

Bibliometric analysis of gender issues in fundamental physics

Alessandro Strumia

Abstract

We extract bibliometric data about gender issues in fundamental physics. We do not find significant gender differences in hiring rates, hiring timing, career gaps, citation patterns. We find the well known gender difference in representation, mildly affected by a higher abandonment rate of female authors. 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 shape that can be fitted by different variabilities.

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, that accumulated bibliometric data about fundamental physics world-wide from ~ 1970 to now [InSPIRE(2010)]. Fundamental physics is a sub-sector of physics that deals with the fundamental aspects of the field and that presently mostly focuses on particle physics, cosmology and astrophysics, from an experimental and theoretical point of view.

Such bibliometric data are being used to study various aspect of the field. We here focus on gender issues: physics exhibits persisting differences and attracts students with high average grades (see e.g. fig. 9 of [Ceci et al.(2014)] and fig. 1 of [Ginther et al.(2015)]). The bibliometric approach relies on large amounts of objective quantitative data about papers, authors, citations and hires. Having enough statistics sometimes is crucial to reveal effects, and to go beyond simple counting by devising dedicated analyses that target specific questions. We can do this, as we have the full database, not just access to some pre-defined metrics. Having a large amount of new data we will follow a data-driven approach.

A vast literature studied gender differences in STEM (Science, Technology, Engineering, and Mathematics), although no previous studies specifically focused on fundamental physics: the present study will fill this gap. A main theme is understanding why women remain under-represented in STEM fields, a worldwide phenomenon persisting since decades, despite interventions on its alleged social causes [Stoet et al.(2018)].

A limitation of bibliometric indicators is that authors start being scientifically active roughly at PhD level: in physics (as in other STEM fields) a low female representation is already present at this entry level of bibliometric data. Earlier phases need to be explored with other tools. Surveys of occupational plans and first choices of high-school students already reveal a gender difference in the same direction [Xie et al.(2003)], possibly mainly due to gender differences in interests [Ceci et al.(2014), Su et al.(2009), Lippa (2010), Hyde (2014), Su et al.(2015), Thelwall (2018)b, Stoet et al.(2018)], with a sub-leading role played by differences in relative attitudes (girls with high math ability tend to also have high verbal ability) [Wang et al.(2013), Ceci et al.(2014), Stoet et al.(2018)].

Coming to the later phase covered by bibliometrics, initial small-scale experiments and anecdotal reports suggested biases against hypothetical female applicants (see e.g. [Wenneras et al.(1997), Moss-Racusin et al.(2012)]). These findings have not been supported by more recent larger-scale experiments (see the review in [Ceci et al.(2011)] and [Williams et al.(2017), Ceci et al.(2015)]). No significant biases have been found in examined real grant evaluations [Ceci et al.(2014), Marsh et al.(2011), Ley et al.(2008), Mutz et al.(2012)] and referee reports of journals [Borsuk et al.(2009), Ceci et al.(2014), Edwards et al.(2018)]; the gender composition of applicants [Way et al.(2016)] and panels [Abramo et al. (2018)] has little effects. Real hires show a higher success rates among women [National Research Council, Wolfinger et al.(2008), Glass et al.(2010), Ceci et al.(2014)], especially in those STEM fields where women are less represented [Ceci et al.(2014)]. Bibliometric attempts of recognising higher merit [Ceci et al.(2014)] found that male faculties write more papers [Xie et al.(1998), Levin

et al.(1998), Fox (2005), Abramo et al.(2009), Lariviere et al.(2013), Way et al.(2016)] (see also [Holman et al.(2018), Thelwall (2018)]), predominate among first and last authors (prestigious in some fields) and in single-authored papers [West et al. (2013), Jagsi et al. (2006)]. Such gender productivity gap persists after accounting for confounding factors such as seniority [Ceci et al.(2014), Caplar et al.(2017), Moldwin et al. (2018)]. Some studies observed a small group of extremely productive, mostly male, ‘star authors’ [Bordons et al. (2003), Abramo et al.(2009)b, Abramo et al.(2015)]. 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)].¹

This paper is structured as follows.

In section 2 we describe how we identify the gender of authors; how we obtain lists of hires; how we combine citations to define bibliometric indicators which can be used as reliable proxies for scientific merit, being significantly correlated to human evaluations such as scientific prizes.

In section 3 we 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. We find that this entry difference in representation is not significantly affected by hiring, consistently with [Ceci et al.(2014)]. As we use citations, in section 3.1 we first verify that citations encode the common opinion of male (M) and female (F) authors. We 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 we then compare reliable bibliometric indices based on citations finding that F authors are hired with indices which 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 fraction of M authors progressively grows going from average to top-authors, consistently with [Bordons et al. (2003), Abramo et al.(2009)b, Abramo et al.(2015)]. As bibliometric data are influenced by a complicated background of social and historical accidents (in particular the growing fraction of female physicists), in the Appendix we show that the results above persist after taking confounders into account.

In section 4 we conclude discussing possible interpretations of the data.

2 Methods

The public INSPIRE database [InSPIre(2010)] maintained by CERN offers a picture of fundamental physics world-wide from ~ 1970 to now. INSPIRE gives data on about one million of scientific papers, 30 million of references, 71104 identified authors in 7 thousands of institutes. INSPIRE individually identified the main authors, solving the problem of name disambiguation. “Boundaries” of the data-base play a minor role, given that fundamental physics presently is a self-contained specialised subject [Sinatra et al.(2005)]. On the other hand, various authors work on multiple topics within the field, that cannot be sharply sub-divided.

