



Author mentions in science news reveal widespread disparities across name-inferred ethnicities

Hao Peng^{1,2,3} , Misha Teplitskiy^{1,4} , and David Jurgens^{1,5} 

¹School of Information, University of Michigan, Ann Arbor, MI, USA

²Kellogg School of Management, Northwestern University, Evanston, IL, USA

³Northwestern Institute on Complex Systems, Northwestern University, Evanston, IL, USA

⁴Laboratory for Innovation Science at Harvard, Harvard University, Boston, MA, USA

⁵Department of Computer Science & Engineering, University of Michigan, Ann Arbor, MI, USA

an open access  journal



Citation: Peng, H., Teplitskiy, M., & Jurgens, D. (2024). Author mentions in science news reveal widespread disparities across name-inferred ethnicities. *Quantitative Science Studies*, 5(2), 351–365. https://doi.org/10.1162/qss_a_00297

DOI:
https://doi.org/10.1162/qss_a_00297

Peer Review:
https://www.webofscience.com/api/gateway/wos/peer-review/10.1162/qss_a_00297

Supporting Information:
https://doi.org/10.1162/qss_a_00297

Received: 17 February 2023
Accepted: 19 January 2024

Corresponding Authors:
Hao Peng
hao.peng@kellogg.northwestern.edu
David Jurgens
jurgens@umich.edu

Handling Editor:
Vincent Larivière

Copyright: © 2024 Hao Peng, Misha Teplitskiy, and David Jurgens.
Published under a Creative Commons Attribution 4.0 International (CC BY 4.0) license.



Keywords: ethnicity, gender, inequality, media coverage, science of science, visibility

ABSTRACT

Media outlets play a key role in spreading scientific knowledge to the public and raising the profile of researchers among their peers. Yet, how journalists choose to present researchers in their stories is poorly understood. Using a comprehensive data set of 223,587 news stories from 288 US outlets reporting on 100,486 research papers across all areas of science, we investigate whether authors' ethnicities, as inferred from names, are associated with whether journalists explicitly mention them by name. We find substantial disparities in mention rates across ethnic names. Researchers with non-Anglo names, especially those with East Asian and African names, are significantly less likely to be mentioned in their news stories, even with extensive controls for author prestige, semantic content, news outlets, publication venues, and research topics. The disparities are not fully explained by affiliation locations, suggesting that pragmatic factors play only a partial role. Furthermore, among US-based authors, journalists more often use authors' institutions instead of names when referring to non-Anglo-named authors, suggesting that journalists' rhetorical choices are also key. Overall, this study finds evidence of ethnic disparities in how often researchers are described in the media coverage of their research, likely affecting thousands of non-Anglo-named scholars in our data alone.

1. INTRODUCTION

Scientific breakthroughs often attract media attention, which serves as a key mechanism for public dissemination of new knowledge (Brossard & Scheufele, 2013; Scheufele, 2013). Science media coverage not only distills research insights but also puts a face on who was responsible for the research. The media coverage can then feed back into researchers' careers (Fanelli, 2013). Furthermore, science news reporting may over time shift the public's perception of *who* a scientist is (Miller, Nolla et al., 2018). Underrepresenting particular demographic groups can perpetuate the view that scientists are White men (Banchefsky, Westfall et al., 2016; Turner, González, & Wood, 2008), and potentially weaken the pipeline of recruiting diverse students into academic careers (Hill, McQuillan et al., 2018; Hofstra, Kulkarni et al., 2020; Reuben, Sapienza, & Zingales, 2014).

Science media coverage can be separated into two stages: the *likelihood of coverage*—whose paper gets reported—and the *quality of coverage*—given that a paper is reported,

how is the depth of coverage? For example, high-quality coverage may describe the sophistication of the research, eminence of the scholars, or credit them in quotes and name attributions.

Recent research has examined disparities in the first stage by comparing the coverage distribution to the overall distribution of published scholarship (i.e., P(paper reported | paper published) (Chapman, Hemingway et al., 2022; Davidson & Greene, 2024; Vászrhelyi, Zakhlebin et al., 2021)). These studies find disparities in the amount of media coverage by author gender and country. Yet coverage can differ not only in likelihood but also in its depth and quality. The contribution of this paper is to examine ethnicity-related disparities in coverage quality among scientific papers receiving at least some coverage. There are substantial trade-offs in this choice of research design. On the one hand, focusing on the set of papers already deemed newsworthy sidesteps some of the potential confounders affecting the association between ethnicity and coverage, in particular research topic, newsworthiness, and quality. On the other hand, the design leaves outside its purview the many other ways in which scientists' ethnicity may affect media coverage, including potential biases in the selection of individuals into topics (Kozlowski, Larivière et al., 2022), institutions, publishing outlets (Peng, Lakhani, & Teplitskiy, 2021), coverage amount, and so on. It is possible that many of these disparities compound across stages, making those found in any one stage a severe underestimate of the overall extent. We highlight that this paper contributes only a piece to a much larger puzzle.

Here, we focus on one highly salient feature related to the quality of coverage—whether the author is mentioned or credited by name in the stories covering their research (i.e., P(name mentioned | paper reported)). By focusing on *how* rather than *whether* reporters choose to report a scientific paper, our research design can precisely control for papers' newsworthiness and analyze important factors related to author demographics.

In writing about specific scientific advances, journalists face choices over how much attention to devote to each relevant researcher, and whom to ignore altogether. Empirical and theoretical literature motivates the possibility that ethnic disparities exist in journalists' choices of who to feature and the nature of the resulting coverage (Callison & Young, 2019; Robinson & Culver, 2019; Sui, Paul et al., 2018).

Empirically, a number of studies have established gender and ethnic disparities in conventional scientific outcomes, such as funding (Ginther, Schaffer et al., 2011; Hoppe, Litovitz et al., 2019; Oliveira, Ma et al., 2019), publications (Peng et al., 2021; Way, Larremore, & Clauset, 2016), and citations (Huang, Gates et al., 2020; Larivière, Ni et al., 2013), as well as scientists' online visibility (Peng, Teplitskiy et al., 2022; Vasarhelyi et al., 2021). Furthermore, research points to demographic disparities and the stereotyped media coverage of the general population (Behm-Morawitz & Ortiz, 2013; Devitt, 2002; Jia, Lansdall-Welfare, & Cristianini, 2015; Jia, Lansdall-Welfare et al., 2016; Merullo, Yeh et al., 2019; Smith, 1997). The presence of abundant ethnic bias in traditional media suggests that the disparities may appear in how science is covered at the very latter stage as research disseminates to the public.

