



# Reply to commentaries about “Gender issues in fundamental physics: A bibliometric analysis”

Alessandro Strumia 

Physicist, Fauglia (Italy)

an open access  journal

Citation: Strumia, A. (2021). Reply to commentaries about “Gender issues in fundamental physics: A bibliometric analysis”. *Quantitative Science Studies*, 2(1), 277–287. [https://doi.org/10.1162/qss\\_c\\_00120](https://doi.org/10.1162/qss_c_00120)

DOI:  
[https://doi.org/10.1162/qss\\_c\\_00120](https://doi.org/10.1162/qss_c_00120)

Corresponding Author:  
Alessandro Strumia  
[astrumia@icloud.com](mailto:astrumia@icloud.com)

My bibliometric study about fundamental physics and gender (Strumia, 2021) received four commentaries: Andersen, Nielsen, and Schneider (2021), Ball, Britton et al. (2021), Hossenfelder (2021), and Thelwall (2021). One commentary replicated the analysis: Sabine Hossenfelder (the only author of the commentaries who works in fundamental physics) writes: “His findings are significant and robust. My collaborators Tobias Mistele, Tom Price, and I have been able to reproduce the bibliometric results with the same database and with a different database of the same disciplines” (Hossenfelder, 2021). Some other commentaries performed partial checks using my data, which I made fully public. They raised doubts about specific scientific points. I first clarify such doubts, confirming my results. I next reply to generic sociological comments that, in my opinion, denote a lack of insider knowledge of fundamental physics. I conclude by addressing comments that go beyond the scientific issue, touching on political topics.

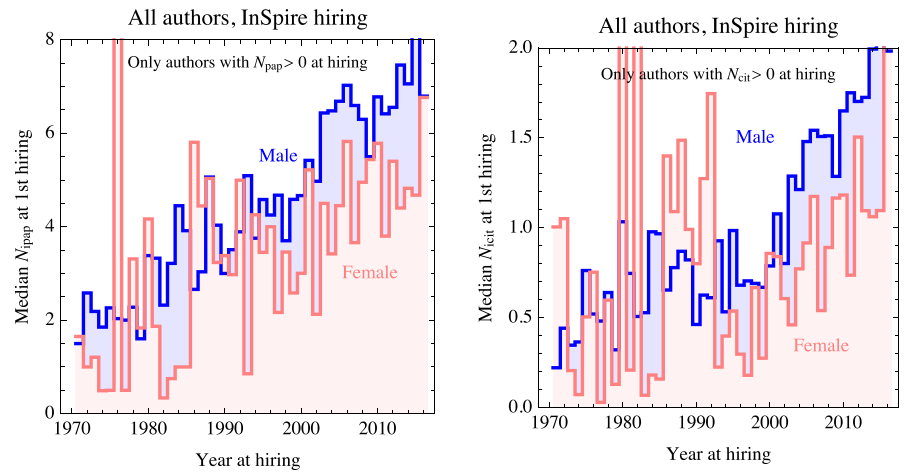
## 1. DOUBTS ABOUT INSPIRE HIRES

Andersen et al. (2021) raise a doubt about Figure 4 of Strumia (2021) (bibliometric indices for hires), viewing as suspect that, according to the InSpire database, about 20% of hired authors had no papers or citations in fundamental physics at the moment of their first hire. I see no solid reason to exclude this fraction, after looking at its geographic and time dependence: The fraction of suspect cases was higher 50 years ago and became small in recent times, consistently with anecdotal reports. To clarify the doubt raised by Andersen et al. (2021), the new Figure R1 shows the result obtained by omitting suspect hires with no citations or papers. Additionally, following another suggestion by Andersen et al. (2021), Figure R1 shows the median (rather than the mean) of the bibliometric indices at hiring. Despite these changes, the gender gap found in Figure 4 persists in Figure R1.

Furthermore, Andersen et al. (2021) argue that pseudohires are not valid because I have not used them in Figure 4. The real reason for this choice is explained in Section 2.2 of my original paper: Pseudo-hires computed from affiliations are an independent sample that gives larger coverage than InSpire hires but less accurate hiring times: Figure 4 has been computed using InSpire hires because timing is here more important than coverage. To clarify this issue, the new Figure R2 shows the same plot recomputed using pseudohires: Once more the same gender gap shows up, as in other checks already presented in Figure S2. Finally, it is worth mentioning that a recent independent analysis Madison and Fahlman (2020) found a similar gap studying hires of professors in Sweden.

