



# Intersectional inequalities in science

Diego Kozłowski<sup>a</sup>, Vincent Larivière<sup>b,c,d</sup>, Cassidy R. Sugimoto<sup>d,e</sup>, and Thema Monroe-White<sup>f,1</sup>

<sup>a</sup>Faculty of Science, Technology and Medicine, University of Luxembourg, Esch-Sur-Alzette 4364, Luxembourg; <sup>b</sup>École de Bibliothéconomie et des Sciences de l'information, Université de Montréal, Montréal, QC H3T 1N8, Canada; <sup>c</sup>Observatoire des Sciences et des Technologies, Université du Québec à Montréal, Montréal, QC H3C 3P8, Canada; <sup>d</sup>Department of Science and Innovation-National Research Foundation Centre of Excellence in Scientometrics and Science, Technology and Innovation Policy, Stellenbosch University, Stellenbosch 7602, South Africa; <sup>e</sup>School of Public Policy, Georgia Institute of Technology, Atlanta, GA 30313; and <sup>f</sup>Campbell School of Business, Berry College, Mount Berry, GA 30149

Edited by Douglas Massey, Office of Population Research, Princeton University, Princeton, NJ; received July 16, 2021; accepted November 19, 2021

**The US scientific workforce is primarily composed of White men. Studies have demonstrated the systemic barriers preventing women and other minoritized populations from gaining entry to science; few, however, have taken an intersectional perspective and examined the consequences of these inequalities on scientific knowledge. We provide a large-scale bibliometric analysis of the relationship between intersectional identities, topics, and scientific impact. We find homophily between identities and topic, suggesting a relationship between diversity in the scientific workforce and expansion of the knowledge base. However, topic selection comes at a cost to minoritized individuals for whom we observe both between- and within-topic citation disadvantages. To enhance the robustness of science, research organizations should provide adequate resources to historically underfunded research areas while simultaneously providing access for minoritized individuals into high-prestige networks and topics.**

intersectionality | science of science | bibliometrics | race | gender

Strong disparities are observed in the composition of the scientific workforce. At the global level, women account for less than a third of scientists and engineers (1); a percentage that is similar to their proportion of scientific authorships (2). In the United States, women represent 28.4% of the scientific workforce, and this percentage varies by domain, with a high of 72.8% in psychology and a low of 14.5% in engineering (3). Disparities are also observed at the intersection of race and gender, with White men comprising a disproportionate amount of the US workforce (4). Although the trend is changing—faculty of color increased from 20% of the scientific workforce in 2005 (5) to 25% in 2018 (6)—increases have not been observed equally across all racially minoritized groups. For instance, the proportion of Black (5 to 6%) and American Indian/Alaska Native (1%) scholars remained relatively stable, while Latinx representation nearly doubled (3.5 to 6%), and Asian representation increased from 9.1 to 11%. Gender differences are also observed within racial categories: men account for a higher share than women, especially for White and Asian/Pacific Islanders (6). Likewise, the presence of minoritized groups varies substantially by discipline. Science, Technology, Engineering/Computer Science, and Mathematics (STEM) disciplines exhibit less demographic diversity than non-STEM fields (7). For example, in Biology, only 0.7% of faculty identify as Black, despite representing 12.2% of the US population (7).

These differences characterize the unequal representation of populations within the scientific community. Such disparities are often a manifestation of inequality—unequal outcomes—and inequity—the degree to which these outcomes are a result of impartiality or bias in judgement. Women and other minoritized populations are underrepresented in scientific publishing (8, 9), for example, and this can be associated with unequal outcomes in peer review (10, 11). Inequalities have been observed at several other pivotal evaluation points in science, including applications for laboratory manager positions (12), grant submissions (13, 14), and scholarly impact (2). Several forms of implicit bias may contribute to these inequities: from perceptions of brilliance (15) to

gendered scripts on women's commitment to science (16). Overt forms of discrimination found in other spheres of society are also observed within the scientific community, such as stereotypes about gender and race (17), anti-Black institutional policies (18), and structural racism (19, 20).

Studies that examine inequities and inequalities at the individual level are often anchored in a justice perspective (21), whereby scientific principles such as universalism (22) are tested against the current system, thus challenging the conception that science is a meritocracy. In contrast with studies of individual-level success, utilitarian studies focus on collective gains and test whether higher equity improves the robustness of science. Extant studies have demonstrated that racial diversity leads to increased productivity (i.e., sales and profits) in industry (23) and that diverse groups outperform homogeneous ones in cognitive tasks (24). In science, diversity in the composition of scientific teams has been linked to higher citations (25) and tied to gains in innovation (26). This emergent body of literature suggests that there are scientific and societal benefits to increasing diversity in science. Studies should, therefore, consider the rich interplay between social identities and scientific work.

A growing body of work examines the affinity between social identities and topic selection. For example, in medical research, decades of male dominance led to little attention to sex differences in medicine (27, 28). The changing demographics of the research community improved the situation, as women are more likely to include female subjects (29) and to report sex as

## Significance

**The US scientific workforce is not representative of the population. Barriers to entry and participation have been well-studied; however, few have examined the effect of these disparities on the advancement of science. Furthermore, most studies have looked at either race or gender, failing to account for the intersection of these variables. Our analysis utilizes millions of scientific papers to study the relationship between scientists and the science they produce. We find a strong relationship between the characteristics of scientists and their research topics, suggesting that diversity changes the scientific portfolio with consequences for career advancement for minoritized individuals. Science policies should consider this relationship to increase equitable participation in the scientific workforce and thereby improve the robustness of science.**

Author contributions: D.K., V.L., C.R.S., and T.M.-W. designed research; D.K., V.L., C.R.S., and T.M.-W. performed research; D.K. and V.L. contributed new reagents/analytic tools; D.K. analyzed data; and D.K., V.L., C.R.S., and T.M.-W. wrote the paper. The authors declare no competing interest.

This article is a PNAS Direct Submission.

This open access article is distributed under [Creative Commons Attribution License 4.0 \(CC BY\)](https://creativecommons.org/licenses/by/4.0/).

<sup>1</sup>To whom correspondence may be addressed. Email: [tmonroewhite@berry.edu](mailto:tmonroewhite@berry.edu).

This article contains supporting information online at <http://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2113067119/-/DCSupplemental>.

Published January 4, 2022.

an analytical variable (30). Women are also more likely to produce scientific discoveries that lead to women's health patents and to contribute to patenting in this area (31). Funding—one of the main drivers of research activity—is similarly affected by researchers' social identities and align with topic selection. For example, funding outcomes at the NIH were found to be lower for Black and African American applicants. This was largely explained in topic selection: These investigators were more likely to propose research on topics with lower success rates (e.g., human subjects research and research on health disparities) relative to White and Asian investigators (32). These outcomes have implications for innovation and scientific competitiveness: Racialized and gendered groups are more likely to contribute novel scientific contributions, yet their work is often neglected by other scientists (26). Taken together, these studies suggest that unequal representation in science leads to underinvestigation of particular topics and may serve to stymie innovation. This motivates a more nuanced understanding of barriers to success for minoritized populations and how these observed disparities intersect with complex social identities, fine-grained topic selection, and the reward structure of science.