Theoretically, we hypothesize a number of mechanisms that may produce ethnic disparities in which authors are mentioned in media coverage of science, and test them where possible. First, US-based journalists may face pragmatic difficulties in interviewing researchers in distant time zones and possibly with country-specific adoptions of different communication technologies (Lin, Frey, & Wu, 2022). Second, even for authors located within the same geographical region (e.g., in the United States), certain authors may have limited proficiency in speaking English. Furthermore, journalists may rely on their professional networks to contact sources. Analyses of the media landscape in the United States (Clark, 2018; Grieco, 2018) and other markets (Nielsen, Selva, & Andi, 2020) show that the demographics of journalists and editors

are highly unrepresentative of the broader populations. The demographics of journalists are likely to correlate with that of individuals in their professional networks (McPherson, Smith-Lovin, & Cook, 2001), suggesting that the researchers journalists can reach most readily are also unrepresentative. To the extent that these pragmatic factors—interviewing difficulties and professional networks—correlate with the perceived ethnicities of names, certain researchers may be more or less frequently mentioned.

Third, although science journalists aim to write stories that appear credible to their audiences (Sundar, 1998), they may lack direct information on the credibility of authors of the relevant research papers and may not have the time to acquire such information. Facing unfamiliar names and time constraints, journalists may rely on stereotypes, inferring for example that some researchers are less competent or authoritative on some topics than others, or expecting their audiences to harbor such perceptions. Prior research has found such stereotyping in the context of researcher gender and gender-typical research topics (Knobloch-Westerwick, Glynn, & Huges, 2013). Inferences of competence and authoritativeness can lead journalists to choose some names over others, which is a form of statistical discrimination (Lang & Lehmann, 2012; Neumark, 2018). Fourth, journalists may not be the relevant actors at all. Some news coverage originates from press releases created by in-house public relations staff at universities. News outlets often reprint these press releases in part or in full, and any disparities therein may thus be passed on directly to the outlets and their audiences.

Here, we present the first large-scale and science-wide analysis of ethnic disparities in author mentions in science news covering research papers and explore the mechanisms producing them based on a computational analysis of 223,587 news stories mentioning 100,486 published papers. *By focusing on papers that already were deemed newsworthy, our research design sidesteps the question of whose research is covered in the news in the first place, choices which may themselves be associated with ethnicity.*

We use the term *ethnicity* rather than nationality or race for two reasons: An author's nationality is largely masked by their affiliations and is very fluid, especially in the US context; and journalists only have access to author names upon reading the paper, and name reflects cultural origin that is more related to ethnicity and signals a richer set of information than race.

Lacking the information about authors' self-identity, we based our study on the *perceived* ethnicity inferred from names to distinguish it from authors' true ethnicity. This research choice entails substantial trade-offs. In fact, authors' self-identities may differ from their perceived ones, and some authors may self-identify with more than one ethnicity. In some cases, journalists know authors' self-identified ethnicity. Nevertheless, in many cases, journalists will not know how authors self-identify and rather infer them from names (Crabtree, Gaddis et al., 2022; Crabtree, Kim et al., 2023). In these cases, using authors' self-identities would be problematic, as it would misrepresent the actual perceptions journalists form and possibly use when they write their stories.

Our adoption of the perceived-ethnicity construct and the operationalization of it via names has three merits: Using the perception of ethnicity enables us to measure disparities in the imperfect information environment that journalists actually face where the ethnicity of the author is not known upon first seeing the paper, and is thus more likely to illuminate their decision processes; the construct of perceived ethnicity inferred via names has been widely used for decades in audit studies that use names to signal ethnicity or race to evaluators (Bertrand & Mullainathan, 2004; Gaddis, 2015, 2017a; Gaddis & Ghoshal, 2020) and is shown to be a highly effective proxy for studying disparities at scale; and the literature suggests that self-identified ethnicity and the perceived one are highly correlated, and that humans

can and do infer ethnicity from names fairly accurately (Gaddis, 2017a, 2017b; Sood & Laohaprapanon, 2018; Sweeney, 2013).

2. DATA AND ANALYSIS DESIGN

We constructed a multidisciplinary data set by combining news stories with metadata of the scientific papers they cover, and then inferred the demographic attributes of the papers' authors based on their names. Our final corpus of news-reported papers was sourced from [Altmetric.com](https://www.altmetric.com), which consists of 223,587 news stories from 288 US-based outlets reporting on 100,486 scientific papers, with 276,202 story-paper mention pairs. For all these new stories, we have the textual content, the mention date, and the outlet information.

We adopted a regression model to examine whether the paper's authors are mentioned by name in the news story. Because journalists can mention and feature several authors when covering a paper in their news story, we treated each (story, paper, author) triplet as an observation in the regression. As the first author and the last author often contribute most to the work and are recognized as such in science journalism guidelines (Blum, Knudson, & Henig, 2006), we included them in our analysis by default. We also included any additional corresponding author of a paper. Thus, for each paper, we focused on authors at the highest "risk" of being mentioned in the story including first author, last author, and all authors designated as the corresponding author. As a result, there are 524,052 observations in our regression analyses (see the detailed data set description in the [Supplementary material](#)).

We additionally obtained papers' metadata from the Microsoft Academic Graph and the Web of Science databases (Peng, Ke et al., 2021; Wang, Shen et al., 2019), including author name, author rank, corresponding author status, authorship position, affiliation rank and location, number of authors, publication year and venue, and paper's abstract and research topics (see the detailed data set description in the [Supplementary material](#)). We would like to note that there was considerable data reduction in our data collection pipeline mainly due to missing content for the news stories and papers, but the data filtering affected different ethnic groups equally based on a chi-square test (Table S8, [Supplementary material](#)).

To quantify disparities in author mentions, we developed a computational method to identify three types of mentions in terms of author attributions, including name mentions, quotes, and institution mentions (see details in the [Supplementary material](#)).

We algorithmically inferred the perceived ethnicity and gender from authors' names, which mirrors how a reader might perceive social identities based on regularities in where the name originates. This choice may introduce bias because algorithmic inference may not be perfectly aligned with human perceptions (Kozłowski, Murray et al., 2021), a limitation we return to in the discussion.

Overall, we do not measure the true ethnicity, but rely on the *perceived* identity inferred from names. Therefore our conclusions should be interpreted as reflecting disparities among scientists with name-inferred ethnicities rather than self-identified ethnicities directly. However, we obtained consistent results when coding author names with racial self-identities using the US census data or an alternate definition of ethnicity using the Wikipedia data (Figure S4 in the [Supplementary material](#)).

We used mixed-effects logistic regression models to control for a broad range of plausible confounding factors, including corresponding author status, authorship position, author rank and popularity, name complexity, affiliation rank and location, abstract readability, team size, research topics, and news features, such as year of coverage, article length, and journalist's

demographics (see details in the [Supplementary material](#)). The model also adds random effects for publication venues and news outlets, enabling us to measure differential mentions within a particular news outlet covering a particular academic journal on a particular research topic. This helps ensure that we are comparing media mentions of scholars doing comparable work. Nevertheless, our model cannot provide causal evidence of ethnic discrimination.

3. WHO GETS MENTIONED LESS OFTEN IN MEDIA COVERAGE?

We find substantial disparities in author mentions across name-inferred ethnicities—that is, given that an author’s research paper is being reported in the news, authors whose names are associated with certain ethnicities are much less likely to be mentioned by name. These disparities are robust to the inclusion of increasingly stringent controls of external factors that may influence the likelihood of being mentioned (Model 5 in [Table S5](#), [Supplementary material](#)). Specifically, compared to Anglo-named authors, most authors with minority ethnicity¹ names are significantly less likely to be mentioned, with European names disadvantaged the least and East Asian and African names disadvantaged the most.

