



COMMENTARY

Selective referencing and questionable evidence in Strumia's paper on "Gender issues in fundamental physics"

Jens Peter Andersen¹ , Mathias W. Nielsen² , and Jesper W. Schneider¹ 

¹Danish Centre for Studies in Research and Research Policy, Aarhus University

²Department of Sociology, Copenhagen University

an open access  journal



Citation: Andersen, J. P., Nielsen, M. W., & Schneider, J. W. (2021). Selective referencing and questionable evidence in Strumia's paper on "Gender issues in fundamental physics". *Quantitative Science Studies*, 2(1), 254–262. https://doi.org/10.1162/qss_a_00119

DOI:
https://doi.org/10.1162/qss_a_00119

Corresponding Author:
Jesper W. Schneider
jws@ps.au.dk

We have accepted the Editor's invitation to comment on Alessandro Strumia's paper in the current issue of *Quantitative Science Studies*. Strumia is a controversial figure. His biologicistic accounts of the persistent gender gap in science have been subject to heated debate—both in print and on social media. Some researchers argue that Strumia's viewpoints should be ignored. We disagree.

Despite overwhelming evidence of gender-related disadvantages, discrimination, and harassment (e.g., Brower & James, 2020; Budden, Tregenza, et al., 2008; Carli, Alawa, et al., 2016; Edmunds, Ovseiko, et al., 2016; El-Alayli, Hansen-Brown, & Ceynar, 2018; Guarino & Borden, 2017; Ilies, Hauserman, et al., 2003; Jagsi, Griffith, et al., 2016; Kabat-Farr & Cortina, 2014; Knobloch-Westerwick, Glynn, & Huges, 2013; Krawczyk & Smyk, 2016; Lerchenmueller & Sorenson, 2018; MacNell, Driscoll, & Hunt, 2015; National Academies of Sciences, Engineering, and Medicine, 2018; Reuben, Sapienza, & Zingales, 2014; Rivera, 2017; Rivera & Tilcsik, 2019; Sheltzer & Smith, 2014; Smyth & Nosek, 2015), Darwinist beliefs that science's gender gap is best explained by a natural selection of the best and the brightest still echo in the corridors of many research institutions.

We find it crucial to expose the questionable evidence used to promote such beliefs. Strumia's paper offers a case in point.

We structure our critique of Strumia's paper in four parts. First, we document practices of selective citing and reporting in the study's framing and conclusions. Second, we expose the questionable bibliometric assumptions guiding the empirical analysis. Third, we highlight data limitations and methodological flaws in Strumia's analysis, and fourth we take issue with the bold and far-fetched interpretations presented in the study's conclusion.

1. SELECTIVE CITING AND REPORTING

Misrepresenting previous research by leaving out relevant evidence that contradicts one's personal views ("cherry picking") or highlighting only those results that fit into one's own argument is at best questionable research practice. In his paper, Strumia does both of those things. Table 1 lists examples of what we believe are cases of selective citing and biased reporting. The left column displays the references in question, the middle column summarizes Strumia's account of these references, and the right column specifies what we see as problematic about Strumia's representation of the literature. Obviously, we may interpret the studies in question somewhat differently from Strumia, but in this case, the account of the literature seems surprisingly skewed

Copyright: © 2021 Jens Peter Andersen, Mathias W. Nielsen, and Jesper W. Schneider. Published under a Creative Commons Attribution 4.0 International (CC BY 4.0) license.



Table 1. Selected examples of selective citing and biased reporting in Strumia's paper

