

Evolution and Behaviour

Race and gender inequalities in citations and research topics in the US

by **Diego Kozlowski**¹ | PhD Student and **Thema Monroe-White**¹ | Assistant Professor

¹: University of Luxembourg, Berry College

This Break was edited by Quentin Laurent, *Senior Scientific Editor* - TheScienceBreaker

ABSTRACT

The diversity of the US population is not well represented in its scientific workforce. In our research, we use large scale bibliometric databases to understand the relation between scientist identities, their research topics, and their impact. Our findings suggest that inclusion is key for a diversified research portfolio.



Image credits: RF_Studio – Pexels

Science is often perceived as a meritocratic landscape, where the “best” authors receive a larger share of citations, grants, and funding to continue their research lines, and subsequent prestige within the scientific community. However, this conceptualization ignores the fact that science is a social phenomenon, occurring within inequality-ridden societies where an individuals' participation in different fields depends highly on demographic markers. Race and gender are two identity markers that shape today's landscape of

inequality in the United States. Our research focuses on how the intersectional identities of authors – considering both race and gender identity in our analysis – affect the attention their work receives and the topics they study.

Our research aims at understanding how authors of different race and gender identities are represented in the scientific workforce, within specific disciplines and research topics. To do so, we used a bibliometric

database with over five million articles published by more than a million first authors from the United States between 2008 and 2019. We used the authors names to infer their race and gender, and the text of their titles and abstracts to infer their research topics. We then calculated the probability that each intersectional identity group has of publishing by discipline and, for Social Sciences and Health, by specific research topics. We also analyzed the number of citations each group and research topic receive. Our work provides new quantitative evidence supporting a mostly qualitative literature about race and gender biases in science. Our results can also be explored on [this website](#).

In aggregate, we found that White men are overrepresented in science publications, with respect to their proportion in the US population. Women are underrepresented, as well as marginalized racial groups, like Black and Latinx. At the intersection of these identities, we find that Black and Latinx women are the most underrepresented groups in science.

When studying the over and under representation of these groups with respect to their proportion in science overall, we found a large correlation between gender and research fields. Black, Latinx, and White women tend to publish more about Health and Psychology, and are particularly underrepresented in Physics, Mathematics, and Engineering. This speaks to the influence of traditional gender roles in society on science knowledge production, where care and reproductive labor are seen as the responsibility of women. It also demonstrates the need to dismantle stereotypes such that women have more space to freely follow their interests and passions.

This analysis provides a broad strokes picture of these patterns; however, in order to gain further insight, we decided to study specific research topics within the fields of Health and Social Science. In general, we found a large correlation between the research interests and the racial and gender identities of researchers. For example, Black authors publish more

about topics of racial discrimination and Afro-American culture, while Latinx authors publish relatively more on migration and language difficulties. The fields of gender-based violence and learning in Social Science show more publications from women, as well as nursing and pregnancy in Health. When it comes to research specialization, what influences topic space is more than simply traditional gender stereotypes and roles. The life experiences that accompany gender and racial identities also have an impact on the research interests and the questions authors pose and seek to answer.

This finding demonstrates how important demographic diversity is for science. Combined with the previously mentioned underrepresentation of marginalized groups in science, we found that all these topics – that are mainly researched by these groups – are understudied. There are knowledge gaps in society precisely on issues that affect the lives of marginalized groups. If we consider the counterfactual scenario in which the proportion by race and gender would match the general population, we would see 29% more articles in public health, 26% more on gender-based violence, or 25% more about gynecology and gerontology. The diversification of the scientific workforce is therefore necessary to create a scientific system whose results benefit us all equally.

Our study provides new evidence on race and gender inequality in science. Our findings suggest that the socially constructed identities of the producers of scientific knowledge affects what science is produced, with implications for the whole of society. Given the method used to infer race and gender, we were only able to identify authors' gender in a binary way. Our method was also unable to identify Native American authors. These are important limitations that need to be addressed with survey-based and qualitative methods. We also focused only on US authors. We hope that this work opens the path for future work in other countries, as well as on different dimensions through which intersectional inequalities take place.