In contrast to ethnicity, we find no disparity in author mentions across genders. However, when fixed effects for paper keywords are not considered, the author gender variable appears to have a significant effect (Model 3 in [Table S5](#), [Supplementary material](#)). As gender representation varies widely across academic disciplines (Handelsman, Cantor et al., 2005; Xie & Shauman, 2003), this result suggests that gender differences in mention rates are likely to be explained by different mention rates across different fields.

To quantify ethnic disparities in mentions, we calculated the average marginal effects for the author ethnicity and gender variables using the fullest model (Model 5 in [Table S5](#), [Supplementary material](#)). As shown in [Figure 1](#), the estimated probability of being mentioned is 1.2–6.0 percentage points lower for most ethnicities compared to the Anglo group. As the average mention rate is only 41.2% (see “[Detecting Author Name Mentions](#)” in the [Supplementary material](#)), these absolute drops represent significant disparities: The 4.3–6.0 percentage points marginal decrease for East Asian and African names represents a 10.4–14.6% relative decrease in media representation for authors with those names. This result reveals that US mainstream media outlets exhibit profound disparity against non-Anglo-named authors in mentioning them by name in science news: Given the current disparities, *we estimate that about six thousand minority-ethnicity scholars should have been mentioned in our data alone if they had Anglo names.*

4. LARGE DISPARITIES STILL EXIST FOR US-BASED AUTHORS WITH EAST ASIAN AND AFRICAN NAMES

In science reporting, journalists often directly seek out the authors by phone or email to contextualize and explain their results. For authors at non-US institutions, journalists from US-based outlets could be less likely to reach out due to time-zone differences and communication costs associated with distance (Lin et al., 2022), potentially resulting in a lower rate of being mentioned.

¹ We refer to “Anglo” as the majority group based on the number of observations in our data (where all papers received media coverage). [Table S2](#) in the [Supplementary material](#) shows that Anglo-named authors have their research covered more than twice as much as that of Western & Northern European named authors, the second largest group.

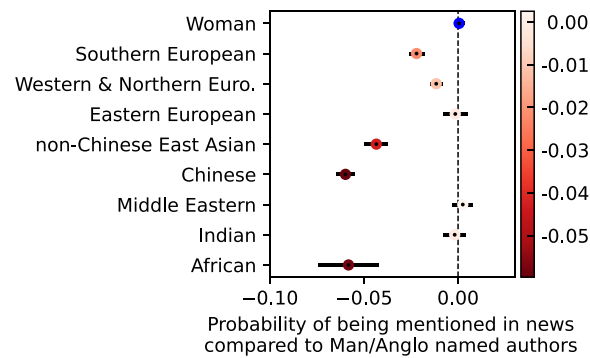


Figure 1. The average marginal effects of author's gender and ethnicity on the probability of being mentioned by name in science news reporting their research. Estimations are based on our full model (Model 5) fitted to all 524,052 observations in our data. Authors with names associated with minority ethnicities, especially East Asian and African names, are much less likely to be mentioned than Anglo-named authors after controlling for corresponding author status, affiliation rank and location, authorship position, author rank and popularity, last name complexity, abstract readability, team size, research topics, and news features such as year of coverage, article length, and journalist's demographics, as well as random effects for publication venues and news outlets. Colors are proportional to absolute probability changes (legend is shown vertically for space consideration). *Woman* is colored blue to reflect its difference from ethnicity identities. The error bars indicate 95% bootstrapped confidence intervals.

Indeed, our regression model shows that international scholars are significantly less likely to be mentioned compared with their US domestic counterparts of the same ethnicity (see the negative coefficient for affiliation location in Table S5 in the Supplementary material). However, the significant negative coefficients for ethnicity in the same regression (Table S5 in the Supplementary material) suggest that location does not explain all disparities in mentions, as disparities between minority ethnicities and Anglo ethnicity still exist conditioning on authors being located in the same geographical region (whether inside or outside the United States).

We next quantified the disparities across different locations by measuring the size of disparities separately for the subset of our data where the authors are all from US-based institutions, and that for all non-US authors.

The coefficients in Table 1 indicate that, *among international authors, those with non-Anglo names are significantly less likely to be mentioned than those with Anglo names, despite that they are all distant to US-based journalists*. Compared to international researchers, the mention disparities are much smaller for US-based authors, suggesting that being affiliated with a US institution does decrease the disparity for each minority ethnicity, and for some groups, including Indian, Middle Eastern, and Eastern European, the mention rate is even higher than for Anglo-named authors. Nevertheless, close proximity between journalists and authors does not eliminate all disparities in who is mentioned, *as the disparities are still large and significant for African and East Asian ethnicities among all authors affiliated with US institutions* (Table 1).

5. DISPARITIES ACROSS THREE TYPES OF MENTIONS AMONG US-BASED AUTHORS

Our prior result shows that ethnic disparities in mentions are observed even among authors based in the United States, where scheduling difficulties and other pragmatic factors should be minimized or, at least, not associated with ethnicity. Focusing on US-based authors, we further separated mentions into different types to better understand the mechanisms driving these disparities. We first quantified the average marginal effects in mention rates based on a Model 5 fitted to the US subset (Table 1). Figure 2(a) shows that US-based Chinese, non-Chinese East

Table 1. The regression coefficients of author’s gender and ethnicity in predicting the author’s name being mentioned in a news story reporting their research. Disparities between minority and Anglo-named scholars are significant when they are all affiliated with international institutions, with each ethnicity reaching statistical significance. The disparities are reduced when scholars are all affiliated with US institutions. However, even among US-based authors, there are significant disparities for authors with East Asian and African names, suggesting that location does not explain all disparities. A separate Model 5 is trained for the subset of US-based institutions and the subset of non-US institutions. When fitting a model for the US subset (or non-US subset), we omitted the *affiliation location* variable. Significance level: *** $p < 0.001$, ** $p < 0.01$, and * $p < 0.05$. The p -values are based on the statistical test of differences in coefficients between two regression models using the equation provided in Clogg, Petkova, and Haritou (1995).

Gender/ethnicity	US-based	non-US	p -value
Woman	-0.01	0.01	0.254
Southern European	-0.02	-0.33***	0.000
Western & Northern European	-0.02	-0.19***	0.000
Eastern European	0.12***	-0.24***	0.000
non-Chinese East Asian	-0.24***	-0.36***	0.003
Chinese	-0.31***	-0.57***	0.000
Middle Eastern	0.07**	-0.11***	0.000
Indian	0.04*	-0.16***	0.000
African	-0.29***	-0.52***	0.034

Asian, and African-named authors experience 4.8, 3.8, and 4.6 percentage points drops in mention rates, respectively, compared to their Anglo-named counterparts.