Cited reference in question	Strumia's interpretations	Problems with Strumia's interpretations
Caplar et al. (2017)	"For example, Caplar, Tacchella, and Birrer (2017) claim (consistent with my later findings) that papers in astronomy written by F authors are less cited than papers written by M authors, even after trying to correct for some social factors." (p. 233).	This is an example of imprecise reporting: In five astronomy journals, papers first-authored by males, on average, were cited approximately 6% more than papers first-authored by women.
Milkman et al. (2015)	"[L]ooking at gender in isolation (rather than at "women and minorities"), female students received slightly more responses from public schools (the majority of the sample) with respect to men in the same racial group." (p. 226).	This is an example of selective reporting. Milkman et al. (2015) report that "faculty were significantly more responsive to White males than to all other categories of students, collectively, particularly in higher-paying disciplines and private institutions." Private universities accounted for 37% percent of the sample.
Witteman et al. (2019)	"found that female grant applications in Canada are less successful when evaluations involve career-level elements" (p. 226)	This is an example of selective reporting. Witteman and colleagues (2019) also found that the sex differences in success rates (in grant obtainment) were marginal when reviewers were asked to rate the proposals independent of track record.
Xie and Shauman (1998), Levin and Stephan (1998), Abramo et al. (2009), Larivière et al. (2013), Way et al. (2016), Holman et al. (2018)	"Bibliometric attempts to recognize higher merit [...] found that male faculty members write more papers." (p. 226).	<p>This is an example of imprecise reporting. Xie and Shauman (1998) observe a 20% gap in research productivity in the late 1980s and early 1990s. However, they also find that "most of the observed sex differences in research productivity can be attributed to sex differences in personal characteristics, structural positions, and marital status."</p> <p>Levin and Stephan (1998) investigate gender differences in publication rates in four disciplines (Physics, Earth science, Biochemistry, and Physiology) and conclude that "in every instance, except the earth sciences, women published less than men, although the difference is statistically significant only for biochemists employed in academe and physiologists employed at medical schools" (p. 1056). The study did not adjust for scientific rank.</p> <p>In Abramo and colleagues' (2009) study of Italian researchers, female professors and associate professors in the physical sciences had higher publication rates than their male counterparts, while male assistant professors had higher publication rates than female counterparts (see Tables 7–9 in Abramo et al., 2009).</p> <p>Larivière et al. (2013) do not compare the average publication rates of women and men.</p>

Table 1. (continued)

Cited reference in question	Strumia's interpretations	Problems with Strumia's interpretations
Aycock et al. (2019)	"Various studies focused on discrimination as a possible source of gender differences. Small samples of female physics students were interviewed by Barthelemy, McCormick, and Henderson (2016) and Aycock, Hazari et al. (2019)." (p. 225).	Way et al. (2016) study publication productivity in computer science from 1970 to 2010 and find that "Productivity scores do not differ between men and women. This is true even when we consider only men and women who moved up the ranks and, separately, men and women who moved down ($p > 0.05$, Mann–Whitney)" (see Table 2 in Way et al., 2016). However, they find that in the cohort hired after 2002 men have higher average publication rates than women. Holman and colleagues' (2018) data set does not allow them to directly compare the publication rates of women and men.
Thelwall, Bailey et al. (2018)	"Large gender differences along the people/things dimension are observed in occupational choices and in academic fields: Such differences are reproduced within sub-fields (Thelwall et al., 2018). In particular, female participation is lower in sub-fields closer to physics, even within fields with their own cultures, such as 'physical and theoretical chemistry' within chemistry (Thelwall et al., 2018). This suggests that the people/things dimension plays a more relevant role than the different cultures of different fields." (p. 248).	This is an example of biased reporting: Aycock et al. (2019) report results from a survey of 455 undergraduate women in physics. Seventy-five percent of these had experienced at least one type of sexual harassment in a context associated with physics. The analysis by Thelwall and colleagues (2018) does not offer any substantial evidence that interest plays a greater role than culture.
Gibney (2017), Guarino and Borden (2017)	"Furthermore, psychology finds that females value careers with positive societal benefits more than do males: (...). Indeed Gibney (2017) finds that women in UK academia report dedicating 10% less time than men to research and 4% more time to teaching and outreach, and Guarino and Borden (2017) finds that women in U.S. non-STEM fields do more academic service than men." (p. 248).	Here, Strumia links women's extra burdens with respect to teaching obligations and academic service to an argument about a female propensity to value careers with positive societal benefits. However, none of these factors are highlighted or examined as potential confounders in his own gender comparisons of publication and citation rates.
Handley et al. (2015)	"Furthermore, fields that study bias might have their own biases: Stewart-Williams, Thomas et al. (2019) and Winegard, Clark et al. (2018) found that scientific results exhibiting male-favoring differences are perceived as less	This is an example of biased reporting. Handley et al. (2015) also found that men evaluated an abstract showing gender bias in research evaluations less favorably than a moderated version of the same abstract indicating no gender bias. This latter result (left out of Strumia's paper) counters his argument on this matter.

