Check for updates

The MIT Press

RESEARCH ARTICLE

Gender issues in fundamental physics:
A bibliometric analysis

Alessandro Strumia iD

Physicist, Fauglia (Italy)

ABSTRACT

I analyze bibliometric data about fundamental physics worldwide from 1970 to now, extracting quantitative data about gender issues. I do not find significant gender differences in hiring rates, hiring timing, career gaps and slowdowns, abandonment rates, citation, and self-citation patterns. Furthermore, various bibliometric indicators (number of fractionally counted papers, citations, etc.) exhibit a productivity gap at hiring moments, at career level, and without integrating over careers. The gap persists after accounting for confounding factors and manifests as an increasing fraction of male authors going from average to top authors in terms of bibliometric indices, with a quantitative shape that can be fitted by higher male variability.

1. INTRODUCTION

This paper originates from an observational opportunity: For the first time sociological issues in fundamental physics can be studied using the public InSpire database, which has accumulated bibliometric data about fundamental physics worldwide from around 1970 to now (InSpire, 2010). Fundamental physics is a subdiscipline of physics that deals with the fundamental aspects of the field and that presently focuses mostly on particle physics, cosmology, and astrophysics, from an experimental and theoretical point of view.

Such bibliometric data are being used to study various aspects of the field. Like other science, technology, engineering, and mathematics (STEM) fields, physics exhibits persistent gender differences, which I try to characterize and understand in the present paper. The bibliometric approach relies on large amounts of objective quantitative data about papers, authors, citations, and hires. Having a large amount of new data I will follow a data-driven approach. Enough statistics is sometimes needed to reveal effects, and to go beyond simple counting by devising dedicated analyses that target specific questions. I can do this, as I have the full database, not just access to some predefined metrics.

While a vast literature has studied gender differences in STEM, no previous studies have specifically focused on fundamental physics: The present study will fill this gap.[1] A main theme is

---

[1] 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). The NASEM Report (2018) focused on the University of Texas System, finding that 17% (13%) of female (male) students in science reported "sexist hostility." However, the NASEM Report (2018) also found that a higher 45% rate of "sexist hostility" is reported by female students in medicine, a field with a negligible gender gap in participation. Only a few percent of male and female STEM students reported more serious problems (NASEM Report 2018, Figure 3.3). As "evidence of direct discrimination is limited," alternative interpretations of the gender representation difference in STEM have been considered and "many scholars now emphasize the role of gender differences in preferences, self-concept and attitudes" (Breda & Napp, 2019).

understanding why women remain underrepresented in STEM fields, a worldwide phenomenon that has persisted for decades, despite interventions on its alleged social causes (Stoet & Geary, 2018). A limitation of bibliometric analyses is that authors start being scientifically active roughly at PhD level: In physics (as in other STEM fields) low female representation is already present at this entry level of bibliometric data. Earlier phases need to be explored with other tools.

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., 2019). Gender differences in relative attitudes (girls with high mathematical ability tend to also have high verbal ability) also contribute to student choices (Ceci et al., 2014; Stoet & Geary, 2018; Wang, Eccles, & Kenney, 2013): Most of the gender gap in student intentions to study math disappears after taking into account their mathematics versus reading difference in PISA scores, while absolute results (boys outperform girls in mathematics, girls outperform boys in reading) are much less able to explain the gender gap (Breda & Napp, 2019).

Coming to the later phase that can be studied by experiments and by bibliometrics, initial small-scale experiments and anecdotal reports suggested biases against hypothetical female applicants (see, for example, Moss-Racusin, Dovidio et al. [2012] and Wennerås and Wold [1997]); see also Eaton, Saunders et al. [2019]. These findings have not been supported by more recent larger-scale experiments (see the review in Ceci and Williams [2011] and also Ceci and Williams [2015] and Williams and Ceci [2017]). Milkman, Akinola, and Chugh (2015) sent letters by fictional students seeking research opportunities to professors and measured their response rate. The result of this social experiment performed in the United States was that, looking 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.

Experiments about hypothetical applicants often miss key elements of real applications, mostly based on scientific results reported in publications and evaluated by scientists who work in the same subfield. No significant biases have been found in examined real grant evaluations (Ceci et al., 2014; Ley & Hamilton, 2008; Marsh, Jayasinghe, & Bond, 2011; Mutz, Bornmann, & Daniel, 2012) and referee reports of journals (Borsuk, Aarssen et al., 2009; Ceci et al., 2014; Edwards, Schroeder, & Dugdale, 2018); the gender composition of applicants (Way, Larremore, & Clauset, 2016) and panels (Abramo, D'Angelo, & Rosati, 2015) has little effect. Real hires show a higher success rate among women (Ceci et al., 2014; Glass & Minnotte, 2010; National Research Council, 2009; Wolfinger, Mason, & Goulden, 2008), especially in those STEM fields where women are less represented (Ceci et al., 2014). Bibliometric attempts to recognize higher merit (Ceci et al., 2014) found that male faculty members write more papers (Abramo, D'Angelo, & Caprasecca, 2009a; Fox, 2005; Larivière, Ni et al., 2013; Levin & Stephan, 1998; Way, Larremore, & Clauset, 2016; Xie & Shauman, 1998) (see also Holman, Stuart-Fox et al. [2018] and Thelwall [2018]), predominate among first and last authors (prestigious in some fields) and in single-authored papers (Jagsi, Guancial et al., 2006; West, Jacquet et al., 2013). Such gender productivity gap persists after accounting for confounding factors such as seniority (Caplar, Tacchella, & Birrer, 2017; Ceci et al., 2014; Moldwin & Liemohn, 2018). Consistent with these results, Witteman, Hendricks et al. (2019) found that female grant applications in Canada are less successful when evaluations involve career-level elements. Some studies observed a small group of extremely productive, mostly male, "star authors" (Abramo, D'Angelo, & Caprasecca, 2009b; Abramo, Cicero, & D'Angelo, 2015; Bordons, Morillo et al., 2003). A smaller "leaky pipeline" rate of female authors is observed in STEM fields than in other

fields with higher female representation (Ceci et al., 2014). Looking at PLOS medical journals, where each author declares his or her role (analysis, design, material, perform, write), Macaluso, Larivière et al. (2016) found that women were more involved in performing experiments, and men more involved in the other roles.

This paper is structured as follows.

In Section 2 I describe how I identify the gender of authors; how I obtain lists of hires; how I combine citations to define bibliometric indicators that can be used as reliable proxies for scientific merit, being significantly correlated to human evaluations such as scientific prizes; and how I deal with the confounding factor due to the gender evolution of the field.

In Section 3 I present findings that exhibit interesting gender differences. New authors today appear with roughly 4:1 male:female proportion, with order one variations in different countries. I find that this entry difference in representation is negligibly affected by hiring, consistent with Ceci et al. (2014). As I use citations, in Section 3.1 I first verify that male (M) and female (F) authors cite in the same way. I achieve this by defining a gender asymmetry in citations sensitive only to a differential gender bias, not to gender differences in number or productivity of authors. In Section 3.2 I then compare reliable bibliometric indices based on citations finding that F authors are hired with indices that are, on average, not higher than those of M authors. Section 3.3 finds a productivity gap consistent with previous studies. This new difference is quantitatively studied in Section 3.4 finding that the M fraction progressively grows going from average to top authors, consistent with Abramo et al. (2009b, 2015), and Bordons et al. (2003). In Section 3.5 I consider self-references, finding no new gender differences. Data are made available as discussed in the "Data availability" section. As bibliometric data are influenced by a complicated background of social and historical accidents in the supplementary Section S1 I show that the results above persist after taking confounding variables into account. Statistical details are presented in the supplementary Section S2. Interpretations of the data are discussed in Section 4.

2. METHODS

The public InSpire database (InSpire, 2010) maintained by CERN and other institutions offers a picture of fundamental physics worldwide from around 1970 to now. InSpire gives data on about 1.3 million scientific papers, 30 million references, 71,104 identified authors in over 7,000 institutes and over 6,000 collaborations. InSpire individually identified all authors (except occasional authors), solving the problem of name disambiguation. I expect that the database is negligibly affected by direct or indirect gender bias. Indeed, the database provides essentially full coverage of the scientific literature within "fundamental physics." For decades this has been a self-contained, highly specialized subject (Sinatra, Deville et al., 2005), so that the "boundaries" of the database play a minor role. Papers in InSpire include all those published in some categories of the preprint bulletin *arXiv* and in some journals with topics considered relevant for fundamental physics. Human intervention is minor, such as adding extra papers considered relevant. Only a minor fraction of authors occasionally work in other fields or arrive from other fields. A possible gender difference in multidisciplinary attitudes is thereby expected to have negligible impact on the subsequent discussion. On the other hand, various authors work on multiple topics within the field, which cannot be sharply subdivided.

InSpire does not provide gender information: In Section 2.1 I describe the procedure to infer gender from full names and nationality of the authors. In Section 2.2 I describe how I obtain lists of hires in fundamental physics worldwide. Section 2.3 motivates the bibliometric index that I will use to indicate scientific merit.

2.1. Name–Gender Association

I need to infer gender from names in an accurate and complete way.[2] Three main problems are encountered. First, the InSpire database provides only name initials for about 13% of the authors. These are mostly authors with little impact, as defined by any index. Second, some names such as Nicola are "ambiguous": They correspond to different genders in different countries. Third, some authors have unusual names. The `Mathematica` machine learning function `Classify` (Wolfram Mathematica, n.d.) uses information about the first name only and leaves about 40% of authors with unclassified gender.

I tested two approaches to determine their strengths and to choose the best combination:

1. First, I run the online Ethnea (Torvik & Agarwal, 2016) tool, which uses the full name (first and family name) to infer both gender and ethnicity. Ethnea leaves 26% of the authors with unclassified gender.

2. Second, for each author I extract a "guessed" nationality from the earlier affiliations in his or her papers and use it to disambiguate "ambiguous" names. The obtained list of first names and nationalities is matched to a database of names and countries from the Worldwide Gender-Name Dictionary (WGND) (Raffo, 2016). This database contains 175,917 names with their associated countries. About 70% of authors have "unambiguous" names that are present in the WGND. Authors with "ambiguous" names present in the WGND are matched using the nationality inferred from their earliest affiliations. The size of this subset of authors is approximately 3% of the total, and the uncertainty induced by this procedure is below the per cent level. About 0.1% of the authors have "ambiguous" names and no nationality information: I match them to the most common gender corresponding to their name, defined as the one used in the largest number of countries. Twenty-three percent of the authors remain unclassified.

The results discussed in the following are affected in a minor way using the Ethnea or WGND classifications. By comparing them it can be seen that the Ethnea classification is less complete, leaving unclassified more authors with unusual names. On the other hand, the WGND classification leads to some authors with misidentified gender, typically arising due to a misidentification of their nationality. Different genders are found for 1.8% of all identified authors; the percentage rises to 5% among Chinese,[3] Indian and Korean authors, and falls to 1% among European authors.

As a best choice, I adopt the Ethnea classification whenever available, and the WGND classification otherwise. Furthermore, I selected 1,000 of the top-cited authors in different time periods and systematically verified and correctly assigned their gender with no errors, using information available on the internet.

2.2. Hiring

InSpire is integrated with HepNames, a database with biographical information about the various authors, including papers, affiliation history, experiments they participated in, PhD advisor, and graduate students. As an example, the internet page http://inspirehep.net/author/profile/A

---

[2] Some U.S. astronomers "discourage" adopting this "quantitative methodology," seen as "epistemically violent" and "discriminatory" (Rasmussen, Maier et al., 2019).

[3] Gender can be reliably extracted from Chinese names only when they are written in Chinese characters: This information is not always provided by InSpire.

.Strumia.1 shows the profile of the present author. A user interface allows researchers to create and update HepNames records for themselves and for other authors, providing precise career information on a voluntary basis (to be validated by the InSpire team). Furthermore, large collaborations systematically provide complete author information upon submission of documents through a dedicated format.

From HepNames I obtained a database of about 10,000 first hires in fundamental physics worldwide, including dates and disambiguated institutions. Such InSpire hires might be biased if those F and M authors who need to self-report their data tend to do this differently. While InSpire is a widely used tool in the community, integrated with a job announcement system, funded by multiple official institutions and endorsed in *Reviews of Particle Physics*, occasional authors might not use it.

I therefore complement InSpire hires by computing unbiased "pseudohires," defined as follows. For each paper, there is a list of disambiguated affiliations of each author. I consider an author as $p$yr-hired when he or she starts writing papers with the same affiliation for at least p years. Using this definition I obtain a database of about 40,000/19,000 5/10yr-first hires from 1960 to 2013/2008 (64,000/23,000, including multiple hires for the same author). However, in this way it is not possible to obtain a precise hiring date for the subset of authors hired by the same institution to which they were previously affiliated.

Therefore, InSpire hires will be used when a precise hiring date is more important than increased statistics, and pseudohires in the opposite situation, when full coverage is more important than precise timing. In each case, the other sample will be used as a control sample.

2.3. Bibliometrics

I here motivate the use of appropriate bibliometric indicators as a proxy for what is commonly considered as scientific merit.

Various authors have studied what citation counts do measure.

At theoretical level, two main models have been proposed. According to the normative interpretation, scientists primarily cite to give credit. A bibliometric index then provides a valid proxy of scientific merit, especially when highly correlated with scientific prizes or other human evaluations of scientific merit. According to the social-constructivist interpretation, citations are instead primarily a social persuasion tool; the concept of scientific merit itself is questioned: as reviewed in Bornmann and Daniel (2008) "scientific knowledge is socially constructed through the manipulation of political and financial resources and the use of rhetorical devices." According to this point of view, citation counts could be correlated to prizes simply because both reflect social status. Some prizes in physics require established scientific results; others are awarded following rules that leave more space for sociological distortions.

