCI), Journal of Personality and Social Psychology (JPSP), and Journal of Experimental Psychology: Learning, Memory, and Cognition (JEP: LMC)” (Open Science Collaboration, 2015, p. 1). Central to this approach is the claim that there are core, universal, reproducible patterns that psychologists should be able to quantify and observe repeatedly. Notably, this approach does not consider the human input into both sides of the research process: both the decision-making of the researchers (e.g., decisions about what to study, whom to study, how to quantify and measure constructs, what statistical tests are appropriate, and how to interpret findings) and the participants (e.g., the individuality, social identity, and social and historical context).

Recently, researchers have noted that the movement to improve science includes two main approaches: open science, which aims to share research assets, such as data and code, and reproducibility, which aims to reproduce and replicate...
previous research (Murphy et al., 2020). Scholars have noted that replication approaches may include several aims, including verifiability (testing the same claims on the original data), robustness (reexamining original data with alternative specifications), repeatability (collecting new data to examine whether previous results are obtained using the same methods), and generalizability (whether similar results hold across different groups and settings; Freese & Peterson, 2017). The Reproducibility Project in psychology focused primarily on repeatability of previous research and the generalizability of the findings; as noted by the Open Science Collaboration (2015), “even research of exemplary quality may have irreproducible empirical findings because of random or systematic error” (p. 1). Interestingly, the authors indicate with this statement that the expectation is that exemplary quality research will reproduce the same findings, except for when there is error in the research design.

This approach assumes that when science is sufficiently rigorous, enduring, and universal, patterns of human psychology can be uncovered and replicated. However, mainstream psychological research tends to ignore the human qualities of the research process and the meaningful variation that may occur among both the researchers and participants. For example, research that focuses on the experiences of the dominant groups tends to disproportionately represent White, upper middle class, male, able-bodied, heterosexual cisgender adults, and often defaults to representing these experiences as universal and defining of human experience.

A growing body of work critiques the pattern in psychological research of prioritizing WEIRD participants (people from Western, Educated, Industrialized, Rich, and Democratic nations) and generalizing from this data as if WEIRD samples represent all of human experience (Arnett, 2016; Burns et al., 2019; Hendriks et al., 2019; Henrich et al., 2010; Roberts et al., 2020). Differences within and between social groups likely produces a high amount of variation when it comes to popular psychological constructs. For example, a review of behavioral science data, including but not limited to popular constructs psychological constructs. For example, a review of behavioral research produces a high amount of variation when it comes to popular constructs that may occur among both the researchers and participants. For example, research that focuses on the experiences of the dominant groups tends to disproportionately represent White, upper middle class, male, able-bodied, heterosexual cisgender adults, and often defaults to representing these experiences as universal and defining of human experience.

A growing body of work critiques the pattern in psychological research of prioritizing WEIRD participants (people from Western, Educated, Industrialized, Rich, and Democratic nations) and generalizing from this data as if WEIRD samples represent all of human experience (Arnett, 2016; Burns et al., 2019; Hendriks et al., 2019; Henrich et al., 2010; Roberts et al., 2020). Differences within and between social groups likely produces a high amount of variation when it comes to popular psychological constructs. For example, a review of behavioral science data, including but not limited to popular constructs such as visual perception, spatial reasoning, categorization and inferential induction, moral reasoning, and self-concept showed significant variability across populations, and suggested that members of WEIRD countries are one of the least representative for generalizing human experience (Henrich et al., 2010).

Psychological scientists suggest that even exemplary research may be irreproducible because of “random or systematic error” (Open Science Collaboration, 2015, p. 1). However, the Reproducibility Project fails to acknowledge that research is bound by time and place and reexamine the lived experiences of the participants in the study (e.g., treating meaningful error as random error); participants who are not genderless, raceless bodies and brains, but people with a complex web of multidimensional and intersecting social identities (Fine, 2008; Harding, 2004; Harding & Norberg, 2005; McCormick-Huhn et al., 2019). We argue that because open science, in particular reproducibility, fails to recognize the intersectional positionality of research participants, as well as the broader sociohistorical structures that influence psychology, the movement leaves much to be explained regarding the non-reproducibility of many key findings in the field. We propose that an intersectional feminist psychological perspective can be applied to identify how WEIRD characteristics of the majority of participants and consideration of relevant intersecting social identity factors and social forces in people’s lives may assist in interpreting failed replication attempts. Further, we suggest that engaging with intersectionality theory in the context of reproducibility provides the opportunity for feminist psychologists (and the field at large) to problematize the significance placed on reproducibility. In the current research, we draw on a feminist framework using an intersectional lens to demonstrate how the Reproducibility Project is limited in its scope by failure to consider intersecting social identities and structural factors that influence psychology.

Feminist psychology has a long history of encouraging researchers to consider the social and historical context and the multidimensionality of participants in research studies (Cole, 2009; Grzanka, 2018; Magnusson & Marecek, 2017; McCormick-Huhn et al., 2019; Rutherford & Davidson, 2019; Shields, 2008; Warner, 2008). Despite the call for inclusivity and transparency by the larger open science movement, by focusing on repeatability and generalizability the Reproducibility Project continues the exclusionary pattern of examining dominant and privileged groups and assuming these patterns are universal. The researchers behind the Reproducibility collaboration lack a clear definition of what counts as “inclusive” in the open science approach, and their efforts toward transparency are more focused on data sharing and procedural aspects of method than around contextual factors and decision-making regarding the inclusion and representation of research participants from non-dominant social groups.

Feminist psychologists have called for an intersectional perspective or lens to critically assess patterns of inclusion and representation in research and have developed questions for critical analysis that uncover broader patterns in psychological research. To assess the representation of different social groups in research produced for the Reproducibility Project, and the extent to which findings are generalizable, our method includes a critical analysis of the replication reports of previously published psychological studies. We aim to do this by systematically addressing a series of questions generated by the authors to investigate (a) who is represented in the studies (e.g., Cundiff, 2012), (b) whether critical demographic and identity factors (i.e., positionality) were considered in the research method and whether these were considered through a multidimensional or intersectional lens, (c) how authors explained non-replication results, and (d) whether authors considered the socio-cultural context.

The current article is the first to our knowledge to systematically analyze the 100 replication reports carried out between 2011 and 2014 (Center for Open Science, 2015) using an intersectional framework. Utilizing intersectionality as a lens through which we view this collection of research...
reports, we explore potential limitations of the reproducibility approach. In doing so, we highlight the ways in which intersectionality theory may explain some of the reproducibility rates that open science advocates deem unsatisfactory. Intersectional theorists have implored researchers to consider the participants in the research, as well as the research design, from a multidimensional perspective (Bowleg, 2008; Cole, 2009; Warner, 2008). This approach calls for investigating both multidimensional and intersecting identities of the participants and study targets, and for examining the structural causes of inequality and difference that consider power, privilege, resources, and policies that shape an individual’s psychology.