Ceci et al. (2014), Su et al. (2009), Lippa (2010), Hyde (2014), Su et al. (2015), Thelwall (2018b), Stoet et al. (2018)	<p>credible and more offensive. Handley, Brown et al. (2015) found that men (especially among STEM faculty) evaluate gender bias research less favorably than women.” (p. 247).</p> <p>“An important clue is that a similar gender difference already appears in surveys of occupational plans and first choices of high-school students (Ceci, Ginther et al., 2014; Xie & Shauman, 2003). This is possibly mainly due to gender differences in interests (Ceci et al., 2014; Hyde, 2014; Lippa, 2010; Stoet & Geary, 2018; Su & Rounds, 2015; Su, Rounds, & Armstrong, 2009; Thelwall, Bailey et al., 2018).” (p. 226).</p>	<p>This is an example of selective citing. Here, Strumia leaves out a vast literature on how prevalent gendered assumptions at play in cultural socialization and upbringing operate to divert men towards and women away from STEM careers. See, for example, Zwick and Renn (2000), Eccles and Jacobs (1990), Jacobs and Eccles (1992), and Jones and Wheatley (1990).</p>
Su et al. (2009), Diekman et al. (2010), Lippa (2010), Su et al. (2015), Thelwall (2018)	<p>“This suggests extending my considerations from possible sociological issues to possible biological issues.</p> <p>It is interesting to point out that the gender differences in representation and productivity observed in bibliometric data can be explained at face value (one does not need to assume that confounders make things different from what they seem), relying on the combination of two effects documented in the scientific literature: differences in interests (Diekman, Johnson, & Clark, 2010; Lippa, 2010; Su, Rounds, & Armstrong, 2009; Su & Rounds, 2015; Thelwall, Bailey et al., 2018)” ... (p. 247–248).</p>	<p>This is an erroneous interpretation of the literature. With the exception of Lippa (2010), none of the studies listed here directly relate their findings to biological sex differences. Indeed, Su and Rounds (2015) argue that “while the literature has consistently shown the influence of social contexts (e.g., parents, schools) on students’ interest development, particularly the development of differential interests for boys and girls (...), little is known about the link between biological factors (e.g., brain structure, hormones) and interest development.”</p>

in the direction of Strumia's underlying agenda. The list of omitted references that could have added nuance to Strumia's review is too comprehensive to be covered in this comment.

2. MISGUIDED ASSUMPTIONS

Strumia's questionable citing practices serve as an illustrative example of what sociologists and scientometricians refer to as "referencing as persuasion" (Gilbert, 1977; Latour, 1987)¹. Paradoxically, Strumia's own empirical analysis builds on a completely different, and more normative conception of what a citation is. In his paper, he claims that citation indicators represent a reliable proxy of scientific merit (i.e., "referencing as rewards": Kaplan, 1965; Merton, 1968). By so doing, Strumia disregards the vast literature demonstrating the drawbacks of using citations as quality indicators (for a recent review, see Aksnes, Langfeldt, & Wouters, 2019). There are very good reasons why Martin and Irvine (1983) chose to equate citations with impact, not merit or quality. Citations are noisy, social measures and their distributions are skewed, not least due to cumulative effects (Merton, 1968). Many references are perfunctory (Moravcsik & Murugesan, 1975) and citing practices often have a social and persuasive function (as illustrated in Strumia's own paper). They are interesting as indices of symbolic capital in the science system (Bourdieu, 1988). In a tautological sense, they may be indicative of "high performance" to some, and they are certainly (mis)used in evaluative contexts, but it is a major delusion to use citation indicators as a direct measure of merit. But Strumia's use of citations is quite unusual as he makes an unsubstantiated chain of reasoning from citations to merit and from merit to nonsensical claims about biological sex differences in physicists' cognitive capacities.

3. METHODOLOGICAL FLAWS

A foundation for Strumia's analysis is his strong belief in the value of what he calls "large amounts of objective quantitative data about papers, authors, citations, and hires" (p. 225). We are not so impressed with the amounts of data or their quality, let alone what they may be a proxy for.

Bibliometric data are not objective *per se*, as Strumia implies. They are generally noisy (i.e., faulty, biased, and incomplete; Schneider, 2013). Noise is additive. Thus, citation linkages introduce errors, and so do author and gender disambiguation. There is no reason to assume that such errors are random. Noisy data are rife in the social sciences, especially in areas where data are "big" and processed algorithmically. We should always interpret results derived from such data with caution, especially when the observed differences are small. Large samples may give "precise" estimates, but precise estimates can be systematically biased when the analysis builds on noisy data. Having worked with author and gender disambiguation ourselves (Andersen, Schneider, et al., 2019; Nielsen, Andersen, et al., 2017), we would be more cautious than Strumia in declaring supremacy of data quantity over data quality.