One plausible mechanism generating mention disparities for US-based minority authors is journalists’ perceptions of, or actual differences in, authors’ fluency in speaking English. Although authors’ actual fluency is not available to journalists, they may make assumptions

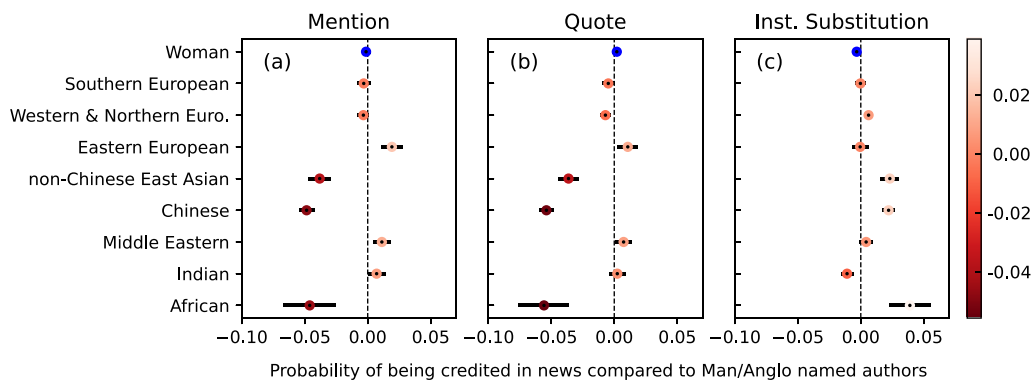


Figure 2. US-based authors with minority ethnicity names are less likely to be mentioned by name (a) or quoted (b), and are more likely to be substituted by their institution (c). The average marginal effects are estimated based on our finest model (Model 5) fitted to 317,626 observations where the author is from US-based institutions. A negative (positive) marginal effect indicates a decrease (increase) in the probability of being credited compared to authors with Man (for gender) or Anglo (for ethnicity) names. The colors are proportional to the absolute probability changes. *Woman* is colored blue to reflect its difference from ethnicity identities. The error bars indicate 95% bootstrapped confidence intervals.

about an author's speaking fluency based on ethnicity or other factors. For example, among US-based authors, journalists may assume that authors with minority ethnicity names are more likely to be foreign-born with less English fluency (Chiswick & Miller, 1998; Shields & Price, 2002). If so, journalists may be less willing to contact the author to ask them to explain the findings in cases where they need additional information to understand the paper, which could result in fewer quotations of these authors.

To measure disparities in quotation rates, we identified authors who are named as part of quotations (a subset of name mentions; see "Author-Quote Detection" in the Supplementary material) and applied the same regression model to US-based authors in predicting whether the author is quoted in the news story reporting their research. Figure 2(b) shows that there are substantial disparities in quotation rates for authors with East Asian-associated and African-associated names. *We note that this result suggests, but does not prove, that the perceived fluency could be a driving mechanism, as other mechanisms, such as the rhetorical value of names, may also produce this result.*

To more directly test the rhetorical mechanism, we examined "institution-substitution" where the author is mentioned by their institution but not by name (see "Detecting Institution Mentions" in the Supplementary material), for example, being named as "researchers at the University of Michigan." Among US-based authors, this mention type should not depend on pragmatic factors such as scheduling difficulties or perceived English fluency. Thus, this substitution effect likely reveals the rhetorical value that journalists place on authors' names vs. institutions.

Figure 2(c) shows the probability of institution-substitution of minority ethnicity authors relative to those with Anglo names, revealing that US-based authors with African and East Asian names are more likely to have their names substituted by their institutions (Figure S3 in the Supplementary material shows similar results of this analysis using the full data). Analyzing three different types of author mentions thus reveal that although some mention disparities may be explained by perceived English fluency or other pragmatic factors, journalists' rhetorical choices are also key.

6. CONSISTENT DISPARITIES ACROSS THREE TYPES OF NEWS OUTLETS

News outlets vary in the depth and breath of their reporting; for example, Science & Technology outlets write about 650 words per story on average, whereas General News outlets write about 900 words (Figure S2 in the Supplementary material). These differences suggest potentially important variability in the nature of journalists' day-to-day work and backgrounds. To explore the discrepancy of disparities in author mentions across different types of outlets, we fitted the specification of Model 5 separately for three outlet types in our full data and quantified the average marginal effects.

Figure 3 shows that ethnic disparities in mention rates surprisingly remain consistent across all outlet types, with authors of non-Anglo names less likely to be mentioned. Larger disparities are found for ethnic categories that are more culturally distant from Anglo (e.g., East Asian and African). Although the three outlet types have similar sizes of absolute disparities, they vary substantially in the relative scale, as the average mention rates of Science & Technology outlets and General News outlets are 34.0–61.9% less than Press Releases outlets (Table S4 in the Supplementary material).

The disparity in Press Releases outlets is particularly notable, as stories in these outlets typically reuse content from university press releases, suggesting that universities' press offices

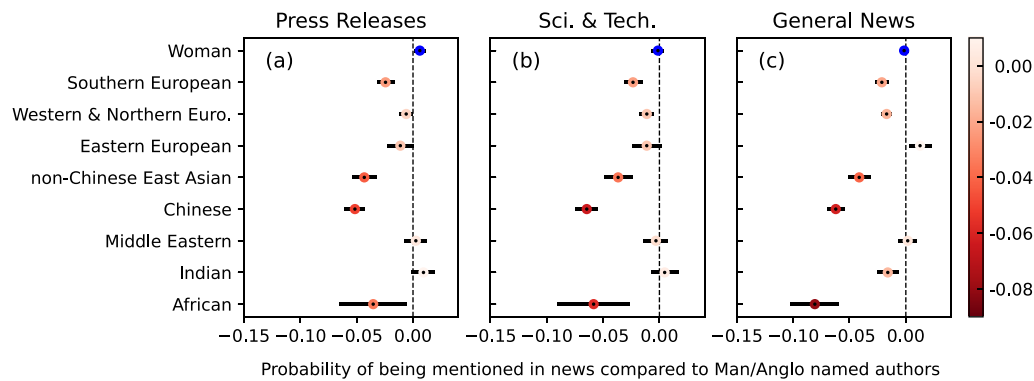


Figure 3. Ethnic disparities in mention rates are consistent across three types of news outlets. The similar sizes of absolute disparities in three outlet types reflect the starkly different magnitude of relative effects, as the average mention rates in Press Releases, Science & Technology, and General News outlets are 63.5%, 41.9%, and 24.2%, respectively. In particular, the 6–8 percentage points drop in mention rates for Chinese and African named scholars reduces nearly one-third of their expected media representation in General News outlets. We fitted a separate model (Model 5) to the subset of all observations in each type of outlet. The colors are proportional to the absolute probability changes. Error bars represent 95% bootstrapped confidence intervals.

themselves, although less biased than other outlet types, still prefer to mention scholars with Anglo names. This result is unexpected because local press offices are expected to have greater direct familiarity with their researchers, reduce the misuse of stereotypes, and be more responsible for representing minority researchers equitably.