An Intersectional Lens for Conducting Scientific Research

African American scholar-activists have long articulated the intersection of race and gender in their lives, from the writings of 19th century African American feminists (see May, 2015) to scholars of color writing in the 1970s and 1980s (see Carastathis, 2016). The term intersectionality was coined by critical race theorist Kimberlé Crenshaw, a legal scholar, in 1989 to illuminate how legal structures define social identities as isolated and mutually exclusive, which effectively erases the experiences of individuals who experience subordination based on multiple identities, such as Black women (Crenshaw, 1989). This term has been widely adopted and utilized in many fields as well as popular media to analyze the causes and consequences of structures of inequality (Harris & Patton, 2019). This lens calls for scholars to examine the intersection of racial, ethnic, class, ability, age, sexuality, and gender disparities (Thornton-Dill & Zambrana, 2009).

In the 21st century, feminist psychologists recognized the importance of this approach, and numerous theoretical and empirical arguments emerged that framed intersectionality as critical in addressing core issues in psychological research. This includes illustrating who has been included in and left out of research in psychology (Cole, 2009; Purdie-Vaughns & Eibach, 2008; Settles et al., 2020; Warner et al., 2018) and expanding beyond looking at differences to examine similarities between groups with different backgrounds or experiences (Cole, 2009; Cole & Stewart, 2001; Shields & Dicecco, 2011). Further, scholars have called for integrating more meaningful measures of constructs like gender and race into research (Helms et al., 2005; Richman & Zucker, 2019; Sabik, 2016). Another key aspect is the recognition of the multidimensionality of people’s identities (Bowleg, 2013; McCormick-Huhn et al., 2019; Shields, 2008; Warner, 2008; Warner & Shields, 2013), as well as remedying the erasure of those not identified as prototypical members of their social groups (Purdie-Vaughns & Eibach, 2008; Sesko & Biernat, 2016). Last, a critical and understudied issue raised by intersectional theorists is exploring larger structural issues such as policy, access, language, and representation among minority groups (Bowleg, 2020; Versey et al., 2019).

Consideration of the aforementioned issues is critical in order to ground psychological research in people’s actual lived experiences, rather than a theoretical and abstract approximation of what we might expect to see in a genderless, raceless body.

The pattern of ignoring gender, race, and other social markers in psychological research is considerable (Cortina et al., 2012; Cundiff, 2012). Despite substantial evidence that race, gender, age, cultural background, and other identity factors influence social judgments central to psychological research, Hester and Gray (2020) point out that moral psychology tends to rely on controlled experimental scenarios in which the identities of the characters are obscured, rendering them as genderless, raceless, ageless, and devoid of identity. Similarly, a thorough review by Bailey et al. (2019) demonstrated how androcentrism (i.e., the assumption of man as universally representative) pervades both past and present approaches to research in the psychological and health sciences. This pattern reflects a broader assumption that “universal” experiences can be identified if we strip away identity and accept the presumed prototypicality of dominant groups, and yet this approach is not ecologically valid or meaningful. Rather, by failing to identify or name identity factors in research design, we default to the assumption that the dominant group identities (White, male, heterosexual, cisgender, able-bodied, young adult, and residing in WEIRD countries) are unmarked and thus considered neutral. For example, samples based in the United States (US) are likely to be unmarked in article titles, whereas samples emerging from non-WEIRD populations are often explicitly named to qualify the research findings within the article title (e.g., “the effect of X on Y among Chinese consumers;” Cheon et al., 2020, p. 1). As Purdie-Vaughns and Eibach (2008) pointed out, dominant groups’ perspectives become defined as the societal standard. Due to ethnocentrism, heterocentrism, and androcentrism, the experiences of White, cisgender, heterosexual men become coded as the norm or standard standpoint and as a result, non-prototypical group members are invisible (Purdie-Vaughns & Eibach, 2008).

Intersectional theorists ask us to question “who has been included in a category” (Cole, 2009, p. 171), as this draws attention to diversity within groups, and asking this question reveals important patterns in samples of mainstream psychology research. Representation matters in terms of participant identity characteristics, but also in terms of sub-discipline and the theoretical/methodological approaches used in the research. We argue that traditional methodological approaches most often employed in research published in mainstream psychology journals tend to prioritize research based on data that are quick and convenient to collect, based on multiple studies, and include significant findings. These practices overrepresent undergraduate convenience samples (Henrich et al., 2010) and research on non-dominant groups is often pushed toward “specialty” journals (Roberts et al., 2020), including journals publishing feminist research, lesbian, gay, bisexual, transgender, queer, intersex, and asexual
(LGBTQIA) research, research on aging, and research on racial/ethnic minorities.

While the open science movement seeks to address the bias of only publishing significant results by valuing pre-registration and reproducibility, this approach fails to grapple with the broader issue of representation and whether reproducibility is a feasible or realistic goal in all areas of psychological science. Our aim in this paper was to analyze the 100 replication reports to show how the intersectional lens can bring into focus issues of representation and internalized disciplinary norms within studies, how “open” science does not represent the majority of the world’s population, and how the Reproducibility Project, in particular, is ignorant of the social and structural constructs that it perpetuates.

The research team collaboratively developed a set of questions to analyze each of the 100 replication reports included in the Open Science Collaboration’s Reproducibility Project (2015). Our research team carefully considered key sources in intersectional psychological research (see method for more details), and generated the following four major areas of inquiry to address representation from an intersectional perspective:

1. Representation: Who is represented in the samples derived for the replication reports?
2. Research Design: Are demographic and identity factors of participants and targets considered in the research design? If so, were these considered through a multidimensional or intersectional lens?
3. Interpretation of Findings: How were non-replication results explained?
4. Context: Was the socio-cultural context considered in either the study design or the interpretation of the results?

**Method**

To analyze aspects of the reports that addressed each of these areas, specific detail-oriented questions were generated by the research team to gather information from the reports (see Appendix for the full list of questions). The research team consisted of four psychologists who are all trained in feminist theory and scholarship, hold graduate degrees and/or faculty positions in social/personality psychology and women’s studies, and have previously published research based on the intersectionality framework. Questions were derived from theoretical and methodological calls for intersectionality to be integrated into psychological research. In particular, we reviewed and drew on the following resources for this process: Bowleg (2017), Cole (2009), Else-Quest and Hyde (2016a, 2016b), Grzanka (2018), McCormick-Huhn et al. (2019), Moradi and Grzanka (2017), Rosenthal (2016), Shields (2008), Warner (2008), Warner et al. (2018), and Warner and Shields (2013). Dialogue among team members over email involved multiple exchanges of generating and revising questions about the nature of intersectionality, and specifically drawing on these articles, of which the author team are experts (e.g., we routinely teach and research this material). We all reviewed the generated list of questions independently and suggested changes and additions that were then reviewed by the group. Once the team of researchers agreed on the set of questions, each researcher independently reviewed the questions and agreed that the list was representative of intersectional approaches to research. Next, we read through the first five replication reports while considering the generated questions, and everyone on the research team was able to weigh in with changes or additions. The research team agreed on the final list of questions and felt that this captured critical information regarding an intersectional perspective for the analysis. For a list of the questions coded, see Table 1.