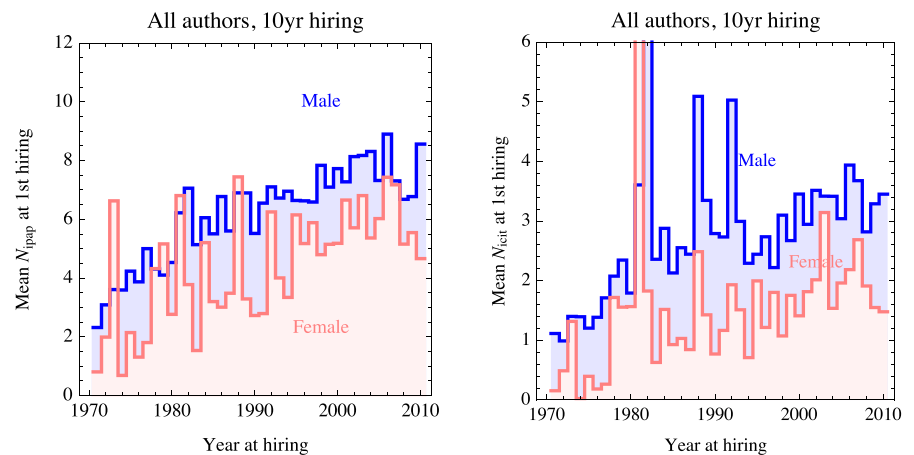
## 2. A POSTERIORI ADJUSTMENT OF METHODOLOGIES?

Andersen et al. (2021) raise the issue that methodologies might have not been chosen *a priori* before designing the analysis and collecting data. While social experiments can be performed



**Figure R1.** Variants of Figure 4 suggested by Andersen et al. (2021). The left and right panels show the median number of fractionally counted papers  $N_{\text{pap}}$  and of individual citations  $N_{\text{cit}}$  respectively, of authors in fundamental physics at the moment of their first hire, as a function of the hiring year, omitting authors who were hired with zero papers or citations in the InSpire database. The gender differences remain similar to Figure 4 in my original paper.

in ideal (possibly unrealistic) conditions, this is more difficult when studying reality. Let me clarify that no *a posteriori* adjustment of methodologies has been made. Indeed, having in mind this possible concern, I followed the same methodologies as in my previous publication (Strumia & Torre, 2019), which focused on exploring and developing bibliometric indices that provide good results based on simple observables with no free adjustments or data manipulations. Some results were already in use in bibliometrics, but not by physicists. Only later did I apply the same data and methodologies to gender, as by chance I was in a physics institution that started focusing on gender. My publication about gender (Strumia, 2021) contains extra



**Figure R2.** Figure 4 recomputed using, instead of InSpire hires, the pseudohires of Section 2.2 (10 consecutive years with the same affiliation). The gender gap in Figure 4 persists. Five-year pseudohires give similar results.

indices and tests that give consistent results (I thank the referees for their suggestions). In the future the same methodologies could be applied to other sectors of physics and to other fields, about which I have no bibliometric data.

Andersen et al. (2021) generically claim that I left out relevant evidence (“cherry picking”); in reality I analyzed all of the data about fundamental physics worldwide from 1970 to 2018, as available in the InSpire database. Nothing has been left out. Andersen et al. (2021) are “not so impressed with the amounts of data,” warn that “estimates can be systemically biased,” and worry that I declared “supremacy of data quantity over data quality.” So let me clarify that I just mean that a larger amount of data (data quantity) helps to test and reduce systematic uncertainties (data quality). Concretely, the large amount of data was used to probe individually various confounders, without applying any data manipulation, by checking that gender differences seen in the full data set persist inside slices of data (restricted based on scientific fields, number of authors, hiring status, countries, time periods, etc). As Hossenfelder (2021) comments, “Strumia’s analysis collects biographic and bibliometric data from about 70,000 scientists and is therefore statistically far more informative than most of the existing studies on gender bias in physics and related disciplines, which recruit on the order of 50 or so participants.”

### 3. A COCKTAIL OF CONFOUNDERS?

Ball et al. (2021) remark that “extreme care must be taken when arguing for causal relationships.” As care should be exercised in all directions, rather than selectively, I used bibliometric data to check the mainstream view according to which gender gaps in physics are causally attributed to biases. The dominant pattern emerging from data is not characteristic of biases, and pointed towards the different interpretation mentioned in my conclusions. If one believes that such a pattern could be produced by confounders, the same reasoning applies *a fortiori* to biases, as they are a possibly smaller effect that has not emerged from the data.

The commentaries do not identify relevant specific overlooked confounders, in addition to the confounders that have already been considered: Gender differences in seniority have been compensated for by the reweighting in Eq. 2, while other confounders turned out to be individually small. As pointed out in my original paper, this leaves open the possibility that the main pattern found in the data might be produced by a combination of small confounders. This possibility is often considered (e.g., in social sciences) through regression analyses, mostly done in linear approximation because it provides a simple model with a manageable number of free parameters. Ball et al. (2021) notice that I have not followed this practice. The reason is that most small confounders are nonlinear: A linear analysis would be invalid, introduce arbitrariness and allow cherry picking (Section 4 shows an example of this). Given that no model of citation practice is available to reliably combine small nonlinear confounders, I preferred presenting simple raw observables and using the large statistics to test confounders by slicing data.