Like many other bibliometric studies, Strumia's analysis is data driven, and nowhere do we get the impression that a preanalysis plan has been specified or followed. A careful preanalysis plan will decrease "researcher degrees of freedom" (Simmons, Nelson, & Simonsohn, 2011) in planning, running, analyzing, and reporting opportunistically, and reassure that the findings are not just the outcome of extensive data mining. Ruling out data mining and data-dependent analysis is essential when studies pretend to be confirmatory with causal-like statements. Strumia's study is exploratory and this has implications for what can be made of the results. The flexibility

¹ While Gilbert's "persuasion" concerns the use of "acknowledged" references to boost one's own work, Latour argues that citing authors often deliberately misrepresent and distort the works they allude to by twisting the meaning to suit their own ends. We believe that Strumia practices both forms of persuasion.

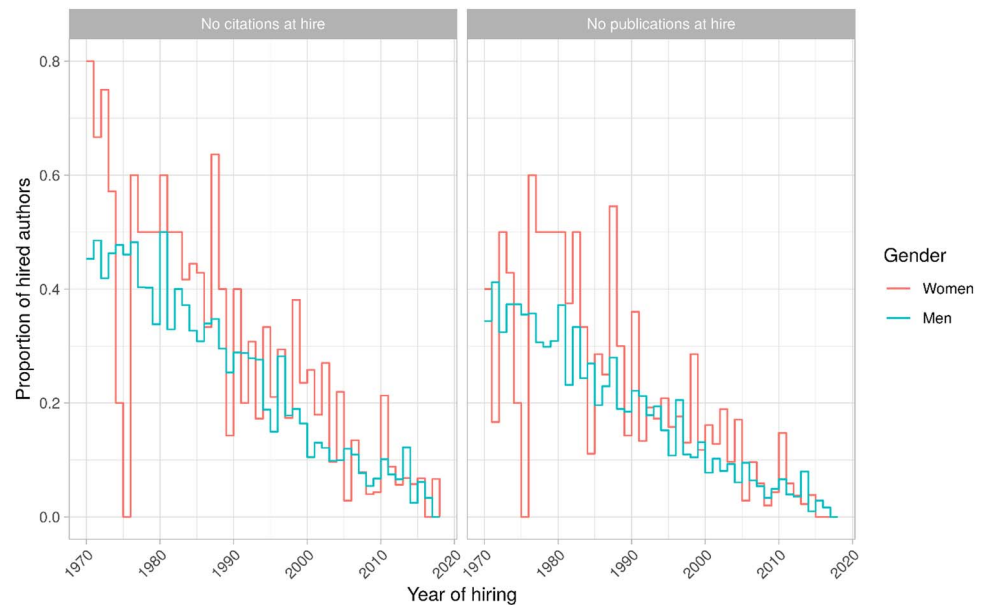


Figure 1. Proportion of hired authors with no citations and no publications in the hiring year.

in sampling, processing, and analytical choices obviously implies that the results are conditional and that different choices could have produced different results. Without a preanalysis or a multiverse of potential alternative analyses (Steege, Tuerlinckx, et al., 2016), selective reporting and confirmation bias seem likely. In such situations, statistical significance tests are uninterpretable (Gelman & Loken, 2013).

The hiring analysis presented in Strumia’s Section 3.2 is based on “big” longitudinal data of questionable quality. Strumia claims that the HepNames database used in this part of the analysis offers “precise career information” (p. 229). However, a quick-and-dirty lookup of five renowned Danish fundamental physicists returned no useful information about “first hirings” that could go into such an analysis². Strumia seems aware that his data are flawed. He ends up with a sample of “about 10,000 first hires” and supplements these with a sample of “unbiased ‘pseudo-hires’” (p. 229). The first of these samples is clearly a convenience sample plagued by selection bias. The “pseudo-hires” are indeed pseudo; if not, then we would assume that Strumia would have used only the “pseudo-hires” as a proxy.