Once a final list of questions was determined and agreed upon, the full set of 100 reports were analyzed using content coding. All replication reports are publicly available via the open science framework (see https://osf.io/ezcuj/ for all information about and access to replication reports). The final report for each replication study was downloaded by the first author and the information from the questions was recorded in a spreadsheet shared with the research team. Two additional researchers on the team also coded 40% of the reports total (20% each) to assess inter-rater agreement. Key questions representing each section of the analysis were selected to be analyzed for inter-rater agreement to ensure that the researchers agreed upon the interpretation of the conceptual framework and similarly rated the data. Cohen’s $κ$ was calculated for each pair of coders to determine if there was agreement on rating each replication report. There was strong agreement between the coders ($κ$s ranged from .737 to 1.00, all $p < .001$). Disagreements between the coders were resolved by reviewing the data for the question, assessing the discrepancy, and determining which assessment was the best fit for the data.

This approach allowed us to gather information about representation and replication patterns in the data and to identify broader patterns across the replication reports. The first author did the initial analysis of the reports using the generated questions, and all of the authors reviewed the generated data and contributed additional comments, questions, and information from the replication reports. Last, we recognized a number of themes that emerged related to the intersectional analysis and we noted examples to illustrate these issues.

**Results**

**Representation**

Sample sizes for the studies ranged from 5 to 455,326. The samples skewed toward smaller sizes, with 59% of the reports obtaining a sample size of 100 participants or less. Twenty-one percent of the reports had sample sizes between 100 and 199,
11% included sample sizes between 200 and 300 participants, and 9% included samples larger than 300 participants.

Overall, the replication reports primarily represented cognitive psychology (57%) and social psychology (36%), with other sub-fields of these areas represented by a small number of studies (e.g., personality psychology, biopsychology, and cross-cultural psychology each = 2%, and positive psychology = 1%). Other areas of psychology, such as clinical and counseling psychology, developmental psychology, industrial/organizational psychology, school/educational psychology, and health psychology, were not represented in this replication effort, largely due to the replication project’s focus on only three psychology journals.

Of the demographic information included, gender (60 reports), age (55 reports), and undergraduate student status (50 reports) were most commonly reported. However, 13 reports included no demographic information on the sample recruited and 12 reports only reported that they recruited university students. Much less frequently reported were race/ethnicity (15 reports), sexual orientation (2 reports), relationship status (2 reports), socioeconomic status (1 report), and no studies reported gender identity (e.g., cisgender, transgender, and non-binary). The lack of or partial reporting of relevant demographic information was not consistent with recommendations by the American Psychological Association (APA, 2008, p. 842) for reporting on sample composition and recruitment.

We also noted that demographic information was not reported in a uniform way, and there was a high amount of variability in what was reported as well as how it was presented. For example, some studies reported the mean age of participants, some reported an age range, and some reported no information about age. Some studies reported the percentage of their samples that was White versus non-White, whereas others reported percentages of different ethnic groups, and some reported nothing about participants’ ethnicity or race. Thus, it is not possible to accurately discern the full demographic profiles of the participants from all of the reports, as numerous reports were missing this information.

We utilized all available data from the reports to assess the age, gender, and racial/ethnic composition of the samples. In total, the reports that included data on gender presented data on 3937 men and 6985 women. Proportionally, women were overrepresented in the available data, as 63.95% of the participants. Only three reports noted that they included the option for participants to not disclose their gender, and this...
accounted for seven individuals in total in the reports that included information on gender.

The mean age reported across all of the reports that included information on age was 21.62; thus, on average, participants in the studies were young adults, and many were undergraduates. Of the reports that included information in the demographic section or the participant recruitment information regarding college or university enrollment, 8336 of the participants in the reports were identified as college students. The total sample size for all reports was 15,293 after we excluded one report as an outlier because of its large sample size (sample size = 455,326; Soderberg, 2016). Thus, from the available data, we can definitively conclude that at least 54.51% of the participants were undergraduate students. However, the total number of undergraduate students represented in this research is likely much higher, as a number of reports failed to present any demographic data and did not identify their sampling procedures (Arnett, 2008; Henrich et al., 2010). Twenty-two of the reports did not include information on their sample with regard to recruitment or university student status, and this accounted for 6488 participants. It is possible, and in fact may be likely, that many of these were undergraduate students but this information was not reported. Three of the reports were on samples of children (217 participants) and four of the studies recruited MTurk samples (321 participants).

Only 11 reports provided an explicit breakdown of the racial/ethnic populations represented in the sample. The data provided represented 5,462 participants in total, and analysis showed that of the total, White people accounted for 81.78% (n = 4476), Black/African American people accounted for 7.67% (n = 419), Asian/Asian American people accounted for 5.79% (n = 316), Hispanic/Latinx people accounted for 1.28% (n = 70), Middle Eastern people accounted for .26% (n = 14), and people who were identified as “other” accounted for 3.01% (n = 167). This compares to US population estimates of people identifying as White 76.3%, Black 13.4%, Asian 5.9%, and Hispanic/Latinx 18.5% (US Census, 2019). We note as well that no participants were indicated as Native American. Additionally, one report (n = 204) collected data from Serbians and did not otherwise identify racial/ethnic information about the sample (Lazarevic & Knezevic, 2016). The non-uniformity of reporting demographic information presented challenges in aggregating the data, and it is noteworthy that only a small (11%) proportion of the reports included any information about race/ethnicity at all.

Samples for the replication reports had a high representation from WEIRD countries. Ninety reports collected data from WEIRD countries, two collected data from non-WEIRD countries (Serbia, and Uruguay), two collected data online and did not disclose the country of the participants, and seven reports noted that this information was not available. When considering the information provided in the replication reports about the original study upon which the replication was based, 39 of the reports indicated that both the original sample and the replication sample were from WEIRD countries. Sixty-one reports did not disclose where the data from the original study originated or did not provide enough information to determine if the data were collected in a WEIRD country.

In sum, we concluded that demographic information in these reports tended to be underreported (e.g., race/ethnicity, sexual orientation, gender identity, and socioeconomic status) or inconsistently reported, and certain groups (e.g., women, college students, and people from WEIRD nations) were overrepresented in researchers’ attempts to reproduce previous psychological findings.

Analysis of Original Study Samples

The demographic information for the samples from the original studies upon which the replication reports were based was gathered by the authors for analysis. The provided demographic information from the original samples was compared with the demographic information from the replication to determine whether these matched on key characteristics. Because sample information was not presented in a uniform way, we noted the three most commonly reported demographic characteristics among study populations (undergraduates, the gender, and country) and compared the replication to the original study to determine whether these matched.