Intersectionality was initially introduced as an analytic framework for understanding how interrelated and mutually shaping categories of race and gender served to compound inequalities for minoritized women (33–36). These studies emphasize women's racialized and gendered experiences by explicitly situating minoritized women as central actors in power struggles and social inequalities (23, 33, 37–41). The intersectional framework has since been expanded and used to frame the marginalization experienced by minoritized groups at the intersection of race, gender, sexual orientation, class, and other identities (42). While gender inequities and inequalities have been the focus of several recent large-scale analyses (2, 10, 43–46), very few studies have focused on the racial and ethnic composition of authors (46–49). Studies from an intersectional perspective (e.g., women of color) have been predominantly qualitative, based on self-reports, or focused on a particular field or set of subfields (50–52). These studies provide rich evidence of the impact of structural biases on career trajectories through valuable storytelling and suggest a need for an intersectional lens to large-scale studies. Furthermore, these studies reveal that a failure to disaggregate at the intersection of race and gender may obfuscate novel findings and lead to the generation of overly simplistic insights and policy recommendations (47).

Therefore, we seek to interrogate the space between the composition of the scientific workforce and the topical profile of science from an intersectional perspective. This study extends investigations of gender disparities in science by providing a macrolevel of view of the phenomena that accounts for the intersection of race, gender, and topic. Our focus on the United States enables us to contend with the unique contextual factors that have led to disparate representation between genders and racial groups in science. Despite the acknowledged importance of race and gender as factors of inequality and decades-long policy interventions, there remains a paucity of evidence on how the selection of fields and topics is scattered across groups at a detailed level and the relation between topics and scientific impact. This paper attempts to demonstrate that the election of the object of study is related to race and gender, with implications for scientific progress and the evaluation of scientists.

## Materials and Methods

We examine the publication patterns of US-affiliated first authors between 2008 and 2019. Our data consist of 5,431,451 articles indexed in the Web of Science (WOS) database and 1,609,107 distinct US first authors. We focus on first authors, as they are generally those who have contributed the most to an article (53, 54) and represent the most visible name in bibliographic references. The metadata includes authors' given and family names, which are used to infer race and gender. Authors were disambiguated using the algorithm

developed by Caron and van Eck (55). The gender disambiguation algorithm builds on the method presented in Larivière et al. (2), which uses census data and country-specific lists of men and women names to assign probable gender to given names and, in the case of certain countries (e.g., Russia and Ukraine), family names. Gender is considered in a binary way, as other genders can only be assigned through self-identification. This is an acknowledged limitation of the study.

Racial categories are a country-dependent social construct, and not all countries have such categorizations. Therefore, our analysis focuses on the specific cultural construct of race found in the United States. For the inference of race/ethnic origin, we use the 2010 US Census information on family names and racial groups (56). Racial groups considered in the US Census are the following: 1) Non-Hispanic White Alone (White), 2) Non-Hispanic Black or African American Alone (Black), 3) Non-Hispanic Asian, Native Hawaiian, and Other Pacific Islander Alone (Asian),\* 4) Non-Hispanic American Indian and Alaska Native Alone (AIAN), 5) Non-Hispanic two or more races (two or more), and 6) Hispanic or Latino origin (Latinx).<sup>†</sup> Given that AIAN and two or more account for only 0.69% and 1.76% of WOS authors, respectively, they were removed from the analysis. Census data provides the number of people that identify with each racial group for the 162,253 most common family names. Using family names, we compute each author's associated probability to each racial group, instead of assigning the most probable group, and using these probabilities to compute weighted aggregates, in which each author contributes to each group's aggregate as a function of the racial group distribution associated with its family name. This means, for example, that when computing the average citations by race, we assign the citations of an article fractionally to each group according to the corresponding distribution. In other words, we do not assign authors to a unique racial category. In previous work (58), we have shown that, given the overlap of Black and White family names (59, 60), the use of a threshold—filtering those names with a probability for a single group above a threshold and assigning all authors with that name to that single category—underestimates the proportion of Black authors. This distinction is critical: We do not aim to identify each author's self-perceived racial category but to build aggregates of racial group disparities. For those names that do not appear in the census, we impute the mean distribution in the subset of authors used at each point in the analysis. For a detailed description of the racial inference methods, see ref. 58, *SI Appendix*, as well as the accompanying website.<sup>‡</sup>

Fields and subfields are defined according to the journal classification developed for the US NSF (61). Following ref. 62, we used articles' abstract, title, and keywords to train a latent Dirichlet allocation (LDA) model to infer the topics within a corpus of papers and the distribution of topics within each article. LDA is an unsupervised model that assumes that there is a fixed number of topics within the corpus that correctly describes its content. Each topic is defined as a distribution over words, and we use the top five words from each topic to infer its semantic content. Given an article's topic, there are some words that are more likely to repeat than others; LDA provides the list of most repeated words for each topic, and we use those to infer topicality. The objective of this model is to create research topics as detailed as the sampling allows, implying a trade-off between granularity and repetition of topics. LDA models are performed on groups of disciplines to identify topics with an independent meaning (63). Given the sample size in Social Science and Health and the interpretability of results across different experiments, we found the optimal number of topics for our analysis to be 200 for Health and 300 for Social Sciences. Using manual inspection, a higher number of topics in each case led to the repetition of words between topics, while fewer topics led to less detailed results. For a selected group of topics, we defined a single label based on these top words. See *SI Appendix* for an explanation of the robustness analysis of the LDA model.

Scholarly impact is assessed through field- and year-normalized citations (64), using an open citation window covering publication years through the end of 2019. For each article, we infer the first authors' gender and distribution over racial categories as well as over topics. Each article then has a probability distribution over racial categories, a binary classification over gender, and a probability distribution over topics. Aggregate results are obtained using fractional counting over these three dimensions. For example, an article can have a first author whose name has a 0.7 racial classification probability of

\*Per 2010 US Census classifications, "Asian" refers to a person with origins East Asia, Southeast Asia, or the subcontinent of India; meanwhile, Native Hawaiian and Other Pacific Islander refers to a person having origins in Hawaii, Guam, Samoa, or other Pacific Islands.

<sup>†</sup>The authors use the term Latinx as a gender-neutral term, consistent with its perceived meaning, according to the 2019 National Survey of Latinos (57).

<sup>‡</sup><https://sciencebias.uni.lu/app/>.