The largest disparities are seen in General News outlets, such as the *New York Times* and the *Washington Post*, where again scholars with Chinese- and African-associated names have a 6.0–8.0 percentage points drop in mention rates. This significant drop reduces nearly one third of the deserved media representation of a large community of scientists, as General News outlets mention authors with a 24.2% chance on average (Table S4 in the Supplementary material). As General News outlets have well-trained editorial staff and science journalists dedicated to accurately reporting science and tend to publish longer stories that have room to mention and engage with authors, this result is alarming. Historically, these ethnic minorities have been stereotyped and underrepresented in US media and leadership roles (Behm-Morawitz & Ortiz, 2013; Lu, Nisbett, & Morris, 2020), which has continued in objective science reporting across all outlet types. The mechanisms of this variation deserve further investigation.

7. IS THE SITUATION GETTING MORE EQUITABLE?

The longitudinally rich nature of our data set allows us to examine how author mentions in science news have changed over the last decade. Mention rates are on average decreasing over time, as shown by the coefficient of the *mention year* variable in Model 5 (Table S5 in the Supplementary material). To examine the time trends across demographic categories, a separate Model 5 was trained to quantify the marginal change in mention rate per year increase for each gender and ethnicity in our full data. Note that demographic attributes not under study were still included in each model; for example, when examining the temporal changes in mention rates for men and women, ethnicity was still included in the model, and vice versa.

As shown in Figure 4, the mention year has a negative association with author mention rates for all gender and ethnic groups, and the larger decrease for Anglo indicates that their

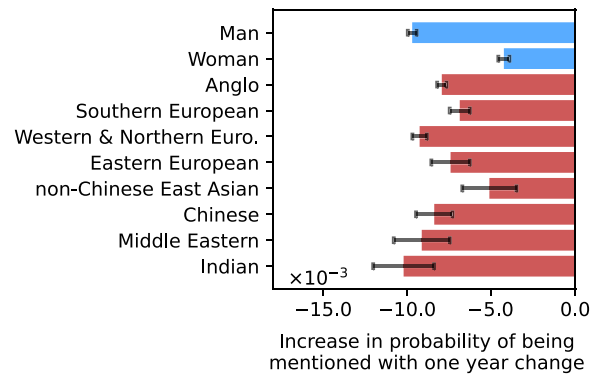


Figure 4. Average marginal effects on the mention probability for a one-unit increase in the mention year variable for each gender (blue) and ethnicity (red). A separate model was fitted using all observations for each demographic group. The African ethnicity is not shown due to insufficient data for fitting the model. Error bars show 95% bootstrapped confidence intervals.

overall advantages are shrinking. Indeed, authors with non-Chinese East Asian names, one of the most disadvantaged groups in this study, have the lowest decreasing rate.

However, the estimated rates of change are relatively small for most ethnic groups, suggesting that the existing disparities are unlikely to disappear in the short term without intentional behavior change. We refrain from making predictions as to when mention equality will be reached eventually, or adopting sophisticated time series models to forecast the trajectory of mention rates in the long run, because such extrapolation will be of little practical use, especially given that the long-term changes in academia and media practices remain unforeseeable.

8. DISCUSSION

Our analyses reveal that the attention researchers get in science news reporting is strongly related to the ethnicities associated with their names. The effects are robust to a variety of plausible confounders, and even appear when controlling for the particular news outlet, particular scientific venue, and particular research topic. Although we cannot claim that the reported effects are causal, this unusually strong observational evidence deserves further attention.

8.1. Ethnicity and Gender

Authors with most non-Anglo names are mentioned substantially less when their research is covered in US science news. Mention rates are especially low for East Asian and African names, less pronounced for European names, and are even less pronounced for Indian and Middle Eastern names. The large disparity for East Asian ethnicities and its sharp contrast with South Asian categories (such as Indian) is consistent with the “bamboo ceiling” phenomenon observed in leadership roles in the United States (Lu et al., 2020). As science becomes more global and increasingly driven by authors of non-Western ethnicities, the way English-language media respond to non-Anglo-named scholars will only grow in importance. In contrast to ethnicity, we do not find gender disparities in mentions of scholars once the research fields are controlled for. One possible reason is that fields vary in their overall level of mention rates and in their gender representation (Handelsman et al., 2005). Looking within fields masks gender disparity that may exist between them.

8.2. Ruling In and Out Different Mechanisms

Our analyses point to a multicausal generation of ethnic disparities, in which both pragmatic difficulties of interviewing researchers (location and possibly perceived fluency) and journalists' tastes regarding names' rhetorical values play key roles.

In support of pragmatic difficulties, we find that international location (which hosts more scholars with non-Anglo names) has a negative effect on mention rates. However, location is not the driving mechanism, as disparities persist among both international authors and US-based authors, which would disappear if location was the decisive factor. In support of English fluency, we find that ethnic disparities for East Asian and African ethnicities appear in quotations among US-based authors, who are unlikely to suffer from pragmatic difficulties in scheduling interviews, but may differ in their actual or perceived fluency. However, we note that these disparities may be produced by other factors such as assertiveness (Lu, Nisbett, & Morris, 2022).

In addition to these pragmatic factors, journalists' rhetorical choices are key. In support of this mechanism, journalists are more likely to "substitute" a direct name mention with the researcher's institution for authors with East Asian and African names, suggesting that the context of discovery is important, but the institution serves the journalists' rhetorical goals better than the name. Additional evidence comes from outlet types: When journalists' role in the news articles is minimal—when the outlet simply republishes a university press release—the relative disparities are also minimal; when the news stories are written by journalists themselves, the relative disparities are the largest. However, we note that the disparities in Press Releases outlets also suggest that journalists are not the only actor behind the inequality.

The data do not allow us to fully explain journalists' rhetorical choices. For example, we hypothesize that their choices may be driven by journalists' own perceptions of author's authoritative-ness or by the expected tastes of their audiences. However, we observe that ethnic disparities in mentions do not vary substantially across Science & Technology and General News outlets, although the two likely differ in their audiences. This observation suggests again that journalists' personal preferences play an important role. Furthermore, in Section IV.B of the Supplementary material, we examine whether the interaction between authors' and journalists' name-inferred identities are associated with the mention rate, but do not find clear evidence. Disentangling the source of journalists' choices is an important avenue for future work.

8.3. Limitations

Although the scale and the breath of our data set enable the use of unusually fine-grained controls, the analysis is not without limitations. First, the observational nature of the data precludes strong causal statements. Second, the analysis was conducted with perceived ethnicities, which do not reflect self-identities accurately, nor account for multiethnicity identities. We hope our work stimulates the collection of such data where possible, to enable more accurate and fine-grained conclusions (Wu, 2020). Thus, a key limitation of our design and the voluminous audit study literature must be acknowledged: Such types of studies do not measure whether journalists actually form an inference of ethnicity when seeing names. We believe assuming that they do form such inferences is very reasonable and supported by the large empirical disparities we observe here. More direct evidence on journalists' decision processes is a fruitful direction for future research. Besides, we inferred the perceived ethnicity via a name-based classifier, *Ethnea*. Although journalists, like the classifier, may have no information about authors except their names, the inference will undoubtedly not match all actual

human perceptions about the authors. Furthermore, the classifier is unable to identify key demographic groups, such as African American scholars. Nevertheless, as an exploratory test, we repeated our analysis using a classification of race based on the US census data (Figure S4 in the Supplementary material), which includes “Black” as one of the labels. The result does not show statistically significant underrepresentation of Black scholars relative to “White.” Note that African-named authors (based on *Ethnea*) are not necessarily classified as “Black” based on the census data (Tables S6 and S7 in the Supplementary material).

