Scientometrics
https://doi.org/10.1007/s11192-021-03911-4

Digital technology helps remove gender bias in academia

Julie Fortin[1,5] · Bjarne Bartlett[2] · Michael Kantar[3] · Michelle Tseng[4] · Zia Mehrabi[1,5]

Received: 1 August 2020 / Accepted: 8 February 2021

Abstract

Science attempts to be a meritocracy; however, in recent years, there has been increasing evidence for systematic gender bias against women. This bias is present in many metrics commonly used to evaluate scientific productivity, which in turn influences hiring and career success. Here we explore a new metric, the Altmetric Attention Score, and find no evidence of bias across many major journals (Nature, PNAS, PLOS One, New England Journal of Medicine, Cell, and BioRxiv), with equal attention afforded to articles authored by men and women alike. The exception to this rule is the journal Science, which has marked gender bias against women in 2018, equivalent to a mean of 88 more tweets or 11 more news articles and a median of 20 more tweets or 3 more news articles for male than female first authors. Our findings qualify Altmetric, for many types and disciplines of journals, as a potentially unbiased measure of science communication in academia and suggest that new technologies, such as those on which Altmetric is based, might help to democratize academic evaluation.

Keywords Altmetric · Gender bias · Academia · Science communication

Julie Fortin and Bjarne Bartlett have contributed equally to this manuscript.

✉ Michael Kantar
mbkantar@hawaii.edu

✉ Zia Mehrabi
ziamehrabi@gmail.com

1 Institute for Resources, Environment, and Sustainability, University of British Columbia, Vancouver, BC V6T 2K9, Canada

2 Department of Molecular Biosciences and Bioengineering, University of Hawai'i, 3190 Maile Way, MānoaHonolulu, HI 96822, USA

3 Department of Tropical Plant and Soil Sciences, University of Hawai'i, 3190 Maile Way, MānoaHonolulu, HI 96822, USA

4 Departments of Botany and Zoology, Biodiversity Research Centre, University of British Columbia, Vancouver, BC, Canada

5 School of Public Policy and Global Affairs, University of British Columbia, Vancouver, BC V6T 1Z2, Canada

Published online: 09 March 2021

Introduction

A large body of recent literature has uncovered unconscious or conscious biases in several metrics used to evaluate the proficiency of natural/social science and humanities scholars. For example, impact factors, h-indices, granting outcomes, and reference letters are repeatedly shown to present biases against women (see our review in Table 1). Because these metrics are all used in decision making within academic institutions, these biases present a severe impediment for equal opportunity among genders throughout their scientific careers.

The emergence of big data and new technologies has the potential to democratize career evaluation and social mobility. For example, machine learning algorithms have been used to reduce biased decisions in hiring, evaluation, and promotion (Raghavan et al. 2020). Additionally, social media has reduced employment barriers, as witnessed by the rise of professional influencers in advertising, and has also increased accountability of corporations in countries with otherwise highly censored traditional media (De Veirman et al. 2019; Enikolopov et al. 2018). While the benefits of these technological advances are significant, there is also concern that new digital technologies may also amplify existing biases, as seen in selective content exposure in social media platforms like Facebook (Bakshy et al. 2015) and in the way algorithms trained on biased data can default to male-gendered pronouns (Zou and Schiebinger 2018).

Scientists now use digital media as a critical platform for disseminating scientific findings, and the field of altmetrics (short for "alternative metrics") has emerged as a tool for quantifying the digital attention received by scientific papers (Erdt et al. 2016). The term "altmetric" refers to a variety of available metrics that differ according to how they aggregate 'mentions' of scientific output on various digital media (e.g., blogs vs. Twitter vs.