¹For simplicity, we avoided summarising how the cited studies are limited within countries and/or STEM fields. The discussion above does not provide a complete panoramic; see [Ceci et al.(2014)] for a recent exhaustive review.

As no gender information is provided, in section 2.1 we describe our procedure to infer gender from names and nationality of the authors. In section 2.2 we describe how we obtain lists of hires in fundamental physics world-wide. Section 2.3 motivates the bibliometric index that we will use to indicate scientific merit.

2.1 Name-gender association

We need to infer gender from names in an accurate and complete way. 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 like Nicola are “ambiguous”: they correspond to different genders in different countries. Third, some authors have unusual names. For example, the `Mathematica` machine learning function `Classify` [Wolfram et al.] uses information about the first name only and leaves about 40% of authors with unclassified gender.

We tested two approaches, in order to determine their strengths and choose the best combination:

1. First, we run the on-line ETHNEA [Torvik et al.(2016)] tool, which uses the full name (first and family name) to infer both gender and ethnicity, and leaves 26% of the authors with unclassified gender.
2. Second, for each author we extract a “guessed” nationality from the earlier affiliations in his/her list of 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) [WGND]. This database contains 175917 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 $\sim 3\%$ of the total, and the uncertainty induced by this procedure is below the percent level. About 0.1% of the authors have “ambiguous” names and no nationality information: we match them to the most common gender corresponding to their name, defined as the one used in the largest number of countries. 23% of the authors remain unclassified.

The results discussed in the following are affected in a minor way using the ETHNEA or the WGND classification. By comparing them we see that that the ETHNEA classification is less complete, leaving unclassified more authors with unusual names. On the other hand, the WGND classifications leads to some authors with misidentified gender, typically arising due to a misidentification of their nationality. We find different genders for 1.8% of all identified authors; with a 5% percentage among Chinese,² Indian and Korean authors, and a 1% percentage among European authors.

²Gender can be reliably extracted from Chinese names only when they are written in Chinese characters: this information is not always provided by INSPIRE.

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

2.2 Hiring

INSPIRE is integrated with HEPNAMES, a data base with profiles of the various authors. As an example the internet page inspirehep.net/author/profile/A.Strumia.1 shows the profile of the present author. From HEPNAMES we obtain a data-base of about 10000 first hires in fundamental physics world-wide, including dates and disambigued institutions. While INSPIRE is a widely used tool in the community, precise career information is only added by authors on a voluntary basis. If F and M authors tend to do this differently, INSPIRE hires might be biased.

We therefore complement INSPIRE hires by computing unbiased “pseudo-hires” defined as follows: we consider an author as p yr-hired when he/she starts writing papers with the same affiliation for at least p years. Using this definition we obtain a database of about 40000/19000 5/10yr-first hires from 1960 to 2013/2008 (64000/23000 including multiple hires for the same author). However, in this way we cannot obtain a precise hiring date for the sub-set of authors hired by the same institution to which they were previously affiliated.

Thereby we will use INSPIRE hires when a precise hiring date is more important than increased statistics, and pseudo-hires in the opposite situation, when a full coverage is more important than a precise timing. In any case, the other sample will be used as a control sample.

2.3 Bibliometrics

It is important to adopt a bibliometric index that is a valid proxy for what is commonly considered as scientific merit: for example it must give top authors highly correlated to scientific prizes.

Traditional metrics (such as citation counts, h index, paper counts) fail to fulfil this role in fundamental physics [Strumia et al.(2018)]. The main reason is, by far, the presence of very productive (up to 6000 papers) large collaborations (up to 3000 authors), mostly in high-energy experimental physics. Signing more papers than what one can read stretches the concept of authorship [Birnholtz et al.(2006)]. The practical problem is that the contribution of big collaborations overwhelms the data-base, if 6000 papers are counted as 3000×6000 .

This situation can be corrected by “fractional counting” [Hooydonk (1997), Perianes-Rodriguez et al.(2016), Leydesdorff et al.(2016)]: a fraction $1/N_{\text{aut}}$ of each paper is equally attributed to its N_{aut} authors (as appropriate for an intensive quantity), rather than summed over them. All authors, including first and last authors, are treated on equal footing because authors are usually sorted alphabetically in fundamental physics, unlike what happens in other fields.³ Adopting

³The first author is not alphabetically sorted in 6% of the papers in the hep-th arXiv bulletin, 13% in hep-ph and hep-ex, 18% in hep-lat, 25% in gr-qc, 44% in astro-ph. In more papers the author highlighted as first might

fractional counting, the total bibliometric output of collaborations scales, on average, as their number of authors [Rossi et al.(2019)], suggesting that large collaborations form when needed and that gift authorship does not play a large role.

Fractional counting of citations provides a list of top-authors in fundamental physics highly correlated with those who received scientific prizes [Strumia et al.(2018)]. Thereby fractionally-counted citations are one simple reliable indicator.

We will actually use “individual citations” $N_{\text{icit}} = N_{\text{cit}}/N_{\text{aut}}N_{\text{ref}}$ (summed over all citing papers) which give less value to citations coming from papers with a larger number N_{ref} of references. This extra refinement only makes minor differences, such as mildly penalising sectors (like phenomenology) where papers tend to have more references [Strumia et al.(2018)]. The resulting list of top authors is shown in table 6 of [Strumia et al.(2018)].

The use of bibliometric indices based on citation counts as a proxy for scientific merit comes with limitations and dangers. On short time-scales 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... Individual citations are not boosted by the first two strategies. As we are concerned with gender differences, it is reassuring that fig. A.1 shows no gender differences in self-referencing.