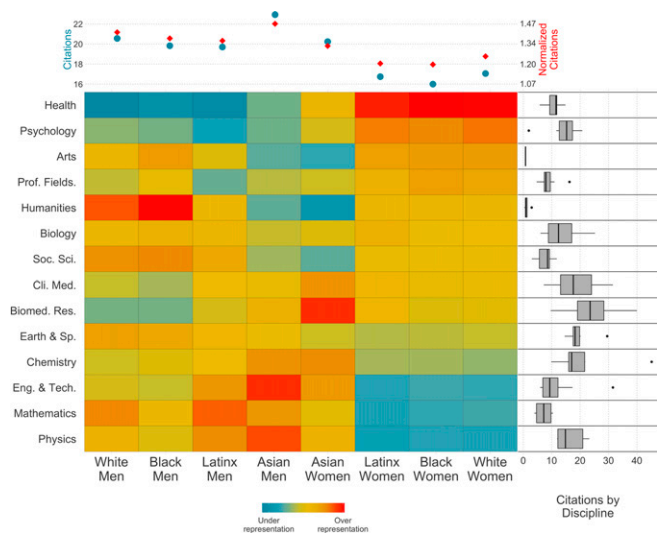
being a Black author, a 0.3 of being a White author, and whose gender is inferred to be woman. It also has a 0.8 probability on topic A, 0.1 on topic B, and so on. Therefore, this article contributes an additional 0.56 ( $0.7 \times 0.8$ ) authors to the group of Black women in topic A, 0.07 ( $0.7 \times 0.1$ ) authors to the group of Black women in topic B, and so on across all topics and racial groups for women. The weighted sum over these dimensions, plus the citations, gives us the aggregate results of the distribution over topics by race and gender and the average number of normalized citations per topic, race, and gender. Over- and underrepresentation in topics of racial groups and genders are based on the overall proportion that each group represents across all articles combined.<sup>§</sup>

## Results

Comparison of race and gender demographics of US first authors with that of the US population shows that White and Asian populations are overrepresented among US authors, while Black and Latinx populations are underrepresented (SI Appendix, Fig. S1). Relative representation varies by field (Fig. 1). Black, Latinx, and White women exhibit similar representation: They are highly underrepresented in Physics, Mathematics, and Engineering and overrepresented in Health (SI Appendix, Figs. S2 and S3), Psychology, and Arts. Asian women follow a different pattern, with underrepresentation in Arts, Humanities, and Social Sciences and overrepresentation in Biomedical Research, Chemistry, and Clinical Medicine. Black, Latinx, and White men are underrepresented in Psychology and Health, together with Asian men, but this latter group is also underrepresented in Humanities and Social Sciences and overrepresented in Physics, Engineering, Math, and Chemistry. Men first authors are generally more cited than women, and Asian authors are more cited than Black, Latinx, and White authors, both in raw citations and field-normalized citations. For White, Black, and Latinx women the citation gap reduces when considering normalized citations, showing that they are more present in lower-cited fields. Nevertheless, even when considering field-normalized citations, the gap remains.

To better understand and explain intersectional differences in citations, we explore the role of research topics for disciplines in the Humanities, Social Sciences, Professional Fields, and Health.<sup>¶</sup> Fig. 2 presents feminization—i.e., the proportion of women authors of each topic (y-axis)—by racialization—i.e., the proportion of authors from a racial group in each topic (x-axis)—for Social Sciences (Fig. 2A)<sup>#</sup> and Health (Fig. 2B).<sup>||</sup> The color of each node (topic) provides the mean number of citations, while size represents relative importance in the dataset.

In the Social Sciences, topics with the highest proportion of Asian authors are related to topics in economics and logistics, like stocks, consumers, firms, and market. These topics are also those where White and Black authors are least represented. Black authors are highly represented on topics of racial discrimination, African American culture, and African studies and communities. Religion is one of the few topics in which both Black and White men are overrepresented. Latinx authors are highly represented in topics related to immigrants, political identity, and racial discrimination, the latter of which is also shared with Black authors. Black and Latinx authors perform research on topics specific to language literacy, as well as on African and Latin-American countries, respectively. Latinx authors publish on topics associated with Latin-American issues and those that redefine the Latinx identity within the United States. Of particular interest is the topic of language literacy, which is both highly feminized and highly Latinx and constitutes a mixture of a traditional



**Fig. 1.** Scholarly impact and distribution of race and gender of authors by field. Average number of raw and field-normalized citations by group. Over- and underrepresentation of groups by discipline, with respect of their average proportion in all fields. Distribution of the average number of raw citations by specialty within each discipline; the hinges of the box correspond to the first and third quartiles, the whiskers extend to the lowest values no further than 1.5 times the interquartile range (IQR) from the hinge, and dots represent values further than 1.5 times IQR. Data consist of US first authors within the WOS from 2008 to 2019. Racial categories from the census corresponding to AIAN and two or more were excluded from the racial inference because of lack of data. On the vertical axis, fields are sorted by the relative over/underrepresentation of Black women authors. On top, we show the average number of citations by group, while on the right, each boxplot summarizes the distribution of citations of all papers published in those fields.

gender role (teaching, related with reproductive labor) and the learning of a second language, a topic that is highly relevant to migrant communities.

Fig. 2 also shows the coefficient of variation (CV) for each racial group's proportion on topics. A high CV means that the group has a high participation on some topics and a small participation on others, relative to its average proportion. Asian authors present the highest CV, while White authors exhibit lowest. This suggests that Asian authors are highly specialized, focusing on certain topics, while White authors are present in a wider range of topics. Black and Latinx authors show greater specialization, focusing on a smaller number of topics. White authors are more evenly distributed among topics; however, this is expected, as they account for the majority of the author population. The most highly feminized topics include gender-based violence, families, learning, and lesbian, gay, bisexual, and transgender (LGBT) studies. These results implicate a relationship between traditional gender roles, and topics that relate specifically with gender-based identity and inequality, and to nonhegemonic gender representations (i.e., gender expressions that do not correspond to binary male/female categorizations).

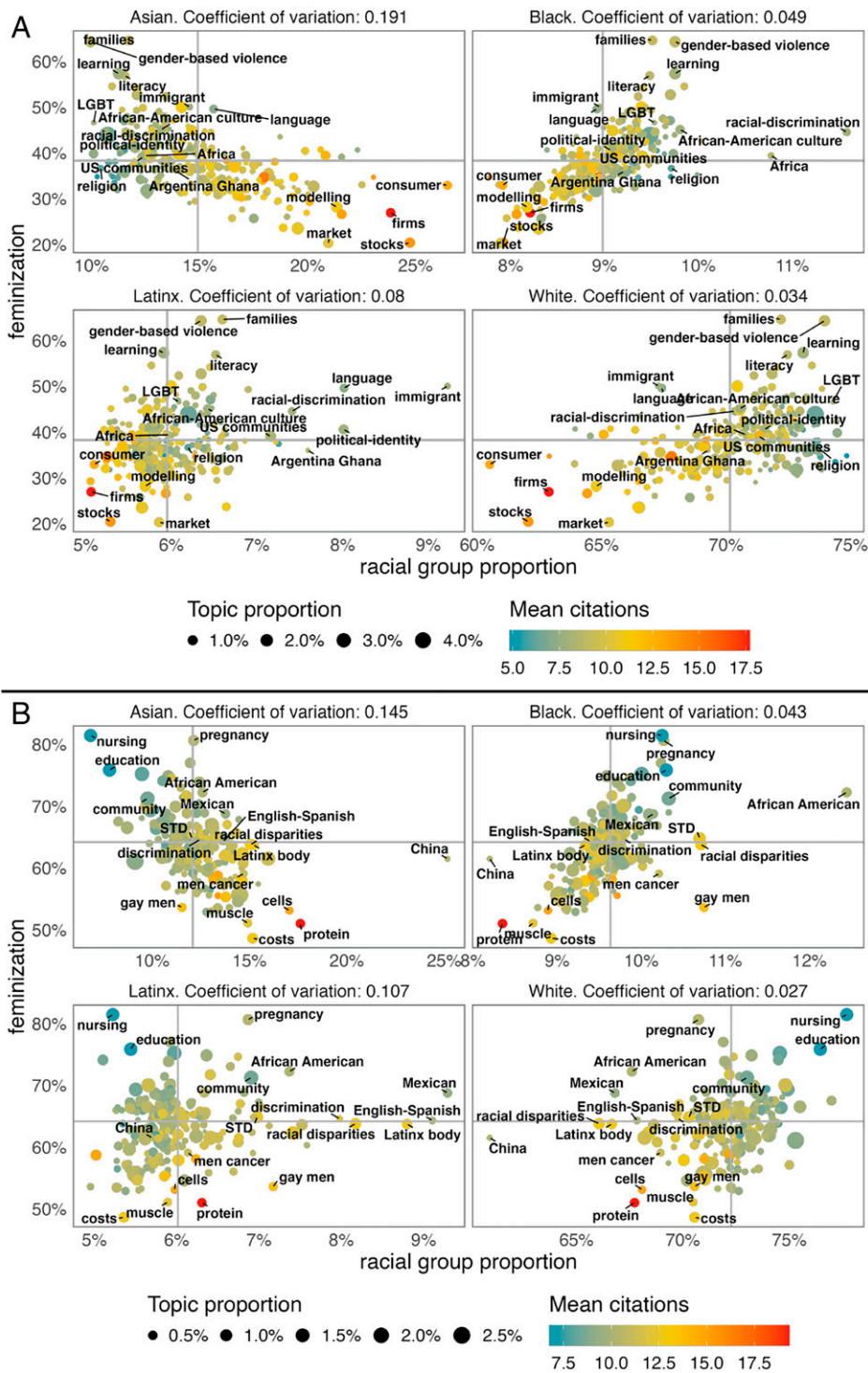
Several important topics appear at the intersection of race and gender. Because of space constraints, it is not possible to assign a label for all topics in Fig. 2; however, the accompanying website presents an interactive visualization displaying topics along the diagonal of Quadrant I that are both highly feminized and racialized. Among Black women authors, for example, we find topics such as black women violation (topic 122), equality promotion (topic 149), and social identity (topic 210). Among Latinx women authors, in addition to topics on language, we find residential segregation (topic 103), gender gap and international migration (topic 300), social class (topic 64), and global south (topic 225), among others.