The analysis revealed that 56 of the pairs of studies (original and replication) were matched in recruiting undergraduate populations; nine pairs of studies were not matched (i.e., one study used a different population than the other with regard to undergraduate status); seven studies were matched in where they recruited their participants, but this was not at a university; and 28 pairs of studies were missing information on whether they were matched on sampling undergraduates. With regard to gender, 22 of the studies were approximately matched on the ratio of women to men in the study; 13 of the pairs of studies were not matched on gender ratio (i.e., one of the studies had a different ratio of men to women); and for 65 pairs of studies this was unclear due to missing information about the gender composition of at least one of the samples. With regard to identifying the country where the data collection took place via either the researchers identifying where the data collection took place or the nationality of the participants in the sample, nine pairs of studies were explicitly matched on country, 29 pairs of studies were not matched on country, and this was unclear for 62 pairs of studies (i.e., not enough information was provided to assess the country from the demographic information provided). Of the 29 pairs of studies that were not matched on country, all except one were both WEIRD. While it may be possible to infer where the data was collected from other information (e.g., the university affiliation of the authors), this is not always clear. For example, if the study has multiple authors at different universities, it is not always evident where the data were collected. To ensure uniformity and consistency, we did not extrapolate about where samples may have been collected.
The most commonly reported information about the samples from the original studies was whether these were based on undergraduate populations. Of the original studies, 77 were based on undergraduate samples; 12 were not (these included children, community-based samples, or online samples), and 10 failed to provide enough information to determine whether their samples were based on college students.

Next, we examined whether results that failed to replicate or partially replicated could be due to a mismatch between sample demographics (see Table 2). For unsuccessful or partial replications, none were clearly and intentionally matched on all three criteria examined here (country, university students, and gender proportion). In the majority of cases, missing information about the demographics makes comparing the sample demographics unclear, and in some cases these factors did not match (e.g., data was collected from a different country for the replication). Overall, we were unable to draw independent conclusions about whether replications failed as a function of whether or not they matched the original samples on demographic factors because this information is not consistently presented and was often missing. Isolating demographic information will not necessarily explain non-replication results, but missing demographic data prevents independent consideration of these factors.

### Issues in Reported Demographic Data

Assumptions about gender and other social categories were embedded in how information about participant demographics were reported. For example, a number of reports reported gender by identifying the total sample size, and reporting how many men were in the sample, implying that all other participants identified as women. For example, one report stated that participants were undergraduates at Princeton, and that the sample included “126 participants (43 men, \( \text{Mage} = 19.72, SD = 1.14 \))” (Chandler, 2016, p. 2). When reported in this format, the reader is left to assume that the non-default group comprises women. This lack of information does not lend the research to feeling “open” in terms of who is included and causes transparency issues, as critical information about identity is often not assessed and/or not reported. Moreover, this form of reporting presumes no other genders exist. When examining the data on racial/ethnic identification, categories were not clearly defined, and varied across studies. For example, one study included “Indians” under the label “Asian” but did not clarify the meaning of this term (LeBel, 2017). Studies that classified participants as Hispanic/Latinx did not acknowledge that Hispanic/Latinx individuals can be of any race. Other studies may have differently identified what “counts” in different racial/ethnic groups.

In another report, the authors noted that they did not collect demographic information because it was not part of the replication materials provided by the original author (Galak, 2016). While the original publication reported no demographic information, it was not clear why the failure to gather these data was included in the replication. In another replication study focused on stereotypes and interracial contexts (e.g., Kelso et al., 2016), a diverse sample was obtained, but only White-identified male participants were included in the replication analyses. This choice was made to mirror the sample from the original study, wherein the focus was on whether White men created physical distance from a Black man in a conversation to appear less racist depending on the framing of the interaction (e.g., as a learning experience). While the intention to matching the original sample characteristics is a critical feature of replication seeking to reproduce the results in a similar sample, and the replication authors included exploratory analyses on the three largest ethnic groups (White, Hispanic/Latino, and Black/African American) in the sample, the underlying decision to focus only on men was not acknowledged. In other words, focusing on

### Table 2. Comparing Original Study Samples to Replication Samples.

<table>
<thead>
<tr>
<th>University Characteristic</th>
<th>Replication Success</th>
<th>Number of Studies</th>
</tr>
</thead>
<tbody>
<tr>
<td>Matched Students</td>
<td>Yes</td>
<td>20</td>
</tr>
<tr>
<td>Matched</td>
<td>No</td>
<td>1</td>
</tr>
<tr>
<td>Matched Partial</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>Not matched</td>
<td>Yes</td>
<td>30</td>
</tr>
<tr>
<td>Not matched</td>
<td>No</td>
<td>5</td>
</tr>
<tr>
<td>Not matched Partial</td>
<td>15</td>
<td></td>
</tr>
<tr>
<td>Not clear</td>
<td>Yes</td>
<td>13</td>
</tr>
<tr>
<td>Not clear Partial</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Matched</td>
<td>Yes</td>
<td>7</td>
</tr>
<tr>
<td>Matched</td>
<td>No</td>
<td>4</td>
</tr>
<tr>
<td>Matched Partial</td>
<td>18</td>
<td></td>
</tr>
<tr>
<td>Not matched</td>
<td>Yes</td>
<td>13</td>
</tr>
<tr>
<td>Not matched</td>
<td>No</td>
<td>8</td>
</tr>
<tr>
<td>Not matched Partial</td>
<td>29</td>
<td></td>
</tr>
<tr>
<td>Not clear</td>
<td>Yes</td>
<td>3</td>
</tr>
<tr>
<td>Not clear No</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Not clear Partial</td>
<td>17</td>
<td></td>
</tr>
<tr>
<td>Country</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Matched</td>
<td>Yes</td>
<td>3</td>
</tr>
<tr>
<td>Matched</td>
<td>No</td>
<td>8</td>
</tr>
<tr>
<td>Matched Partial</td>
<td>18</td>
<td></td>
</tr>
<tr>
<td>Not matched</td>
<td>Yes</td>
<td>6</td>
</tr>
<tr>
<td>Not matched</td>
<td>No</td>
<td>10</td>
</tr>
<tr>
<td>Not matched Partial</td>
<td>34</td>
<td></td>
</tr>
<tr>
<td>Not clear</td>
<td>Yes</td>
<td>2</td>
</tr>
<tr>
<td>Not clear No</td>
<td>7</td>
<td></td>
</tr>
<tr>
<td>Not clear Partial</td>
<td>12</td>
<td></td>
</tr>
</tbody>
</table>

Note: Nosek et al., (2015) noted that no single standard or indicator adequately captures replication success. For the purposes of this analysis, replications were classified as successful, partial, or unsuccessful based on the replication author’s determination in the replication report.
specific sub-groups (e.g., White men) may be valuable and can add to our knowledge about the psychology of this group, but a clear rationale for these methodological choices contextualizes and strengthens this approach. It would be useful if replication authors included some form of sample justification in the introductions of the replication reports because, by addressing specific sample characteristics, they would help to reinforce the idea that sampling practices can be critical for understanding the contextual boundaries of results.

Research Design

Few of the reports explicitly considered the participant’s identity in the study design (19%). A small number of reports (15%) focused on specific demographic groups or included stimuli that represented particular groups, yet overall, demographic and identity characteristics were not relevant to the constructs measured in the reports. None of the reports measured identity or demographic characteristics in a meaningful way (e.g., questions about gender ideology or identification, ethnic or racial identity, sexual identity, or assessment of the meaning behind group membership). None of the studies explicitly considered the culture or location in the selection of participants for the replication, and the majority of the reports (96%) did not justify why they chose the sample they did for replication. For a full report of the research questions, see Table 1.