Since “when a measure becomes a target, it ceases to be a good measure”, in the Appendix we consider a metric more different from common targets, which is not enhanced by the latter three strategies. The ‘CitationCoin’ \mathcal{C} is defined as the difference between the number of received and given individual citations (up to a factor that prevents systematically negative contributions from recent papers), such that it is not affected by self-citations not by networks of circular citations [Strumia et al.(2018)]. Authors that write too many small papers can even get a negative \mathcal{C} score.

Bibliometric indicators measure the average opinion of the community: while all opinions can be wrong, a better possibility could be relying on the opinion of top-authors. This is done by metrics based on the PageRank algorithm (such as those discussed in [Pinski et al.(1976), Chen et al.(2007), Ma et al.(2008), Radicchi et al.(2009), West et al.(2014), Strumia et al.(2018)]). This is studied in the Appendix, together with 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 indeces among authors are so large that log-scale plots will be appropriate and refined metrics only make minor differences. We use individual citations N_{icit} because this metric is simpler and closer to the commonly used number of citations N_{cit} , while allowing 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.

accidentally be also alphabetically first.

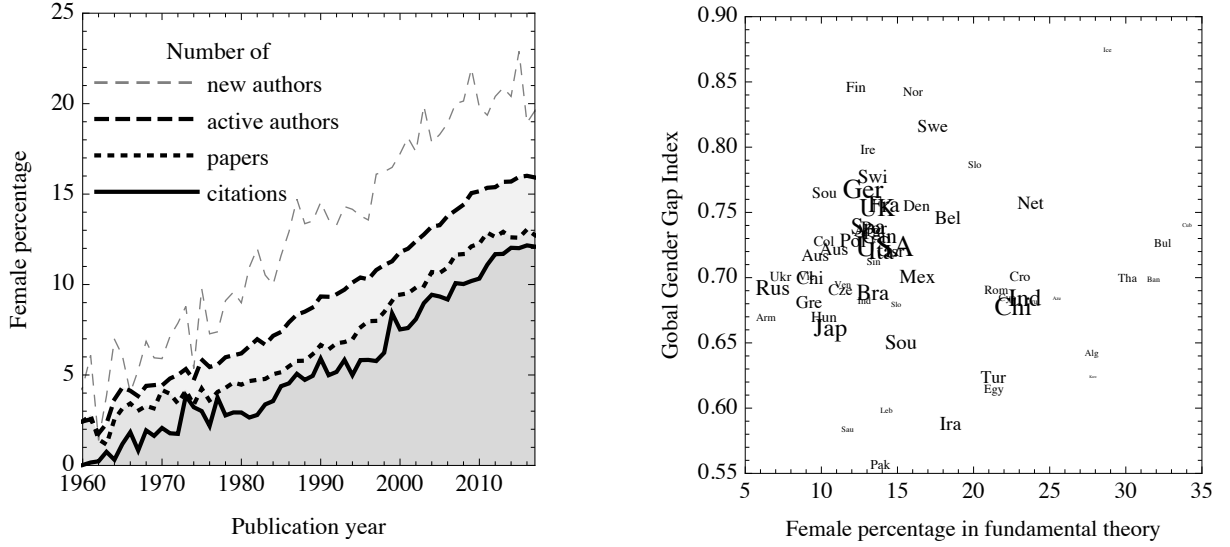


Figure 1: **Left:** percentage female contribution to the number of new authors, of authors that wrote at least a paper during the year, to the number of fractionally-counted papers, and to the number of received individual citations, considering the papers written each year. **Right:** the percentage of F authors in fundamental theory is not positively correlated with the Global Gender Gap Index of the country [World Economic Forum].

3 Results

Among the authors in fundamental physics listed in the INSPIRE data-base, 16% of those 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 et al.(2018)] the field significantly expanded: about half of citations have been given after 2000, so that metrics based on (individual) citations favour recent authors. Next, the F percentage grew with time as shown by the raw data in the left panel of fig. 1.

The right panel shows a mild gender geographical variation within countries that most contributed to fundamental physics. It is interesting to explore if the female fraction is correlated with the Global Gender Gap Index (GGGI) of the countries [World Economic Forum] which measures the gap between women and men in education, politics, health, 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 favour of women (as the GGGI ignores imbalances to the advantage of women). The right panel of fig. 1 shows that the female fraction is not positively correlated with the GGGI, as similarly observed among students in STEM [Stoet et al.(2018)].

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

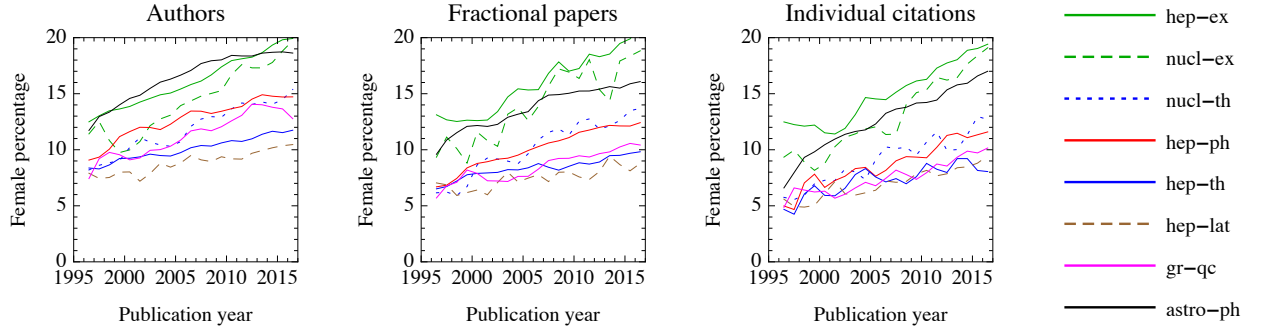


Figure 2: As in fig. 1, showing after 1995 the result within the main arXiv categories, plotted as colored curves: experimental categories include *hep-ex* (high-energy experiments) and *nucl-ex* (nuclear experiments). Theoretical categories include *hep-ph* (high-energy phenomenology), *hep-th* (high-energy theory), *hep-lat* (lattice), *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.