Strumia has granted us access to the raw data used in the hiring analysis. Our inspection of this data set reveals that a large share of the listed authors do not have any publications or citations prior to being hired (see Figure 1). We estimate that first-hires without any publications or citations account for up to 40% of the listed authors in the early period of the sample. In other words, the hiring analysis is based on questionable data. Strumia’s own hiring analysis also includes suspicious yearly fluctuations in the average number of fractionally counted papers and individual citations at the first hiring moment. Further, it does not provide any annual baseline of how many men and women were hired (see Figure 4 and Figures S2 and S3 in Strumia’s paper). The longitudinal perspective is also misleading given that a large share of the authors hired in the

² For example, Nils Overgaard Andersen, Jens Hjorth, Flemming Besenbacher, Lene Vestergaard Hau, and Sune Lehmann. We also looked up some where information was present, albeit not in standardized form (e.g., Benny Lautrup and Andrew Jackson).

early period have no registered publications or citations in the database. Given the volatility and skewed nature of the data, we find it peculiar that Strumia only reports mean scores in Figures 4 and 8. Median scores and variances would have underscored the fragility of the data.

Confounding is a major challenge in bibliometric research, and especially so in observational studies of hiring and selection. Strumia's analysis is no exemption. The analytical approach is overly simplistic and atheoretical, and Strumia does not offer any convincing solutions for how to deal with the many potential confounders that plague the analysis. Indeed, inelegant attempts are made to rule out the influence of selected confounders (including institutional prestige, continent, and scientific age), but all of these confounders are examined in isolation (see Figures S2–S4 in Strumia's paper). In a social science perspective, this makes the hiring analysis unavailing.

4. FAR-FETCHED CONCLUSIONS

We do not as such reject all of Strumia's empirical findings. The slight gender variations observed in the citation and publication distributions are compatible with the results of other bibliometric gender comparisons³. Note here that in observational settings, such aggregate findings are extremely vulnerable to selection bias. What we *do* reject is Strumia's far-fetched interpretations of these findings. Here, we present selected statements from the study's conclusion and take issue with the most preposterous and nonsensical claims.

While many social phenomena could produce different averages, producing different variances would need something that specifically disadvantages research by top female authors. Just to take one example of a social nature, a gender gap in research productivity could arise if better female authors receive more honors and leadership positions that drive them away from research. However, data also show an excess of young authors among those who produced top-cited papers: The excess is observed among both M and F authors. This suggests extending my considerations from possible sociological issues to possible biological issues. [p. 247]

It is interesting to point out that the gender differences in representation and productivity observed in bibliometric data can be explained at face value (one does not need to assume that confounders make things different from what they seem), relying on the combination of two effects documented in the scientific literature: differences in interests and in variability." [p. 248]

The claims made here are speculative, empirically unsubstantiated, and founded on twisted assumptions. First, there is no reason to believe that differences in averages are more likely to stem from social factors than differences in variability, at least not when it comes to scientific performance. The argument presented here does not follow logically from the results. Second, Strumia's biologicistic reading of the literature on gender differences in interests is misguided (see Table 1). Third, Strumia does not measure intelligence in his analysis. Thus, his assertion that sex differences in variability "explain" gender differences in productivity is both unreasonable and unwarranted. Fourth, extant research on intelligence and scientific productivity is scarce, and does not suggest any direct relationship between the two (Bayer & Folger, 1966; Cole & Cole, 1974). Fifth, Strumia's speculations of a higher male variability in fundamental physics have no empirical basis in the peer-reviewed scientific literature.

³ Further, Sabine Hossenfelder and colleagues (2018) seem to corroborate Strumia's aggregate findings in a comparison of Inspire and arXiv data albeit with smaller, average, gender differences and diverging results on the question of gender homophily in citing practices.

In summary, what Strumia's gender analysis contributes is (a) a strongly biased representation of the existing literature, (b) a superficial, exploratory citation and publication analysis based on misguided assumptions, (c) an overly simplistic hiring analysis plagued by confounding and noisy data, and (d) concluded by highly speculative explanations based on twisted assumptions and with little or no empirical basis.