Observational works supported the normative interpretation at high aggregation level (Bornmann & Daniel, 2008; Tahamtan & Bornmann, 2019); personal factors important at individual level average out when considering many authors. My data provides extra evidence in this direction: For example, authors of top-cited papers tend to be younger, rather than powerful senior scientists.

Citation counts are surely influenced by some confounding factors (Bornmann & Daniel, 2008): the citation intensity depends on time and field (my indicator will compensate for this), on language (essentially all physics literature is in English), on accessibility (physics literature has been freely available on the preprint bulletin *arXiv* since 1995), and on collaboration size.

Collaboration size is a big issue in fundamental physics, due to the presence of very large (up to 3,000 authors) and very productive (up to 6,000 papers) collaborations, mostly in high-energy experimental physics. Because of this main reason, traditional metrics (such as citation counts, *h*-index, paper counts) now fail to provide reasonable proxies for scientific merit in fundamental physics (Strumia & Torre, 2019). Signing more papers that one can read stretches the concept of authorship (Birnholtz, 2006). At a quantitative level, the problem is that the contribution of one big collaboration overwhelms the database (1.3 million papers) if 6,000 papers are counted as $3,000 \times 6,000 = 1.8$ million.

This situation can be corrected by "fractional counting" (Hooydonk, 1997; Leydesdorff & Park, 2016; Perianes-Rodriguez, Waltman, & van Eck, 2016): A fraction $1/N_{aut}$ of each paper (rather than the full paper) is equally attributed to its N_{aut} authors, as appropriate for an intensive quantity. All authors, including first and last authors, are treated on equal footing because authors are usually sorted alphabetically in fundamental physics, unlike in other fields.[4] Thereby there is no way of telling who contributed what to multiauthored papers. When huge collaborations are involved there is no warranty that each author contributed to each paper. Despite this, the data show that the total fractionally counted bibliometric output of collaborations scales, on average, as their number of authors (Rossi, Strumia, & Torre, 2019), suggesting that large collaborations form when scientifically needed and that gift authorship does not play a large role.

Fractional counting of citations already provides one simple acceptable indicator. I improve on it by using the closely related number of "individual citations"

$$N_{icit} = N_{cit}/N_{aut}N_{ref} \tag{1}$$

(summed over all citing papers, as precisely defined in Strumia and Torre (2019), which gives reduced weight to citations coming from papers with a larger number N_{ref} of references. This refinement addresses the issue of normalization between different fields and times (Zitt and Small (2008); see Waltman (2016) for a review and extra references): Papers in sectors with a higher rate of publication (such as phenomenology in fundamental physics) tend to receive more citations; for the same reason these papers also tend to have more references. Thereby, dividing by the number of references tends to give a common normalization to different fields, without needing a field classification system. Indeed, the average number of individual citations received by papers in any field disconnected from other fields is 1. As a test that this concept works in practice in the InSpire database, I computed the average number of citations, of references, and of individual citations of papers within the main theoretical fields defined by *arXiv* (hep-th, hep-ph, gr-qc, nucl-th, astro-ph, hep-lat[5]), finding that the dispersion in N_{icit} among different fields is reduced to 8%, more than twice as small as the dispersion in N_{cit} or N_{ref}. Similar results are found considering different times: The field grows with time, such that newer papers receive more citations and have more references, in roughly proportional amounts.

Individual citations have the following meaning: An author who wrote N_{icit} fractionally counted papers of average impact in his or her field received N_{icit} individual citations. Table 7 of Strumia and Torre (2019) lists the 50 physicists who received the most individual citations, together with their scientific prizes. Physicists can read their names and consider whether N_{icit} is dominantly influenced by scientific achievements or by social constructivism. For practical

---

[4] The first author is not alphabetically sorted in 6% of the multiauthored papers in the hep-th *arXiv* bulletin, 13% in hep-ph and hep-ex, 18% in hep-lat, 25% in gr-qc, and 44% in astro-ph. In more papers the author highlighted as first might accidentally be also alphabetically first.

[5] Experimental papers form a separate category, as they tend to have many coauthors and to receive more citations.

purposes, an indicator provides an acceptable proxy of scientific merit if scientific merit positively affects the index more than confounding variables. A full or large correlation with scientific merit improves the sensitivity of the analysis, but some effects can be large enough that fine sensitivity is not needed to reveal them.

The use of bibliometric indices based on citation counts as a proxy for scientific merit comes with limitations and dangers. Some citations are given for negative reasons. On short timescales citations are more influenced by visibility, and some authors engage in boosting their citation counts in various ways: large collaborations, many references, self-references, citation networks, salami slicing into minimum publishable units, etc. Individual citations are not boosted by the first two strategies. As this paper is concerned with gender differences, it is reassuring that Section 3.5 will find no extra significant gender differences in self-referencing.

As "when a measure becomes a target, it ceases to be a good measure," in the supplementary Section S1 I consider a metric based on citations more different from common targets, which is not enhanced by the latter three strategies. The CitationCoin $\cancel{C}$ is defined as the difference between the number of received and given individual citations (up to a correction factor that prevents systematically negative contributions from recent papers), such that it is not affected by self-citations or by networks of circular citations (Strumia & Torre, 2019). Authors who write too many poorly cited papers can even have a negative $\cancel{C}$ score.

Bibliometric indicators measure the average opinion of the community: While all opinions can be wrong, a better possibility could be to rely on the opinion of top authors. This is done by metrics based on the PageRank algorithm (such as those discussed in Chen, Xie et al. (2007), Ma, Guan, and Zhao (2008), Pinski and Narin (1976), Radicchi, Fortunato et al. (2009), Strumia and Torre (2019), and West, Jensen et al. (2014)). This is studied in the supplementary Section S1, where for completeness I also consider the widely used but naive bibliometric indicators based on paper counting and on the average number of citations per paper.

In practice, the differences in bibliometric indices among authors are so large that log-scale plots will be appropriate and refined metrics only make minor differences. Individual citations N_{icit} are used because this metric is simpler and closer to the commonly used number of citations N_{cit}, while allowing us to meaningfully deal with experimentalists, theorists, and astrophysicists by compensating for the vastly different typical number of co-authors N_{aut} of papers produced by these communities.

2.4. The Age Confounder

The fraction of female authors in fundamental physics has significantly increased with time, producing demographic gender differences (female authors are on average younger than male authors) that act as a confounding factor to my later analyses. Apparent gender differences can just be age differences. Career-integrated indices tend to favor senior authors, and indices based on single papers tend to favor younger authors. As age is a significant confounder, I will compensate for the different time evolution $N_{F,M}^{\mathrm{start}}(t)$ (number of F and M authors that produced their first paper during year t) by assigning to each author A a weight proportional to

$$\frac{N_F^{\mathrm{start}}(t_A) + N_M^{\mathrm{start}}(t_A)}{2N_G^{\mathrm{start}}(t_A)}, \tag{2}$$

where t_A is the date of his or her first paper and G is his or her gender. This is equivalent to selecting every year a random subset of new authors (respecting the time evolution of the total number of authors) such that M and F authors are numerically equal, and averaging over the possible choices.

3. RESULTS

Among the 71,104 authors in fundamental physics listed in the InSpire database, 49,860 male and 9,205 female authors were identified. 16% of authors with identified gender are classified as female, and wrote 10% of the fractionally counted papers receiving 7% of the individual citations. These raw numbers, meant only to give a first rough idea of the field, are affected by a variety of historical accidents.

As documented in Strumia and Torre (2019), the field has expanded significantly: Due to increased publication intensity, about half of citations have been given after 2000, so that metrics based on citations favor recent authors (the metric N_{icit} automatically compensates for publication intensity). Furthermore, the F percentage grew with time, as shown by the raw data in the left panel of Figure 1.

The right panel shows that, within the countries that most contributed to fundamental physics, the female fractions range between 7% and 23%. It is interesting to explore if the female fraction is correlated with the Global Gender Gap Index (GGGI) of the countries (World Economic Forum, 2016), which measures the gap between women and men in education, politics, health, and economy, as this is a possible cause of the low female representation. The GGGI ranges between 0 and 1, with 1 indicating parity or a gap in favor of women (as the GGGI ignores imbalances to the advantage of women). The right panel of Figure 1 shows that the female fraction is not positively correlated with the GGGI, as similarly observed among students in STEM (Stoet & Geary, 2018).

Figure 2 shows that the female percentage is a factor of 2 higher in subfields dominated by large experimental collaborations than in theoretical fields.

Clearly, the field and its gender composition have evolved in the past 50 years. While describing such changes from a bibliometric point of view is an interesting subject, I try to focus on the general features that emerge from the complicated background of social factors. This will need to take into account possible confounding variables, by studying subperiods and subtopics or by trying to compensate for the above variations.
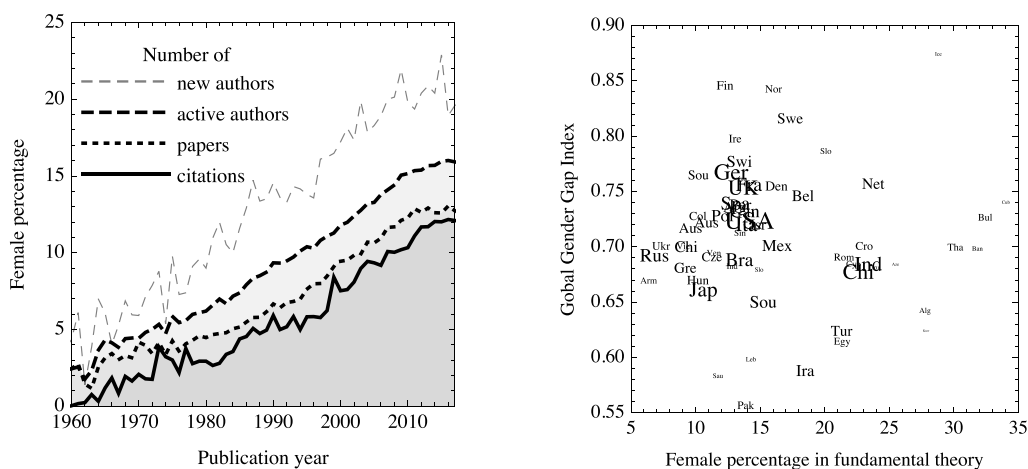

Figure 1. Left panel: Percentage female contribution to the number of new authors, of authors that wrote at least a paper during the year, to the number of fractionally counted papers, and to the number of received individual citations, considering the papers written each year. Citations are counted based on the year of the cited publication. Right panel: The percentage of *F* authors in fundamental theory is not positively correlated with the Global Gender Gap Index of the country (World Economic Forum, 2016).
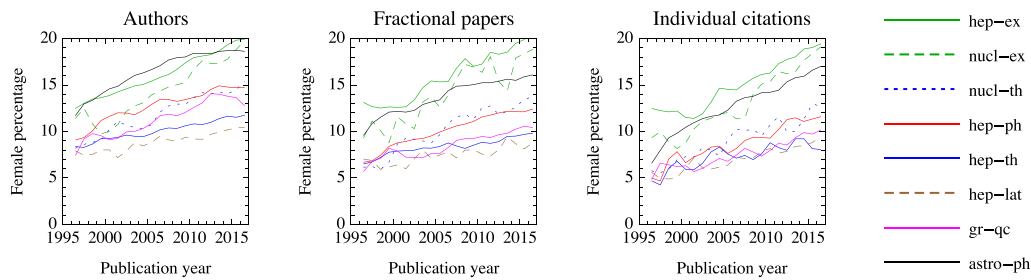
Figure 2. As in Figure 1, showing after 1995 the result within the main *arXiv* categories, plotted as colored curves: Experimental categories include hep-ex (high-energy experiments) and nucl-ex (nuclear experiments). Theoretical categories include hep-ph (high-energy phenomenology), hep-th (high-energy theory), hep-lat (lattice), and nucl-th (nuclear theory); gr-qc (general relativity and quantum cosmology) is mostly theoretical, although it includes some experiments. Finally, astro-ph contains astrophysics and cosmology.

3.1. Citations

I want to investigate if citations are influenced by the gender of the cited authors, searching for a possible different tendency of the two genders to cite a given gender more often.

In principle, complete information could be extracted by comparing "how citations are" with "how citations would be" in the absence of gender discrimination. In practice, this strategy needs a theoretical model of citations, but such models are affected by questionable systematic issues. One can try controlling for the main factors (such as different numbers of M and F authors, different average seniorities, regional differences), but reality can contain more complicated effects, such as different scientific qualities. 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. After considering attributing the remaining difference to gender bias, Caplar et al. (2017) conclude "of course we cannot claim that we have actually measured gender bias."

I will follow a different strategy, which is often more useful in the presence of backgrounds that cannot be reliably modeled: I construct an asymmetry such that it is not affected by the backgrounds. The extracted information encoded in the asymmetry is reliable but partial, as I give up on the attempt to model the full citation process.