<sup>§</sup> See the SI Appendix for how we operationalize over- and underrepresentation.

<sup>¶</sup> The analysis was also carried out for all fields: <https://sciencebias.uni.lu/app/>.

<sup>#</sup> SI Appendix, Table S1 provides the top words of each of these topics. An interactive version of this plot, with labels for all topics is available at <https://sciencebias.uni.lu/app/>.

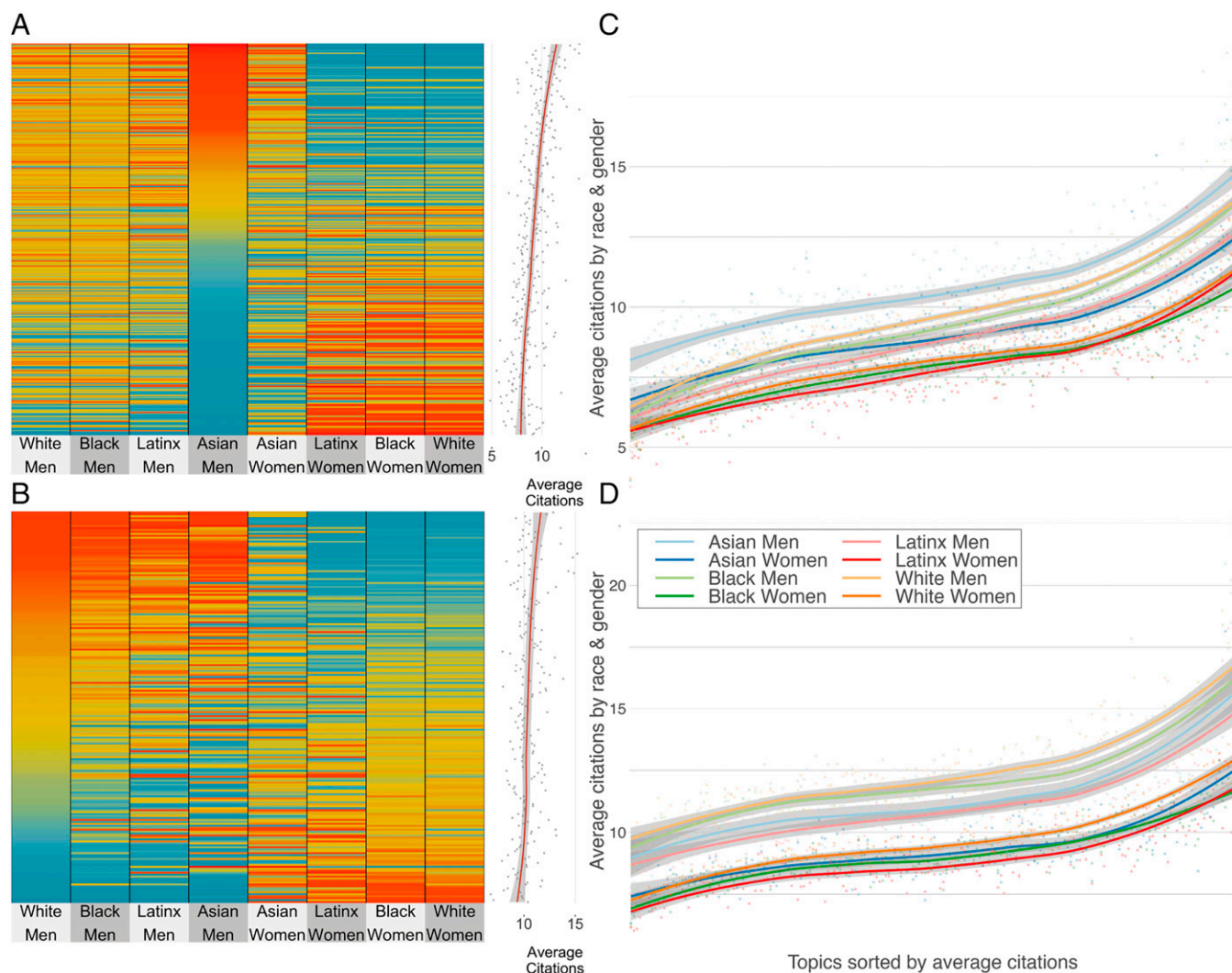
<sup>||</sup> SI Appendix, Table S2 provides the top words of each of these topics. An interactive version of this plot, with labels for all topics is available at <https://sciencebias.uni.lu/app/>.



**Fig. 2.** Distribution of topics by racial group and gender participation. (A) Social Sciences, Humanities, and Professional Fields, (B) Health. For Social Sciences, Humanities, and Professional Fields ( $n = 283,589$  articles), we train an LDA model for 300 topics. For Health ( $n = 142,032$ ), we train the LDA model for 200 topics. The vertical axis shows the proportion of women, while the horizontal axis shows the proportion by each racial group. In color, there is the mean number of citations by topic. The CV, as a standardized measure of variability, is provided for each racial group. Topics with the highest proportion in each race and in each gender are highlighted (labeled). Racial categories from the census corresponding to AIAN and two or more were excluded from the racial inference because of lack of data. Minimum number of average citations: 4.37, max. 15.74, and mean 9.25. STD, sexually transmitted disease.