### 4. THE REANALYSIS BY BALL ET AL. (2021)

Ball et al. (2021) claim that my analysis is “ideologically motivated” because I presented data without correcting for “direct discrimination” and “tendency to overcite well-known authors.” On the contrary, what can lead to ideologically biased analyses is doing this kind of data manipulations Ball et al. (2021). provide an example in this direction, presenting an alternative reanalysis of a small subset of my database where they add corrections they view as necessary

and conclude that “female-authored papers are actually cited more than male-authored papers.” Their reanalysis is flawed because of the multiple reasons listed below:

1. Ball et al. (2021) do not average citations but use the inverse hyperbolic sine of citations. This does not satisfy sum rules that allow us to measure groups as the sum of their parts. For example

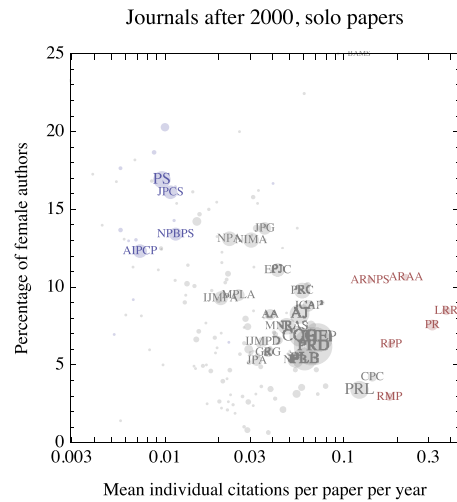
$$2 + 2 = 4 \neq \operatorname{arcsinh}(\sinh 2 + \sinh 2) \approx 2.7. \quad (1)$$

Intensive quantities (such as densities in physics) satisfy sum properties that make them useful observables, unlike random functions. In the bibliometric context, this was mentioned in Section 4 of Strumia and Torre (2019). Ball et al. (2021) justify their arcsinh choice by claiming that “male authors disproportionately cluster at the very top and very bottom of the citation distribution”. This means that Ball et al. (2021) see higher male variance in data, and suppress it by replacing citations with a function that artificially reduces the gap between poorly cited and top-cited papers<sup>1</sup>.

2. Next, Ball et al. (2021) “adjust for... authors’ research age and their lifetime fame.” In practice, this means that they artificially penalize citations when the cited author is older and/or has published many past papers in high-quality journals. Their “adjustments” bias the gender averages because male authors are currently older on average and the adjustments are wrongly done in linear approximation. The linear approximation is not valid because the phenomena are nonlinear. Indeed, bibliometric data show that the average scientific output of authors does not increase linearly with their age. Approximating it with a linear function is wrong, although a monotonically growing function might be motivated by the postmodernist view of science as a social hierarchy where elder authors are rewarded for power. As discussed in footnote 13 of Strumia (2021), data about fundamental physics show instead that top-cited papers tend to be produced by younger authors, possibly because cognitive abilities decline after middle age. While this might motivate a correction opposite to Ball et al. (2021), my analysis avoided questionable data adjustments.
3. Furthermore Ball et al. (2021), “adjust for ... journal of publication.” In practice, this essentially means that they compute the average citations received by male and female authors within each journal (out of five good journals selected in their analysis). This is a logical mistake, because journals aim to select papers according to their scientific value: Good journals primarily select good papers independently of author gender/age etc. So, when analyzing subsamples of papers sliced at roughly fixed quality, a gap in citations is hidden by looking only at within-journal arcsinh averages, as Ball et al. (2021) do. The correct implication is that the publishing system is fairly doing its expected job. Then one would expect the gender gap to be visible by looking at the gender ratio of the total number of papers in different journals. Indeed, female authors produced 3.8% (i.e., less than their representation) of the solo papers in a few good journals considered by Ball et al. (2021). By extending the analysis to all solo papers in all journals in the full database, Figure R3 shows a pattern consistent with my results<sup>2</sup>.

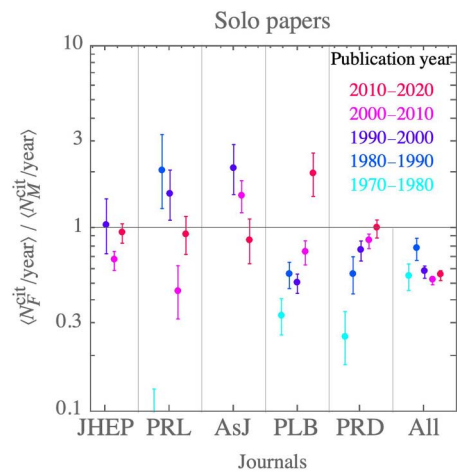
<sup>1</sup> Ball et al. (2021) claim that introducing the arcsinh has a small effect, but the gender gap they claim is also small and arcsinh averages are lower than averages by an amount comparable to standard deviations, which are large and show mild gender differences. Their additional motivation “raw citation counts are truncated below at zero” does not apply to my analysis, based on fractionally counted citations. Fractional counting gives an intensive observable and does not introduce an artificial scale, while an arcsinh misses both features.