We observed that some reports made assumptions about and did not clearly report relevant demographic factors within the experimental materials. One report (Talhelm et al., 2016) focused on whether “naïve” undergraduates would associate positive impressions of Fortune 500 corporate executive officer’s (CEO) appearance with company profits. However, the researchers did not disclose relevant information about the photos of CEO’s they used, such as gender or race; age was the only relevant demographic factor disclosed by the original study authors, and both the original study and replication considered age, attractiveness, and the affect displayed in the CEO photos. Given that the pool of CEO’s is generally composed of White, middle-age, and older men, we can assume that those were demographic features of the faces being rated which would be consistent with prototypicality effects (see Bailey et al., 2019; Coles & Pasek, 2020; Purdie-Vaughns & Eibach, 2008). However, this was never stated in the context of the study, and it also was not considered as relevant to the research design.

Another study used movie clips to prime feelings of social isolation, fear, or a neutral condition. Notably, fear was induced using a clip from the movie Silence of the Lambs, wherein the protagonist (a woman) is alone with a suspected serial killer and is chasing him through a house (Dunn & Sandstrom, 2016). The gender and race/ethnicity of the protagonists in the movie clips were not mentioned in the descriptions, and it is quite possible that participants’ own identities may shape how they experience the task, in which they were instructed to empathize with the main character. Given people’s social location on axes of gender and racial oppression differently place them at heightened risk for violence, their experiences may impact the degree to which they emotionally identify with characters in the study materials. It was not uncommon for researchers to omit information in the study design that may have reflected an assumed “universal” participant experience and failed to consider gender, race/ethnicity, age, and other socio-demographic factors that may impact how study materials were experienced by the participants.

Finally, we recognized that there was very little to no acknowledgment of basic tenets of intersectionality for understanding human psychology: Individuals simultaneously hold multiple identities and their experiences are characterized by multiple interlocking systems of advantage and disadvantage ascribed to their social identities (Collins, 1990). Though the uptake of intersectionality in psychology has been relatively slow (Grzanka, 2018), the reports’ failures to consider how lived experiences influence people’s psychological responses and behaviors presents a missed opportunity for making sense of non-replicable findings. For example, one study assessed optimism with regard to a series of events, including decisions about a financial investment, an academic award application, a surgical procedure, and a dinner party (Lasseter et al., 2016). While these study materials may align with the study participants’ experiences and identities (i.e., students for the original study and replication were recruited from two elite, private universities in the US), because the social location of the participants was not discussed or explicitly considered, the dominant identities and experiences are invisible in the research design. An individual’s outlook on these experiences may be significantly different depending on their relative privilege. Participants’ identities and their location within social structures that produce and maintain institutionalized inequalities were not considered. An intersectional approach to this study might consider how race/ethnicity, gender, social class status, and other relevant aspects of identity may influence participants’ responses to these scenarios, all of which have information about social roles and access to resources implicitly embedded within. Research materials likely need to be sample and context dependent to “work,” and researchers need to consider, acknowledge, or analyze how the study materials they have created reflect the experiences of the dominant groups.

Interpretation of Results

The reports presented their findings in relation to the studies they were attempting to replicate, and overall, 50 (out of 100) of the reports failed to replicate the original results. Of the studies that were considered successful, 21 had a partial replication of the original findings. More importantly, we were interested in how the authors discussed non-replication results. Of the 50 non-replications, only 15 reports included a
discussion of potential identity, demographic, cultural, or contextual factors that may have accounted for non-replication. Further, nearly all of these discussions focused on one aspect of identity or culture (e.g., ethnic differences, contextual differences) without regard to other categories or identities that may have been relevant to the study design or participants.

One example is a replication of a study on perception that was originally done in the US and was reproduced in Australia; however, the report authors were confident that there would be no differences between their samples, noting “It is not anticipated that these differences will affect the outcome of the study” (Goodbourn, 2016, p. 5). In cases where both samples (the original study and the replication) were WEIRD, there seemed to be a tendency to assume or to state that potential differences between the samples were not likely to affect the study outcomes.

One aspect of repeatability addresses whether different scientists are able to obtain the same results with new data, implying that robust findings will repeat regardless of who conducts the study. However, an intersectionality framework with its attention to the structural causes of inequality and difference, including power, privilege, and resources, suggests that how science is received might be importantly shaped by the identities of the scientists. Along these lines, and beyond the scope of our initial investigation, we noted that in some cases, interpretation of the findings included a deference to the original study authors.

One successful replication report commended the original study authors as being warm, accommodating, engaged in “full-hearted cooperation...[and] should be commended as good scientists” (Selterman et al., 2016, p. 11). Characterizing the original researchers as “good scientists” is moralized judgment that is outside the scope of the replication study and indicates that researchers, not just the research itself, were under scrutiny. In other cases, report authors would consider the methodological aspects of the study but would fail to consider any significant difference in the samples as a potential explanation for non-replication. For example, a report that did not replicate the main findings from the original study but did replicate minor findings spent considerable space arguing for minor aspects of the study that did replicate and gave less attention to the non-replication findings (Callahan, 2016). This approach suggests that in some cases replication can include a bias toward minimizing non-replication effects, including non-significant hypotheses, through explanation. The replication project is based on the presumption that other scientists should be able to replicate a study and get the same results because they are value neutral actors. However, because the judgment of scientists who practice replication can be colored by social norms of deference and the authority of precedent, there is a risk that the entire enterprise may serve to maintain epistemic injustice that privileges scholars from dominant social identities (Settles, et al., 2020).

Context

Very few of the reports considered context when presenting their results. First, 91 of the reports did not report the year in which data was collected. Second, even though 12 reports had study characteristics that relied on social power differences or the assessment of constructs that may be related to power, discrimination, or disadvantage, none of the replication reports explicitly addressed the roles of sexism, racism, and other forms of oppression. While it is notable that the reports were utilizing the same materials as the original studies, the inclusion of contextual information grounds the research and provides necessary framing for understanding the potential impact of the results. For example, one replication included only White participants and used Black people as targets in the study (Frank & Holubar, 2016; Frank & Mathur, 2016), and another examined prejudice among participants but did not identify the racial/ethnic composition of the sample (Johnsen-Grey et al., 2015). This approach demonstrates how social psychology has adopted a script for positioning certain people as targets (e.g., people of color) and aligning White participants as vantage-less perceivers (e.g., using Black people as objects and White people as actors; Shelton, 2000).

None of the reports considered the multidimensionality of participants’ identities in a meaningful way, and typically only single-axis social group memberships were reported. In some reports, this was notably missing. For example, one study examined the different needs of victims versus perpetrators and hypothesized that victims need their status and power to be restored in order to reconcile, whereas perpetrators need their sense of moral and social acceptability to be restored (Gilbert, 2016). This research design is ripe for an intersectional analysis based on how the survey questions might be read and interpreted by people in different categories or groups (e.g., with different intersecting identities). Whether we view someone as a victim or perpetrator is likely influenced by their social identities. However, race/ethnicity was not reported in the demographic information for the sample, and gender was not considered in the analyses. In other words, the analytic approach included the assumption that all individuals will respond the same way to the research questions without regard to social identities.