In Health, the topics with the highest representation of Asian authors are China, proteins, cells, and the economics of health (i.e., costs). Black authors publish on topics about racial disparities and sexually transmitted diseases—with a special

emphasis on African Americans among Black women and gay men for Black men. This latter topic is also relevant for Latinx men. Latinx authors publish more on topics that mention the Latinx population, racial disparities, (a topic that is shared with



**Fig. 3.** Scholarly impact by topic. Over- and underrepresentation of groups by topic, in Social Sciences (A) and Health (B). Topics sorted by the participation of the most cited group in each case. On the right margin, the figure shows the number of citations by topic, and loess smoothing. Distribution of average topic citations by race and gender in Social Sciences (C) and in Health (D). Topics are sorted by average number of citations, and a smoothing function is drawn for each group using loess to model the evolution of the expected number of citations, as the topics become more cited on average. The gray shadow in each model represents the 95% confidence level, and therefore, when these shadows do not overlap, the differences between groups are significant. Racial categories from the census corresponding to AIAN and two or more were excluded from the racial inference because of lack of data.

Black authors), and English-Spanish, a topic similar to that which was previously found in the Social Sciences. The CV between topics by racial group also shows that Asian authors are the most specialized, followed by Black and Latinx authors, and White authors are the most ubiquitous across topics. The most feminized topics are about nursing, pregnancy, and education, reinforcing the association of women with care- and service-related research (65, 66).

**Scholarly Impact by Topic.** Our results demonstrate that macrolevel differences in citation rates are observed at the intersection of race and gender—even when controlling for disciplines (Fig. 1)—and that topic selection is related to author race and gender (Fig. 2). It stands to reason, therefore, that there might be a relationship between the populations engaged in certain topics and the citation density of these topics. Fig. 3 presents the over- and underrepresentation of race and gender of authors by topic, sorted by the participation of Asian men in Social Sciences (Fig. 3A) and by White men in Health (Fig. 3B), the two most highly cited groups in each discipline. The average number of citations of a topic is positively

correlated with the presence of Asian and White men. Fig. 3C and D provides the average number of citations by race and gender within each topic\*\* for each discipline, respectively. In the Social Sciences, Asian men have a higher number of citations; they tend to be more present within highly cited topics and are more cited than other groups within lower-cited topics. All other groups start with a relatively similar number of citations, which later split into three branches. White and Black men increase their relative number of citations to equalize those of Asian men in the highest cited topics. Latinx men and Asian women follow a similar course but yield fewer citations for the highest cited topics. Black, White, and Latinx women form a block with systematically fewer citations than all other groups. Health presents a stronger gender split: Men, regardless of racial categorizations, are significantly more cited along the distribution of topics, with White and Black men having slightly more citations than Asian and Latinx men. Women from all racial groups present a lower number of citations, with White

\*\*An interactive version of this visualization with information on each topic and group is available at <https://sciencebias.uni.lu/app/>.

women presenting a slightly higher number of citations for highly cited topics than Asian, Latinx, and Black women. This provides evidence of the intersectional between- and within-topic disadvantage for minoritized groups: 1) minoritized groups are overrepresented in lowly cited topics and underrepresented in highly cited topics; and 2) their work is less cited within and across topics, especially where they are underrepresented.

## Discussion

Inequalities in science have been studied for a century (67, 68), and several analyses have shown that these inequalities are the consequence of a nonmeritocratic scientific system (69–71). Our results show that minoritized authors tend to publish in scientific disciplines and on research topics that reflect their gendered and racialized social identities. Specifically, we have shown a contrast between the topic specialization of US Asian, Black, and Latinx first authors, reflected via a higher CV, and the ubiquity of White authors. The even participation of White authors across topics shows that the relation between race and research topic operates primarily on minoritized authors. In other terms, there is a privilege of choice in scientific knowledge production, wherein research on a particular topic is influenced by scientist's race and gender. As Bourdieu explains, the amount of scientific capital possessed by a researcher defines the strategies they can follow (72). The ubiquity of White men in science and across topics implies that this demographic group has a wider range of possible strategies to follow and an advantage in the way their scientific capital can be invested, reinforcing inequalities in scholarly outcomes.

We found that differences in research impact can be at least partially explained by topics' citation density but that within-topic differences remain. The compound effect of different citation rates of topics and unequal distribution on topics by race and gender leads to negative effects for marginalized groups and for science itself, as some topics become systematically less studied. The history of science is ripe with examples of understudied topics, such as female genitalia, which had direct implications on the life expectancy of women (73). Assuming constant productivity (74) and considering the career age of authors, we can estimate the cumulative loss in particular topics over the last 40 y, assuming that researchers with 20 y of publication activity produced 20 times that of incoming researchers. If the author distribution over the last 40 y would have matched the 2010 US Census, there would have been 29% more articles in public health, 26% more on gender-based violence, 25% more in gynecology and in gerontology, 20% more on immigrants and minorities, and 18% more on mental health (*SI Appendix, Figs. S4–S6*). While this counterfactual scenario is coarse, it highlights the fact that a different body of knowledge would be produced in the absence of inequalities and that this body would more closely reflect the spectrum of topics relevant across society. The diversification of the scientific workforce is necessary to create a scientific system whose results benefit all of society.

This paper has provided evidence of the relation between race, gender, research topic, and research impact and contributes to the wider dialogue on intersectional inequalities in science. However, race and gender are not the only spaces of inequality in science; several other variables should be included to create a fully intersectional understanding of inequalities in science. Socioeconomic status, when intersected with race, gender, and topic, is likely to have large effects: A recent study suggested that the estimated median childhood income among faculty is 23.7% higher than that of the general population (75). Inequities have also been observed on the basis of disability (76) and sexual orientation (77)—variables that are often excluded or underreported in studies of the scientific workforce. Attrition and career age (78, 79) may also play an

important role here, as well as the prestige of institutional affiliations (80, 81). Causal modeling that considers topic choice, along with markers of prestige, would be germane in understanding the different mechanisms through which systemic inequalities are mediated. Finally, racial categories used in this research are only meaningful in the context of the US academic workforce; further research should be performed to understand general patterns across the globe and provide insights on the role of science policy in mitigating disparities.

Discrimination defies notions of objective, apolitical, and meritocratic ideals in scientific discourse (21); a perception that serves to reinforce and mask race and gender biases in science (81). Structural racism (82) remains a persistent source of mental and physical strain on minoritized groups in the United States (83–87), whose calls for justice across socioeconomic (e.g., health-care, housing, education, finance, and criminal justice) and professional domains are intermittently elevated (and subsequently ignored) in accordance with the ebbs and flows of American racial discourse (88). Academia is no different in this regard (89, 90). The underrepresentation in science is similar to other sectors and may be attributed to the pervasive legacy of US federal- and state-sanctioned campaigns of systemic, racialized exclusion aimed to reduce the representation and participation of minoritized race groups in all aspects of human life (12, 17, 19, 91–94). Recent calls for increased transparency and accountability in graduate student recruitment, retention, and faculty hiring and promotion (95, 96) are particularly notable after the marked increase in media attention on anti-Black police violence, the Black Lives Matter movement, and the disproportionate impact of the COVID-19 pandemic on Black and Latinx populations (97–100) and on women academics (101–103). The effect of related policy interventions in response to these events remains to be tested.