Table 1 Identification of performance areas where bias exists from published examples in the literature. Most of the common performance metrics within the Academy show bias in favor of men

Type of evidence	Bias in favor of males	Bias in favor of females	No evidence of bias	Citation
Cite score	x			Dion et al. (2018)
% first authorships	x			Filardo et al. (2016)
h-index	x			García-Pérez et al. (2009)
Number of publications	x (given existing distribution of resources)		x (if controlling for position & funding)	Holliday et al. (2014)
Invited papers	x			Holman et al. (2018)
Reference letters	x			Madera (2019)
Grants	x			Morgan et al. (2018)
Quotes in media	x			Morris (2016)
Invited speakers	x			Nittrouer et al. (2018)
Altmetrics			x (for all journals except Science)	(Present study)
Interviews	x			Quadlin (2018)
% last authorships	x			West et al. (2013)

policy documents). Altmetric.com is one of the largest aggregators of altmetric scores, and hundreds of journals now publish the Altmetric Attention Score (AAS) for individual journal articles (Bornmann et al. 2018). Altmetric scores have been shown to be positively correlated with traditional measures of impact, such as citations, h-indices, or impact factors in some fields (Kunze et al. 2020; Nocera et al. 2019; Thelwall and Nevill 2018). Previous studies have suggested that altmetrics such as the AAS are difficult to interpret due to lack of normalization and standardization, as well as the proprietary nature of the algorithms used for web scraping. Nevertheless, AAS in particular has become an important measure of how articles are perceived and it is commonly shown alongside citation scores and journal impact factor (Gumpenberger et al. 2016).

Given the widespread use of altmetrics, scientists and policymakers have advocated for the incorporation of such scores when evaluating the overall impact of scientific papers (Sopinka et al. 2020). However, whether altmetric scores carry the same gender biases as traditional metrics used in scientific evaluation is unclear. To address this knowledge gap, here we investigated gender bias in AAS in seven major scientific publications for the years 2011–2018. A total of 208,804 journal articles were analyzed, representing ~ 1.6% of the 12 million research works covered by altmetric.com.

Materials and methods

We targeted articles from: Science, Proceedings of the National Academy of Sciences (PNAS), PLOS One, New England Journal of Medicine (NEJM), Nature, Cell, and BioRxiv (Figure S1). These journals represent idiotypic journal types, specifically general interest (Nature, Science, PNAS), open access (PLOS One), disciplinary (NEJM, Cell), and a preprint server (BioRxiv). This sample enabled us to explore bias across a range of journals in which scholars publish their work.

We extracted the following variables from the Altmetric data base: author names, journal, publication date (which we used to create independent variables in our models, see below) and Altmetric score 1 year after publication (which we used to create the dependent variables in our models). While AAS are available at different time intervals after article publication (e.g. 5 days, 1 month, 1 year, all-time), we used 1-year scores in our analysis, to both maintain a controlled exposure time, and to enable reasonable length of time for scores to amount.

We genderized author names using the genderizeR package in R (Wais 2016). We extracted all author names and inputted first names into the genderize.io API, which returns a suggested gender and probability, based on proportion of references in the genderize.io database. Names that can be considered unisex are assigned a gender based on the gender with the highest probability score, with a threshold of > 0.5. Names that do not appear in the genderize.io database are listed as "unknown". The genderize.io database includes names from over 79 countries and 89 languages. The error rate of genderize.io predictions has previously been estimated at 5.02% on a test dataset including names of European, African and Asian origin (Santamaría and Mihaljevic 2018).

We created three variables using this method: first author gender, last author gender and proportion of female authors for each article. Our genderized dataset consists of 31% female-first-authored papers (61% male-first-authored, 7% unknown) and 21% female-last-authored papers (69% male-last-authored, 9% unknown).

As the majority (139,035 out of 208,804 = 66.6%) of articles in our dataset, with 1-year exposure times, had an AAS of 0 (introducing error that cannot be normalized through a ln(x + 1) transform for Altmetric data as has been suggested elsewhere, see Thelwall 2020), we split our analysis into two simple questions: (1) does gender explain whether an article receives an AAS greater than 0; (2) does gender explain the magnitude of the score. We evaluated these two hypotheses using logistic and linear regression models, respectively.