Another pattern that reflects the assumption that social identities are not relevant to the subject matter is that few of the perceptual studies reported or considered demographic or social identities in their research design. First, many of the perceptual tasks relied on samples with no vision impairment, and yet disability is rarely acknowledged as a factor that may affect the generalizability of the results. Second, there may be gender and/or ethnic differences in how people process memory tasks, in part because perceptual models often fail to take into account factors like chronic stress and weathering, which are disproportionately experienced by ethnic minority groups and by those with multiple intersecting minority identities (Geronimus et al., 2006; Richardson & Brown, 2016).
Discussion

This study provides a systematic analysis of the 100 replication reports produced and published by the Open Science Collaboration. By applying questions derived from theoretical and methodological calls for intersectionality to be integrated into psychological research, this analysis shows patterns among these studies regarding representation of groups of people in mainstream psychological research. Through the lens of an intersectional analysis, our findings highlight potential reasons that studies might not replicate. In addition, we consider how in the context of the Reproducibility Project, “open” science refers more to study materials and methods being transparently communicated, rather than extending to inclusion and representativeness of groups historically left out of mainstream research. The Reproducibility Project represents a perspective in psychological science that assumes that there is a universal experience or truth that can be documented and determined through rigorous research. This approach highly values particular tenets of scientific research, such as a lack of ambiguity, verifiability, repeatability, and generalizability (Freese & Peterson, 2017), and assumes that non-replicability of previously established results is due to methodological error, inconsistency, or randomness (Open Science Collaboration, 2015). The current study is a call to recognize the implicit values and perspectives represented in this approach. A critical examination of study design and methodological choices made in mainstream psychological research is in order. Replicating studies that fail to consider the intersectional facets of an individual’s identities, experiences, and socio-cultural context is not likely to move the field closer to establishing a generalizable and verifiable truth about human psychology. Toward this end, we make recommendations for researchers to address or consider when designing a study or a replication based on our findings (see Table 3). These practices allow researchers to leverage intersectional theory to broaden representation and to understand the reasons that key findings do not replicate beyond simply failing to control all of the criteria or the environment in the same way as the original study.

Representation

Our analyses showed that the majority of replication reports were based on WEIRD samples composed of college students and women were disproportionately represented in the samples. In addition, a number of studies were missing key demographic information about samples and recruitment. First, this illustrates that “open” refers primarily to researchers’ access to materials, not open with regard to inclusion and representation.

Table 3. Recommendations for Researchers to Consider.

<table>
<thead>
<tr>
<th>Area of Analysis</th>
<th>Recommendation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Representation</td>
<td>1. Increase transparency to include standard and all relevant demographic and identity information in all reports.</td>
</tr>
<tr>
<td></td>
<td>2. Be explicit about the focus of the reproduction attempt. Is the goal to reproduce the same effect in the same or very similar sample? Or is the goal to test an effect in a new sample or context to gauge the broader applicability of the effect? Explicitly state why this is the goal.</td>
</tr>
<tr>
<td></td>
<td>3. Consider the questions “To whom does this work apply?” and “Who is missing from this work?” when making decisions about sample characteristics. Provide a rationale for these exclusions.</td>
</tr>
<tr>
<td></td>
<td>4. If your research design relies on any aspect of culture, identity, or membership in a demographic group, consider assessing this in a meaningful way rather than as a group membership (e.g., assess gender identity rather than using a group category as a variable in the study; assess within-group differences; shift to consider and name the sources of inequality rather than ascribing this to members of a marginalized group).</td>
</tr>
<tr>
<td></td>
<td>5. Consider how your sample and context may be the same or may differ from studies used as a comparison or a model for your research.</td>
</tr>
<tr>
<td></td>
<td>6. Report all basic contextual information for your sample, including what year data was collected and where it was collected, as well as additional contextual information (e.g., social or cultural) that may be relevant to your study or your participants.</td>
</tr>
<tr>
<td></td>
<td>7. Reflect on and include information, when possible, about power, privilege, and social structures that are relevant to your study or sample (e.g., racism, sexism, and other forms of oppression).</td>
</tr>
<tr>
<td></td>
<td>8. Consider the multidimensionality of participants’ identities when making decisions about who to include, as well as in the interpretation of the results.</td>
</tr>
</tbody>
</table>
representation of non-dominant, privileged populations. Second, to consider replicating a study, detailed and meaningful information about the participant identity, the recruitment process, and the context for the study must be clearly documented and shared. Our finding confirms previous research indicating that the majority of psychological research publications fail to report the racial demographics of their samples or reduce these to oversimplified categories (Delleus et al., 2019). The predominantly WEIRD samples in the replication reports are not representative of the majority of the world’s population (Arnett, 2016), or even of the US. Women are overrepresented in the replication studies, which reflect the pattern that women are overrepresented among psychology majors, the common source of data for psychological studies (Richmond et al., 2015). Although women were overrepresented in the replication samples, this was not noted or discussed in any of the reports. The sampling pattern in the reports reflects inequalities that are reflective of and embedded in the power structures that shape research practices, and this reinforces the epistemological viewpoint of the predominantly White researchers creating this knowledge (Roberts et al., 2020; Saab et al., 2020).

Comparing the demographic information from the original study sample to the replication sample provided insight into the issues of transparency as well as generalizability. Many studies and reports were missing key demographic information about the samples, as is reflected in the pattern of unclear results when examining whether studies matched on key demographic criteria. Most noteworthy in terms of transparency is the lack of consistency in reporting demographic information across the studies and replication reports. These omissions make it impossible to draw independent conclusions about whether replications failed as a function of whether or not they matched the original samples on demographic factors. Thus, our first recommendation is that to increase transparency, researchers should include all relevant demographic and identity information in their reports. Given that science continually sheds light on what factors may be considered relevant, and that socio-demographic categories and identities may change over time, we recognize that researchers may not know currently what factors may be relevant in the future. However, we suggest that a minimum of collecting and reporting demographic information based on the Journal Article Reporting Standards is necessary (e.g., the inclusion of all major demographic information including age, sex, ethnicity, and socioeconomic status, as well as any topic-specific characteristics). In addition, we encourage researchers to think beyond these basic categories and to include any additional contextual and identity information that is available or able to be collected.

In many cases, demographic information was reported but was not considered a feature of the replication. That is, the authors of the replication most often did not explicitly state that they were interested in whether the effect from the original study replicated to another specific population. It is also noteworthy that in many cases, the authors did not provide enough information to determine whether the samples were matched (e.g., over a quarter of the pairs of studies were missing information on whether university students were recruited for the sample; over half of the pairs of studies were missing information on gender and/or the country where the data were collected). Overall, whether the results were generalizable to a different population was not clearly addressed in the replication reports, and the scope of missing information between the pairs of studies (original and replication) made it difficult to assess the generalizability of the findings beyond the initial population.