To start, I restrict my inquiry to the subsample of single-author papers with identified gender G, as these would likely be more strongly affected by a possible gender bias. I count $N^{\text{cit}}_{G \to G'}$, the number of single-author papers with gender G citing single-author papers with gender G'. I compute the proportions $f_{G \to G'} = N^{\text{cit}}_{G \to G'}/N^{\text{cit}}_{G \to}$ dividing by the total numbers $N^{\text{cit}}_{G \to} = \Sigma_{G'} N^{\text{cit}}_{G \to G'}$, so $0 \le f_{G \to G'} \le 1$. From this I define the gender asymmetry as

$$A = f_{M \to M} - f_{F \to M} = f_{F \to F} - f_{M \to F} = \frac{1}{N^{\text{cit}}_{M \to} N^{\text{cit}}_{F \to}} \det \begin{pmatrix} N^{\text{cit}}_{M \to M} & N^{\text{cit}}_{M \to F} \\ N^{\text{cit}}_{F \to M} & N^{\text{cit}}_{F \to F} \end{pmatrix}. \tag{3}$$

The first formula means that A is the proportion in which solo men cite solo male research more than solo women cite solo male research. The second formula means that A is also the proportion in which solo women cite solo female research more than solo men cite solo female research. So the gender asymmetry ranges between $-1 \le A \le 1$. The final formula shows that A is symmetric under $M \leftrightarrow F$ permutations, with a property that makes it useful: A vanishes whenever citations are given without considering gender. $A > 0$ ($A < 0$) signals same-gender (opposite-gender) preference, although more complicated patterns are possible: Only one gender might have a particular preference for citing a given gender, or both might have a preference for opposite genders, or both might
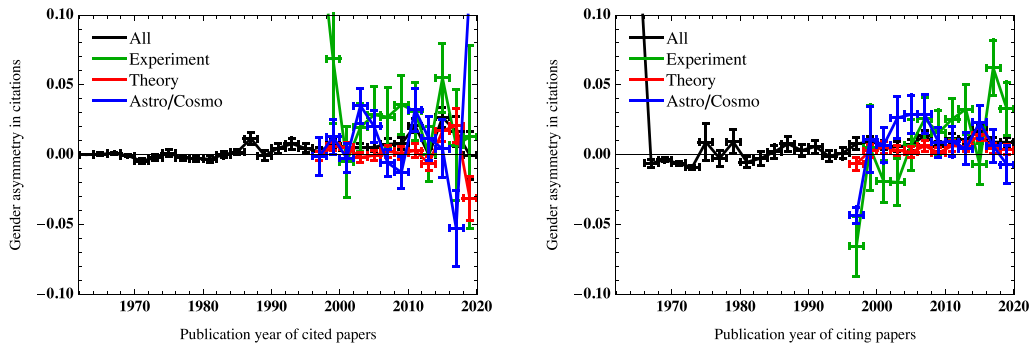
Figure 3. Time evolution of the gender asymmetry defined in Eq. 3; $A > 0$ ($A < 0$) signals same-gender (opposite-gender) preference. Left panel: As a function of the publication year of the cited single-author papers. Right panel: As a function of the publication year of the citing single-author papers. After 1995 I also show the asymmetry in different sectors of fundamental physics, based on their *arXiv* categories: theory (hep-ph, hep-th, hep-lat, nucl-th, and gr-qc), experiment (hep-ex and nucl-ex), and astrophysics (astra-ph). The bin 2018–20 only uses data available up to mid-2018.

have a preference for the same gender, in different amounts. On the other hand, A is insensitive to a difference in the total number and in the average scientific quality of M and F authors (as quantified by the chosen indicator), as well as to a possible collective equal bias of both genders towards one gender, which corresponds to multiplying one column of the matrix above by a fixed constant.[6]

To better understand what the asymmetry measures, it is useful to compute its predicted value in a toy model of citations where N_G^{aut} authors of gender G cite with gender-dependent rates $p_{G \to G'}$ (for simplicity I ignore that some authors are more active than others, so that an effective number would be directly relevant). In this model $N_{G \to G'}^{cit} \propto N_G^{aut} N_{G'}^{aut} p_{G \to G'}$ and the asymmetry equals

$$A \simeq \frac{N_M^{aut} N_F^{aut}}{\left(N_M^{aut} + N_F^{aut}\right)^2} \det \begin{pmatrix} p_{M \to M} & p_{M \to F} \\ p_{F \to M} & p_{F \to F} \end{pmatrix} \tag{4}$$

in the limit where all $p_{G \to G'}$ are close to a common value (otherwise a slightly more cumbersome expression applies).

I extract A from the data by removing self-citations, which introduce a background of same-gender preference not due to an actual gender preference. This removal is done exactly, as I have a list of all references where all authors are identified with a unique code. The removal of self-citations reduces same-gender citations, introducing a small gender discrimination of order $1/N_G^{aut}$ in the asymmetry: When considering many authors this bias is negligibly smaller than the statistical uncertainty of A, which scales as $1/\sqrt{N_G^{aut}}$.[7] Figure 3 shows the time evolution of the gender asymmetry, found to be compatible with zero at all times.[8] Restricting to papers after 2010 gives the results shown in Table 1.[9] The uncertainty is shown as one standard deviation after the ± symbol. A hint of an asymmetry, $A_{other} = (4.8 \pm 1.2)\%$, is observed among about 10^4

---

[6] The subsample of "ambiguous" authors (whose name is associated with different genders in different countries) does not show anomalous features that would support the hypothesis of a collective gender bias.

[7] More precisely, the uncertainty on A equals $[N_{F \to F}^{cit} N_{F \to M}^{cit} / N_{F \to}^{cit3} + N_{M \to F}^{cit} N_{M \to M}^{cit} / N_{M \to}^{cit3}]^{1/2}$ using the usual propagation of statistical fluctuations on each counts, $\sqrt{N_{G \to G'}^{cit}}$.

[8] An analysis performed along the same lines but replacing genders with countries shows an order one preference for citing authors of the same country, especially in some countries. This can be a manifestation of the stronger contacts between nearby authors.

[9] Our results have been reproduced by Hossenfelder (2018), who also tried to go beyond the asymmetry by assuming a model similar to my Eq. 4 (but with N_G^{aut} replaced by N_G^{pap}). As such models introduce questionable systematic uncertainties, I restrict my attention to the model-independent gender asymmetry.

Table 1. Gender asymmetry *A* defined in Eq. 3 computed restricting to single-author papers after 2010, in the *arXiv* categories defined in the caption of Figure 3. The counts are the number of single-author papers in a given *arXiv* category cited by any single-author papers, not necessarily in the same category

Category	hep-ex	hep-ph	hep-th	hep-lat	nucl-ex	nucl-th	gr-qc	astro-ph
Counts	2,755	14,627	15,370	1,762	1,673	1,258	6,706	6,733
A in %	−1.0 ± 1.7	0.5 ± 0.6	0.0 ± 0.7	−0.3 ± 2.2	6.0 ± 2.4	0.7 ± 2.2	0.5 ± 1.2	0.5 ± 1.1

other papers (mostly unpublished) not included in the eight major *arXiv* categories relevant for fundamental physics. As a result, combining all single-author papers citing single-author papers gives an asymmetry $A_{\text{published}} = (1.0 \pm 0.5)\%$ when restricting to published papers, or $A_{\text{all}} = (1.9 \pm 0.4)\%$ when including all papers.

The definition of the gender asymmetry could be extended to multiauthored papers knowing how a hypothetical gender bias would depend on the relative amount of *F* and *M* authors. One simple possibility is just to generalize the definition of $N^{\text{cit}}_{G \to G'}$ into $\Sigma_{\text{citations}} \, f_G f'_G$ where f_G (f'_G) is the fraction of authors with gender *G* in each citing (cited) paper. All self-citations, now defined as whenever the cited and citing paper have at least one author in common, are now dropped. With the new $N^{\text{cit}}_{G \to G'}$ I find the results in Table 2. Uncertainties (not shown) are there about five times smaller than in the single-author sample, if propagation of errors is naively applied to fractional counts.

Taking into account the definition of the asymmetry *A* and the relative number of *F* and *M* authors in my data, I conclude that *A* is so close to zero that a nonzero gender asymmetry in citations within its measured range would not significantly distort the bibliometric indices based on citations discussed in the following.

3.2. Hiring

The lack of a gender asymmetry in citations means that there is no fracture along gender lines in the community about which research in fundamental physics is more relevant/used/visible. In Section 2.3 it was shown that appropriate bibliometric indices based on citations are useful proxies for scientific merit. I here use such indices to search for a possible gender difference in hiring. For each hired or pseudohired author I compute his or her bibliometric indices at the hiring moment, defined as in Section 2.2. From this I extract the mean bibliometric indices of hired *F* and *M* authors.

The left and right panels of Figure 4 show the mean number of fractionally counted papers and of individual citations N_{icit}, respectively, of authors at their hiring date as reported by InSpire. For the sake of clarity I use traditional color codes: blue for male and pink for female authors.

It can be seen that hired *F* authors do not have, on average, bibliometric indicators above those of hired *M* authors. Rather, a tendency in the opposite direction seems present at all times, across the main subfields[10] and most countries (statistical uncertainties become significant when restricting to some countries with not enough authors). This result persists after taking into account the possible confounding variables considered in the supplementary Section S1.1.

I next provide extra information.

---

[10] Experimentalists who work in large collaborations tend to have similar bibliometric indicators. The average N_{icit} at hiring can be lower for *F* authors if they are hired younger than *M* authors.

Table 2. As in Table 1, considering all papers after 2010

Category	hep-ex	hep-ph	hep-th	hep-lat	nucl-ex	nucl-th	gr-qc	astro-ph
Counts/1,000	115	421	270	42	22	44	90	285
A in %	0.0	0.3	0.5	1.0	1.0	0.5	−0.1	0.4

Figure 5 shows the cumulative distribution of hired physicists as a function of their scientific age at hiring. It exhibits no significant gender difference. A gender difference could have been produced in various ways:

1. Some hiring committees might take into account career gaps due to maternity (about which no information is available): This would tend to increase the average scientific age of female hired scientists.
2. A gender discrimination in hiring would tend to reduce the average scientific age at hiring of scientists with the favored gender.
3. A gender difference in abandonment rates would tend to reduce the average scientific age at hiring of scientists with the higher abandonment rate.[11]

A warning is necessary about the two next plots, which extend the analysis to authors who have not been hired. My analysis is restricted to InSpire authors listed in HepNames (described in Section 2.2), which misses many authors who leave the field after writing a few papers. This generates an extra systematic issue, which presumably tends to be gender neutral, such that gender ratios presumably are more reliable than absolute rates. Indeed, information for M and F authors presumably is similarly incomplete, as InSpire does not collect data about gender, especially of unknown authors.

Figure 6 shows the fraction of hired authors among those who started writing papers in given time periods. Significant gender differences are not seen. I used 10-year hiring because coverage is more important here than timing. Therefore the plot stops 10 years ago, and absolute numbers would be different using incomplete InSpire hiring. Furthermore, as warned above, extra un-hired authors not in InSpire would lower the hired fraction.

Figure 7 shows the abandonment rate per year as a function of scientific age. I considered departures during 2000–2015, counting as having left those authors who wrote no further papers up to 2018. Older authors started when the M fraction was higher and the abandonment rate was lower (as hinted by Figure 6): This confounder generates an apparently lower abandonment rate among M authors. I thereby compensate for gender history as described in Eq. 2. I find that the abandonment rate is maximal among older authors who retire, minimal among senior authors,

---

[11] The temporal distribution of 245 hires of astronomers in the United States after 2010 was studied in Flaherty (2018), finding that F authors are hired on average 1.1 ± 0.6 years earlier than M authors (considering the time after receiving their PhD; astronomers are hired on average five years later). I find a difference of 0.95 ± 0.5 yr restricting to astro/cosmo authors (considering the time after the first paper; authors are hired on average nine years later). According to Flaherty (2018), the hiring time distribution is better fitted assuming a three to four times higher F abandonment rate, rather than assuming a 10:1 bias in favor of F astronomers. However this claim is only based on a very simplified model of hiring that neglects important effects (some authors are better than others; quotas would not be overfilled, etc). I do not attempt to model hiring, as I do not see how models can be made realistic. Rather, I have extra data about papers and citations that support neither a 10:1 bias (see Figure 4) nor a 4:1 difference in abandonment rates (see Figure 7). Flaherty's (2018) results have been "firmly ruled out" by Perley (2019), who ruled out gender differences larger than 40% in hiring and abandonment rates.
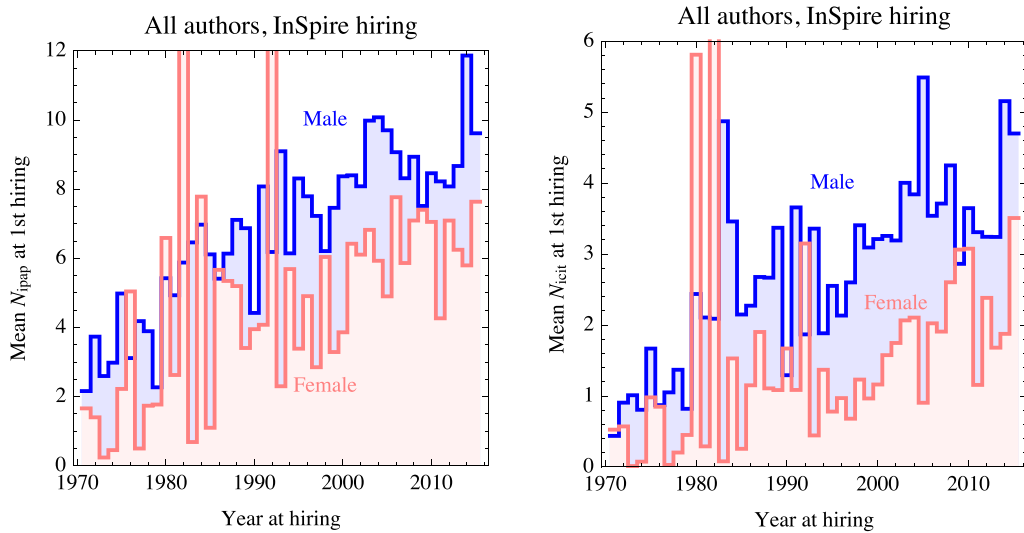
Figure 4. The left and right panels show the mean number of fractionally counted papers N_ipap and of individual citations N_icit, respectively, of authors in fundamental physics at the moment of their first hiring, as a function of the hiring year. Data are shown separately for male (blue) and female (pink) authors, and compensated for gender history as described in Eq. 2.

and intermediate among junior authors (as warned above, the abandonment rate of very young authors who leave the field after writing just a few papers is underestimated). Abandonment rates show no significant gender difference, in agreement with the null result by Perley (2019) and in disagreement with Flaherty (2018) (these authors only considered astrophysics).