The first model (model 1) had a binary dependent variable indicating whether an article received a score greater than 0, and a series of independent variables:

$$y_{\mathrm{i}} = \mathrm{Bin}\left(1, p_{\mathrm{i}}\right)$$

$$p_{\mathrm{i}} = \mathrm{logit}^{-1}(\beta_0 + \beta_{1..m-1}\mathrm{Gender} \times \mathrm{Journal} \times \mathrm{Year}_{\mathrm{i}} + \delta_1 \mathrm{Month}_{\mathrm{i}} + \delta_2 \mathrm{No.\ Auth}_{\mathrm{i}} + \delta_3 \mathrm{Prop.Fem}_{\mathrm{i}}$$

where p_{i} is the fraction of articles that have a score, β_0 is an intercept, Gender × Journal × Year are dummy variables representing m 3rd order interactions between gender of the first or last author, journal and year, Month is the publication month, No.Auth is the total number of authors of the article and Prop.Fem is the proportion of authors in the author list that are female.

The second model (model 2) has the magnitude of an article's 1-year AAS (log-transformed) as a continuous response variable and included the same independent variables as the binomial model above:

$$\log(y_{\mathrm{i}}) = N\left(\mu_{\mathrm{i}}, \sigma^2\right)$$

$$\mu_{\mathrm{i}} = \beta_0 + \beta_{1..m-1}\mathrm{Gender} \times \mathrm{Journal} \times \mathrm{Year}_{\mathrm{i}} + \delta_1 \mathrm{Month}_{\mathrm{i}} + \delta_2 \mathrm{No.\ Auth}_{\mathrm{i}} + \delta_3 \mathrm{Prop.Fem}_{\mathrm{i}}$$

where μ_{i} is $E(\log(y_{\mathrm{i}}))$.

We used these models for inference on the a priori null hypothesis of no difference between male and female first (or last) authors in the probability of obtaining an AAS (model 1), and in the magnitude of the score received (model 2). We computed simultaneous confidence intervals on the differences to account for multiple testing across journals, years, and for first and last authors.

BioRxiv was modeled separately because its available time series was shorter than the other journals in our sample (the repository was launched in late 2013). For all journals, we conducted standard model checking through visual diagnostics.

The overall goodness of fit for the binomial model for journals excluding BioRxiv was a pseudo-R^2 of 0.29, and for the linear model the adjusted R^2 was 0.26; while BioRxiv had a pseudo-R^2 of 0.64 and an adjusted R^2 of 0.06.

We conducted a sensivity analysis to check how the accuracy of the gender assignment impacted the model results. We found that the effect of misclassification error was likely negligible on our results: assuming a genderizing error rate (proportion of males misclassified as females and females misclassified as males) of 5% and a gender bias error rate (difference between misclassification of males and females) of 2% (Santamaría and Mihaljevic 2018), our observed difference in scores would be 10% larger than the true difference in score, indicating our results are not false negative due to inaccuracy in the genderize algorithm. All scripts are available as supplemental files and on Github (https://github.com/bjarnebartlett/AltmetricAnalysis).

Results

The objective of this study was to examine if author gender explains (a) whether a study receives an AAS greater than zero, and (b) the magnitude of the score. No substantial trends in gender bias were found in any journal regarding whether an article received an AAS, the only exception being Science in 2012, which was associated with women first authors obtaining a score (log odds ratio = 0.63, 95% confidence interval [0.09,1.17]) (Figure S2). We found these model results to be robust to the presence of outliers. We further found that there was no clear evidence of gender bias against women in the magnitude of AAS across six of the seven journals for 2011–2018 (Fig. 1). While we identified potential emerging effects in favor of women appearing in 2018 for the first author in Nature (difference in log(AAS) = 0.28, 95% CI [− 0.1, 0.65]) and PNAS (difference in log(AAS) = 0.34, 95% CI [0.01, 0.68]), these effects did not carry high confidence. We conclude, based on our analysis, that in general, there is little current evidence for gender bias in AAS across major academic journals.