Our analysis suggests structural effects that reproduce systemic inequities in terms of value assigned to particular topics, in both scientific evaluation and distribution of resources. Several policy recommendations emerge from this analysis. First, scientific institutions need to recognize the existence of knowledge gaps related with author race and gender segregation and promote topics in which gendered and racially minoritized authors are more present. Funding agencies can take immediate action to allocate increased funding in areas that have been historically underrepresented (31). Such funding will affect the entire academic reward system: Funding is strongly correlated with productivity and impact, both of which are associated with institutional advancement and rewards (104, 105). This has implications for individual scientists but also serves to increase the visibility of and participation in understudied areas. Second, institutions need to promote diverse participation within high-impact topics, taking into account the need for resources and initiatives that provide access for marginalized populations into high-prestige networks. Taken together, these activities will serve to both reduce the variance in impact across topics and reduce the within-topic disparities at the intersection of race and gender, thereby increasing equity in science and expanding the knowledge horizon.

**Data Availability.** Detailed methods and results tables and code and materials data have been deposited in the University of Luxembourg website (<https://sciencebias.uni.lu/app/>) and GitHub ([https://github.com/DiegoKoz/intersectional\\_inequalities](https://github.com/DiegoKoz/intersectional_inequalities)). All the data for this article are available in the supporting information. Restrictions apply to the proprietary bibliometric data, which is used under license from Clarivate Analytics. To obtain the bibliometric data in the same manner as the authors (i.e., by purchasing them), readers can contact Clarivate Analytics (<https://clarivate.com/webofsciencegroup/solutions/web-of-science/contact-us/>).

**ACKNOWLEDGMENTS.** We thank Alexis Bell, Will Hulsey, and Dakota Murray for contributions to an earlier phase of the project.