<sup>2</sup> Adjusting for gender history would negligibly modify the figure. The analysis in my original paper avoids using data about journals, as this would involve unnecessary complications: almost all papers first freely appear on *arXiv* and some top-cited authors avoid publishing their papers.



**Figure R3.** As in Figure S1, restricting to solo papers. For each journal, we plot on the horizontal axis the mean individual citations received per year by all solo papers and on the vertical axis the percentage of female authors among solo papers. Journals in red (blue) publish reviews (proceedings), which tend to receive more (fewer) citations.

Summarizing with a soccer analogy, what Ball et al. (2021) claim is like claiming that “young players in weak teams actually score more than Cristiano Ronaldo” while actually computing the arcsinh of scored goals subtracting past goals and the team average. Figure R4 (analogous to Figure 2 in Ball et al., 2021) shows that their claim no longer holds if data manipulations are avoided. We adjusted for gender history following Eq. 2 in the main paper; this has a minor effect. Figure R4 confirms that the publishing system is doing a fair job.



**Figure R4.** Female/Male ratio of the average number of individual citations received per year by solo papers published in the indicated journals during the indicated time periods. Authors are weighted to compensate for gender history. We compare within periods of publication years, as citations accumulate with time.

## 5. BIBLIOMETRICS AS A PROXY FOR SCIENTIFIC QUALITY

Ball et al. (2021) think that physicists might be top-cited due to “biased citing, which involves a number of considerations including the ‘halo effect’, ... in-house citations, ... and the Matthew effect.” Those who view science primarily as a social hierarchy think that bibliometric indices cannot be a valid proxy for scientific quality because they are too much distorted by sociological factors. Indeed, other fields show that this can be a problem: For example, more citations are received by some research finding gender bias (Jussim, 2019; see also Clark & Winegard, 2020). But such a generic problem loses quantitative significance when applied to fundamental physics: a field far from politically sensitive topics and guided by objective data. Physics accepted relativity and quantum mechanics, despite their conflict with human “bias” about space, time, and realism. Sociological distortions happened in physics but remained local, while the field itself avoided major problems: No person, school, institution ever controlled physics. Occasional divergences between schools of thought were decided by experiments, as “the job of a scientist is to listen carefully listen to nature, not to tell nature how to behave” (Richard Feynman, <https://www.washingtonpost.com/archive/entertainment/books/2005/11/06/richard-feynman-plumbed-the-mysteries-of-life-and-physics-with-no-respect-whatsoever-for-authority/0a4dc009-6287-4f74-995a-96dc37480304/>). Bibliometric indices, being dominated by counts at global level, average out local social distortions and thereby provide a less biased proxy than local evaluations.

Furthermore, sociological effects (such as time available for research) that can produce mild differences in bibliometric output are relatively less important in physics, as this field exhibits the biggest bibliometric differences between top and average authors (as mentioned in footnote 20 of Strumia [2021]). Andersen et al. (2021) doubt that there is any relation between intelligence and scientific productivity: “Extant research on intelligence and scientific productivity is scarce, and does not suggest any direct relationship between the two.” Having read and understood many top-cited physics papers, I appreciate their nontrivial results achieved thanks to the brainpower of their authors, a key feature missed by sociological arguments focused on power.

## 6. UNBALANCED LITERATURE REVIEW IN THE INTRODUCTION?

Ball et al. (2021) claim that my literature review is biased. Balance can be an issue when touching currently controversial topics about which some authors have strong opinions. Indeed, I got interested in “STEM and gender” because a STEM institution hosted a workshop about this topic and, by chance, I had relevant bibliometric data. From the workshop I got the impression that my data were in disagreement with past literature. Only later did I become aware that the literature contains many similar results, which had been ignored in the workshop. In my view this imbalance is a more serious issue than the relative amount of space I gave to both points of view. Selective criticism of my balance seems to reflect the wider imbalance.

Indeed, in order to prove my supposed “lack of balance in citing,” Ball et al. (2021) list some papers that I have not cited. But most of those papers are subsequent to mine (submitted to *QSS* and to *arXiv* in 2018, blocked by *arXiv*, accepted by *QSS* in 2019). Furthermore, unlike what Ball et al. (2021) write, the papers they list do not provide evidence for biases. Indeed, let us go through the bibliometric studies.

- The preprint Dworkin, Linn et al. (2020) studies some journals in neurosciences, finding that male authors are cited slightly more than predicted by a naive citation model that ignores scientific quality and assumes random citations, giving more to older authors. Claiming that this kind of result implies gender bias lacks adequate foundation: Some papers are cited more simply because they contain more scientific results.