Clearly, the field and its gender composition evolved in the past 50 years. While describing such changes from a bibliometric point of view is an interesting subject, we try focusing on general features which emerge from the complicated background of social factors. This will need taking into account possible confounders, by studying sub-periods and sub-topics or by trying to compensate for the above variations.

3.1 Citations

We 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 more often a given gender.

In line of 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 requires relying on models of citations, which are affected by questionable systematic issues. One can try controlling for main factors (such as different numbers of M and F authors, different average seniorities, regional differences, etc), but reality can contain more complicated effects such as different scientific qualities. For example, [Caplar et al.(2017)] claim (consistently with our 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”.

We will follow a different strategy, which is often more useful in the presence of backgrounds that cannot be reliably modelled: restricting the attention to an asymmetry constructed such

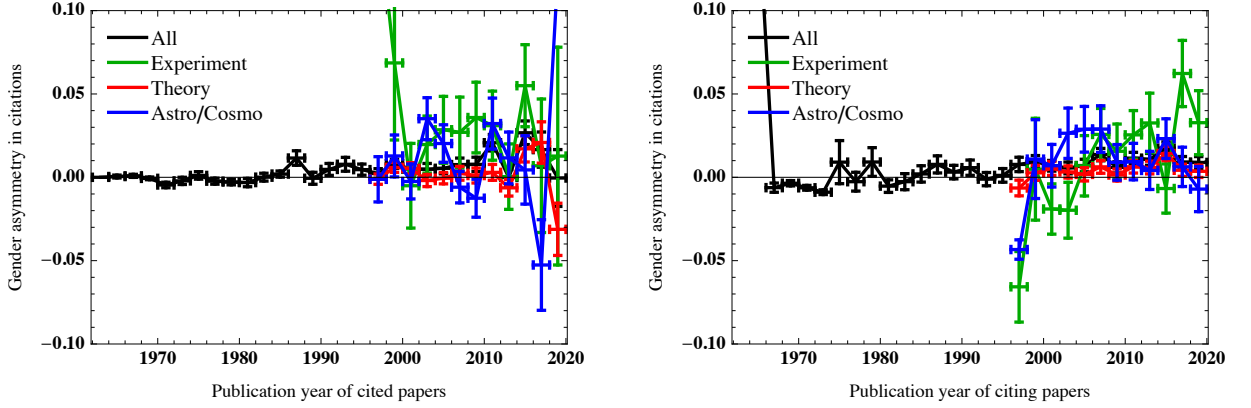


Figure 3: *Time evolution of the gender asymmetry defined in eq. (1); $A > 0$ ($A < 0$) signals same-gender (opposite-gender) preference. **Left:** As function of the publication year of the cited single-author papers. **Right:** As function of the publication year of the citing single-author papers. After 1995 we 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 (astro-ph). The bin 2018-20 only uses data available up to mid 2018.*

that it is not affected by the backgrounds. The extracted information encoded in the asymmetry is reliable but partial, as we give up on the attempt of modelling the full process.

To start, we restrict our inquiry to the sub-sample of single-author papers with identified gender G , as these would likely be more strongly affected by a possible gender bias. We count $N_{G \rightarrow G'}^{\text{cit}}$, the number of single-author papers with gender G citing single-author papers with gender G' . We compute the proportions $f_{G \rightarrow G'} = N_{G \rightarrow G'}^{\text{cit}} / N_{G \rightarrow}^{\text{cit}}$ dividing by the total numbers $N_{G \rightarrow}^{\text{cit}} = \sum_{G'} N_{G \rightarrow G'}^{\text{cit}}$, so $0 \leq f_{G \rightarrow G'} \leq 1$. From this we define the gender asymmetry as

$$A = f_{M \rightarrow M} - f_{F \rightarrow M} = f_{F \rightarrow F} - f_{M \rightarrow F} = \frac{1}{N_{M \rightarrow}^{\text{cit}} N_{F \rightarrow}^{\text{cit}}} \det \begin{pmatrix} N_{M \rightarrow M}^{\text{cit}} & N_{M \rightarrow F}^{\text{cit}} \\ N_{F \rightarrow M}^{\text{cit}} & N_{F \rightarrow F}^{\text{cit}} \end{pmatrix}. \quad (1)$$

The first formula means that A is the proportion at which solo males cite solo male research more than solo females cite solo male research. The second formula means that A also is the proportion at which solo females cite solo female research more than solo males cite solo female research. So the gender asymmetry ranges between $-1 \leq A \leq 1$. The final formula shows that A is a more symmetric quantity, 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 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

category	hep-ex	hep-ph	hep-th	hep-lat	nucl-ex	nucl-th	gr-qc	astro-ph
counts	2755	14627	15370	1762	1673	1258	6706	6733
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

Table 1: Gender asymmetry A defined in eq. (1) 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/1000	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

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

collective equal bias of both genders towards one gender, which corresponds to multiplying one column of the matrix above by a fixed constant.⁴