Third, some plausible covariates are unavailable for inclusion, such as the number of citations a paper received at the time of being mentioned. However, we anticipate the effect of such covariates to be small given current controls. Furthermore, the majority of papers in our data were mentioned within 1 year after publication (Figure S1 in the Supplementary material), which limits the number of citations a paper can accrue in such a short time window. Relatedly, we lacked a direct measure of authors’ English fluency, and our measure of prestige-related factors such as author rank and affiliation rank may not be able to accurately reflect prestige.

Fourth, we did not test other potential mechanisms. For instance, reporters often choose to interview authors listed as the corresponding author and prestigious authors from top institutions. Although our model controls for the corresponding author status, author rank, and affiliation rank, we did not examine how much of the disparities are driven by these prestige-related factors. It is possible that which author of a paper is designated as corresponding and their selection into top institutions are themselves a product of structural discrimination with respect to authors’ demographics. Thus disparities observed in press mentions may be partly driven by decades or centuries of decisions that are ingrained in institution hiring practices (Small & Pager, 2020).

Fifth, our data contain too few examples of some ethnicities (e.g., Polynesian and Caribbean) to accurately estimate disparities; such ethnicities are regrettably omitted, though we recognize that these groups likely experience disparity from their minority status as well.

Sixth, our study has focused solely on US-based news outlets. Many of these outlets are often global in reach and mentions in them often serve as markers of prestige for scholars. However, the behaviors of these outlets may not be representative of broader media reporting practices. At present, the Almetrics data only provide sufficient quantity for US-based outlets that allow us to control for potential confounders related to each outlet (62% of all news mentions are solely from US-based outlets), which is critical for our study design. Nevertheless, bias is likely not unique to one country and additional global-scale data is necessary to move beyond a US focus and study country-specific and global journalistic practices.

Last, this research relies on large-scale data sets and algorithms that may themselves encode systemic social inequalities. For instance, which venues are considered “mainstream” and therefore worthy of tracking by Almetric may be the outcome of racial inequities (Alamo-Pastrana & Hoynes, 2020). Which groups the algorithms choose to identify as distinct groups are choices that may reflect long histories of racialization seen through a “white racial frame” (Feagin, 2020; Tatum, 2017). The availability of data also drove our focus on English-language science and media, thereby accumulating more activity around certain cultures than others. We believe these limitations place substantial scope conditions on the findings.

8.4. Conclusions and Implications

Our work shows that science journalism is rife with disparities in which author receives name attribution, with authors from certain ethnic groups receiving many more name mentions and

quotations than their peers when their comparable research papers are all reported in US news. These ethnic disparities likely have direct negative consequences for the careers of unmentioned scientists, and skew the public perception of who a scientist is—a key factor in recruiting and training new scientists.

Our findings have two implications for science policy and science journalism. First, bringing attention to large-scale ethnic disparities in author mentions in science news, of which journalists may themselves have been unaware, can be an agent of change. Second, decision-makers at US research institutions may take these ethnic disparities into account when making hiring or promotion decisions. More importantly, addressing this problem requires more research to investigate the mechanisms leading to it, which we hope this paper helps stimulate.

Although our study only focuses on the “second stage” of science media coverage—its quality—it is likely that such ethnic disparities would be even larger in the first stage of coverage where media outlets choose whose papers to report on in the first place. Supporting evidence comes from recent empirical studies that find gender and regional disparities in the online attention to scientists’ work (Chapman et al., 2022; Davidson & Greene, 2024; Vasarhelyi et al., 2021). Our work thus suggests that disparities in science media are likely to compound across different aspects of coverage, yielding ultimate differences in outcomes much larger than those shown by studies of any one stage.

ACKNOWLEDGMENTS

We thank [Altmetric.com](https://www.altmetric.com) for sharing the mention data used in this study. This work uses Web of Science data by Clarivate Analytics provided by the Indiana University Network Science Institute and the Cyberinfrastructure for Network Science Center at Indiana University. We thank Aparna Ananthasubramaniam, Jiaxin Pei, Daniel M. Romero for helpful discussions and suggestions.

AUTHOR CONTRIBUTIONS

Hao Peng: Conceptualization, Formal analysis, Methodology, Writing—original draft, Writing—review & editing. Misha Teplitskiy: Conceptualization, Methodology, Writing—original draft, Writing—review & editing. David Jurgens: Conceptualization, Methodology, Writing—original draft, Writing—review & editing.

COMPETING INTERESTS

The authors have no competing interests.

FUNDING INFORMATION

No funding was received for this work.

DATA AVAILABILITY

The Altmetric data is available free of charge to university-affiliated scientometric researchers; see details at <https://www.altmetric.com/research-access/>. The Microsoft Academic Graph data can be publicly accessed at <https://openalex.org/>. All code used in this study is available at https://github.com/haopeng/author_mentions.