- Indeed, Fox and Paine (2019) find similar results studying citation and acceptance rates in ecology journals and correctly warn that “our data do not allow us to test hypotheses about mechanisms underlying the gender discrepancies we observed.”
- Similarly, Royal Society of Chemistry (2019) finds similar results studying chemistry journals and warns that “these results suggest that even when papers authored by women are published, their work is less cited. However, we cannot be sure whether this is due to a true gender bias.”

It's then unclear what the scientific basis is of some other statements in these papers. In view of this confusing situation, I read previous research critically: considering their data without stopping at sentences written by authors. This explains why Andersen et al. (2021) believe that I misinterpreted some previous research. Indeed they provide a table where they simply quote some sentences by authors of past research. To exemplify, I discuss the first few items of their table:

- Caplar, Tacchella, and Birrer (2017) is one more work (focused on astronomy) that finds that papers by female authors are less cited, even with respect to some naive citation model that tries to account for possible social factors. Once again, after discussing gender bias, the authors warn that “of course we cannot claim that we have actually measured gender bias.” This is why I ignored this part, and focused instead on the data, correctly reporting what the data in Figure 6 of Caplar et al. (2017) show.
- My supposed “biased reporting” about Milkman, Akinola, and Chugh (2015) is actually a correct description of the data in Figure 3 of that paper. Indeed, simple mathematics shows that female students received +3%, +4%, +3%, -2%, +8% more responses in public U.S. universities compared to male students in the same racial group (white, black, Hispanic, Chinese, Indian). The corresponding numbers in private universities (a smaller part of the sample) are -6%, -1%, +8%, -13%, +5%.
- Ball et al. (2021) criticize me for not having cited Witteman, Hendricks et al. (2019), while Andersen et al. (2021) criticize how I cited that paper: I mentioned explicitly one result, hinting only implicitly at a second result. Indeed, the second result was less relevant in the context of the citation in my introduction, and Witteman et al. (2019) warn that their second result “does not allow for estimation of the contribution of three possible sources—individual bias, systemic bias, or lower performance.” Once more it's the same issue.

Rather than providing more examples by going through each paper in the lists by Andersen et al. (2021) and Ball et al. (2021), let me draw the general lesson. Some literature seems to exhibit a bias for bias: The amount of evidence decreases when moving from newspaper reports to titles of actual research, to abstract, to text, to data.

## 7. SOCIOLOGY OR BIOLOGY

Many recent studies only consider gender as a social self-identity. In my opinion this is an unjustified limitation that leads these authors to restrict their attention to sociological interpretations, ignoring those biological differences that can arise given that sex is determined by chromosomes present in any cell. Indeed, since the pattern of biases expected on the basis of sociological interpretations did not show up in the data, my conclusions mentioned the possibility of interpreting the data at face value in terms of the combined effect of gender

difference in interests and higher male variance (HMV). In particular I noticed that the HMV suggested by the InSpire data could be interpreted, even at a quantitative level, as due to the HMV seen in biology. This is “an entirely unjustified conflation of correlation and causation” according to Ball et al. (2021), and “highly speculative explanations based on twisted assumptions and with little or no empirical basis” according to Andersen et al. (2021). Such commentaries even raise doubts about the existence of HMV and gender difference in interests. The empirical basis of this science has recently been summarized by Halpern, Benbow et al. (2007), Murray (2020), Pinker and Spelke (2005), Stewart-Williams and Halsey (2021), and Stevens and Haidt (2017), showing that the surrounding controversy is now mostly outside science. HMV is observed in a wide variety of physical and cognitive traits (Lehre, Laake, & Danbolt, 2009; Murray, 2020) and in many dimorphic species. Focusing on human traits that appear more relevant for the present discussion, HMV is seen in subcortical regions, in personality measures (e.g., extraversion, conscientiousness, agreeableness, openness) and in mental tests (e.g., PISA scores worldwide<sup>3</sup>).

Concerning the issue of causation, I did not actually discuss it in my paper. Various commentaries emphasize the difficulty of identifying causal relations in sociology. Indeed, some apparently causal findings in social sciences were later recognized to be correlations due to genetics (Boutwell, 2015). On the other hand, biological factors, by their very nature, tend to act causally. This leads to the question: What are the mechanisms behind the two factors of interest? Gender differences in interests seem significantly shaped by prenatal hormones (Berenbaum & Beltz, 2016), and stable in time and cross-culturally (e.g., Stoet & Geary, 2020; Murray, 2020). The origin of HMV is not yet established and plausible interpretations of biological nature have been proposed (Del Giudice, Barrett et al., 2018; Murray, 2020; Reinhold & Engqvist, 2013; Wyman & Rowe, 2014).