REFERENCES

- Aksnes, D. W., Langfeldt, L., & Wouters, P. (2019). Citations, citation indicators, and research quality: An overview of basic concepts and theories. *SAGE Open*, 9(1), 1–17. DOI: <https://doi.org/10.1177/2158244019829575>
- Andersen, J. P., Schneider, J. W., Jagsi, R., & Nielsen, M. W. (2019). Gender variations in citation distributions in medicine are very small and due to self-citation and journal prestige. *eLife*, 8, e45374. DOI: <https://doi.org/10.7554/eLife.45374>, PMID: 31305239, PMCID: PMC6677534
- Bayer, A. E., & Folger, J. (1966). Some correlates of a citation measure of productivity in science. *Sociology of Education*, 39(4), 381–390. DOI: <https://doi.org/10.2307/2111920>
- Bourdieu, P. (1988). *Homo academicus*. Redwood City, CA: Stanford University Press.
- Brower, A., & James, A. (2020). Research performance and age explain less than half of the gender pay gap in New Zealand universities. *PLOS ONE*, 15(1), e0226392. DOI: <https://doi.org/10.1371/journal.pone.0226392>, PMID: 31967992, PMCID: PMC6975525
- Budden, A. E., Tregenza, T., Aarssen, L. W., Koricheva, J., Leimu, R., & Lortie, C. J. (2008). Double-blind review favours increased representation of female authors. *Trends in Ecology & Evolution*, 23(1), 4–6. DOI: <https://doi.org/10.1016/j.tree.2007.07.008>, PMID: 17963996
- Carli, L. L., Alawa, L., Lee, Y., Zhao, B., & Kim, E. (2016). Stereotypes about gender and science: Women ≠ scientists. *Psychology of Women Quarterly*, 40(2), 244–260. DOI: <https://doi.org/10.1177/0361684315622645>
- Cole, J. R., & Cole, S. (1974). *Social Stratification in Science*. University of Chicago Press. DOI: <https://doi.org/10.1119/1.1987897>
- Edmunds, L. D., Ovseiko, P. V., Shepperd, S., Greenhalgh, T., Frith, P., Roberts, N. W., ... & Buchan, A. M. (2016). Why do women choose or reject careers in academic medicine? A narrative review of empirical evidence. *The Lancet*, 388(10062), 2948–2958. DOI: [https://doi.org/10.1016/S0140-6736\(15\)01091-0](https://doi.org/10.1016/S0140-6736(15)01091-0)
- El-Alayli, A., Hansen-Brown, A. A., & Ceynar, M. (2018). Dancing backwards in high heels: Female professors experience more work demands and special favor requests, particularly from academically entitled students. *Sex Roles*, 79(3–4), 136–150. DOI: <https://doi.org/10.1007/s11199-017-0872-6>
- Gelman, A., & Loken, E. (2013). The garden of forking paths: Why multiple comparisons can be a problem, even when there is no “fishing expedition” or “p-hacking” and the research hypothesis was posited ahead of time. (Unpublished: http://www.stat.columbia.edu/~gelman/research/unpublished/p_hacking.pdf)
- Gilbert, N. G. (1977). Referencing as persuasion. *Social Studies of Science*, 7(1), 113–122. DOI: <https://doi.org/10.1177/030631277700700112>
- Guarino, C. M., & Borden, V. M. (2017). Faculty service loads and gender: Are women taking care of the academic family? *Research in Higher Education*, 58(6), 672–694. DOI: <https://doi.org/10.1007/s11162-017-9454-2>
- Hossenfelder, S. (2018). Do women in physics get fewer citations than men? *Backreaction*, November 30. <http://backreaction.blogspot.com/2018/11/do-women-in-physics-get-fewer-citations.html>.
- Ilies, R., Hauserman, N., Schwochou, S., & Stibal, J. (2003). Reported incidence rates of work-related sexual harassment in the United States: Using meta-analysis to explain reported rate disparities. *Personnel Psychology*, 56(3), 607–631. DOI: <https://doi.org/10.1111/j.1744-6570.2003.tb00752.x>
- Jagsi, R., Griffith, K. A., Jones, R., Perumalswami, C. R., Ubel, P., & Stewart, A. (2016). Sexual harassment and discrimination experiences of academic medical faculty. *JAMA*, 315(19), 2120–2121. DOI: <https://doi.org/10.1001/jama.2016.2188>, PMID: 27187307, PMCID: PMC5526590
- Kabat-Farr, D., & Cortina, L. M. (2014). Sex-based harassment in employment: New insights into gender and context. *Law and Human Behavior*, 38(1), 58. DOI: <https://doi.org/10.1037/lhb0000045>, PMID: 23914922
- Kaplan, N. (1965). The norms of citation behavior: Prolegomena to the footnote. *American Documentation*, 16(3), 179–184. DOI: <https://doi.org/10.1002/asi.5090160305>
- Knobloch-Westerwick, S., Glynn, C. J., & Huge, 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. DOI: <https://doi.org/10.1177/1075547012472684>
- Krawczyk, M., & Smyk, M. (2016). Author's gender affects rating of academic articles: Evidence from an incentivized, deception-free laboratory experiment. *European Economic Review*, 90, 326–335. DOI: <https://doi.org/10.1016/j.euroecorev.2016.02.017>
- Latour, B. (1987). *Science in action*. Cambridge, MA: Harvard University Press.
- Leichenmueller, M. J., & Sorenson, O. (2018). The gender gap in early career transitions in the life sciences. *Research Policy*, 47(6), 1007–1017. DOI: <https://doi.org/10.1016/j.respol.2018.02.009>
- MacNell, L., Driscoll, A., & Hunt, A. N. (2015). What's in a name: Exposing gender bias in student ratings of teaching. *Innovative Higher Education*, 40(4), 291–303. DOI: <https://doi.org/10.1007/s10755-014-9313-4>
- Martin, B. R., & Irvine, J. (1983). Assessing basic research: Some partial indicators of scientific progress in radio astronomy. *Research Policy*, 12(2), 61–90. DOI: [https://doi.org/10.1016/0048-7333\(83\)90005-7](https://doi.org/10.1016/0048-7333(83)90005-7)
- Merton, R. K. (1968). The Matthew effect in science. *Science*, 159(3810), 56–63. DOI: <https://doi.org/10.1126/science.159.3810.56>, PMID: 5634379
- Moravcsik, M. J., & Murugesan, P. (1975). Some results on the function and quality of citations. *Social Studies of Science*, 5(1), 86–92. DOI: <https://doi.org/10.1177/030631277500500106>
- National Academies of Sciences, Engineering, and Medicine. (2018). *Sexual harassment of women: climate, culture, and consequences in academic sciences, engineering, and medicine*. National Academies Press.