In conclusion, the gender gap in representation at the entrance level of research is negligibly affected by "leaky pipeline" effects consistent with Ceci et al. (2014) that finds large gender differences at PhD level in STEM, and mild differences in the subsequent progress; see also Allen-Hermanson (2017) and Miller and Wai (2015).

3.3. Productivity

In this section I study scientific productivity as quantified through bibliometric indices. Of course, such indices say nothing about other activities of researchers that do not result in publications, such as teaching, mentoring, and outreach. Figures 1, 2, 8, and 11 show a possible gender gap in the fractionally counted number of papers: Male authors write, on average, 10% more papers. The gap is consistent with earlier findings in the literature (see, for example,
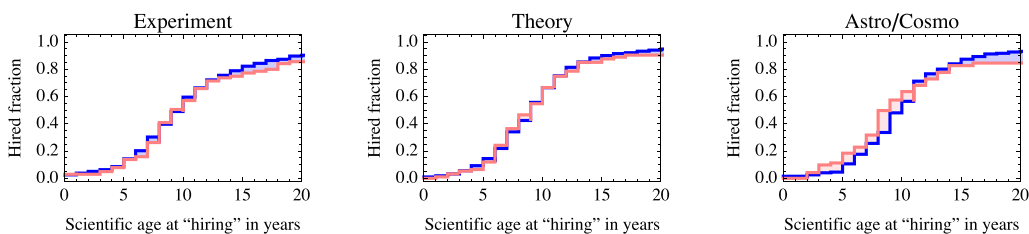

Figure 5. Among all authors first hired after 2000, the cumulative fraction of hired authors as a function of their scientific age is shown, for male (blue) and female (pink) authors in experiment (left), theory (middle), astro/cosmo (right). InSpire hires are used and gender history is compensated for as described in Eq. 2.
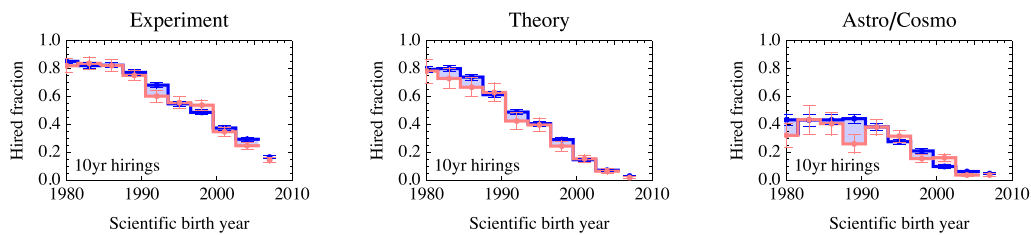
Figure 6. Fraction of authors hired up to now as function of the date of the first paper. Only the statistical uncertainty is shown; see the text for warnings.

Table 2 of Ceci et al. (2014), Abramo et al. (2009a, 2015). A slightly larger gap is found in the number of received individual citations.

Is such a gap due to the different average scientific age of M and F authors? To check this possibility, the left panel of Figure 8 shows the mean number of fractionally counted papers written by M and F authors as a function of their scientific age (time since their earliest paper). The gap persists. The right panel of Figure 8 similarly shows the mean number of received individual citations. In both cases it can be seen that junior M and F authors have similar productivity, and that a gap develops with their scientific age. A higher scientific age means going backwards in time to authors who started earlier, when the field was different and when the F percentage was smaller.

The averages in Figure 8 are shown separately for hired and not-hired authors, using 10-yr-hires in order to have more complete coverage. It can be seen that hiring is not the reason for the gap.

Furthermore, in Figure 8 only scientifically active authors are considered (those who wrote at least one paper after 2013), such that these results would not be affected by a gap in abandonment rates.

Figure 8 does not compensate for possible career gaps, as such gaps do not exhibit significant gender differences. This is shown in Figure 9, where for each author I computed the longest time gap between consecutive papers, using *arXiv* dates to have precise information about publication dates. The distribution of longest gaps among M and F authors shown in Figure 9 does not exhibit significant gender differences. A similar null result is found when restricting to hired authors. Stopping writing papers might, however, be the extremum of a tendency towards reduced productivity (possibly due to maternity issues). I therefore searched for consecutive years of reduced publication intensity: Some authors are more regular, other experience periods of relatively lower productivity, but again the distributions show no significant gender differences (see Figure 10). No significant gender differences are found looking at periods of relatively higher productivity.
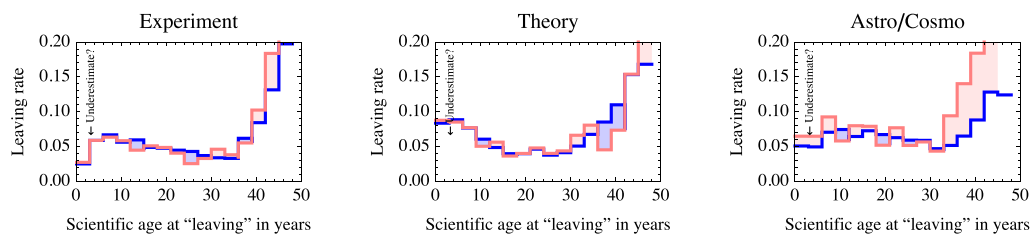

Figure 7. Fraction of active authors that leave research each year, as a function of their scientific age. Departures during 2000–2015 are considered, counting as having left those authors who had written no further papers up to 2018 and gender history is compensated for as described in Eq. 2. See the text for warnings.
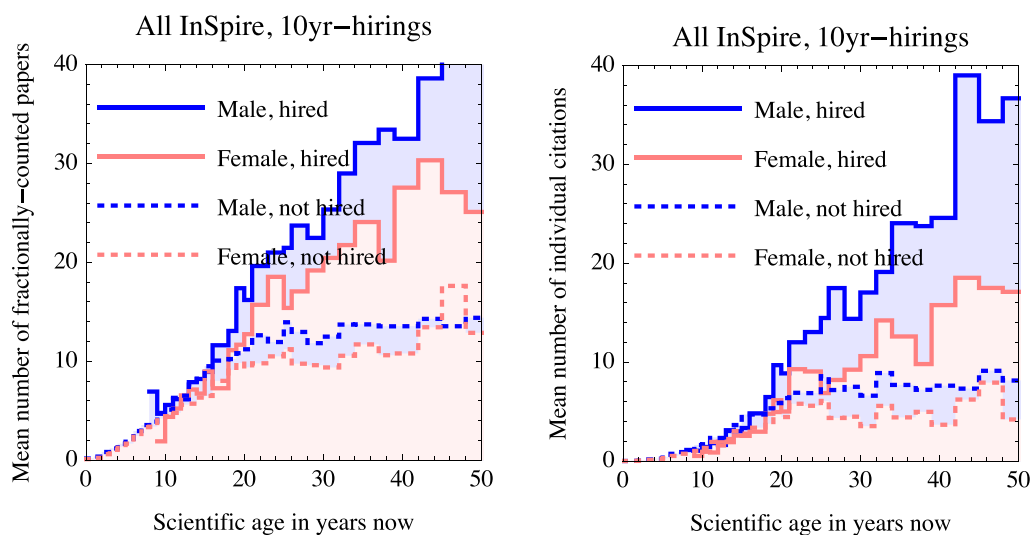
Figure 8. Mean number of fractionally counted papers (left) and of individual citations (right) as a function of scientific age (time after the first paper) of scientifically active authors now.

A gender difference in abandonment rates or career gaps or periods of lower productivity would reduce the cumulative number of papers and of received citations of authors at career level. It is thereby interesting to test whether a gap persists in noncumulative productivity indices that avoid summing over author careers. The procedure is as follows: For each year the subset of scientifically active authors that produced papers is selected, and Figure 11 shows their average productivity, separately for M and F authors. I find that active F authors produce on average roughly 30% fewer papers than active M authors, and receive roughly half the number of citations.[12]

Furthermore, Figure 12 (left panel) analyzes the gap at the level of papers, finding a smaller F percentage among authors of top-cited papers, even when restricting to single-author papers (see also Jagsi et al., 2006; West et al., 2013). The right panel of Figure 12 shows that F authors tend to work in larger collaborations.

The supplementary Section S1.2 discusses other possible confounding variables, without finding anything that can remove the gender gap in productivity.

I now discuss some possible causes of such a gap.

In various countries F authors have earlier retirement ages. But gender differences show up before retirement in Figure 8. Furthermore, many physicists tend to remain scientifically active after retirement (although the productivity of most physicists tends to decline before retirement).

A possible reason for the gender gap observed in various fields is children and maternity. See Ceci et al. (2014) for a recent summary of the literature, which is not univocal. Some studies find no or small effect (Cole & Zuckerman, 1987; Sax, Hagedorn et al., 2002; Stack, 2004; Xie & Shauman, 2003). Other studies find a negative impact on women (Fox, 1995; Ginther & Kahn, 2009), and on men and women equally (Hargens, McCann, & Reskin, 1978), while some studies found a positive impact on men (Ceci et al., 2014), possibly due to selection effects.

---

[12] Fractional counting is used. Using full counting, hyperauthored publications would lead to a recent boom, roughly equal for M and F authors, due to the appearance of experimental collaborations with thousands of authors. Alternatively, the productivity gap can be seen using full counting and restricting to theorists.
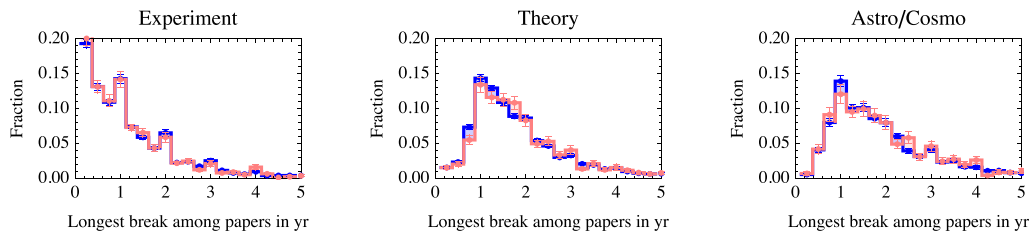
Figure 9. Fraction of authors active between 2000 and now (divided by their main topic) as a function of the longest time break among their papers. Gender history is compensated for as described in Eq. 2.

Results vary depending on field (with physical sciences sometimes being an outlier, possibly a fluctuation) and are mostly focused on the situation in the United States and on the number of papers produced or worked hours. Ceci et al. (2014) conclude: "the presence of children cannot explain the overall gender productivity gaps." While maternity would deserve a dedicated study, the InSpire data do not provide any personal information, so it is only possible to proceed indirectly. As has already been described, timing of publications does not show gender differences in periods of null or reduced productivity. Figure 8 indicates that the productivity gap opens at an age roughly consistent with maternity (but also consistent with the transition to scientific independence), and that it does not close at older ages. A similar situation is found when analyzing the salaries of physicists in the United States: no gender gap just after graduation; a 10% gap after 10–15 years according to Porter and Ivie (2019), who report large differences in personal life choices, in particular that women are four times more likely to have a career break.

As maternity laws are different in different countries, an alternative possible strategy is looking for national differences in the *M/F* gap, which seems stronger in Germany, the United Kingdom, and Italy; weaker in the United States and France, and null in Japan. However single-country statistics are poor and many other national differences can act as confounding factors.

3.4. Distribution of Individual Citations

In the previous section a productivity gap was found. I here characterize its statistical properties. Figure 13 shows the distributions in the number of individual citations N_{icit} received by female and male authors in fundamental physics, considering the whole InSpire database. The bell-shaped distributions spread through a few orders of magnitude in N_{icit}. The dotted curves show that each bell is well approximated, at least on its upper side, by a log-normal as a function of N_{icit} (namely, by a Gaussian as function of log N_{icit}, the variable used on the horizontal axis of
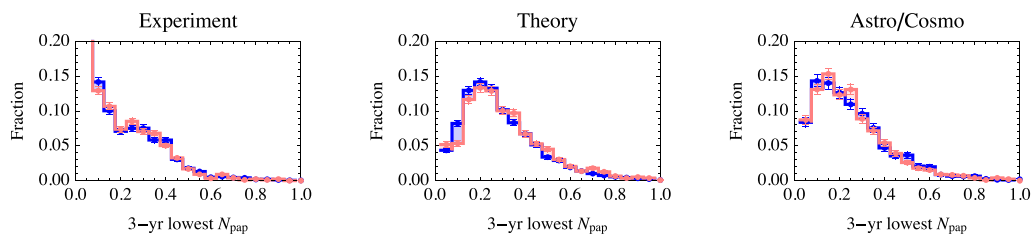

Figure 10. For each author the minimal number of papers he or she produced in a consecutive three-year period is computed. This number is divided by the author average publication rate, obtaining a number $0 \leq r \leq 1$ normalized such that $r = 1$ indicates an author who published in a regular way, while $r = 0$ indicates an author with a three-year period of null productivity. As a function of r the fraction of authors active between 2000 and now is plotted, divided by their main topic. Gender history is compensated for as described in Eq. 2.
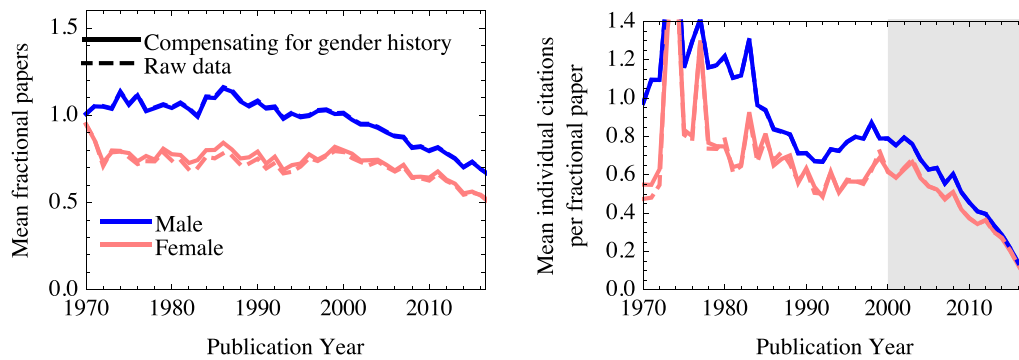
Figure 11. Left panel: Mean number of fractionally counted papers produced each year by M and F authors active that year. Right panel: Mean number of received individual citations divided by mean number of fractionally counted papers. The shading reminds us that citation counts are incomplete for recent papers. The continuous curves show the result compensated for gender history as described in Eq. 2, and the negligibly different dashed curves show raw data.