In my opinion, the controversy surrounding such topics arises because a constructivist attitude in some corners of present sociology and anthropology tends to disregard the role that basic facts of human nature have for social interaction and postulates that everything is totally shaped by the symbols and meanings that people come to develop in society. For example, the Standard Social Science Model relies on the Blank Slate paradigm (see Pinker (2002) for a critical introduction; see also Buss and von Hippel (2018)). While physics relies on mathematics, chemistry on physics, and biology on chemistry, this part of current sociology refuses to rely on its natural root, biology (Boutwell, 2017; Murray, 2020).

In conclusion, having mentioned gender differences in interests and HMV does not make my results wrong. On the contrary, it is scientifically dubious to reject *a priori* such notions corroborated by evidence.

## 8. POLITICAL VALUES

Ball et al. (2021) contains heated criticisms about my paper and others who find related results. For example, according to their paper, “Stoet and Geary’s arguments have been undermined significantly by the many deficiencies in their data analysis,” while a more accurate statement would be “Stoet and Geary (2019) clarified a point.” More precisely, the main point clarified in Stoet and Geary (2019) is that Stoet and Geary (2018) had correctly considered

<sup>3</sup> As reported in Appendix 3 of Murray (2020), distributions of recent PISA math scores show a male-to-female variance ratio equal to 1.14 in Western Europe, the Anglosphere and Scandinavia; 1.12 in Eastern Europe and Latin America; 1.10 in SouthEast Asia; 1.18 in East Asia; 1.16 in Mideast/North Africa. HMV is similarly found in PISA reading scores, where girls outscore boys.

(a function of) gender ratios in STEM relative to gender ratios in the graduate population, rather than in the whole population.

Similarly, Ball et al. (2021) claim that my paper is “merely a flawed, biased, and ideologically motivated analysis. It is also likely to be actively harmful to the progress of women in physics, to the detriment not only of many individuals but of our entire community.” Apart from their understatement style, the problem with Ball et al. (2021) is that their approach is closer to activism than to science: They try to discredit a scientific result not by logic or evidence but by rhetorically attacking its supposed implications. When they go beyond criticizing and try doing science, their alternative reanalysis contains so many problems that it becomes an alternative reality (see Section 4). When they propose that a missed “confounding factor is vocal criticism of women within academia by individuals such as Strumia,” they just confound scientific arguments with insults. Similarly, they claim with no basis that my analysis is “very far from neutral or disinterested.” My paper contained a statement about no competing interests. To remove any doubt, I expand on it: I have never been affiliated to any party, political association, or even academy; I got interested in the topic only accidentally; my research is not financed by anybody (I avoid using my affiliation); and as expected only trouble is courted by presenting data that challenge the dominant political narrative in some academies<sup>4</sup>.

These situations happen when scientific results cast doubts on beliefs that somebody holds as sacred. Such conflicts happen because science is a method for seeking truth. Science emerged after a historical period with divisive moral issues by finding a common ground in empirical data and objectivity. This allowed scientists to agree on facts, at the price of making science an equal opportunity offender (Clark & Winegard, 2020). Centuries ago, science cast doubt on sacred religious beliefs, and the Church took up indefensible positions. Something similar is happening now: Research about human differences challenges the beliefs of major political orientations. Following the commentaries, I discuss the left wing of the political spectrum, where the desire to see all groups thrive equally has become an apodictic belief in absolute equality among groups. By denying any difference, one gets caught in a bind: Tribalism reinforces the conflict (Clark & Winegard, 2020), giving a stronger position within their group to those guardians of their sacred values who try to discredit scientific progress.

Given this present context, the disciplines that study bias in science risk being significantly affected by the very same bias they study. According to Clark and Winegard (2020), “when the majority of scientists in a discipline share the same sacred values, then the checks and balances of peer review and peer skepticism that science relies upon can fail.” The risk that “social science will become another form of covert political activism” (Clark & Winegard, 2020) has been highlighted by gender journals that recently accepted hoaxes for publication (Pluckrose, Lindsay, & Boghossian, 2018), while papers with “controversial” findings get unpublished for scientifically unclear reasons (for recent gender-related examples see AlShebli, Makovi, and Rahwan [2020], Hill [2017], and Hudlicky [2020]). Interpreting any difference as bias leads to wrongly painting physics, other fields with similar representation gaps (e.g., Reges, 2018; Sesardic & De Clerq, 2014), and academia itself as sexist, discriminatory, and hostile. This view, especially when promoted instrumentally, may lead some female researchers to wrongly fear a hostile environment. This is more harmful for progress of women in physics than my bibliometric analysis.

---

<sup>4</sup> The commentary by Thelwall (2021) mentions the web site “Particles for Justice.” Instead of writing a scientific commentary, these authors stopped at personal attacks against me based on false statements, as clarified in my site [alessandrostrumia.home.blog](http://alessandrostrumia.home.blog).