The one exception to these findings was for Science magazine—which showed biases against women in 2017 and 2018. Specifically, women scored lower than men for first authors, with a difference in log(AAS) of − 0.71 (95% CI [− 0.97, − 0.45]) and − 1.88 (95% CI [− 2.19, − 1.56]), respectively. While we did not undertake a separate analysis of the individual contributing components of the AAS, these mean differences for Science equate

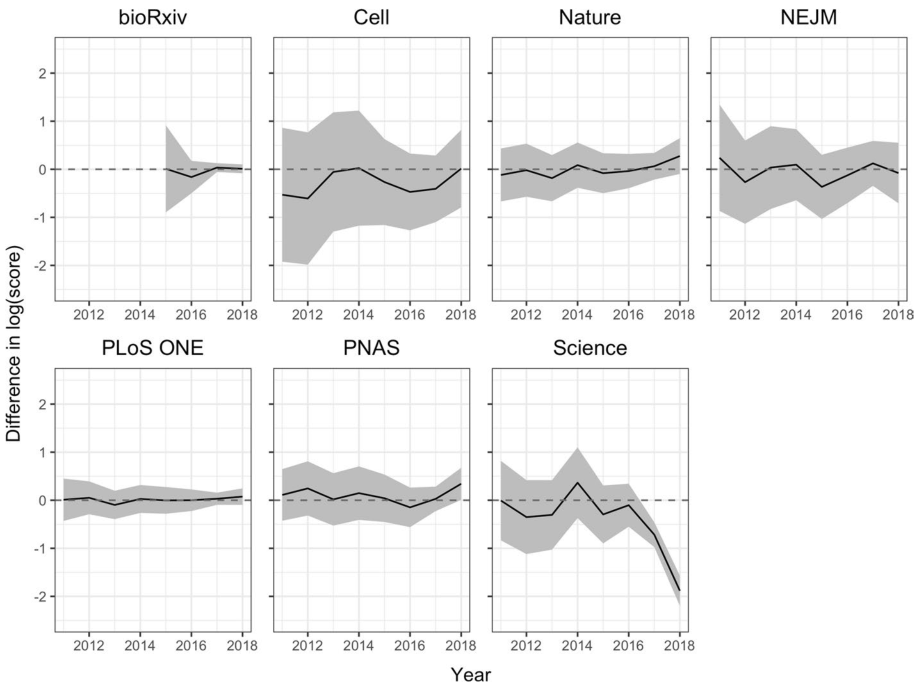

Fig. 1 Gender bias in seven idiotypic journals for Altmetric Attention Scores (AAS) for first authors. The black line represents the mean difference in log(AAS) score between female and male first authors; where positive numbers represent a higher score for females and vice versa for males. The gray shading is the 95% confidence interval. With the exception of Science magazine which shows bias in favor of male authors in 2017 and 2018, all other journals show no clear evidence of bias

to roughly 22 more tweets or 3 more news mentions per article for men than women in 2017, and 88 more tweets or 11 more news mentions per article in 2018. We re-ran this analysis with our inference based on quantile estimates, and found very similar results with respect to the mean differences, albeit with the median bias for first authors in Science in 2018 less prominent than the mean bias (equivalent to 5 more tweets for men than women in 2017, and to 20 more tweets or 2.5 more news articles in 2018) (Figure S3).

The results we found for last authors were generally consisent with those for first authors (Figure S2B and S2C). Beyond these effects of gender of the first and last author, we found no impact of publication month or proportion of female authors on AAS (Figure S4). As would be expected, the year of publication carried an important effect, with recent publications receiving higher scores (the mean 1-year score across all journals in 2011 was 10.5; in 2018, it was 42.1), as did having many authors, with scores increasing as the number of authors increased (each additional author leads to a 0.018 increase in log score for BioRxiv and a 0.005 increase in log score for all other journals) (Figure S4).