REFERENCES

- Alamo-Pastrana, C., & Hoynes, W. (2020). Racialization of news: Constructing and challenging professional journalism as “white media.” *Humanity & Society*, 44(1), 67–91. <https://doi.org/10.1177/0160597618820071>
- Banchevsky, S., Westfall, J., Park, B., & Judd, C. M. (2016). But you don’t look like a scientist! Women scientists with feminine appearance are deemed less likely to be scientists. *Sex Roles*, 75, 95–109. <https://doi.org/10.1007/s11199-016-0586-1>
- Behm-Morawitz, E., & Ortiz, M. (2013). Race, ethnicity, and the media. In *Oxford handbook of media psychology* (pp. 252–266). Oxford: Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780195398809.013.0014>
- Bertrand, M., & Mullainathan, S. (2004). Are Emily and Greg more employable than Lakisha and Jamal? A field experiment on labor market discrimination. *American Economic Review*, 94(4), 991–1013. <https://doi.org/10.1257/0002828042002561>
- Blum, D., Knudson, M., & Henig, R. M. (Eds.). (2006). *A field guide for science writers*. New York: Oxford University Press. <https://doi.org/10.1093/oso/9780195174991.001.0001>
- Brossard, D., & Scheufele, D. A. (2013). Science, new media, and the public. *Science*, 339(6115), 40–41. <https://doi.org/10.1126/science.1232329>, PubMed: 23288529
- Callison, C., & Young, M. L. (2019). *Reckoning: Journalism’s limits and possibilities*. Oxford: Oxford University Press. <https://doi.org/10.1093/oso/9780190067076.001.0001>
- Chapman, C. A., Hemingway, C. A., Sarkar, D., Gogarten, J. F., & Stenseth, N. C. (2022). Altmetric scores in conservation science have gender and regional biases. *Conservation & Society*, 20(3), 195–200. https://doi.org/10.4103/cs.cs_27_21
- Chiswick, B. R., & Miller, P. W. (1998). English language fluency among immigrants in the United States. *Research in Labor Economics*, 17.
- Clark, M. (2018). *ASNE’s 2018 diversity survey results reflect low participation but encouraging shifts*. <https://www.asne.org/diversity-survey-2018>
- Clogg, C. C., Petkova, E., & Haritou, A. (1995). Statistical methods for comparing regression coefficients between models. *American Journal of Sociology*, 100(5), 1261–1293. <https://doi.org/10.1086/230638>
- Crabtree, C., Gaddis, S. M., Holbein, J. B., & Larsen, E. N. (2022). Racially distinctive names signal both race/ethnicity and social class. *Sociological Science*, 9, 454–472. <https://doi.org/10.15195/v9.a18>
- Crabtree, C., Kim, J. Y., Gaddis, S. M., Holbein, J. B., Guage, C., & Marx, W. W. (2023). Validated names for experimental studies on race and ethnicity. *Scientific Data*, 10(1), 130. <https://doi.org/10.1038/s41597-023-01947-0>, PubMed: 36899034
- Davidson, N. R., & Greene, C. S. (2024). Analysis of science journalism reveals gender and regional disparities in coverage. *eLife*, 12, RP84855. <https://doi.org/10.7554/eLife.84855>, PubMed: 38804191
- Devitt, J. (2002). Framing gender on the campaign trail: Female gubernatorial candidates and the press. *Journalism & Mass Communication Quarterly*, 79(2), 445–463. <https://doi.org/10.1177/107769900207900212>
- Fanelli, D. (2013). Any publicity is better than none: Newspaper coverage increases citations, in the UK more than in Italy. *Scientometrics*, 95(3), 1167–1177. <https://doi.org/10.1007/s11192-012-0925-0>
- Feagin, J. R. (2020). *The white racial frame: Centuries of racial framing and counter-framing*. London: Routledge. <https://doi.org/10.4324/9780429353246>
- Gaddis, S. M. (2015). Discrimination in the credential society: An audit study of race and college selectivity in the labor market. *Social Forces*, 93(4), 1451–1479. <https://doi.org/10.1093/sf/sou111>
- Gaddis, S. M. (2017a). How black are Lakisha and Jamal? Racial perceptions from names used in correspondence audit studies. *Sociological Science*, 4, 469–489. <https://doi.org/10.15195/v4.a19>
- Gaddis, S. M. (2017b). Racial/ethnic perceptions from Hispanic names: Selecting names to test for discrimination. *Socius*, 3. <https://doi.org/10.1177/2378023117737193>
- Gaddis, S. M., & Ghoshal, R. (2020). Searching for a roommate: A correspondence audit examining racial/ethnic and immigrant discrimination among millennials. *Socius*, 6. <https://doi.org/10.1177/2378023120972287>, PubMed: 34355061
- Ginther, D. K., Schaffer, W. T., Schnell, J., Masimore, B., Liu, F., ... Kington, R. (2011). Race, ethnicity, and NIH research awards. *Science*, 333(6045), 1015–1019. <https://doi.org/10.1126/science.1196783>, PubMed: 21852498
- Grieco, E. (2018). Newsroom employees are less diverse than US workers overall. *Pew Research Center*. <https://www.pewresearch.org/short-reads/2018/11/02/newsroom-employees-are-less-diverse-than-u-s-workers-overall/>
- Handelsman, J., Cantor, N., Carnes, M., Denton, D., Fine, E., ... Sheridan, J. (2005). More women in science. *Science*, 309(5738), 1190–1191. <https://doi.org/10.1126/science.1113252>, PubMed: 16109868
- Hill, P. W., McQuillan, J., Spiegel, A. N., & Diamond, J. (2018). Discovery orientation, cognitive schemas, and disparities in science identity in early adolescence. *Sociological Perspectives*, 61(1), 99–125. <https://doi.org/10.1177/0731121417724774>, PubMed: 29576677
- Hofstra, B., Kulkarni, V. V., Munoz-Najar Galvez, S., He, B., Jurafsky, D., & McFarland, D. A. (2020). The diversity-innovation paradox in science. *Proceedings of the National Academy of Sciences*, 117(17), 9284–9291. <https://doi.org/10.1073/pnas.1915378117>, PubMed: 32291335
- Hoppe, T. A., Litovitz, A., Willis, K. A., Meseroll, R. A., Perkins, M. J., ... Santangelo, G. M. (2019). Topic choice contributes to the lower rate of NIH awards to African-American/black scientists. *Science Advances*, 5(10), eaaw7238. <https://doi.org/10.1126/sciadv.aaw7238>, PubMed: 31633016
- Huang, J., Gates, A. J., Sinatra, R., & Barabási, A. L. (2020). Historical comparison of gender inequality in scientific careers across countries and disciplines. *Proceedings of the National Academy of Sciences*, 117(9), 4609–4616. <https://doi.org/10.1073/pnas.1914221117>, PubMed: 32071248
- Jia, S., Lansdall-Welfare, T., & Cristianini, N. (2015). Measuring gender bias in news images. In *Proceedings of the 24th International Conference on World Wide Web* (pp. 893–898). <https://doi.org/10.1145/2740908.2742007>
- Jia, S., Lansdall-Welfare, T., Sudhakar, S., Carter, C., & Cristianini, N. (2016). Women are seen more than heard in online newspapers. *PLOS ONE*, 11(2), e0148434. <https://doi.org/10.1371/journal.pone.0148434>, PubMed: 26840432
- Knobloch-Westerwick, S., Glynn, C. J., & Huges, M. (2013). The Matilda effect in science communication: An experiment on gender bias in publication quality perceptions and collaboration interest. *Science Communication*, 35(5), 603–625. <https://doi.org/10.1177/1075547012472684>
- Kozlowski, D., Larivière, V., Sugimoto, C. R., & Monroe-White, T. (2022). Intersectional inequalities in science. *Proceedings of the*