Figure 13; log is the logarithm base 10). A log-normal, already observed in bibliometrics (Thelwall & Wilson, 2014), arises when many positive independent random variables contribute multiplicatively.[13]

The difference between the F and M distributions in Figure 13 is statistically significant. This is better shown by the black curve in Figure 13, that shows the ratio N_F/N_M of female versus male authors (left axis) as function of the number of received individual citations (horizontal axis).[14] The N_F/N_M gender ratio is not constant: The M fraction progressively grows when going from average to top authors in terms of individual citations.

Again it is necessary to study whether such a difference can be a byproduct of confounding factors that affect my composite sample of data. This issue is discussed in supplementary Section S1.3: I do not find any confounder that washes away the trend. As discussed in Section 2.4 I compensate for one significant confounder: Male authors are presently on average more senior than female authors. This is relevant here because senior authors had more time to receive citations (and because younger authors contribute more to top-cited papers). This confounder does not remove the gender difference in N_{icit}, given that it is observed within sub-samples of authors with same scientific age (see supplementary Figure S11). Correcting for the age confounder is, however, needed to precisely quantify the difference. I proceed as described in Eq. 2. The left panel of Figure 13 shows age-corrected results, while the right panel shows raw data. Both averages and variances differ in raw distributions. The difference in variances persists among the upper sides of the age-corrected distributions, while no significant gender differences are seen on the lower sides of age-corrected distributions, more affected by social phenomena (lower sides would be mostly removed by restricting to hired authors).

Different ages can be accounted for by using a complementary strategy: I consider a new bibliometric index $N_{\text{icit}}/\Delta t_A^p$ that approximatively compensates for the scientific age $\Delta t_A = t_{\text{now}} - t_A$ of each author. In order to achieve the compensation I choose $p \approx 1.8$, because N_{icit}

---

[13] The following observations suggest that the dominant random variables are unlikely to be of social type (e.g., the possibility that some authors get more visibility and funds that boost their citation counts (Ruocco, Daraio et al., 2017): I find that the number of citations received by authors in physics tends to grow linearly or quadratically rather than exponentially with their scientific age. Furthermore, when looking at single papers, rather than at author careers, a significant excess of young authors (scientific age below approximatively 15 years) is observed among authors of top-cited papers.

[14] In precise mathematical notation this is dN_F/dN_M, and the bells are $N_G^{-1} dN_G/d \log N_{\text{cit}}$.
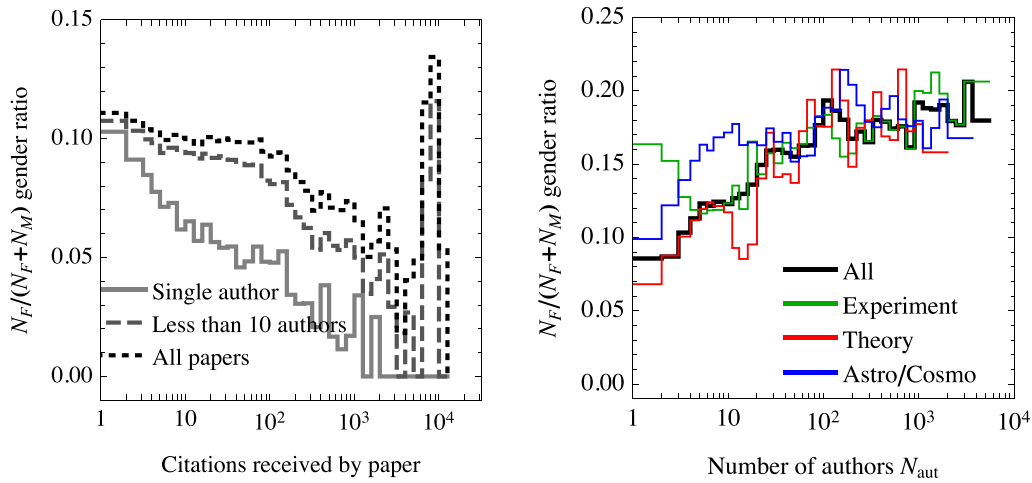
Figure 12. Left panel: Fractional contribution of all *F* authors as function of the number of individual citations received by the paper. The *F* fraction of each paper is determined assuming that each author contributed equally to collaboration papers and compensating for gender history. The same result is also computed restricting to solo papers and to papers with fewer than 10 authors. Right panel: Fractional contribution of all *F* authors to papers with N_{aut} authors.

averaged over authors approximatively scales as $\Delta t_A^{1.8}$ within all main subfields. The result of such a correction is shown in the left panel of Figure 14: Both *M* and *F* distributions become narrower, having removed one source of their variability. I again find a higher male variance among the upper sides. The right panel of Figure 14 shows that applying both age corrections has negligible extra effect. The difference in variances on the upper side is seen in any case.
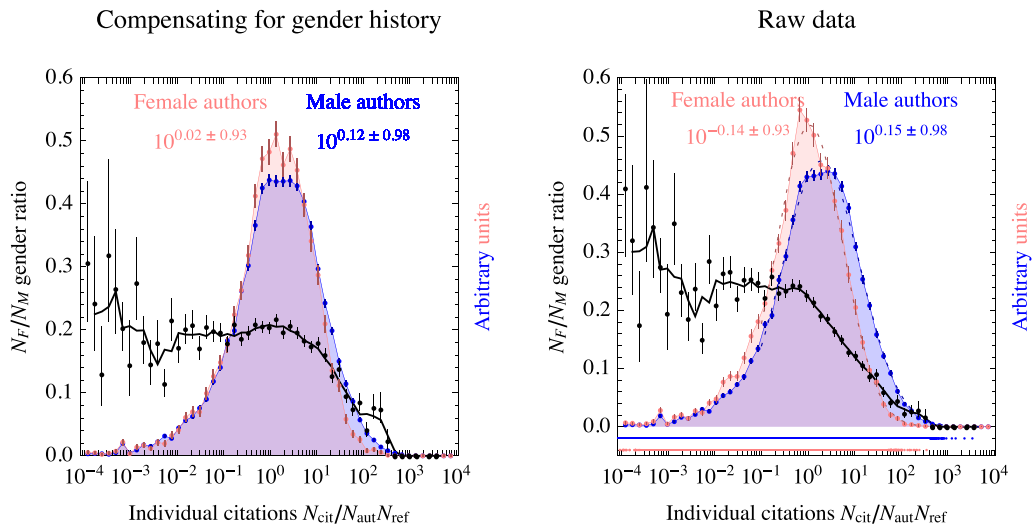
Figure 13. The bells show the probability distributions of authors in fundamental physics as a function of their number of individual citations, separately for Male (blue) and Female (pink) authors and plotted with a common normalization (right axis). The numbers in the figure show the logarithmic average and standard deviation; and the dots below the bells in the right panel show individual authors separately for *M* and *F* authors; the dotted curves show how well a log-normal approximates the upper side of the bells. The black smoothed curve (left axis) shows the ratio N_F/N_M between the absolute numbers of female and male authors who received the amount N_{icit} of individual citations indicated on the horizontal axis. Error bars on data points are one standard deviation statistical uncertainties. Left panel: Data after compensating for the higher average age of male authors, as described in Eq. 2. Right panel: Raw data, biased by the different age distribution.
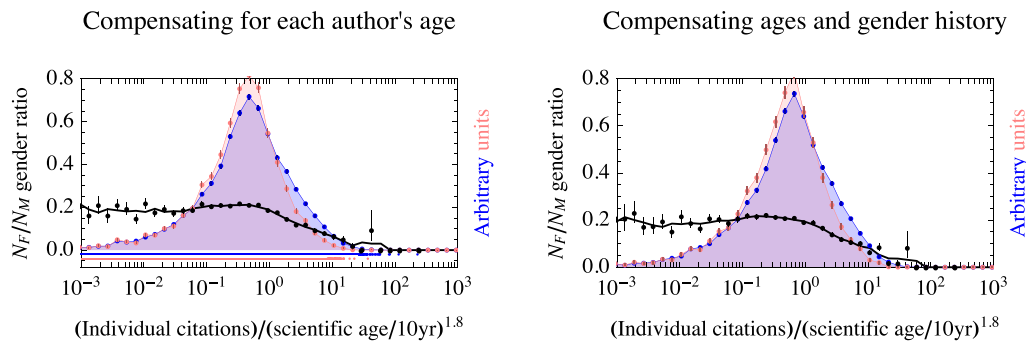
Compensating for each author's age

Compensating ages and gender history

Figure 14.   As in Figure 13, adopting a measure that, on average, does not depend on the scientific age of authors. The M distribution still has a longer upper tail.

The M fraction is largest among top authors, as clear from the dots below the bells in the right panel of Figure 13, which show the individual authors. It is useful to focus on the subset of top authors. A physicist might read their names and conclude that no sociological confounder can wash away most of them. It is thereby interesting to show that the gender difference is statistically significant, even when restricting to top authors. This is done by considering the hypothesis of no gender difference: While being sociologically implausible it has precisely computable consequences that can be compared to data about top authors. Starting from raw data, the F author with most individual citations is in position $F_1 = 69$. Under the hypothesis of no gender difference one can mathematically compute that the probability of observing $F_1 \geq 69$ is $\wp_1 = m^{F_1-1} \approx 3\,10^{-6}$, where $m = 1 - f \approx 0.83$ is the male fraction of the large sample (or $m \approx 0.87$ restricting to theorists). Under the same assumption, the kth F author should be on average position $\langle F_k \rangle = k/f$, with probability distribution $f^k m^{F_k-k} \binom{F_k-1}{k-1}$ (Knapp, 2010). The observed positions are $F_2 = 147$ (the probability of being in this or lower position is $\wp_2 \approx 3\,10^{-11}$), $F_3 = 191$ ($\wp_3 \approx 2\,10^{-13}$), etc., roughly fitted by $F_k \approx 69k$. Such low probabilities mean that gender differences are needed to account for data about top authors. Of course, it is already known that the time evolution of $f = N_F/(N_M + N_F)$ is a significant confounder. I again find that such a confounder is not enough to remove the difference, as the female fraction f is larger than $1/69$ at all relevant times. More precisely, after performing the age correction as in Eq. 2, the kth top female authors shift to $F_k = \{36, 82, 114, 126, \ldots\} \sim 33k$ position. Considering the age-corrected metric $N_{\text{icit}}/\Delta t_A^{1.8}$ the positions are $F_k = \{20, 57, 109, 180, \ldots\} \sim 38k$. After correcting for the age confounder, $\wp$-values remain small because the female fraction $\sim 1/33$ or $\sim 1/38$ found among top authors remains smaller than the fraction f of female authors in the full sample. An excess of male top authors is found even after correcting for the age confounder.[15]

The fact that gender differences are maximal among those top authors who receive up to 3,000 more individual citations than average authors indicates that differences do not predominantly result from sociological factors that could give factors of 2 differences at individual level (such as harder working vs. career gaps, more research vs. more teaching, specialization on physics vs. wider interests, etc).

---

[15] A varying gender fraction that culminates in a small group of extremely productive, mostly male, "star authors" has been observed in Abramo et al. (2009b, 2015) and Bordons et al. (2003) (see also Kwiek, 2016). Ioannidis, Baas et al. (2019) used Scopus data about 6.9 million scientists in all disciplines to compute a "composite" bibliometric indicator, producing a list of top authors. When restricted to fundamental physics, their list (despite minor problematic aspects) is significantly correlated to my list. In their all-fields list the female authors are found in positions $F_k = \{133, 146, 160, \ldots\}$. While this naively seems to extend my findings, I cannot correct their list for the age confounder nor for other confounders.
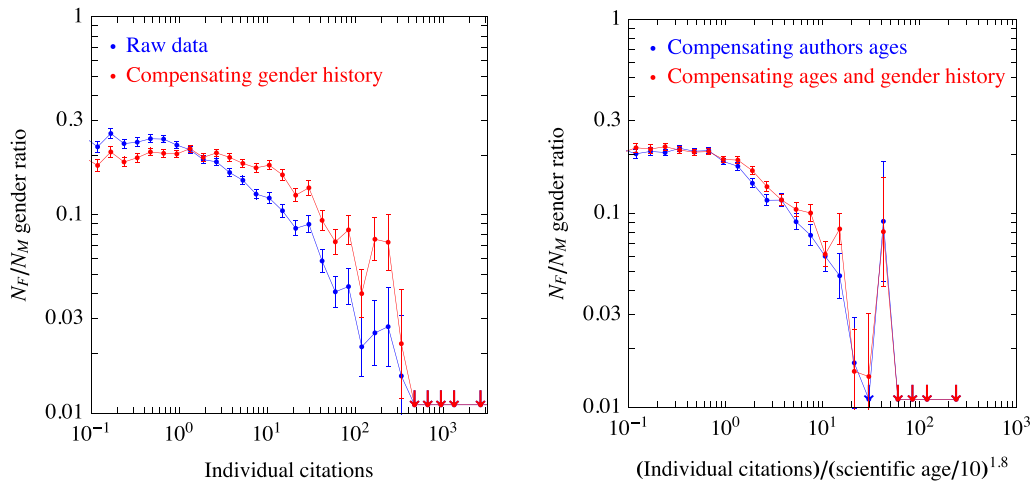
Figure 15. In the left panel the points with Gaussian 1σ statistical errors are data about the N_F/N_M female/male ratio as a function of the number of individual citations received. Points with $N_M > N_F = 0$ are shown as down arrows. The blue points are raw data, the red points are corrected compensating for the different time evolution of the overall number of M and F authors. In the right panel a different bibliometric index is considered, which approximatively compensates also for the scientific age of each author. Data are not well fitted by a linear function in a log-log scale (which corresponds to $p = 1$ in the supplementary Eq. S2) and can be fitted by a quadratic function (which corresponds to $p = 2$). This can be interpreted as different male and female variances, as in Eq. 6, rather than as different averages, as in Eq. 5.