Based on this analysis, we suggest that there is a missing step in the replication process wherein research would benefit if explicit information about sampling choices and how these decisions shape the interpretation of the results were routinely included. If the researcher’s goal is to reproduce the same finding in a closely matched sample in order to demonstrate that the finding holds within that particular population, it would be beneficial to know the demographic and identity features of the population. Thus, our goal is not to criticize whether or not replications choose “matching” samples (though, this needs to be clearly defined by the researchers: matching with regard to what? And why?), it is to show how the data reflect a pattern of non-inclusion of information, and relatedly, a non-inclusion of diverse samples. Researchers tend not to report demographic information when the sample resembles the prototypical/dominant group, and thus stating that the sample in a study consisted of undergraduates carries a number of implicit assumptions about who is included in that group. We are calling for an explicit and thus more transparent statement about who is represented in these studies. Indeed, replication reports should be held to a higher standard with regard to transparency and intention, because whether their goal is replication of a finding with the same population or generalization of a finding to other populations, this goal should be stated and supported clearly. To this end, our second recommendation is that researchers be explicit about the focus of their reproduction attempt: is the goal to reproduce the same effect in the same or a very similar sample? Or is the goal to test an effect in a new sample or context to gauge the broader applicability of the effect? Explicitly state why this is the goal.

The findings also showed a pattern among replication researchers to assume who is represented in categories they reported. For example, some reports would only provide the number of men and overall sample size, which presumes all remaining participants identify as women and ignores the existence of individuals who do not identify with the assumed dominant gender groups. Notably, only three of the reports indicated that they gave an option other than “male” and “female” for participants to report gender. When identities or demographic categories go unnamed or unreported, we tend to cast WEIRD and dominant groups as the default and prototypical experience. This reinforces the invisibility of some
categories and reinforces that others are marked as the “norm” (e.g., Bailey et al., 2019; Coles & Pasek, 2020; Purdie-Vaughns & Eibach, 2008). This thinking also reflects the entrenched and uncritical views that researchers may hold regarding their own social groups. As Harding and Norberg (2005) noted, “Dominant groups are especially poorly equipped to identify oppressive features of their own beliefs and practices” (p. 2010). One consequence of the obfuscation of non-dominant groups is that research based on their experience is often viewed as “specialty” and is relegated to non-mainstream journals. The invisibility of ethnocentrism, androcentrism, and heteronormativity prevents critical analysis of the impact of those identities on research outcomes and undermines the need for work toward equitable inclusion and representation in psychological research.

Thus, our third recommendation is to consider the questions “To whom does this work apply?” and “Who is missing from this work?” when making decisions about sample characteristics (Cole, 2009) and to provide a rationale for these exclusions. Despite the growing criticism of the WEIRD, ethnocentric, androcentric, and heteronormative approach to psychological research (Arnett, 2016; Henrich et al., 2010; Saab et al., 2020), non-dominant social groups, identities, and norms are rarely made visible in the research. The erasure of non-prototypical group members is a central concern of intersectionality (Purdie-Vaughns & Eibach, 2008) and was clearly reflected in the replication reports. To address this, we recommend clarity in how demographic information is collected and coded, including asking more representative questions about how participants self-identify with regard to gender and other identities; reporting both the n and percentage for all demographic groups and patterns; creating (and enforcing) a standard for what is reported about samples for transparency; and to contribute to more “open” research, allowing readers to identify the contextual factors that may be relevant to future research and replication efforts.

Research Design

In our analysis we noted that some reports used research methods that reflected stereotypical and Eurocentric thinking and assumptions in their design. This includes exclusion of Black participants from the study without a rationale, studies focused only on White participants without a rationale, and the failure to include diverse representation in study targets. Perhaps some replication researchers assumed that there was not a need to provide a justification for research design choices made by the researchers who produced the original study, but we maintain that choosing to replicate these results signifies their importance, and that explicitly considering and reporting both the study rationale and inclusion/exclusion criteria should be standard in all studies, especially in replication reports wherein methodological and participant choices carry greater significance with regard to both reproducibility and generalizability.

There are a number of issues raised in these practices. First, there was a lack of theoretical justification for the inclusion and exclusion of groups based on race in these reports, though we argue that a theoretical justification is not enough. Using images or descriptions of Black people as the targets in a study (such as being the target of police harassment and violence), particularly for the purpose of seeing how White people respond, is objectifying, and aligns the researcher and participant with an implicitly White gaze. While some of the original studies may have contextualized this information and considered racism or other forms of discrimination in the rationale for these study choices, this information did not carry over to the replication reports. Second, one of the most striking observations is how infrequently psychologists refer to racism at all, even in the context of studies about White participants engaging with experimental materials that positioned Black people as the targets of police violence. Failing to identify and name racism and other forms of discrimination compounds the potential harm while failing to stop or change these patterns.

Just as the disadvantage of some groups was not noted or considered in research design, the privilege of the participants with regard to their background and identities was not considered, even when this would be relevant to the study. In the results section, we noted that one study assessed optimism with regard to events such as financial investments, academic awards, a surgical procedure, and a dinner party (Lassetter, et al., 2016). However, an individual’s outlook on these may be significantly different depending on their relative privilege. While we noted that the participants in the replication and original study were from elite universities, the replication report did not include information on participant’s socioeconomic status or racial/ethnic group membership. There are well-documented structural barriers to access to wealth (including investment and education) for ethnic minorities (Assari, 2020). There are significant differences in health outcomes based on race, and expectations for a surgical procedure may look drastically different for women of color, whose accounts are often not believed by their healthcare providers, a pattern reflecting systemic and interpersonal racism (Trawalter & Hoffman, 2015; White & Stubblefield-Tave, 2017). Last, dinner parties may reflect and reinforce the social position of those involved. In sum, all of the scenarios in this study may be experienced very differently by people with different socioeconomic, racial, and educational backgrounds. In addition to failure to consider the material resources or privileges inherent in the study design, many of the reports sampled privileged participants but did not consider this as context for interpreting the results.

In order to change this pattern, we call for the reframing of issues regarding privilege and discrimination to consider the source and put the onus on the people and structures that create and maintain systems of inequality. One example of this is in how Versey et al. (2019) reframed “internalized racism” by referring to it as “appropriated racial oppression.” This shifts from a negative or deficit model based on the messages that
people of color internalize to a model in which the racist messages of White supremacy are understood as a process affecting oppressed and dominant groups. This kind of shift is central to the recognition of and reckoning with the assumptions underlying research design. A similar shift to recognize and reframe privilege with regard to socioeconomic status, age, ability, and other social identities, and their intersections, is needed, wherein the onus lies with the systems of inequality and the people that create and maintain them, rather than with their victims. In sum, our fourth recommendation is that researchers consider whether their research design relies on any aspect of culture, identity, or membership in a demographic group, and if so, consider assessing this in a meaningful way rather than as a group membership (e.g., assess gender identity rather than using a group category as a variable in the study; assess within group differences; shift to consider and name the sources of inequality rather than ascribing this to members of a marginalized group).