- National Academy of Sciences, 119(2), e2113067119. <https://doi.org/10.1073/pnas.2113067119>, PubMed: 34983876
- Kozlowski, D., Murray, D. S., Bell, A., Hulse, W., Larivière, V., ... Sugimoto, C. R. (2021). Avoiding bias when inferring race using name-based approaches. *arXiv*. <https://doi.org/10.48550/arXiv.2104.12553>
- Lang, K., & Lehmann, J.-Y. K. (2012). Racial discrimination in the labor market: Theory and empirics. *Journal of Economic Literature*, 50(4), 959–1006. <https://doi.org/10.1257/jel.50.4.959>
- Larivière, V., Ni, C., Gingras, Y., Cronin, B., & Sugimoto, C. R. (2013). Bibliometrics: Global gender disparities in science. *Nature*, 504(7479), 211–213. <https://doi.org/10.1038/504211a>, PubMed: 24350369
- Lin, Y., Frey, C. B., & Wu, L. (2022). Remote collaboration fuses fewer breakthrough ideas. *arXiv*, arXiv:2206.01878. <https://doi.org/10.48550/arXiv.2206.01878>
- Lu, J. G., Nisbett, R. E., & Morris, M. W. (2020). Why East Asians but not South Asians are underrepresented in leadership positions in the United States. *Proceedings of the National Academy of Sciences*, 117(9), 4590–4600. <https://doi.org/10.1073/pnas.1918896117>, PubMed: 32071227
- Lu, J. G., Nisbett, R. E., & Morris, M. W. (2022). The surprising underperformance of East Asians in US law and business schools: The liability of low assertiveness and the ameliorative potential of online classrooms. *Proceedings of the National Academy of Sciences*, 119(13), e2118244119. <https://doi.org/10.1073/pnas.2118244119>, PubMed: 35312365
- McPherson, M., Smith-Lovin, L., & Cook, J. M. (2001). Birds of a feather: Homophily in social networks. *Annual Review of Sociology*, 27, 415–444. <https://doi.org/10.1146/annurev.soc.27.1.415>
- Merullo, J., Yeh, L., Handler, A., Grissom, A., O'Connor, B., & Iyer, M. (2019). Investigating sports commentator bias within a large corpus of American football broadcasts. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)* (pp. 6355–6361). <https://doi.org/10.18653/v1/D19-1666>
- Miller, D. I., Nolla, K. M., Eagly, A. H., & Uttal, D. H. (2018). The development of children's gender-science stereotypes: A meta-analysis of 5 decades of US Draw-A-Scientist studies. *Child Development*, 89(6), 1943–1955. <https://doi.org/10.1111/cdev.13039>, PubMed: 29557555
- Neumark, D. (2018). Experimental research on labor market discrimination. *Journal of Economic Literature*, 56(3), 799–866. <https://doi.org/10.1257/jel.20161309>
- Nielsen, R. K., Selva, M., & Andri, S. (2020). *Race and leadership in the news media 2020: Evidence from five markets*. Oxford: Reuters Institute for the Study of Journalism.
- Oliveira, D. F., Ma, Y., Woodruff, T. K., & Uzzi, B. (2019). Comparison of National Institutes of Health grant amounts to first-time male and female principal investigators. *JAMA*, 321(9), 898–900. <https://doi.org/10.1001/jama.2018.21944>, PubMed: 30835300
- Peng, H., Ke, Q., Budak, C., Romero, D. M., & Ahn, Y.-Y. (2021). Neural embeddings of scholarly periodicals reveal complex disciplinary organizations. *Science Advances*, 7(17), eabb9004. <https://doi.org/10.1126/sciadv.abb9004>, PubMed: 33893092
- Peng, H., Lakhani, K., & Teplitskiy, M. (2021). Acceptance in top journals shows large disparities across name-inferred ethnicities. *SocArXiv*. <https://doi.org/10.31235/osf.io/mjbxg>
- Peng, H., Teplitskiy, M., Romero, D. M., & Horvát, E.-Á. (2022). The gender gap in scholarly self-promotion on social media. *arXiv*. <https://doi.org/10.48550/arXiv.2206.05330>
- Reuben, E., Sapienza, P., & Zingales, L. (2014). How stereotypes impair women's careers in science. *Proceedings of the National Academy of Sciences*, 111(12), 4403–4408. <https://doi.org/10.1073/pnas.1314788111>, PubMed: 24616490
- Robinson, S., & Culver, K. B. (2019). When White reporters cover race: News media, objectivity and community (dis)trust. *Journalism*, 20(3), 375–391. <https://doi.org/10.1177/1464884916663599>
- Scheufele, D. A. (2013). Communicating science in social settings. *Proceedings of the National Academy of Sciences*, 110(Suppl. 3), 14040–14047. <https://doi.org/10.1073/pnas.1213275110>, PubMed: 23940341
- Shields, M. A., & Price, S. W. (2002). The English language fluency and occupational success of ethnic minority immigrant men living in English metropolitan areas. *Journal of Population Economics*, 15, 137–160. <https://doi.org/10.1007/PL00003836>
- Small, M. L., & Pager, D. (2020). Sociological perspectives on racial discrimination. *Journal of Economic Perspectives*, 34(2), 49–67. <https://doi.org/10.1257/jep.34.2.49>
- Smith, K. B. (1997). When all's fair: Signs of parity in media coverage of female candidates. *Political Communication*, 14(1), 71–82. <https://doi.org/10.1080/105846097199542>
- Sood, G., & Laohaprapanon, S. (2018). Predicting race and ethnicity from the sequence of characters in a name. *arXiv*. <https://doi.org/10.48550/arXiv.1805.02109>
- Sui, M., Paul, N., Shah, P., Spurlock, B., Chastant, B., & Dunaway, J. (2018). The role of minority journalists, candidates, and audiences in shaping race-related campaign news coverage. *Journalism & Mass Communication Quarterly*, 95(4), 1079–1102. <https://doi.org/10.1177/1077699018762078>
- Sundar, S. S. (1998). Effect of source attribution on perception of online news stories. *Journalism & Mass Communication Quarterly*, 75(1), 55–68. <https://doi.org/10.1177/107769909807500108>
- Sweeney, L. (2013). Discrimination in online ad delivery. *Communications of the ACM*, 56(5), 44–54. <https://doi.org/10.1145/2447976.2447990>
- Tatum, B. D. (2017). *Why are all the black kids sitting together in the cafeteria? And other conversations about race*. Hachette UK.
- Turner, C. S. V., González, J. C., & Wood, J. L. (2008). Faculty of color in academe: What 20 years of literature tells us. *Journal of Diversity in Higher Education*, 1(3), 139–168. <https://doi.org/10.1037/a0012837>
- Vásárhelyi, O., Zakhlebin, I., Milojević, S., & Ágnes Horvát, E. (2021). Gender inequities in the online dissemination of scholars' work. *Proceedings of the National Academy of Sciences*, 118(39), e2102945118. <https://doi.org/10.1073/pnas.2102945118>, PubMed: 34544861
- Wang, K., Shen, Z., Huang, C.-Y., Wu, C.-H., Eide, D., ... Rogahn, R. (2019). A review of Microsoft Academic Services for science of science studies. *Frontiers in Big Data*, 2, 45. <https://doi.org/10.3389/fdata.2019.00045>, PubMed: 33693368
- Way, S. F., Larremore, D. B., & Clauset, A. (2016). Gender, productivity, and prestige in computer science faculty hiring networks. In *Proceedings of the 25th International Conference on World Wide Web* (pp. 1169–1179). <https://doi.org/10.1145/2872427.2883073>
- Wu, K. J. (2020). Scientific journals commit to diversity but lack the data. *New York Times*. <https://www.nytimes.com/2020/10/30/science/diversity-science-journals.html>
- Xie, Y., & Shauman, K. A. (2003). *Women in science: Career processes and outcomes*. Cambridge, MA: Harvard University Press.