- Nielsen, M. W., Andersen, J. P., Schiebinger, L., & Schneider, J. W. (2017). One and a half million medical papers reveal a link between author gender and attention to gender and sex analysis. *Nature Human Behaviour*, 1(11), 791–796. **DOI:** <https://doi.org/10.1038/s41562-017-0235-x>, **PMID:** 31024130
- 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. **DOI:** <https://doi.org/10.1073/pnas.1314788111>, **PMID:** 24616490, **PMCID:** PMC3970474
- Rivera, L. A. (2017). When two bodies are (not) a problem: Gender and relationship status discrimination in academic hiring. *American Sociological Review*, 82(6), 1111–1138. **DOI:** <https://doi.org/10.1177/0003122417739294>
- Rivera, L. A., & Tilcsik, A. (2019). Scaling down inequality: Rating scales, gender bias, and the architecture of evaluation. *American Sociological Review*, 84(2), 248–274. **DOI:** <https://doi.org/10.1177/0003122419833601>
- Sheltzer, J. M., & Smith, J. C. (2014). Elite male faculty in the life sciences employ fewer women. *Proceedings of the National Academy of Sciences*, 111(28), 10107–10112. **DOI:** <https://doi.org/10.1073/pnas.1403334111>, **PMID:** 24982167, **PMCID:** PMC4104900
- Schneider, J. W. (2013). Caveats for using statistical significance tests in research assessments. *Journal of Informetrics*, 7(1), 50–62. **DOI:** <https://doi.org/10.1016/j.joi.2012.08.005>
- Simmons, J. P., Nelson, L. D., and Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, 22, 1359–1366. **DOI:** <https://doi.org/10.1177/0956797611417632>, **PMID:** 22006061
- Smyth, F. L., & Nosek, B. A. (2015). On the gender–science stereotypes held by scientists: Explicit accord with gender-ratios, implicit accord with scientific identity. *Frontiers in Psychology*, 6, 415. **DOI:** <https://doi.org/10.3389/fpsyg.2015.00415>, **PMID:** 25964765, **PMCID:** PMC4410517
- Steege, S., Tuerlinckx, F., Gelman, A., & Vanpaemel, W. (2016). Increasing transparency through a multiverse analysis. *Perspectives on Psychological Science*, 11(5), 702–712. **DOI:** <https://doi.org/10.1177/1745691616658637>, **PMID:** 27694465