## REFERENCES

- AlShebli, B., Makovi, K., & Rahwan, T. (2020). The association between early career informal mentorship in academic collaborations and junior author performance. *Nature Communications*, *11*, 5855. See *Retraction Watch*, December 21, 2020: *Nature Communications* retracts much-criticized paper on mentorship. <https://retractionwatch.com/2020/12/21/nature-communications-retracts-much-criticized-paper-on-mentorship/>. DOI: <https://doi.org/10.1038/s41467-020-19723-8>, PMID: 33203848, PMCID: PMC7672107
- 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* (this issue).
- Ball, P., Britton, T. B., Hengel, E., Moriarty, P., Oliver, R. A., ... Wade, J. (2021). Gender issues in fundamental physics: Strumia's bibliometric analysis fails to account for key confounders and confuses correlation with causation. *Quantitative Science Studies* (this issue).
- Berenbaum, S. A., & Beltz, A. M. (2016). How early hormones shape gender development. *Current Opinion in Behavioral Sciences*, *7*, 53–60. DOI: <https://doi.org/10.1016/j.cobeha.2015.11.011>, PMID: 26688827, PMCID: PMC4681519
- Boutwell, B. (2015). Why parenting may not matter and why most social science research is probably wrong. *Quillette*, December 1. <https://quillette.com/2015/12/01/why-parenting-may-not-matter-and-why-most-social-science-research-is-probably-wrong/>.
- Boutwell, B. (2017). Sociology's stagnation. *Quillette*, March 5. <https://quillette.com/2017/03/05/sociologys-stagnation/>.
- Buss, D. M., & von Hippel, W. (2018). Psychological barriers to evolutionary psychology: Ideological bias and coalitional adaptations. *Archives of Scientific Psychology*, *6*(1), 148–158. DOI: <https://doi.org/10.1037/arc0000049>
- Caplar, N., Tacchella, S., & Birrer, S. (2017). Quantitative evaluation of gender bias in astronomical publications from citation counts. *Nature Astronomy*, *1*, 0141. DOI: <https://doi.org/10.1038/s41550-017-0141>
- Clark, C. J., & Winegard, B. M. (2020). Tribalism in war and peace: The nature and evolution of ideological epistemology and its significance for modern social science. *Psychological Inquiry*, *31*(1), 1–22. DOI: <https://doi.org/10.1080/1047840X.2020.1721233>
- Del Giudice, M., Barrett, E. S., Belsky, J., Hartman, S., Marteld, M. M., ... Kuzawaf, C. W. (2018). Individual differences in developmental plasticity: A role for early androgens? *Psychoneuroendocrinology*, *90*, 165–173. DOI: <https://doi.org/10.1016/j.psyneuen.2018.02.025>, PMID: 29500952, PMCID: PMC5864561
- Dworkin, J. D., Linn, K. A., Teich, E. G., Zurn, P., Shinohara, R. T., & Bassett, D. S. (2020). The extent and drivers of gender imbalance in neuroscience reference lists. *Nature Neuroscience*, *23*, 918–926. DOI: <https://doi.org/10.1038/s41593-020-0658-y>, PMID: 32561883
- Fox, C. W., & Paine, C. E. T. (2019). Gender differences in peer review outcomes and manuscript impact at six journals of ecology and evolution. *Ecology and Evolution*, *9*(6), 3599–3619. DOI: <https://doi.org/10.1002/ece3.4993>, PMID: 30962913, PMCID: PMC6434606
- Halpern, D. F., Benbow, C. P., Geary, D. C., Gur, R. C., Hyde, J. S., & Gernsbacher, M. A. (2007). The science of sex differences in science and mathematics. *Psychological Science in the Public Interest*, *8*(1), 1–51. DOI: <https://doi.org/10.1111/j.1529-1006.2007.00032.x>, PMID: 25530726, PMCID: PMC4270278
- Hill, T. P. (2017). An evolutionary theory for the variability hypothesis. *arXiv*, arXiv:1703.04184. See *Retraction Watch*, January 18, 2019: What really happened when two mathematicians tried to publish a paper on gender differences? The tale of the emails. <https://retractionwatch.com/2018/09/17/what-really-happened-when-two-mathematicians-tried-to-publish-a-paper-on-gender-differences-the-tale-of-the-emails/>.
- Hossenfelder, S. (2021). Analyzing data is one thing, interpreting it another. *Quantitative Science Studies* (this issue).
- Hudlicky, T. (2020). 'Organic synthesis—Where now?' is thirty years old. A reflection on the current state of affair. *Angewandte Chemie International Edition*, *59*(31), 12576. See *Retraction Watch*, June 5, 2020: Following outrage, chemistry journal makes a paper decrying diversity efforts disappear. <https://retractionwatch.com/2020/06/05/following-outrage-chemistry-journal-makes-a-paper-decrying-diversity-efforts-disappear/>. DOI: <https://doi.org/10.1002/anie.202006167>, PMID: 32497328
- Jussim, L. (2019). Scientific bias in favor of studies finding gender bias. *Psychology Today*, June 23, and references therein. <https://www.psychologytoday.com/us/blog/rabble-rouser/201906/scientific-bias-in-favor-studies-finding-gender-bias>.
- Lehre, A. C., Laake, P., & Danbolt, N. C. (2009). Greater intrasex phenotype variability in males than in females is a fundamental aspect of the gender differences in humans. *Developmental Psychobiology*, *51*(2), 198–206. DOI: <https://doi.org/10.1002/dev.20358>, PMID: 19031491
- Madison, G., & Fahlman, P. (2020). Sex differences in the number of scientific publications and citations when attaining the rank of professor in Sweden. *Studies in Higher Education*. DOI: <https://doi.org/10.1080/03075079.2020.1723533>
- Milkman, K. L., Akinola, M., & Chugh, D. (2015). What happens before? A field experiment exploring how pay and representation differentially shape bias on the pathway into organizations. *Journal of Applied Psychology*, *100*(6), 1678–1712. DOI: <https://doi.org/10.1037/apl0000022>, PMID: 25867167
- Murray, C. (2020). *Human diversity. The biology of gender, race and class*. New York, Boston: Twelve.
- Pinker, S. (2002). *The blank slate, The modern denial of human nature*. London: Penguin.
- Pinker, S., & Spelke, E. (2005). *The science of gender and science*. Debate at Harvard University, April 22. <https://stevenpinker.com/publications/science-gender-and-science-conversation-stein-pinker-and-elizabeth-spelke>.
- Pluckrose, H., Lindsay, J. A., & Boghossian, P. (2018). Academic grievance studies and the corruption of scholarship. *Areo*, October 2. <https://areomagazine.com/2018/10/02/academic-grievance-studies-and-the-corruption-of-scholarship/>.
- Reges, S. (2018). Why women don't code. *Quillette*, June 19. <https://quillette.com/2018/06/19/why-women-dont-code/>.
- Reinhold, K., & Engqvist, L. (2013). The variability is in the sex chromosomes. *Evolution*, *67*(12), 3662–3668. DOI: <https://doi.org/10.1111/evo.12224>, PMID: 24299417
- Royal Society of Chemistry. (2019). *Is publishing in the chemical sciences gender biased?* Royal Society of Chemistry. <https://www.rsc.org/globalassets/04-campaigning-outreach/campaigning/gender-bias/gender-bias-report-final.pdf>.
- Sesardic, N., & De Clerq, R. (2014). Women in philosophy: Problems with the discrimination hypothesis. *Academic Questions*, *27*, 461–473. DOI: <https://doi.org/10.1007/s12129-014-9464-x>
- Stevens, S., & Haidt, J. (2017). The Google memo: what does the research say about gender differences? *Heterodox: The Blog*, August 10.