As the variation of N_F/N_M survives to confounding variables, I try to better understand this effect by investigating its quantitative shape.

A look at Figures 13 or 14 suggests that the M bell has a longer tail of top authors (the raw data also show a difference in averages, mostly due to confounders, that makes the difference in variances less easily visible). The supplementary Section S2 shows that the difference in upper variances is statistically significant. I here provide a simpler, more intuitive, argument through analytical approximations based on the observation that the distributions of individual citations received by M and F authors separately are approximatively log-normal (Gaussian as a function of $\ell = \ln N_{\text{icit}}$) on their upper side. Log-normal distributions with a common standard deviation σ and different averages $\mu_M \neq \mu_F$ for M and F authors would produce a N_F/N_M of the form

$$\frac{N_F}{N_M} \propto \frac{e^{-(\ell-\mu_F)^2/2\sigma^2}}{e^{-(\ell-\mu_M)^2/2\sigma^2}} = \exp\left[-\frac{R_\sigma}{2}\left(\ell - \frac{\mu_M+\mu_F}{2}\right)\right], \qquad R_\sigma = 2\frac{\mu_M-\mu_F}{\sigma^2}. \tag{5}$$

Different standard deviations $\sigma_M \neq \sigma_F$ would produce

$$\frac{N_F}{N_M} \propto \frac{e^{-(\ell-\mu)^2/2\sigma_F^2}}{e^{-(\ell-\mu)^2/2\sigma_M^2}} = \exp\left[-\frac{R_\sigma}{2}(\ell-\mu)^2\right], \qquad R_\sigma = \frac{1}{\sigma_F^2} - \frac{1}{\sigma_M^2}. \tag{6}$$

Thereby a dominant difference in averages (standard deviations) would produce a line (a parabola) when N_F/N_M is plotted as a function of N_{icit} on a log-log scale. Such a plot is shown in Figure 15, again performing the usual corrections for the age confounder. The important point is that, in all cases, N_F/N_M exhibits a parabolic shape along the upper sides of the bells, where the log-normal approximation is accurate enough.[16] In conclusion, the data exhibit a gender difference in the upper variances.

---

16 Adding higher-order terms in the exponent would not change the above conclusion, because fits to the observed distributions find small higher-order terms.

Is this statistically strong preference an artefact of the complexity of the full data sample? To answer, the analysis is repeated within the independent subsamples of supplementary Figure S11: Plotted on a log-log scale they independently tend to show a parabolic (rather than linear) trend in N_F/N_M. The dotted curves in Figure S11 show how well each subsample can be fitted in terms of $R_\sigma \approx 2$ and $p \approx 2$. Furthermore, the probabilities $\wp_i$ that test the hypothesis of no gender difference restricting to top authors are small within the subsamples of Figure S11 (where the top authors are plotted as points), consistent with Abramo et al. (2009b) and Bordons et al. (2003).

A similar difference in upper variances is also found when looking at different bibliometric indicators; see the supplementary Section S1.3.

3.5. Self-References

Gender differences in self-references are an interesting topic on its own, as as well as a possible confounding factor to previous analyses. I verified that my previous results persist when dropping self-references, and I next justify why there is no need to drop self-references.

Bibliometric studies have found that men cite their own papers more than women: Cameron, White, and Gray (2016) focused on six ecology journals; Ghiasi, Larivière, and Sugimoto (2016) on the Web of Science database; King, Bergstrom et al. (2017) on the JSTOR database; and Hossenfelder (2018) on single-authored papers in *arXiv*. In agreement with such studies, restricting to solo papers, I find a $\sim$20–30% higher fraction of self-references among the papers written by M authors (7.3% versus 5.9%).

However, King et al. (2017) mention the possibility that this gender difference in self-references is just a reflection of the fact that male authors tend to write more solo papers, and thereby have more scientific reasons for self-references. King et al. (2017) could not check if this is a significant confounding factor, because authors are not disambiguated in their database. As authors are disambiguated in the InSpire database, I can perform this check finding that this confounding factor removes the gender difference in self-references: A similar fraction of self-references is found when comparing male and female authors who wrote the same number of papers N_{pap}. In other terms, the fraction of self-references can be a misleading indicator because the average number of self-references grows with the number of solo papers following a scale law $N_{\mathrm{cit}}^{\mathrm{self}} \propto N_{\mathrm{pap}}^{p}$ with power $p > 1$. My data suggest $p \approx 1.3$.

As solo papers are a relatively small subsample that might be not representative of the full database, I extend the analysis to multiauthored papers. In order to do so, it is necessary to distinguish self-references from self-citations. I count a citation as self-citation whenever the citing paper has at least one author in common with the cited paper. I count it as a self-reference only for those authors who wrote both the cited and the citing paper. I clarify this with an example. One paper by authors A and B cites a paper by authors B and C: B gives a self-reference; both B and C receive a self-citation (B directly and C indirectly).

As usual, fractional counting is used. For each author I compute the total number of received individual citations N_{icit}, of received individual self-citations $N_{\mathrm{icit}}^{\mathrm{received}}$, of given individual self-references $N_{\mathrm{icit}}^{\mathrm{given}}$, and of fractionally counted papers N_{pap} (equal to the number of given individual references).

I consider the mean fraction of given individual self-references $N_{\mathrm{icit}}^{\mathrm{given}}/N_{\mathrm{icit}}$ (left panels of Figure 16) and the mean fraction of received individual self-citations $N_{\mathrm{icit}}^{\mathrm{received}}/N_{\mathrm{icit}}$ (right panels of Figure 16). The upper row shows some gender difference: Male authors tend to give
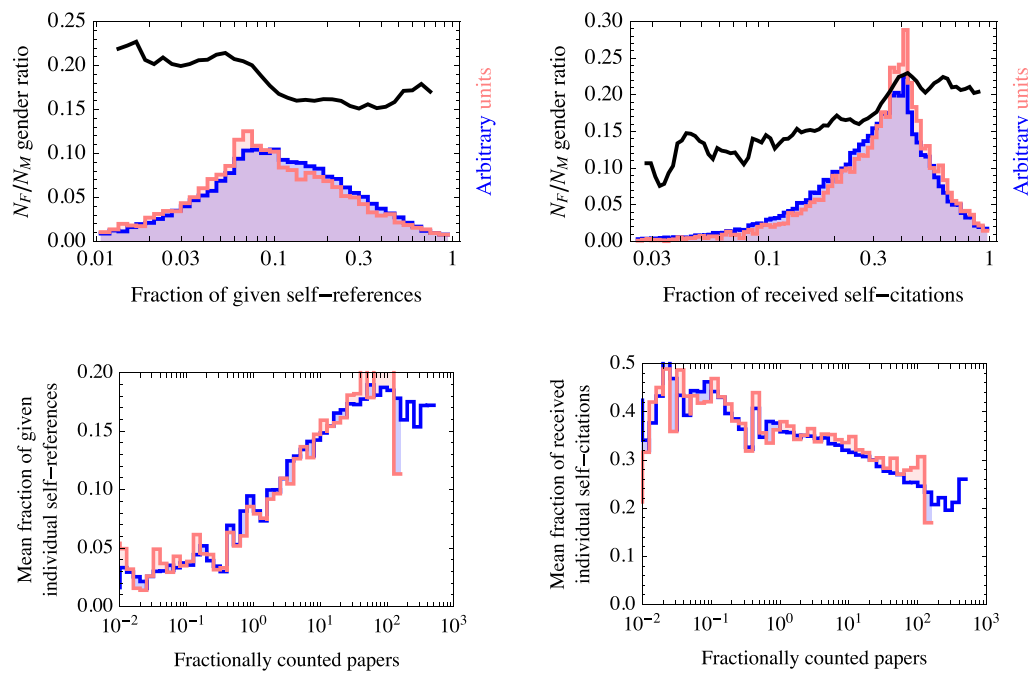
Figure 16. Left panel: Distribution of authors as a function of their self-reference rate. Right panel: Distribution of authors as a function of their self-citation rate. Bottom row: Mean fraction of given self-references (left) and of received self-citations (right) among M (blue) and F (pink) authors with the number of fractionally counted number of papers indicated on the horizontal axis.

themselves a higher fraction of self-references (13.2% instead than 12.0%); female authors tend to receive a higher fraction of indirect self-citations (23.5% instead t hen 19.8%). Such differences again disappear in the lower row of Figure 16, where the self fractions are computed as a function of the number N_{pap} of fractionally counted papers written by each author. Similarly to what we found in the solo sample, male authors tend to cite themselves and their collaborators more just because male authors tend to have more past papers. Similar results are found restricting within the main subfields.

4. CONCLUSIONS

I performed a bibliometric analysis of gender issues in fundamental physics worldwide from approximately 1970 to now. Bibliometrics gives quantitative data on activities of researchers that result in publications and tells nothing about other possible activities, such as teaching, mentoring, and outreach. Nevertheless, research is an interesting area in which individual talent can be expressed, as confirmed by the large differences between authors found in data. Concerning gender I find the following results:

1.   First, the well-known initial gender difference in representation is seen: There are roughly four males for each female among new authors that appear at PhD-level. The initial female fraction is not positively correlated with the Global Gender Gap Index of the countries,[17] and negligibly evolves in the subsequent career stages (Figure 7).

---

[17] An anticorrelation known as "gender equity paradox" (Stoet & Geary, 2018) has been observed among students, who are less mobile than researchers.

2. When citing works by others, authors exhibit no or small gender difference: Male and female authors have the same average opinion about which research in fundamental physics deserves to be cited (Figure 3).[18] Furthermore, *M* and *F* authors give to their own papers a similar fraction of self-references (Figure 16), taking into account that *M* authors tend to write more papers, especially solo papers (Figure 12).

3. Female authors do not have, at hiring moments, higher average bibliometric indicators based on individual citations or fractionally counted papers than male authors (Figures 4 and 5).

4. Among authors identified by InSpire HepNames profiles (which misses authors who write very few papers), I do not find a gender difference in hired percentages (Figure 6), in abandonment rates (Figure 7), in longest breaks between papers (Figure 9), in periods of reduced activity relatively to their average (Figure 10).

The above results are in line with the literature, as summarized in Ceci et al. (2014): "the overall picture is one of gender neutrality"; "no evidence of women having harder time getting tenure"; postbachelor gender differences in attrition rates are significantly smaller in STEM than in life science, psychology, and social science. The literature finds a second gender difference, in productivity: "women on average publish fewer papers than men"; "there are no sex differences in citations per article" (Ceci et al., 2014). I find:

5. A productivity gap both in the fractionally counted number of publications and in their citational impact (Figure 8), which does not appear to be predominantly concentrated in specific countries, topics, periods, bibliometric indicators, journals (supplementary Figure S1), etc. The gap is also found without integrating over careers (see Figures 11 and 12).

6. A gradually increasing male fraction when going from average to top authors in terms of individual citations (or other indices, such as fractionally counted publications). The quantitative shape of this trend appears predominantly due to a higher variance on the upper side of the *M* distributions (see Figure 15 and Eqs. 5 and 6), rather than due to a difference in averages.

I verified that my results still hold when ignoring hyperauthored papers (possibly affected by guest authorship) or restricting to single-authored papers.

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 honours 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.

---

[18] Some authors discuss the possibility that physicists are collectively affected by an unconscious gender bias—a concept that has received recent attention in the United States following the development of Implicit Association Tests (IAT) that claimed to reveal such biases. Even if such tests were scientifically valid (see, however, Oswald, Mitchell et al. (2013) for a metareview), reading a scientific paper involves different mental processes than those probed by IAT. Some of my results are based on citations: The same conclusions are reached when removing from my analysis citations to single-authored research, which would be more affected by a hypothetical collective gender bias. 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 credible and more offensive. Handley, Brown et al. (2015) found that men (especially among STEM faculty) evaluate gender bias research less favorably than women.

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, Johnston, & Clark, 2010; Lippa, 2010; Su, Rounds, & Armstrong, 2009; Su & Rounds, 2015; Thelwall, Bailey et al., 2019) and in variability (Deary, Irwing et al., 2007; Halpern, Benbow et al., 2007; Hyde, 2014; Stevens & Haidt, 2017; Wang et al., 2013).

Greater male interest in things and greater female interest in people is observed consistently across cultures and time and is large ($d \approx 1$, i.e., distributions differ by about one standard deviation): Such a difference in interests predominantly accounts for the initial difference in representation.[19] Difference in variability accounts for the difference in productivity. This is consistent with O'Dea, Lagisz et al. (2018), which confirms the difference in variabilities looking at grades, and observes that this difference alone cannot reproduce the representation gap.

The amount of higher male variability suggested by bibliometric data in fundamental physics is at the 10% level, roughly consistent with independent observations of presumably relevant traits. While such psychometrics observations predominantly probe the central, most populated, part of the distributions, I reasonably expect that physicists probe the upper tail[20] and that top-cited physicists reach the far-end upper tail. I estimate to reach about five standard deviations above the mean, given that this is the maximal deviation statistically expected from a pool of $\sim 10^9$ persons in a Gaussian approximation.