**Interpretation of Findings**

We noticed a trend in the reports that authors framed non-replication explanations in a way that assumed the original findings were “correct” even though we know there is publication bias toward significant results (Bakker et al., 2012). When interpreting results, there was an overall failure to consider identity, demographic, or cultural factors, or to meaningfully explore differences between study samples. For example, some reports (e.g., Frank & Holubar, 2016; Frank & Mathur, 2016; Humphries et al., 2016) noted that demographic differences between participant populations may be a reason for non-replication but did not explain what aspects of identity or culture might account for the different findings. When findings fail to replicate, there are a number of reasons why this might occur, including that there may have been a problem with the original results; there may be a problem with the replication; or there may be a subtle difference in the studies, which in itself may be a significant scientific discovery (Freese & Peterson, 2017). In other words, when replication fails, it is not clear why this is the case, and we can only be certain a replication has worked when it is successful. Freese and Peterson (2017) pointed out that “if replications are only interpreted as evidence when successful, this would imply that they cannot be properly considered to be testing anything. Thus, we need some way of judging the quality of a replication that is independent of its outcome” (p. 149). Toward this end, our fifth recommendation is that when evaluating the results of a replication, researchers must consider how the sample and context may be the same or may differ from studies used as a comparison or a model for the replication.

**Context**

Contextual factors were largely missing from the reports. Including information such as the year that data were collected can help contextualize the data. Missing relevant details like demographic and cultural information limits the interpretability and the ability to replicate the studies and draw conclusions about generalizability. This pattern reflects the findings of Van Bavel et al., (2016), who found that attention to contextual information in the replication reports was associated with greater chances of successful replication and suggests that researchers need to be more attentive to the contextual factors that influence study outcomes. Our sixth recommendation is to report all basic contextual information for samples, including when and where data were collected, as well as additional contextual information (e.g., social or cultural) that may be relevant to the study or the participants.

The current study builds upon this finding by suggesting that conceptualizing context in the framework of intersectionality includes attention to additional layers, such as structural oppression and privilege. We noticed that the majority of the reports ignored structural factors, such as sexism, racism, and other forms of discrimination, even in studies that were explicitly about gender, race, or other relevant groups. Recognizing the surrounding sociohistorical contexts of demographic characteristics provides a more thorough analysis of how identities are experienced (Bowlé, 2008). Consistent with intersectionality theory, we suggest researchers not only account for identity categories in their research but also the power dynamics and contexts for which category labels are used as proxies. Thus, our seventh recommendation is to reflect on and include information, when possible, about power, privilege, and social structures that are relevant to the study or sample (e.g., racism, sexism, and other forms of oppression).

Our last recommendation reflects that lack of engagement with multidimensional aspects of identity in the studies that were replicated. While it is noteworthy that none of the replication reports considered multiple identities or the intersection of identity factors in a meaningful way, this reflects the patterns observed in mainstream psychological research (McCormick-Huhn et al., 2019). Fully reporting demographic and identity information and taking steps to address the previous recommendations (e.g., assessing the meaning of group membership, reflecting on power, privilege, oppression, and social structures) will create a foundation for considering identity as multidimensional. Considering participants’ multiple and intersecting identities that are relevant in the context of the study may uncover new and different information and provide a lens through which to re-evaluate study materials as well as the interpretation of results. Participants’ identities are not irrelevant or secondary in the research process; they are a central part of the context to consider in research design and interpretation. Thus, our final recommendation is that researchers consider the multidimensionality of participants’ identities when making decisions about who to include as well as in the interpretation of the results.
Conclusion

The goals of the open science movement are to increase transparency, access, and collaboration in science changing scientific practices. Our analysis demonstrates a lack of transparency in the studies that comprise the Reproducibility Project, a lack of transparency rooted in failure to attend to critical demographic, identity, and contextual information. Moreover, we found that these studies failed to offer minority populations in the US, and non-WEIRD populations globally, access to participation. In that sense, such studies cannot claim to be representative. Drawing on questions framed by intersectionality helps bring into focus some of the missing information and invisible assumptions that guide this work. Critically, when demographic information is missing from studies, readers and researchers are unable to determine whether replication efforts have re-created these aspects of the original study. More importantly, participant identity and context is a central part of the method that needs to be considered in research design. Because this aspect of methodology has largely been made invisible, particularly in social and cognitive psychology, the norm has become for researchers to not collect or report data when samples represent the dominant group. Including questions of representation, invisibility, bias, power, and structural discrimination should guide the research in order to ensure that transparency and openness go beyond replication of problematic methods and reporting.

We recognize that the reports we analyzed occurred in an era of science reform in which open science discourse was nascent and perhaps there has since been growth in theoretical and methodological rigor in reproducibility efforts. If so, we are eager to see psychologists take intersectionality seriously (McCormick-Huhn et al., 2019) in such efforts and it is our hope that an intersectionality lens can assist in the pursuit to build a better psychological science.

Appendix

Questions generated to analyze the replication reports from an intersectional lens

Background Information

1. Replication authors and report name?
2. Hypothesis?
3. Main topic of the study?
4. Area of psychology represented?

Representation (Who is represented in the samples derived for the replication reports)

1. What was the sample size?
2. What demographic information was provided about the sample?
3. Who was represented in the sample for the replication? (i.e., what demographic or identity information was reported, including gender, age, race/ethnicity, SES, sexuality, or anything noted by the authors)
4. What was the location/country of origin for the replication?
5. Was the location/country of origin for the original study noted? If so, where was it?
6. If both sample locations are reported (original and replication) are they both WEIRD?
7. Do the participants in the replication fit the WEIRD (Western, educated, industrialized, rich, democratic) criteria? (If not, in what way do they differ from this norm?)

Research Design (Are demographic and identity factors considered in the research design? If so, were these considered through a multidimensional or intersectional lens?)

1. Did the research design consider the participant’s demographic characteristics or identity characteristics?
2. Did the research design include a manipulation that relied in some way on demographic categories, like race/ethnicity, age, gender, etc.?
3. Are any aspects of culture identified or acknowledged in the rationale for the study or for the participants selected for the replication?
4. Did the researchers of the replication justify why they chose the sample they did to replicate the effect?

Interpretation of Findings (How were non-replication results explained?)

1. Was the replication successful?
2. If the replication failed, did the authors acknowledge demographic, identity, or cultural factors that may have made a difference?

Context (What context was provided for the study, results, and interpretation?)

1. Was the year the data was collected for the replication reported?
2. Did any of the original study characteristics rely on power differences or the assessment of constructs that may be related to power, discrimination, disadvantage?
3. If gender, race, or demographic categories are explicitly addressed in the paper, do the authors contend with the roles of sexism, racism, and other forms of oppression?
4. Did the researchers consider the multidimensionality of participants’ identities (e.g., were only single-axis social group memberships reported)?
5. Did the replication consider or analyze within-group variation with regard to identity or demographic group membership?

Declaration of Conflicting Interests
The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding
The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iDs
Natalie J. Sabik  https://orcid.org/0000-0003-1420-4242
Jes L. Matsick  https://orcid.org/0000-0003-4368-3211
Kaitlin McCormick-Huhn  https://orcid.org/0000-0003-3156-336X

References


Murphy, M. C., Mejia, A. F., Mejia, J., Yan, X., Cheryan, S., Dasgupta, N., Destin, M., Fryberg, S. A., Garcia, J. A., Haines,