- <https://heterodoxacademy.org/the-google-memo-what-does-the-research-say-about-gender-differences/> (accessed March 18, 2020).
- Stewart-Williams, S. & Halsey, L. G. (2021). Men, women and STEM: Why the differences and what should be done? *European Journal of Personality*, 35(3). **DOI:** <https://doi.org/10.1177/0890207020962326>
- Stoet, G., & Geary, D. C. (2018). The gender-equality paradox in science, technology, engineering, and mathematics education. *Psychological Science*, 29(4), 581–593. **DOI:** <https://doi.org/10.1177/0956797617741719>, **PMID:** 29442575
- Stoet, G. & Geary, D. C. (2019). Corrigendum: The gender-equality paradox in science, technology, engineering, and mathematics education. *Psychological Science*, 31(1), 110–111. **DOI:** <https://doi.org/10.1177/0956797619892892>, **PMID:** 31809229
- Stoet, G., & Geary, D. C. (2020). Sex-specific academic ability and attitude patterns in students across developed countries. *Intelligence*, 81, 101453. **DOI:** <https://doi.org/10.1016/j.intell.2020.101453>
- Strumia, A. (2021). Gender issues in fundamental physics: A bibliometric analysis. *Quantitative Science Studies* (this issue).
- Strumia, A., & Torre, R. (2019). Biblioranking fundamental physics. *Journal of Informetrics*, 13(2), 515–539. **DOI:** <https://doi.org/10.1016/j.joi.2019.01.011>
- Thelwall, M. (2021). Female contributions to high-energy physics in a wider context: Commentary on an article by Strumia. *Quantitative Science Studies* (this issue).
- Witteman, H. O., Hendricks, M., Straus, S., & Tannenbaum, C. (2019). Female grant applicants are equally successful when peer reviewers assess the science, but not when they assess the scientist. *The Lancet*, 393(10171), 531–540. **DOI:** [https://doi.org/10.1016/S0140-6736\(18\)32611-4](https://doi.org/10.1016/S0140-6736(18)32611-4)
- Wyman, M. J., & Rowe, L. (2014). Male bias in distributions of additive genetic, residual, and phenotypic variances of shared traits. *American Naturalist*, 184(3), 326–337. **DOI:** <https://doi.org/10.1086/677310>, **PMID:** 25141142