When dealing with complex systems, any simple interpretation can easily be incomplete, including a hypothetical gender discrimination. In any case, it is interesting that data can be explained without invoking such a hypothesis.

I conclude by addressing ethical and social values, given that a gender difference in variances is seen by some as offensive, like other ideas originally proposed by Darwin (Hill, 2017) (modestly keeping things in proportion in this comparison). The interpretation in terms of different variabilities implies that one should keep giving gender-neutral equal opportunities to everybody by considering each person based on his or her individual qualities, not as member of a demographic group (gender, nationality, or whatever). The refusal to consider population level differences in distributions when trying to understand gaps in representation can lead to discrimination allegedly aimed at establishing equal outcomes (see, for example, Strumia (2019) for a more extensive discussion of such issues).

---

[19] Large gender differences along the people/things dimension are observed in occupational choices and in academic fields: Such differences are reproduced within subfields (Thelwall et al., 2019). In particular, female participation is lower in subfields closer to physics, even within fields with their own cultures, such as "physical and theoretical chemistry" within chemistry (Thelwall et al., 2019). This suggests that the people/things dimension plays a more relevant role than the different cultures of different fields.

Furthermore, psychology finds that females value careers with positive societal benefits more than do males: Some authors propose that women tend more to opt out of STEM because "women tend to endorse communal goals more than men" (Diekman et al., 2010; Evans & Diekman, 2009). Indeed Gibney (2007) 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. Concerning fundamental physics, old discoveries gave huge societal benefits, but no practical applications are resulting from contemporary explorations of very small and very large scales (such as production at colliders of particles that decay in 10^{-25} seconds, or cosmological observations of objects billions of light years far away).

[20] Physics attracted students with high average grades; see, for example, Figure 9 of Ceci et al. (2014) and Figure 1 of Ginther and Kahn (2015). Looking at career-integrated citations, physics (and especially fundamental physics) shows the largest ratio between high (90th) and low (25th) percentile (Tables 1 and S3 of Ioannidis et al. (2019)).

ACKNOWLEDGMENTS

COMPETING INTERESTS

FUNDING INFORMATION

DATA AVAILABILITY

This study is based on bibliometric data extracted from the InSpire database around mid-2018. While data are already public from the InSpire web site (InSpire, 2010), I selected the key data and converted them into user-friendly Mathematica format. The code used to extract data is available at GitHub (https://github.com/RiccardoTorre/InSpireTools) and the resulting data at Zenodo (https://doi.org/10.5281/zenodo.3482884). The data consist of a table with more than one million entries, one for each paper. Each entry contains the following information:

1. an integer number identifying the paper;
2. publication date;
3. publication date on *arXiv* when available;
4. number of authors;
5. number of references;
6. number of citations;
7. PageRank with self-citations;
8. PageRank without self-citations;
9. list of authors with each author identified by an integer number;
10. list of references, with each paper identified by the integer in item 1);
11. list of citations;
12. title;
13. *arXiv* main category when available;
14. 1 if published, 0 otherwise;
15. list of lists of affiliations of each author, with each institute identified by an integer number;
16. *arXiv* subcategory when available;
17. *arXiv* number when available;
18. journal, with each journal identified by an integer number;
19. collaboration(s), with each collaboration identified by an integer number;
20. number of individual citations.

The correspondence between integer numbers and papers, authors, institutes, journals, and collaborations is available from InSpire (for privacy reasons I avoid making my tables public).

REFERENCES

Abramo, G., Cicero, T., & D'Angelo, C. A. (2015). Should the research performance of scientists be distinguished by gender? *Journal of Informetrics*, 9, 25–38. **DOI:** https://doi.org/10.1016/j.joi.2014.11.002

Abramo, G., D'Angelo, C. A., & Caprasecca, A. (2009a). Gender differences in research productivity: A bibliometric analysis of the Italian academic system. *Scientometrics*, 79, 517–539. **DOI:** https://doi.org/10.1007/s11192-007-2046-8

Abramo, G., D'Angelo, C. A., & Caprasecca, A. (2009b). The contribution of star scientists to sex differences in research productivity. *Scientometrics*, 81, 137. **DOI:** https://doi.org/10.1007/s11192-008-2131-7

Abramo, G., D'Angelo, C. A., & Rosati, F. (2015). Selection committees for academic recruitment: Does gender matter? *Research Evaluation*, 24(4), 392–404. **DOI:** https://doi.org/10.1093/reseval/rvv019

Allen-Hermanson, S. (2017). Leaky pipeline myths: In search of gender effects on the job market and early career publishing in philosophy. *Frontiers in Psychology*, 8, 953. **DOI:** https://doi.org/10.3389/fpsyg.2017.00953, **PMID:** 28659843, **PMCID:** PMC5469890

Aycock, L. M., Hazari, Z., Brewe, E., Clancy, K. B. H., Hodapp, T., & Goertzen, R. M. (2019). Sexual harassment reported by undergraduate female physicists. *Physical Review Physics Education Research*, 15, 010121. **DOI:** https://doi.org/10.1103/PhysRevPhysEducRes.15.010121

Barthelemy, R. S., McCormick, M., & Henderson, C. (2016). Gender discrimination in physics and astronomy: Graduate student experiences of sexism and gender microaggressions. *Physical Review Physics Education Research*, 12, 020119. **DOI:** https://doi.org/10.1103/PhysRevPhysEducRes.12.020119

Birnholtz, J. P. (2006). What does it mean to be an author? The intersection of credit, contribution, and collaboration in science. *Journal of the American Society for Information Science and Technology*, 57(13), 1758–1770. **DOI:** https://doi.org/10.1002/asi.20380

Bordons, M., Morillo, F., Fernandez, M. T., & Gomez, I. (2003). One step further in the production of bibliometric indicators at the micro level: Differences by gender and professional category of scientists. *Scientometrics*, 57(2), 159–173. **DOI:** https://doi.org/10.1023/A:1024181400646

Bornmann, L., & Daniel, H. D. (2008). What do citation counts measure? A review of studies on citing behavior. *Journal of Documentation*, 64(1), 45–80. **DOI:** https://doi.org/10.1108/00220410810844150

Borsuk, R., Aarssen, L. W., Budden, A. E., Koricheva, J., Leimu, R., … Lortie, C. J. (2009). To name or not to name: The effect of changing author gender on peer review. *BioScience*, 59(11), 985–989. **DOI:** https://doi.org/10.1525/bio.2009.59.11.10

Breda, T., & Napp, C. (2019) Girls' comparative advantage in reading can largely explain the gender gap in math-related fields. *Proceedings of the National Academy of Sciences of the United States of America*, 116(31), 15435–15440. **DOI:** https://doi.org/10.1073/pnas.1905779116, **PMID:** 31308226, **PMCID:** PMC6681723

Cameron, E. Z., White, A. M., & Gray, M. E. (2016). Solving the productivity and impact puzzle: Do men outperform women, or are metrics biased? *BioScience*, 66(3), 245–252. **DOI:** https://doi.org/10.1093/biosci/biv173

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

Ceci, S. J., Ginther, D. K., Kahn, S., & Williams, W. M. (2014). Women in academic science: A changing landscape. *Psychological Science in the Public Interest*, 15(3), 75–141. **DOI:** https://doi.org/10.1177/1529100614541236, **PMID:** 26172066

Ceci, S. J., & Williams, W. M. (2011) Understanding current causes of women's underrepresentation in science. *Proceedings of the National Academy of Sciences of the United States of America*, 108(8), 3157–3162. **DOI:** https://doi.org/10.1073/pnas.1014871108, **PMID:** 21300892, **PMCID:** PMC3044353

Ceci, S. J., & Williams, W. M. (2015). Women have substantial advantage in STEM faculty hiring, except when competing against more-accomplished men. *Frontiers in Psychology*, 6, 1532. **DOI:** https://doi.org/10.3389/fpsyg.2015.01532, **PMID:** 26539132, **PMCID:** PMC4611984

Chen, P., Xie, H., Maslov, S., & Redner, S. (2007). Finding scientific gems with Google's PageRank algorithm. *Journal of Informetrics*, 1(1), 8–15. **DOI:** https://doi.org/10.1016/j.joi.2006.06.001

Cole, J. R., & Zuckerman, H. (1987). Marriage, motherhood, and research performance in science. *Scientific American*, 256(2), 119–125. **DOI:** https://doi.org/10.1038/scientificamerican0287-119, **PMID:** 3810117

Deary, I. J., Irwing, P., Der, G., & Bates, T. C. (2007). Brother-sister differences in the *g* factor in intelligence: Analysis of full, opposite-sex siblings from the NLSY1979. *Intelligence*, 35(5), 451–456. **DOI:** https://doi.org/10.1016/j.intell.2006.09.003

Diekman, A. B., Johnston, A. M., & Clark, E. K. (2010). Seeking congruity between goals and roles: A new look at why women opt out of science, technology, engineering, and mathematics careers. *Psychological Science*, 21(8), 1051–1057. **DOI:** https://doi.org/10.1177/0956797610377342, **PMID:** 20631322

Eaton, A. A., Saunders, J. F., Jacobson, R. K., & West, K. (2019). How gender and race stereotypes impact the advancement of scholars in STEM: Professors' biased evaluations of physics and biology post-doctoral candidates. *Sex Roles*, 82, 127–141. **DOI:** https://doi.org/10.1007/s11199-019-01052-w

Edwards, H. A., Schroeder, J., & Dugdale, H. L. (2018). Gender differences in authorships are not associated with publication bias in an evolutionary journal. *PLOS ONE*, 29, e0201725. **DOI:** https://doi.org/10.1371/journal.pone.0201725, **PMID:** 30157231, **PMCID:** PMC6114708

Evans, C. D., & Diekman, A. B. (2009). On motivated role selection: Gender beliefs, distant goals, and career interest. *Psychology of Women Quarterly*, 33(2), 235–249. **DOI:** https://doi.org/10.1111/j.1471-6402.2009.01493.x

Flaherty, K. (2018). The leaky pipeline for postdocs: A study of the time between receiving a PhD and securing a faculty job for male and female astronomers. *arXiv*, arXiv:1810.01511.

Fox, M. F. (1995). Women in scientific careers. In S. Jasanoff, G. E. Markle, J. C. Peterson, & T. Pinch (Eds.), *Handbook of science and technology studies*. Thousand Oaks, CA: Sage.

Fox, M. F. (2005). Gender, family characteristics, and publication productivity among scientists. *Social Studies of Science*, 35(1), 131–150. **DOI:** https://doi.org/10.1177/0306312705046630

Ghiasi, G., Larivière, V., & Sugimoto, C. R. (2016). Gender differences in synchronous and diachronous self-citations. *21st International Conference on Science and Technology Indicators*, Valencia.

Gibney, E. (2007). Teaching load could put female scientists at career disadvantage. *Nature Trend Watch*. **DOI:** https://doi.org/10.1038/nature.2017.21839

Ginther, D. K., & Kahn, S. (2009). Does science promote women? Evidence from Academia 1973–2001. In R. B. Freeman & D. L. Goroff (Eds.), *Science and engineering careers in the United States*. National Bureau of Economic Research Conference Report. Chicago, IL: University of Chicago Press.

Ginther, D. K., & Kahn, S. (2015). Comment on expectations of brilliance underlie gender distributions across academic disciplines. *Science*, *349*(6246), 391. **DOI:** https://doi.org/10.1126/science.aaa9632, **PMID:** 26206926

Glass, C., & Minnotte, K. (2010). Recruiting and hiring women in STEM fields. *Journal of Diversity in Higher Education*, *3*(4), 218–229. **DOI:** https://doi.org/10.1037/a0020581

Guarino, C. M., & Borden, V. M. H. (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

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

Handley, M., Brown, E. R., Moss-Racusin, C. A., & Smith, J. L. (2015). Quality of evidence revealing subtle gender biases in science is in the eye of the beholder. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(43), 13201–13206. **DOI:** https://doi.org/10.1073/pnas.1510649112, **PMID:** 26460001, **PMCID:** PMC4629390

Hargens, L. L., McCann, J. C., & Reskin, B. F. (1978). Productivity and reproductivity: Fertility and professional achievement among research scientists. *Social Forces*, *57*(1), 154–163. **DOI:** https://doi.org/10.2307/2577631

Hill, T. P. (2017). An evolutionary theory for the variability hypothesis. *ArXiv*, arXiv:1703.04184.

Holman, L., Stuart-Fox, D., Hauser, C. E., & Sugimoto, C. (2018). The gender gap in science: How long until women are equally represented? *PLOS Biology*, *16*, e2004956. **DOI:** https://doi.org/10.1371/journal.pbio.2004956, **PMID:** 29672508, **PMCID:** PMC5908072

Hooydonk, G. V. (1997). Fractional counting of multiauthored publications: Consequences for the impact of authors. *Journal of the American Society for Information Science*, *48*(10), 944–945. **DOI:** https://doi.org/10.1002/(SICI)1097-4571(199710)48:10<944::AID-ASI8>3.0.CO;2-1

Hossenfelder, S. (2018). *Do women get fewer citations than men?* Talk at Chapman University, November 11, 2018.

Hyde, J. S. (2014). Gender similarities and differences. *Annual Review of Psychology*, *65*, 373–398. **DOI:** https://doi.org/10.1146/annurev-psych-010213-115057, **PMID:** 23808917

InSpire. (2010). *High-energy physics literature INSPIRE database* (A. Holtkamp, S. Mele, T. Simko, and T. Smith (INSPIRE collaboration)).