The meaning of the asymmetry can be clarified giving its predicted value in a toy model where N_G^{aut} authors of gender G (for simplicity we ignore that some authors are more active than others, so that an effective number would be directly relevant) cite with gender-dependent rates $p_{G \rightarrow G'}$, such that $N_{G \rightarrow G'}^{\text{cit}} \propto N_G^{\text{aut}} N_{G'}^{\text{aut}} p_{G \rightarrow G'}$ and

$$A \simeq \frac{N_M^{\text{aut}} N_F^{\text{aut}}}{(N_M^{\text{aut}} + N_F^{\text{aut}})^2} \det \begin{pmatrix} p_{M \rightarrow M} & p_{M \rightarrow F} \\ p_{F \rightarrow M} & p_{F \rightarrow F} \end{pmatrix} \quad (2)$$

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

We extract A from data removing self-citations (also known as self-references), which introduce a background of same-gender preference not due to an actual gender preference. This removal is done exactly, as we have a list of all references where all authors are identified with a unique code. The removal introduces a small bias of order $1/N_G^{\text{aut}}$ in the asymmetry: we neglect this bias since it is smaller than our statistical uncertainty, which scales as $1/\sqrt{N_G^{\text{aut}}}$. Fig. 3 shows the time evolution of the gender asymmetry, found to be compatible with zero at all times.⁵ Restricting to papers after 2010 we find the result shown in table 1.⁶ The un-

⁴The sub-sample of “ambiguous” authors (whose name is associated to different genders in different countries) does not show anomalous features that would support the hypothesis of a collective gender bias.

⁵An 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.

⁶Our results have been reproduced by [Hossenfelder et al.(2018)], who also try to go beyond the asymmetry by assuming a model similar to our eq. (2) (but with N_G^{aut} replaced by N_G^{pap}). As such models introduce questionable systematic uncertainties, we restrict our attention to the model-independent gender asymmetry.

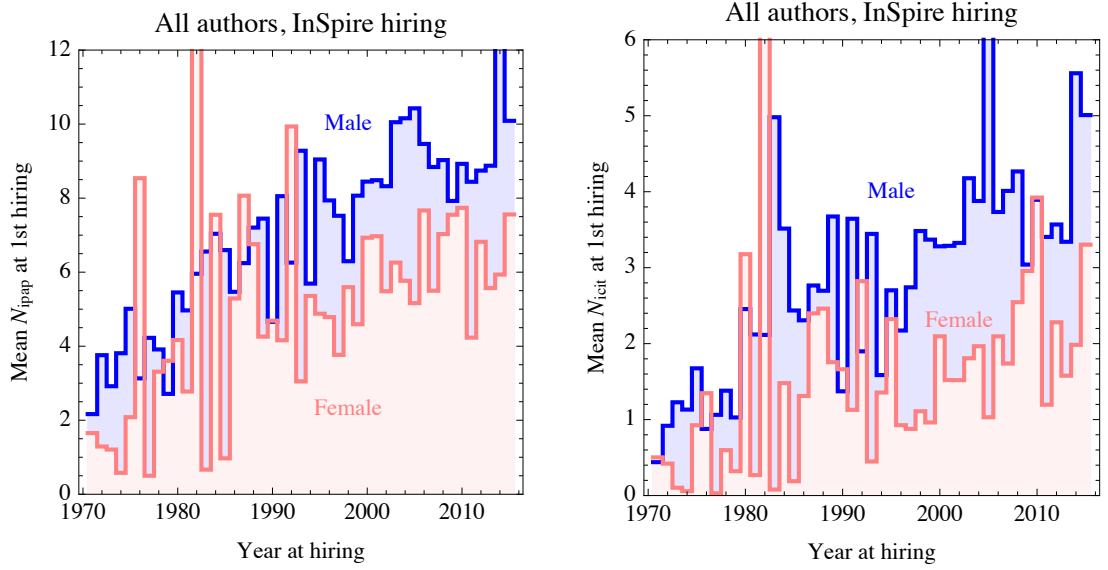


Figure 4: The left (right) panel shows the mean number of fractionally-counted papers N_{ipap} (of individual citations N_{icit}) of authors in fundamental physics at the moment of their hiring, as function of the hiring year. Data are shown separately for male (blue) and female (pink) authors.

certainty (shown as one standard deviation after the \pm symbol) has been computed with the usual propagation of errors assuming statistical $\sqrt{N_{G \rightarrow G'}^{\text{cit}}}$ fluctuations on the counts, such that the uncertainty on A is $[N_{F \rightarrow F}^{\text{cit}} N_{F \rightarrow M}^{\text{cit}} / N_{F \rightarrow}^{\text{cit}3} + N_{M \rightarrow F}^{\text{cit}} N_{M \rightarrow M}^{\text{cit}} / N_{M \rightarrow}^{\text{cit}3}]^{1/2}$. A hint of an asymmetry, $A_{\text{other}} = (4.8 \pm 1.2)\%$, is observed among other about 10^4 papers (mostly unpublished) not included in the 8 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 multi-authored papers knowing how a hypothetical gender bias would depend on the relative amount of F and M authors. One simple possibility is just generalizing the definition of $N_{G \rightarrow G'}^{\text{cit}}$ into $\sum_{\text{citations}} f_G f_{G'}$ where f_G ($f_{G'}$) is the fraction of authors with gender G in each citing (cited) paper. We drop all self citations, now defined as whenever the cited and citer paper have at least one author in common. With the new $N_{G \rightarrow G'}^{\text{cit}}$ we find the result in table 2. Uncertainties (not shown) are there about 5 times smaller than in the single-author sample, if propagation of errors is naively applied to fractional counts.