1. UNESCO, "Women in Science" (UNESCO Fact Sheet No. 55, 2019). [uis.unesco.org/sites/default/files/documents/fs55-women-in-science-2019-en.pdf](https://uis.unesco.org/sites/default/files/documents/fs55-women-in-science-2019-en.pdf). Accessed 21 December 2021.
2. V. Larivière, C. Ni, Y. Gingras, B. Cronin, C. R. Sugimoto, Bibliometrics: Global gender disparities in science. *Nature* **504**, 211–213 (2013).
3. US National Science Foundation, "Women, Minorities, and Persons with Disabilities in Science and Engineering" (Tech. Rep. NSF 17-310, National Center for Science and Engineering Statistics, Arlington, VA, 2017). <https://www.nsf.gov/statistics/2017/nsf17310/digest/occupation/women.cfm>. Accessed 21 December 2021.
4. L. Davis, R. Fry, College faculty have become more racially and ethnically diverse, but remain far less so than students. *Pew Research Center*, 31 July 2019. <https://www.pewresearch.org/fact-tank/2019/07/31/us-college-faculty-student-diversity/>. Accessed 17 December 2021.
5. O. Taylor, C. B. Apprey, G. Hill, L. McGrann, J. Wang, Diversifying the faculty. *Peer Rev.* **12**, Summer (2010).
6. B. Hussar et al., *The Condition of Education 2020. NCES 2020-144* (National Center for Education Statistics, 2020).
7. D. Li, C. Koedel, Representation and salary gaps by race-ethnicity and gender at selective public universities. *Educ. Res.* **46**, 343–354 (2017).
8. J. Huang, A. J. Gates, R. Sinatra, A.-L. Barabási, Historical comparison of gender inequality in scientific careers across countries and disciplines. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 4609–4616 (2020).
9. M. A. Bertolero et al., Racial and ethnic imbalance in neuroscience reference lists and intersections with gender. *bioRxiv* [Preprint] (2020). <https://www.biorxiv.org/content/10.1101/2020.10.12.336230v1> (Accessed 17 December 2021).
10. D. Murray et al., Gender and international diversity improves equity in peer review. *bioRxiv* [Preprint] (2019). <https://www.biorxiv.org/content/10.1101/400515v2> (Accessed 17 December 2021).
11. E. A. Eroshva et al., NIH peer review: Criterion scores completely account for racial disparities in overall impact scores. *Sci. Adv.* **6**, eaaz4868 (2020).
12. C. A. Moss-Racusin, J. F. Dovidio, V. L. Brescoll, M. J. Graham, J. Handelsman, Science faculty's subtle gender biases favor male students. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 16474–16479 (2012).
13. D. K. Ginther et al., Race, ethnicity, and NIH research awards. *Science* **333**, 1015–1019 (2011).
14. D. K. Ginther, S. Kahn, W. T. Schaffer, Gender, race/ethnicity, and National Institutes of Health R01 research awards: Is there evidence of a double bind for women of color? *Acad. Med.* **91**, 1098–1107 (2016).
15. S.-J. Leslie, A. Cimpian, M. Meyer, E. Freeland, Expectations of brilliance underlie gender distributions across academic disciplines. *Science* **347**, 262–265 (2015).
16. L. A. Rivera, When two bodies are (not) a problem: Gender and relationship status discrimination in academic hiring. *Am. Sociol. Rev.* **82**, 1111–1138 (2017).
17. A. A. Eaton, J. F. Saunders, R. K. Jacobson, K. West, How gender and race stereotypes impact the advancement of scholars in STEM: Professors' biased evaluations of physics and biology post-doctoral candidates. *Sex Roles* **82**, 127–141 (2020).
18. J. B. Mustaffa, Mapping violence, naming life: A history of anti-Black oppression in the higher education system. *Int. J. Qual. Stud. Educ.* **30**, 711–727 (2017).
19. E. A. Odekunle, Dismantling systemic racism in science. *Science* **369**, 780–781 (2020).
20. J. Kaiser, NIH apologizes for 'structural racism,' pledges change. *Science* **371**, 977 (2021).
21. S. E. Cozzens, Distributive justice in science and technology policy. *Sci. Public Policy* **34**, 85–94 (2007).
22. R. K. Merton, *The Sociology of Science: Theoretical and Empirical Investigations* (University of Chicago Press, 1974).
23. C. Herring, Does diversity pay? Race, gender, and the business case for diversity. *Am. Sociol. Rev.* **74**, 208–224 (2009).
24. R. B. Freeman, W. Huang, Collaborating with people like me: Ethnic co-authorship within the United States. *J. Labor Econ.* **3**, S289–S318 (2015).
25. B. K. AlShebli, T. Rahwan, W. L. Woon, The preeminence of ethnic diversity in scientific collaboration. *Nat. Commun.* **9**, 5163 (2018).
26. B. Hofstra et al., The diversity–innovation paradox in science. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 9284–9291 (2020).
27. J. A. Clayton, F. S. Collins, Policy: NIH to balance sex in cell and animal studies. *Nature* **509**, 282–283 (2014).
28. S. L. Klein et al., Opinion: Sex inclusion in basic research drives discovery. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 5257–5258 (2015).
29. C. R. Sugimoto, Y.-Y. Ahn, E. Smith, B. Macaluso, V. Larivière, Factors affecting sex-related reporting in medical research: A cross-disciplinary bibliometric analysis. *Lancet* **393**, 550–559 (2019).
30. M. W. Nielsen, J. P. Andersen, L. Schiebinger, J. W. Schneider, One and a half million medical papers reveal a link between author gender and attention to gender and sex analysis. *Nat. Hum. Behav.* **1**, 791–796 (2017).
31. R. Koning, S. Samila, J.-P. Ferguson, Who do we invent for? Patents by women focus more on women's health, but few women get to invent. *Science* **372**, 1345–1348 (2021).
32. T. A. Hoppe et al., Topic choice contributes to the lower rate of NIH awards to African-American/black scientists. *Sci. Adv.* **5**, eaav7238 (2019).
33. K. Crenshaw, "Demarginalizing the intersection of race and sex: A black feminist critique of antidiscrimination doctrine, feminist theory, and antiracist politics" in *Feminist Legal Theory*, K. T. Bartlett, R. Kennedy, Eds. (Routledge, 1989), pp. 57–80.
34. G. Gutiérrez y Muhs, Y. F. Niemann, C. G. González, A. P. Harris, *Presumed Incompetent: The Intersections of Race and Class for Women in Academia* (Utah State University Press, 2012).
35. E. O. McGee, L. Bentley, The troubled success of Black women in STEM. *Cogn. Instr.* **35**, 265–289 (2017).
36. P. H. Collins, S. Bilge, *Intersectionality* (John Wiley & Sons, 2020).
37. K. Crenshaw, Mapping the margins: Intersectionality, identity politics, and violence against women of color. *Stanford Law Rev.* **43**, 1241–1299 (1990).
38. S. M. Lord et al., Who's persisting in engineering? A comparative analysis of female and male Asian, black, Hispanic, Native American, and white students. *J. Women Minor. Sci. Eng.* **15**, 167–190 (2009).
39. P. H. Collins, Intersectionality's definitional dilemmas. *Annu. Rev. Sociol.* **41**, 1–20 (2015).
40. N. E. Brown, S. A. Gershon, Examining intersectionality and symbolic representation. *Polit. Groups Identities* **5**, 500–505 (2017).
41. L. Hong, S. E. Page, Groups of diverse problem solvers can outperform groups of high-ability problem solvers. *Proc. Natl. Acad. Sci. U.S.A.* **101**, 16385–16389 (2004).
42. V. May, "Intellectual genealogies, intersectionality, and Anna Julia Cooper" in *Feminist Solidarity at the Crossroads: Intersectional Women's Studies for Transracial Alliance*, K. M. Vaz, G. L. Lemons, Eds. (Routledge, 2012), pp. 59–71.
43. L. Holman, D. Stuart-Fox, C. E. Hauser, The gender gap in science: How long until women are equally represented? *PLoS Biol.* **16**, e2004956 (2018).
44. B. Macaluso, V. Larivière, T. Sugimoto, C. R. Sugimoto, Is science built on the shoulders of women? A study of gender differences in contributorship. *Acad. Med.* **91**, 1136–1142 (2016).
45. J. D. West, J. Jacquet, M. M. King, S. J. Correll, C. T. Bergstrom, The role of gender in scholarly authorship. *PLoS One* **8**, e66212 (2013).
46. H. O. Witteman, M. Hendricks, S. Straus, C. Tannenbaum, Are gender gaps due to evaluations of the applicant or the science? A natural experiment at a national funding agency. *Lancet* **393**, 531–540 (2019).
47. C. B. Leggon, Women in science: Racial and ethnic differences and the differences they make. *J. Technol. Transf.* **31**, 325–333 (2006).
48. L. D. Cook, Violence and economic activity: Evidence from African American patients, 1870–1940. *J. Econ. Growth* **19**, 221–257 (2014).
49. L. Smith-Doerr, S. N. Alegria, T. Sacco, How diversity matters in the US science and engineering workforce: A critical review considering integration in teams, fields, and organizational contexts. *Engag. Sci. Technol. Soc.* **3**, 139–153 (2017).
50. A. Johnson, J. Brown, H. Carlone, A. K. Cuevas, Authoring identity amidst the treacherous terrain of science: A multiracial feminist examination of the journeys of three women of color in science. *J. Res. Sci. Teach.* **48**, 339–366 (2011).
51. R. Kachaf, L. Ko, A. Hodari, M. Ong, Career–life balance for women of color: Experiences in science and engineering academia. *J. Divers. High. Educ.* **8**, 175–191 (2015).
52. K. Owens, Colorblind science? Perceptions of the importance of racial diversity in science research. *Spontaneous Generations*, **8**, 13–21 (2016).
53. V. Larivière et al., Contributorship and division of labor in knowledge production. *Soc. Stud. Sci.* **46**, 417–435 (2016).
54. V. Larivière, D. Pontille, C. R. Sugimoto, Investigating the division of scientific labor using the Contributor Roles Taxonomy (CRediT). *Quant. Sci. Stud.* **2**, 111–128 (2021).
55. E. Caron, N. J. van Eck, "Large scale author name disambiguation using rule-based scoring and clustering" in *Proceedings of the 19th International Conference on Science and Technology Indicators*, E. Noyons, Ed. (CWTS-Leiden University, Leiden, 2014) pp. 79–86.
56. US Census Bureau, Frequently Occurring Surnames from the 2010 Census. (USCB, 2016). [https://www.census.gov/topics/population/genealogy/data/2010\\_surnames.html](https://www.census.gov/topics/population/genealogy/data/2010_surnames.html). Accessed 21 December 2021.
57. L. Noe-Bustamante, L. Mora, M. H. Lopez, About one-in-four US Hispanics have heard of Latinx, but just 3% use it. *Pew Research Center*, 11 August 2020. <https://www.pewresearch.org/hispanic/2020/08/11/about-one-in-four-u-s-hispanics-have-heard-of-latinx-but-just-3-use-it/>. Accessed 17 December 2021.
58. D. Kozłowski et al., "Avoiding bias when inferring race using name-based approaches" in *Proceedings of the 18th International Conference on Scientometrics & Informetrics*, W. Glänzel, S. Heeffer, P. S. Chi, R. Rousseau, Eds. (International Society for Scientometrics and Informetrics, Belgium, 2021) pp. 597–608.
59. R. G. Fryer Jr. S. D. Levitt, The causes and consequences of distinctively black names. *Q. J. Econ.* **119**, 767–805 (2004).
60. H. Girma, Black names, immigrant names: Navigating race and ethnicity through personal names. *J. Black Stud.* **51**, 16–36 (2020).
61. K. S. Hamilton, "Subfield and level classification of journals" (CHI Research No. 2012-R, CHI Research, Haddon Heights, NJ, 2003).
62. D. M. Blei, A. Y. Ng, M. I. Jordan, Latent Dirichlet allocation. *J. Mach. Learn. Res.* **3**, 993–1022 (2003).
63. D. Kozłowski, V. Semeshenko, A. Molinari, Latent Dirichlet allocation model for world trade analysis. *PLoS One* **16**, e0245393 (2021).
64. L. Waltman, A review of the literature on citation impact indicators. *J. Informetrics* **10**, 365–391 (2016).
65. C. Cockburn, *Machinery of Dominance: Women, Men, and Technical Know-how* (Northeastern University Press, Boston, MA, 1988).