Ioannidis, J. P. A., Baas, J., Klavans, R., & Boyack, K. W. (2019). A standardized citation metrics author database annotated for scientific field. *PLOS Biology*, *17*, e3000384. **DOI:** https://doi.org/10.1371/journal.pbio.3000384, **PMID:** 31404057, **PMCID:** PMC6699798

Jagsi, R., Guancial, E. A., Worobey, C. C., Henault, L. E., Chang, Y., … Hyek, E. M. (2006). The gender gap in authorship of academic medical literature—A 35-year perspective. *New England Journal of Medicine*, *355*(3), 281–287. **DOI:** https://doi.org/10.1056/NEJMsa053910, **PMID:** 16855268

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*. **DOI:** https://doi.org/10.1177/2378023117738903

Knapp, M. (2010). Are participation rates sufficient to explain gender differences in chess performance? *Proceedings of the Royal Society B: Biological Sciences*, *277*(1692), 2269–2270. **DOI:** https://doi.org/10.1098/rspb.2009.2257, **PMID:** 20335203, **PMCID:** PMC2894898

Kwiek, M. (2016). The European research elite: A cross-national study of highly productive academics in 11 countries. *Higher Education*, *71*, 379–397. **DOI:** https://doi.org/10.1007/s10734-015-9910-x

Larivière, V., Ni, C., Gingras, Y., Cronin, B., & Sugimoto, C. R. (2013). Bibliometrics: Global gender disparities in science. *Nature*, *504*, 211–213. **DOI:** https://doi.org/10.1038/504211a, **PMID:** 24350369

Levin, S. G., & Stephan, P. E. (1998). Gender differences in the rewards to publishing in academe: Science in the 1970s. *Sex Roles*, *38*, 1049–1064. **DOI:** https://doi.org/10.1023/A:1018882711314

Ley, T., & Hamilton, B. (2008). The gender gap in NIH grant applications. *Science*, *322*(5907), 1472–1474. **DOI:** https://doi.org/10.1126/science.1165878, **PMID:** 19056961

Leydesdorff, L., & Park, H. W. (2016). Full and fractional counting in bibliometric networks. *arXiv*, arXiv:1611.06943.

Lippa, R. A. (2010). Gender differences in personality and interests: When, where, and why? *Social and Personality Psychology Compass*, *4*(11), 1098–1110. **DOI:** https://doi.org/10.1111/j.1751-9004.2010.00320.x

Ma, N., Guan, J., & Zhao, Y. (2008). Bringing PageRank to the citation analysis. *Information Processing & Management*, *44*(2), 800–810. **DOI:** https://doi.org/10.1016/j.ipm.2007.06.006

Macaluso, B., Larivière, V., Sugimoto, T., & Sugimoto, C. R. (2016). Is science built on the shoulders of women? A study of gender differences in contributorship. *Academic Medicine*, *91*(8), 1136–1142. **DOI:** https://doi.org/10.1097/ACM.0000000000001261, **PMID:** 27276004

Marsh, H. W., Jayasinghe, U. W., & Bond, N. W. (2011). Gender differences in peer reviews of grant applications: A substantive-methodological synergy in support of the null hypothesis model. *Journal of Informetrics*, *5*(1), 167–180. **DOI:** https://doi.org/10.1016/j.joi.2010.10.004

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

Miller, D. I., & Wai, J. (2015). The bachelor's to Ph.D. STEM pipeline no longer leaks more women than men: A 30-year analysis. *Frontiers in Psychology*, *6*, 37. **DOI:** https://doi.org/10.3389/fpsyg.2015.00037, **PMID:** 25741293, **PMCID:** PMC4331608

Moldwin, M. B., & Liemohn, M. W. (2018). High-citation papers in space physics: Examination of gender, country, and paper characteristics. *Journal of Geophysical Research*, *123*(4), 2557–2565. **DOI:** https://doi.org/10.1002/2018JA025291

Moss-Racusin, C. A., Dovidio, J. F., Brescoll, V. L., Graham, M. J., & Handelsman, J. (2012). Science faculty subtle gender biases favor male students. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(41), 16474–16479. **DOI:** https://doi.org/10.1073/pnas.1211286109, **PMID:** 22988126, **PMCID:** PMC3478626

Mutz, R., Bornmann, L., & Daniel, H. D. (2012). Does gender matter in grant peer review? An empirical investigation using the example of the Austrian Science Fund. *Zeitschrift für Psychologie*, *220*(2), 121–129. **DOI:** https://doi.org/10.1027/2151-2604/a000103, **PMID:** 23480982, **PMCID:** PMC3414231

NASEM Report. (2018). *Sexual harassment of women*. US National Academies of Science, Engineering and Medicine.

National Research Council. (2009). *Gender differences at critical transitions in the careers of science, engineering and mathematics faculty*. National Academies Press.

O'Dea, R. E., Lagisz, M., Jennions, M. D., & Nakagawa, S. (2018). Gender differences in individual variation in academic grades fail to fit expected patterns for STEM. *Nature Communications*, *9*(1), 3777. **DOI:** https://doi.org/10.1038/s41467-018-06292-0, **PMID:** 30254267, **PMCID:** PMC6156605

Oswald, F. L., Mitchell, G., Blanton, H., Jaccard, J., & Tetlock, P. (2013). Predicting ethnic and racial discrimination: A meta-analysis of IAT criterion studies. *Journal of Personality and Social Psychology*, *105*(2), 171–192. **DOI:** https://doi.org/10.1037/a0032734, **PMID:** 23773046

Perianes-Rodriguez, A., Waltman, L., & van Eck, N. J. (2016). Constructing bibliometric networks: A comparison between full and fractional counting. *Journal of Informetrics*, *10*(4), 1178–1195. **DOI:** https://doi.org/10.1016/j.joi.2016.10.006

Perley, D. A. (2019). Gender and the career outcomes of PhD astronomers in the United States. *Publications of the Astronomical Society of the Pacific*, *131*(1005), 114502. **DOI:** https://doi.org/10.1088/1538-3873/ab0cc4

Pinski, G., & Narin, F. (1976). Citation influence for journal aggregates of scientific publications: Theory, with application to the literature of physics. *Information Processing & Management*, *12*(5), 297–312. **DOI:** https://doi.org/10.1016/0306-4573(76)90048-0

Porter, A. M., & Ivie, R. (2019). *Women in physics and astronomy*. American Institute of Physics.

Radicchi, F., Fortunato, S., Markines, B., & Vespignani, A. (2009). Diffusion of scientific credits and the ranking of scientists. *Physical Review*, *E80*, 056103. **DOI:** https://doi.org/10.1103/PhysRevE.80.056103, **PMID:** 20365039

Raffo, J. (2016). *Worldwide gender-name dictionary*. WIPO Economics & Statistics Related Resources 10, World Intellectual Property Organization.

Rasmussen, K. C., Maier, E., Strauss, B. E., Durbin, M., Riesbeck, L., … Erena, A. (2019). The nonbinary fraction: Looking towards the future of gender equity in astronomy. *arXiv*, arXiv:1907.04893.

Rossi, P., & Strumia, A., & Torre, A. (2019). Bibliometrics for collaboration works. *arXiv*, arXiv:1902.01693.

Ruocco, G., Daraio, C., Folli, V., & Leonetti, M. (2017). Bibliometric indicators: The origin of their log-normal distribution and why they are not a reliable proxy for an individual scholar's talent. *Palgrave Communications*, *3*, 17064. **DOI:** https://doi.org/10.1057/palcomms.2017.64

Sax, L. J., Hagedorn, L. S., Arredondo, M., & Dicrisi, F. A. (2002). Faculty research productivity: Exploring the role of gender and family-related factors. *Research in Higher Education*, *43*, 423–446. **DOI:** https://doi.org/10.1023/A:1015575616285

Sinatra, R., Deville, P., Szell, M., Wang, D., & Barabasi, A. L. (2005). A century of physics. *Nature Physics*, *11*, 791–796. **DOI:** https://doi.org/10.1038/nphys3494

Stack, S. (2004). Gender, children and research productivity. *Research in Higher Education*, *45*, 891–920. **DOI:** https://doi.org/10.1007/s11162-004-5953-z

Stevens, S., & Haidt, J. (2017). The Google memo: What does the research say about gender differences? *Heterodox: The Blog*, August 10. https://heterodoxacademy.org/blog/the-google-memo-what-does-the-research-say-about-gender-differences/.

Stewart-Williams, S., Thomas, A. G., Blackburn, J. D., & Chan, C. Y. M. (2019). Reactions to male-favoring vs. female-favoring sex differences: A preregistered experiment. *British Journal of Psychology*. **DOI:** https://doi.org/10.1111/bjop.12463, **PMID:** 32701171

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

Strumia, A. (2019). Why are women under-represented in physics? *Quillette*, April 16. https://quillette.com/2019/04/16/why-are-women-under-represented-in-physics/.

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

Su, R., Rounds, J., & Armstrong, P. I. (2009). Men and things, women and people: A meta-analysis of sex differences in interests. *Psychological Bulletin*, *135*(6), 859–884. **DOI:** https://doi.org/10.1037/a0017364, **PMID:** 19883140

Su, R., & Rounds, J. (2015). All STEM fields are not created equal: People and things interests explain gender disparities across STEM fields. *Frontiers in Psychology*, *6*, 189. **DOI:** https://doi.org/10.3389/fpsyg.2015.00189, **PMID:** 25762964, **PMCID:** PMC4340183

Tahamtan, I., & Bornmann, L. (2019). What do citation counts measure? An updated review of studies on citations in scientific documents published between 2006 and 2018. *arXiv*, arXiv:1906.04588. **DOI:** https://doi.org/10.1007/s11192-019-03243-4

Thelwall, M. (2018). Do females create higher impact research? Scopus citations and Mendeley readers for articles from five countries. *Journal of Informetrics*, *12*(4), 1031–1041. **DOI:** https://doi.org/10.1016/j.joi.2018.08.005

Thelwall, M., Bailey, C., Tobin, C., & Bradshaw, N.-A. (2019). Gender differences in research areas, methods and topics: Can people and thing orientations explain the results? *Journal of Informetrics*, *13*(1), 149–169. **DOI:** https://doi.org/10.1016/j.joi.2018.12.002

Thelwall, M., & Wilson, P. (2014). Regression for citation data: An evaluation of different methods. *Journal of Informetrics*, *8*, 963. **DOI:** https://doi.org/10.1016/j.joi.2014.09.011

Torvik, V. I., & Agarwal, S. (2016). ETHNEA—An instance-based ethnicity classifier based on geo-coded author names in a large-scale bibliographic database. *International Symposium on Science of Science*, March 22–23.

Waltman, L. (2016). A review of the literature on citation impact indicators. *Journal of Informetrics*, *10*(2), 365–391. **DOI:** https://doi.org/10.1016/j.joi.2016.02.007

Wang, M. T., Eccles, J., & Kenney, S. (2013). Not lack of ability but more choice: Individual and gender differences in choice of careers in science, technology, engineering, and mathematics. *Psychological Science*, *24*(5), 770–775. **DOI:** https://doi.org/10.1177/0956797612458937, **PMID:** 23508740

Way, S. F., Larremore, D. B., & Clauset, A. (2016). Gender, productivity, and prestige in computer science faculty hiring networks. *Proceedings of the 25th International Conference on World Wide Web* (pp. 1169–1179). New York: Association for Computing Machinery. **DOI:** https://doi.org/10.1145/2872427.2883073

Wennerås, C., & Wold, A. (1997). Nepotism and sexism in peer-review. *Nature*, *387*, 341–343. **DOI:** https://doi.org/10.1038/387341a0, **PMID:** 9163412

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*, e66212. **DOI:** https://doi.org/10.1371/journal.pone.0066212, **PMID:** 23894278, **PMCID:** PMC3718784

West, J. D., Jensen, M. C., Dandrea, R. J., Gordon, G. J., & Bergstrom, C. T. (2014). Author-level eigenfactor metrics:

Evaluating the influence of authors, institutions, and countries within the social science research network community. *Journal of the American Society for Information Science and Technology*, *64*(4), 787–801. **DOI:** https://doi.org/10.1002/asi.22790

Williams, W. M., & Ceci, S. J. (2017). National hiring experiments reveal 2:1 faculty preference for women on STEM tenure track. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(17), 5360–5365. **DOI:** https://doi.org/10.1073/pnas.1418878112, **PMID:** 25870272, **PMCID:** PMC4418903

Winegard, B., Clark, C. J., Hasty, C., & Baumeister, R. (2018). Equalitarianism: A source of liberal bias. *SSRN Electronic Journal*. **DOI:** https://doi.org/10.2139/ssrn.3175680

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. *Lancet*, *393*(10171), 531–540. **DOI:** https://doi.org/10.1016/S0140-6736(18)32611-4

Wolfinger, N. H., Mason, M. A., & Goulden, M. (2008). Problems in the pipeline: Gender, marriage, and fertility in the ivory tower. *Journal of Higher Education*, *79*(4), 388–405. **DOI:** https://doi.org/10.1080/00221546.2008.11772108

Wolfram Mathematica. (n.d.). *Determine the gender of given name*. https://www.wolfram.com/mathematica/new-in-10/built-in-classifier-collection/determine-the-gender-of-given-name.html.

World Economic Forum. (2016). *The global gender gap report*. World Economic Freedom.

Xie, Y., & Shauman, K. A. (1998). Sex differences in research productivity: New evidence about an old puzzle. *American Sociological Review*, *63*(6), 847–870. **DOI:** https://doi.org/10.2307/2657505

Xie, Y., & Shauman, K. A. (2003). *Women in science: Career processes and outcomes*. Cambridge, MA: Harvard University Press.

Zitt, M., & Small, H. (2008). Modifying the journal impact factor by fractional citation weighting: The audience factor. *Journal of the American Society for Information Science and Technology*, *59*(11), 1856–1860. **DOI:** https://doi.org/10.1002/asi.20880