In conclusion, taking into account the definition of the asymmetry A and the relative number of F and M authors in our data, a non-zero gender asymmetry in citations within its uncertainty range would not significantly distort the bibliometric indices based on citations discussed in the following.

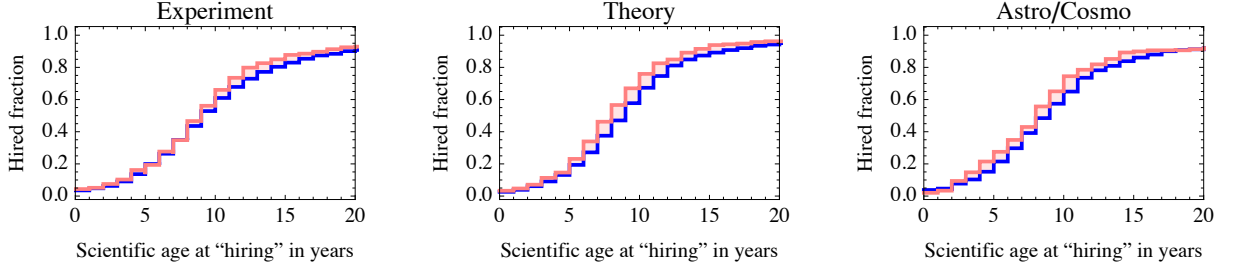


Figure 5: Among all hired authors after 2000, we show the cumulative fraction of hired authors as function of their scientific age, for male (blue) and female (pink) authors in experiment (left), theory (middle), astro/cosmo (right). We use INSPIRE hires.

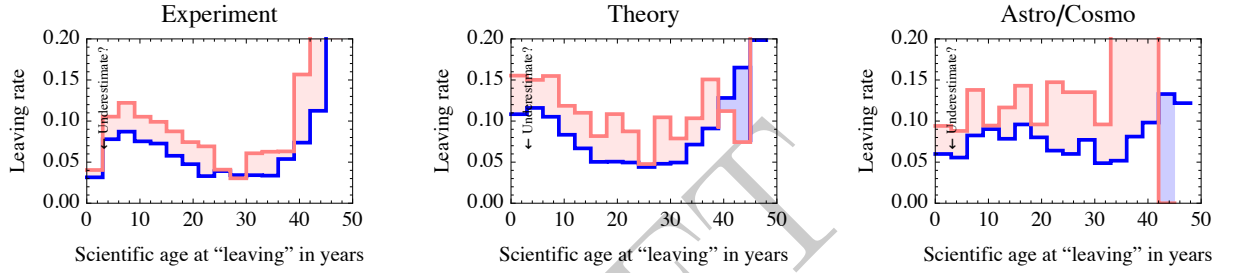


Figure 6: Fraction of active authors that each year leave the field, as function of their scientific age. We considered leavings during 2000-2015, counting as left those authors who no longer wrote a paper up to now (2018). See the text for warnings.

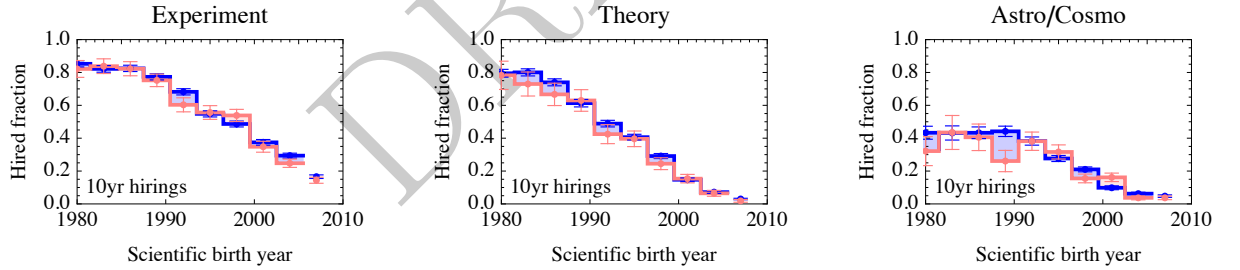


Figure 7: Fraction of authors hired up to now as function of the date of the first paper. Only the statistical uncertainty is shown; see the text for warnings.

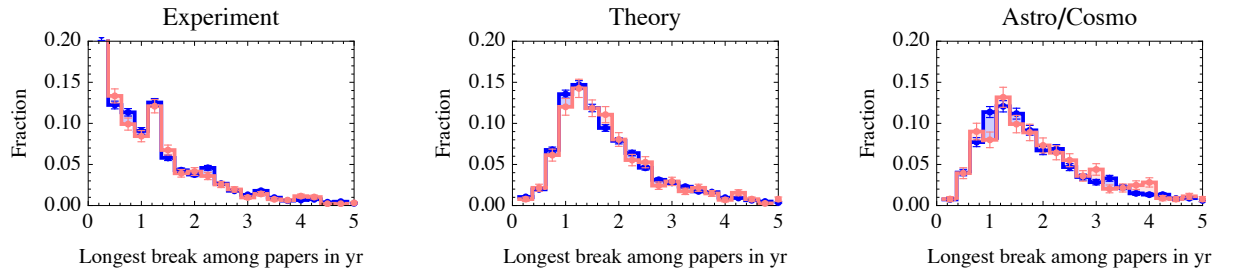


Figure 8: Fraction of active authors (divided by their main topic) as function of the longest time break among their papers.