Discussion

Our results indicate a promising shift in new digital metrics relative to traditional metrics. Alternative metrics, like those consolidated by altmetric.com, leverage digital technology to attempt to quantify scientific reach, and aggregating mentions from vast user bases may serve to democratize the evaluation of scholarly outputs. However, there are arguments on either side that this digitization of evaluation could favor either gender. For instance, men are quoted more frequently in the news (Morris 2016) and are more likely to self-promote on platforms like Twitter (Duggan et al. 2015; Mancuso et al. 2017). Yet some communication patterns favor women: working on questions of interest to the public (Milkman and Berger 2014), attracting more student readers (Thelwall 2018), having higher crowdfunding success rates (PwC 2017). Our findings corroborate that the balance does not appear to be tipped in either direction; men and women authors of peer-reviewed articles get similar reach on digital platforms.

Gender plays a contributory role in professional advancement; moving toward less biased metrics is essential to create equitable workplaces. Traditional citation metrics, such as the h-index and i10-index, favor the male academic community (King et al. 2017). Other factors within the academy also favor the male population: differences in salary, space, awards, and resources have resulted in the marginalization of women faculty with women receiving less despite professional accomplishments equal to those of their male colleagues (A Study on the Status of Women Faculty in Science at MIT 1999). Some communication patterns favor women, as discussed above, but many of these factors are most present in digital spaces outside the academy and have little influence on tenure and promotion.

Citation metrics are particularly relevant to tenure and promotion in fields within science, technology, engineering, and math (STEM). Obstacles within STEM fields synergize to form an achievement disparity between men and women working in STEM, where less-qualified men are retained over more qualified women (Cimpian et al. 2020). Our analysis highlights that, while gender bias is rampant in academia, new digital metrics, carrying the balancing effect of a large, diverse group, have the potential to be democratizing and might provide less biased ways of assessing the impact of an academic article (Harambam et al. 2018). However, bias is not totally absent, as the worrying trend in Science shows. It is unclear where this bias arises from for Science, but one possibility is that the media team at Science themselves are biased in their promotion of men over women authors. This and other possible causes should be investigated further.

There are some potential limitations to our study. First, our methods rely on being able to ascribe binary gender based on names, which is limited according to references in the genderize.io database. Second there may be additional heterogeneity that we did not capture. For instance, field, article topic or open access status may have an influence on AAS in a way that was not adequately encapsulated in our selection of idiotypic journals. Third, we recognize the importance of recognizing non-binary genders in STEM and that current methods for genderizing names do not reflect this. The third issue in particular warrants further attention and study as it implies that our approach does not answer the question of whether authors' gender identity influences AAS, but rather whether the likely binary perception of authors' gender by others influences AAS.

Conclusion

Our results suggest that while traditional metrics used by academic institutions exacerbate bias, new metrics, such as the AAS, may present a levelling of the playing field. Unlike the findings for most indices used to measure competency in academia, we found no clear gender bias in AAS in six of the seven journals. While the AAS is not a definitive measure of the quality of the research or the researcher, our results present the first quantitative assessment of whether bias exists in this new metric, highlight its value as a complementary index, and suggest that new technologies, such as those on which altmetrics is based, may indeed help to reduce the institutional biases reflected in traditional metrics used in academic evaluation.

Supplementary file 1 : Supplementary figures S1-S4 3539 kbSupplementary file 2 : Python script for parsing Altmetric.com data 8 kbSupplementary file 3 : R script for genderizing author names 25 kbSupplementary file 4 : R script for modelling and figure creation 38 kbSupplementary file 5: Input data set used in modelling and figures 32433 kbSupplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s11192-021-03911-4.