66. A. Witz, *Professions and Patriarchy* (Routledge, New York, 1992).
67. A. J. Lotka, The frequency distribution of scientific productivity. *J. Wash. Acad. Sci.* **16**, 317–323 (1926).
68. J. R. Cole, S. Cole, *Social Stratification in Science* (University of Chicago Press, Chicago, 1973).
69. A. C. Johnson, Unintended consequences: How science professors discourage women of color. *Sci. Educ.* **91**, 805–821 (2007).
70. A. J. Rodriguez, Busting open the meritocracy myth: Rethinking equity and student achievement in science education. *J. Women Minor. Sci. Eng.* **4**, 195–216 (1998).
71. A. Zivony, Academia is not a meritocracy. *Nat. Hum. Behav.*, **3**, 1037 (2019).
72. P. Bourdieu, The specificity of the scientific field and the social conditions of the progress of reason. *Soc. Sci. Inf.* **14**, 19–47 (1975).
73. M. Ah-King, A. B. Barron, M. E. Herberstein, Genital evolution: Why are females still understudied? *PLoS Biol.* **12**, e1001851 (2014).
74. D. Fanelli, V. Larivière, Researchers' individual publication rate has not increased in a century. *PLoS One* **11**, e0149504 (2016).
75. A. Morgan, A. Clauset, D. Larremore, N. LaBerge, M. Galesic, Socioeconomic roots of academic faculty, Socarxiv [Preprint] (2021). <https://osf.io/preprints/socarxiv/6wjxc> (Accessed 17 December 2021).
76. J. J. Yerbury, R. M. Yerbury, Disabled in academia: To be or not to be, that is the question. *Trends Neurosci.* **44**, 507–509 (2021).
77. E. Gibney, Discrimination drives LGBT+ scientists to think about quitting. *Nature* **571**, 16–17 (2019).
78. A. M. Petersen, M. Riccaboni, H. E. Stanley, F. Pammolli, Persistence and uncertainty in the academic career. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 5213–5218 (2012).
79. National Science Board, "Science and Engineering Indicators 2018 (Tech. Rep. NSB-2018-1, National Science Foundation, Alexandria, VA, 2018). <https://nsf.gov/statistics/2018/nsb20181/>. Accessed 21 December 2021.
80. D. Crane, The gatekeepers of science: Some factors affecting the selection of articles for scientific journals. *Am. Sociol.* **2**, 195–201 (1967).
81. A. Clauset, S. Arbesman, D. B. Larremore, Systematic inequality and hierarchy in faculty hiring networks. *Sci. Adv.* **1**, e1400005 (2015).
82. Z. D. Bailey et al., Structural racism and health inequities in the USA: Evidence and interventions. *Lancet* **389**, 1453–1463 (2017).
83. J. S. Jackson et al., Racism and the physical and mental health status of African Americans: A thirteen year national panel study. *Ethn. Dis.* **6**, 132–147 (1996).
84. J. A. Lewis, M. G. Williams, E. J. Peppers, C. A. Gadson, Applying intersectionality to explore the relations between gendered racism and health among Black women. *J. Couns. Psychol.* **64**, 475–486 (2017).
85. D. R. Williams, J. A. Lawrence, B. A. Davis, Racism and health: Evidence and needed research. *Annu. Rev. Public Health* **40**, 105–125 (2019).
86. D. S. Curtis et al., Highly public anti-Black violence is associated with poor mental health days for Black Americans. *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2019624118 (2021).
87. M. Davis, Anti-Black practices take heavy toll on mental health. *Nat. Hum. Behav.* **5**, 410 (2021).
88. R. Kennedy, The ebb and flow of racial progress: Is this one of those rare breakthrough moments like Lincoln's emancipation and the civil rights era of the 1960s—Or prologue to more disappointment? *The American Prospect*, 3 December 2020. <https://prospect.org/culture/books/the-ebb-and-flow-of-racial-progress/>.
89. E. O. McGee, Interrogating structural racism in STEM higher education. *Educ. Res.* **49**, 633–644 (2020).
90. A. H. Wingfield, Systemic racism persists in the sciences. *Science* **369**, 351 (2020).
91. M. I. Norton, S. R. Sommers, Whites see racism as a zero-sum game that they are now losing. *Perspect. Psychol. Sci.* **6**, 215–218 (2011).
92. A. Morris, *The Scholar Denied: W. E. B. Du Bois and the Birth of Modern Sociology* (University of California Press, 2015).
93. K. S. Brown, K. Kijakazi, C. Runes, M. A. Turner, "Confronting structural racism in research and policy analysis" (Publication 99852, Urban Institute, 2019, [https://www.urban.org/sites/default/files/publication/99852/confronting\\_structural\\_racism\\_in\\_research\\_and\\_policy\\_analysis\\_0.pdf](https://www.urban.org/sites/default/files/publication/99852/confronting_structural_racism_in_research_and_policy_analysis_0.pdf)).
94. Z. D. Bailey, J. M. Feldman, M. T. Bassett, How structural racism works—Racist policies as a root cause of US racial health inequities. *N. Engl. J. Med.* **384**, 768–773 (2021).
95. M. P. Bell, D. Berry, J. Leopold, S. Nkomo, Making Black Lives Matter in academia: A black feminist call for collective action against anti-blackness in the academy. *Gen. Work Organ.* **28**, 39–57 (2020).
96. M. Jindal et al., Getting "Our House" in order: Re-building academic pediatrics by dismantling the Anti-Black Racist Foundation. *Acad. Pediatr.* **20**, 1044–1050 (2020).
97. C. T. Laurencin, A. McClinton, The COVID-19 pandemic: A call to action to identify and address racial and ethnic disparities. *J. Racial Ethn. Health Disparities* **7**, 398–402 (2020).
98. A. v. Dorn, R. E. Cooney, M. L. Sabin, COVID-19 exacerbating inequalities in the US. *Lancet* **395**, 1243–1244 (2020).
99. M. A. Garcia, P. A. Homan, C. Garcia, T. H. Brown, The color of COVID-19: Structural racism and the disproportionate impact of the pandemic on older Black and Latinx adults. *J. Gerontol.: Ser. B* **76**, e75–e80 (2021).
100. E. McGee, Y. Fang, Y. Ni, T. Monroe-White, How an antisience president and the COVID-19 pandemic altered the career trajectories of STEM PhD students of color. *AERA Open* **7**, 23328584211039217 (2021).
101. A. K. Ribarovska, M. R. Hutchinson, Q. J. Pittman, C. Pariente, S. J. Spencer, Gender inequality in publishing during the COVID-19 pandemic. *Brain Behav. Immun.* **91**, 1–3 (2021).
102. K. R. Myers et al., Unequal effects of the COVID-19 pandemic on scientists. *Nat. Hum. Behav.* **4**, 880–883 (2020).
103. P. Vincent-Lamarre, C. R. Sugimoto, V. Larivière, The decline of women's research production during the coronavirus pandemic. *Nature Index*, 19 May 2020. <https://www.natureindex.com/news-blog/decline-women-scientist-research-publishing-production-coronavirus-pandemic>. Accessed 17 December 2021.
104. B. A. Jacob, L. Lefgren, The impact of research grant funding on scientific productivity. *J. Public Econ.* **95**, 1168–1177 (2011).
105. P. Mongeon, C. Brodeur, C. Beaudry, V. Larivière, Concentration of research funding leads to decreasing marginal returns. *Res. Eval.* **25**, 396–404 (2016).