3.2 Hiring

Having established that citations represent the common opinion of M and F authors of the community, we can use citations to test for a possible gender difference in hiring. For each hired or pseudo-hired author we compute his/her bibliometric indices motivated in section 2.3 at the hiring moment, defined as in section 2.2. From this we extract the mean bibliometric indices of hired F and M authors.

The left (right) panel of fig. 4 shows the mean number of fractionally-counted papers (of individual citations N_{cit}) of authors at their hiring date as reported by INSPIRE. For the sake of clarity we use traditional color codes: blue (pink) for male (female) authors.

We see that hired F authors do not have, on average, bibliometric indicators above those of hired M authors. Rather, a tendency in the opposite direction seem present at all times, across the main sub-fields⁷ 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 confounders considered in appendix A.1.

We next show distributions that provide extra information.

Fig. 5 shows the cumulative distribution of hired physicists as function of their scientific age at hiring. It exhibits no significant gender difference, which 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 favoured 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.⁸

A warning is necessary about the two next plots, which extend the analysis to authors which have not been hired. Our analysis is restricted to authors identified by INSPIRE, that 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.

Fig. 6 shows the abandonment rate per year as function of scientific age. It is maximal

⁷Experimentalists who work in large collaborations tend to have similar bibliometric indicators. The average N_{cit} at hiring can be lower for F authors if they are hired younger than M authors.

⁸The temporal distribution of 245 hires of astronomers in the US 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 the PhD; astronomers are hired on average 5 years later). We find a similar difference of 0.95 ± 0.5 yr restricting to astro/cosmo authors (considering the time after the first paper; authors are hired on average 9 years later). According to [Flaherty (2018)] the hiring time distribution is better fitted assuming a 3-4 times higher F abandonment rate, rather than assuming a 10:1 bias in favour 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). We do not attempt modelling hiring, as we do not see how models can be made realistic. Rather, we have extra data about papers and citations which do not support neither a 10:1 bias (see fig. 4) nor a 4:1 difference in abandonment rates (see fig. 6).

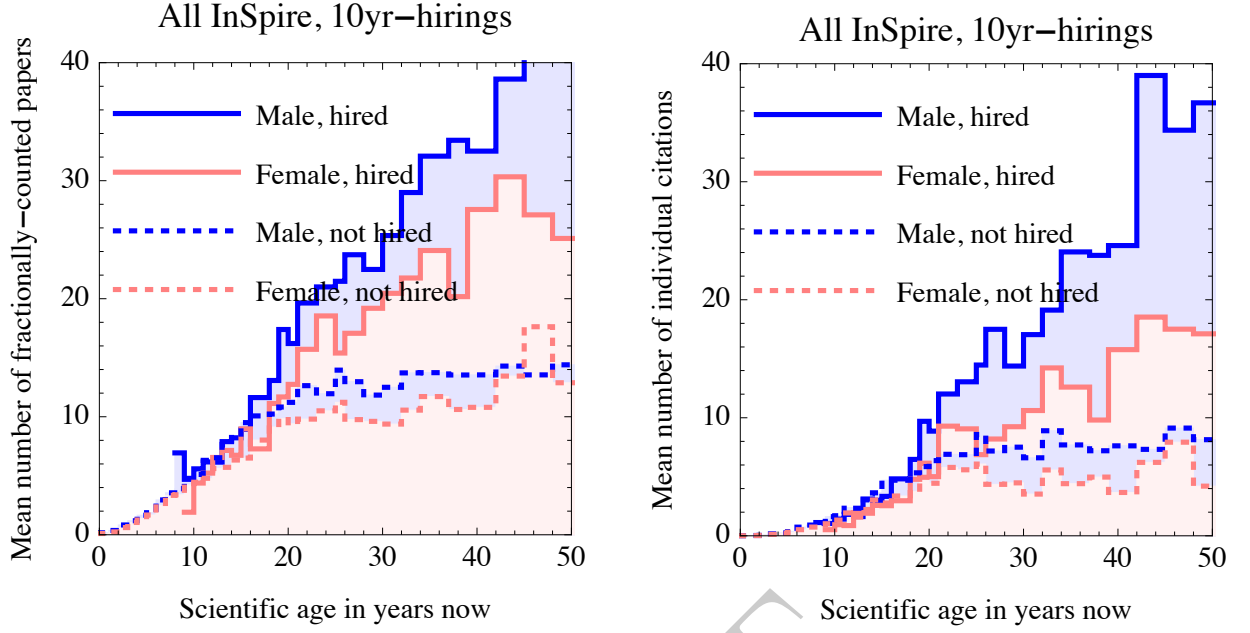


Figure 9: *Mean number of fractionally-counted papers (left) and of individual citations (right) as function of scientific age (time after the first paper) of scientifically active authors now.*

among elder authors that retire, minimal among senior authors, and intermediate among junior authors (as warned above, we under-estimate the abandonment rate of very young authors that leave the field after writing just a few papers). F authors show, on average, an abandonment rate 30% higher than M authors. Such difference is present at all scientific ages and restricting to hired authors. Within uncertainties, the excess abandonment rate at junior ages (possibly due to hiring issues) is equal for M and F authors. In conclusion we see a mild ‘leaky pipeline’ effect apparently unrelated to hiring issues. This is 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 [Miller et al.(2015), Allen-Hermanson (2017)].

Fig. 7 shows the fraction of authors hired among those that started writing papers in given time periods. We do not see significant gender differences. We used 10-yr hiring because coverage is here more important than timing. Thereby the plot stops 10 yr 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.