Acknowledgments We would like to acknowledge the SEED Diversity Grant from UH Manoa for providing support and the KTUH radio station for providing a platform for us to share our research. We thank Navin Ramankutty for early discussion. We thank Stacy Konkiel and the Altmetric data science team for advice on the data. We thank an anonymous reviewer for their comments which helped improve the manuscript.

Author contributions ZM had the idea. BB, JF, ZM, MK MT designed research; BB and JF cleaned and curated data; BB conducted exploratory analyses, JF genderized the data and conducted the statistical modelling with assistance from ZM; BB, JF, ZM, MK, MT wrote the paper.

Funding SEED Diversity Grant from the University of Hawai'i at Manoa.

Data availability All data will be made available as supplemental material.

Code availability All scripts are available as supplemental files and at Github (https://github.com/bjarnebartlett/AltmetricAnalysis).

Compliance with ethical standards

Conflicts of interest The authors declare that they have no competing interests.

References

Bakshy, E., Messing, S., & Adamic, L. A. (2015). Exposure to ideologically diverse news and opinion on Facebook. *Science, 348*(6239), 1130–1132.

Bornmann, L., & Haunschild, R. (2018). Do altmetrics correlate with the quality of papers? A large-scale empirical study based on F1000Prime data. *PLoS ONE, 13*(5), e0197133.

Cimpian, J., Kim, T., & McDermott, Z. (2020). Understanding persistent gender gaps in STEM. *Science, 368*(6497), 1317–1319.

De Veirman, M., Hudders, L., & Nelson, M. R. (2019). What is influencer marketing and how does it target children? A review and direction for future research. *Frontiers in Psychology, 10,* 2685.

Dion, M. L., Sumner, J. L., & Mitchell, S. M. (2018). Gendered citation patterns across political science and social science methodology fields. *Political Analysis, 26*(3), 312–327.

Duggan, M., Ellison, N. B., Lampe, C., Lenhart, A., & Madden, M. (2015). Demographics of key social networking platforms. *Pew Research Center*, 9.

Enikolopov, R., Petrova, M., & Sonin, K. (2018). Social media and corruption. *American Economic Journal Applied Economics, 10*(1), 150–174.

Erdt, M., Nagarajan, A., Sin, S. C. J., & Theng, Y. L. (2016). Altmetrics: an analysis of the state-of-the-art in measuring research impact on social media. *Scientometrics, 109*(2), 1117–1166.

Filardo, G., da Graca, B., Sass, D. M., Pollock, B. D., Smith, E. B., & Martinez, M. A. M. (2016). Trends and comparison of female first authorship in high impact medical journals: observational study (1994–2014). *bmj*, *352*.

García-Pérez, M. A. (2009). The Hirsch h index in a non-mainstream area: methodology of the behavioral sciences in Spain. *The Spanish Journal of Psychology, 12*(2), 833.

Gumpenberger, C., Glänzel, W., & Gorraiz, J. (2016). The ecstasy and the agony of the altmetric score. *Scientometrics, 108*(2), 977–982.

Harambam, J., Helberger, N., & van Hoboken, J. (2018). Democratizing algorithmic news recommenders: how to materialize voice in a technologically saturated media ecosystem. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 376*(2133), 20180088.

Holliday, E. B., Jagsi, R., Wilson, L. D., Choi, M., Thomas, C. R., Jr., & Fuller, C. D. (2014). Gender differences in publication productivity, academic position, career duration and funding among US academic radiation oncology faculty. *Academic Medicine: Journal of the Association of American Medical Colleges, 89*(5), 767.

Holman, L., Stuart-Fox, D., & Hauser, C. E. (2018). The gender gap in science: How long until women are equally represented? *PLoS Biology, 16*(4), e2004956.

King, M. M., Bergstrom, C. T., Correll, S. J., Jacquet, J., & West, J. D. (2017). Men set their own cites high: Gender and self-citation across fields and over time. *Socius, 3,* 2378023117738903.

Kunze, K. N., Polce, E. M., Vadhera, A., Williams, B. T., Nwachukwu, B. U., Nho, S. J., & Chahla, J. (2020). What Is the Predictive Ability and Academic Impact of the Altmetrics Score and Social Media Attention? *The American Journal of Sports Medicine, 48*(5), 1056–1062.

Madera, J. M., Hebl, M. R., Dial, H., Martin, R., & Valian, V. (2019). Raising doubt in letters of recommendation for academia: Gender differences and their impact. *Journal of Business and Psychology, 34*(3), 287–303.

Mancuso, J., Neelim, A., & Vecci, J. (2017). Gender differences in self-promotion: Understanding the female modesty constraint. Available at SSRN: https://ssrn.com/abstract=3039233 or https://doi.org/10.2139/ssrn.3039233.

Milkman, K. L., & Berger, J. (2014). The science of sharing and the sharing of science. *Proceedings of the National Academy of Sciences, 111*(Supplement 4), 13642–13649.

MIT Committee on Women Faculty in the School of Science. A Study on the Status of Women Faculty in Science at MIT. (1999). http://web.mit.edu/fnl/women/women.html.

Morgan, R., Hawkins, K., & Lundine, J. (2018). The foundation and consequences of gender bias in grant peer review processes. *Canadian Medical Association Journal, 190*(16), E487–E488.

Morris, M. (2016). *Gender of sources used in major Canadian media*. Ottawa: Informed Opinions.

Nittrouer, C. L., Hebl, M. R., Ashburn-Nardo, L., Trump-Steele, R. C., Lane, D. M., & Valian, V. (2018). Gender disparities in colloquium speakers at top universities. *Proceedings of the National Academy of Sciences, 115*(1), 104–108.

Nocera, A. P., Boyd, C. J., Boudreau, H., Hakim, O., & Rais-Bahrami, S. (2019). Examining the correlation between Altmetric score and citations in the urology literature. *Urology, 134,* 45–50.

PWC (2017), Women unbound: Unleashing female entrepreneurial potential, accessed 20 December 2019 at https://www.pwc.com/gx/en/diversity-inclusion/assets/women-unbound.pdf

Quadlin, N. (2018). The mark of a woman's record: Gender and academic performance in hiring. *American Sociological Review, 83*(2), 331–360.
Raghavan, M., Barocas, S., Kleinberg, J., & Levy, K. (2020). Mitigating bias in algorithmic hiring: Evaluating claims and practices. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency* (pp. 469–481).
Santamaría, L., & Mihaljević, H. (2018). Comparison and benchmark of name-to-gender inference services. *PeerJ Computer Science, 4,* e156.
Sopinka, N. M., Coristine, L. E., DeRosa, M. C., Rochman, C. M., Owens, B. L., & Cooke, S. J. (2020). Envisioning the scientific paper of the future. *FACETS, 5*(1), 1–16.
Thelwall, M. (2018). Does female-authored research have more educational impact than male-authored research? Evidence from Mendeley. *Journal of Altmetrics, 1*(1).
Thelwall, M. (2020). Measuring societal impacts of research with altmetrics? Common problems and mistakes. *Journal of Economic Surveys*. https://doi.org/10.1111/joes.12381.
Thelwall, M., & Nevill, T. (2018). Could scientists use Altmetric.com scores to predict longer term citation counts? *Journal of informetrics, 12*(1), 237–248.
Wais, K. (2016). Gender Prediction Methods Based on First Names with genderizer. *The R Journal, 8*(1), 17.
West, J. D., Jacquet, J., King, M. M., Correll, S. J., & Bergstrom, C. T. (2013). The role of gender in scholarly authorship. *PLoS ONE, 8*(7), e66212.
Zou, J., & Schiebinger, L. (2018). AI can be sexist and racist—it's time to make it fair. *Nature, 589*, 324–326.