3.3 Productivity

Fig. 1 and 2 show a possible gender gap in the fractionally-counted number of papers at a few 10% level consistent with earlier findings in the literature (see e.g. table 2 of [Ceci et al.(2014)] and [Abramo et al.(2009), Abramo et al.(2015)]), and a larger gap in the number of received individual citations.

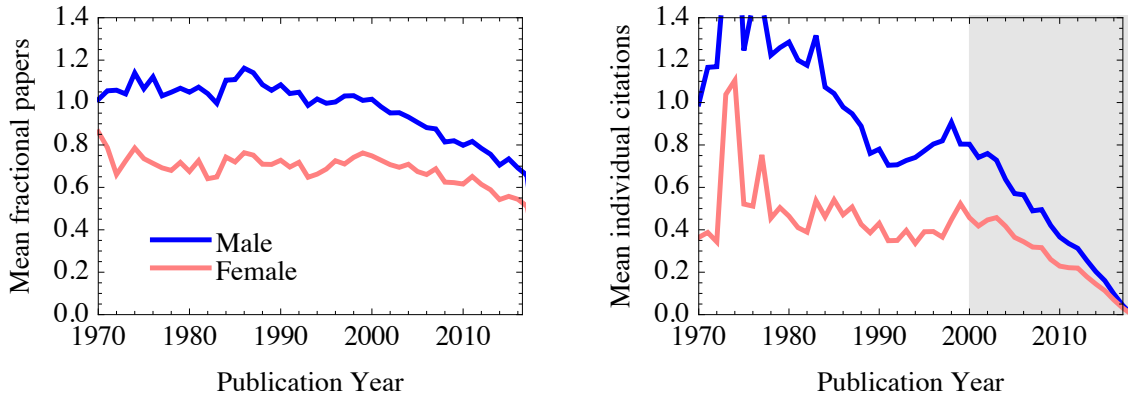


Figure 10: **Left:** mean number of fractionally-counted papers produced each year by M and F authors active that year. **Right:** mean number of received individual citations (the shading reminds that citation counts are largely incomplete for recent papers).

As M and F authors have different average seniorities, the left panel of fig. 9 shows the mean number of fractionally-counted papers written by M and F authors as function of their scientific age (time since their earliest paper). We restricted to scientifically active authors: those who wrote at least one paper after 2013. The average is shown separately for hired and not-hired authors, using 10yr-hires (in order to have a more complete coverage). The right panel of fig. 9 similarly shows the mean number of received individual citations. In both cases we see 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 that started earlier when the field was different and when the F percentage was smaller.

Appendix A.2 discusses other possible confounders, without finding anything that can remove the gender gap in productivity. We then discuss its possible causes.

In various countries F authors have earlier retirement ages. But gender differences show up in fig. 9 before retirement. 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 et al.(1987), Sax et al.(2002), Xie et al.(2003), Stack (2004)], other studies find a negative impact (on women [Fox (1995), Ginther et al.(2009)], on men and women equally [Hargens et al.(1978)]), other studies found a positive impact (on men [Ceci et al.(2014)], possibly due to selection effects). Results vary depending on field (with physical sciences sometimes being an outlier, possibly a fluctuation) and are mostly focused on the situation in the US and on the number of produced papers 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, our INSPIRE data do not provide any personal information so

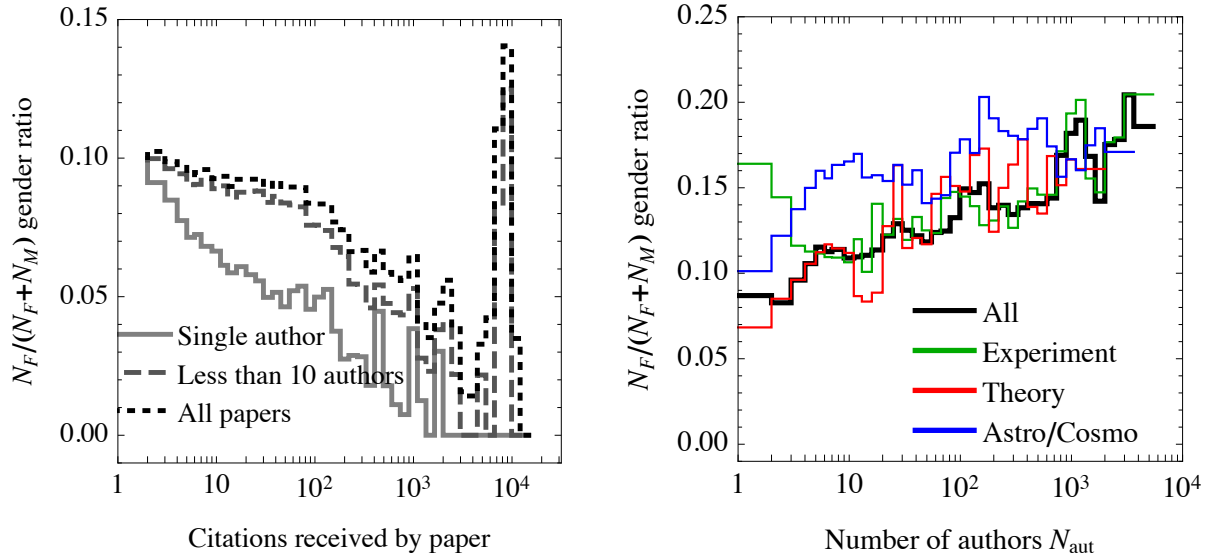


Figure 11: **Left:** fractional contribution of all F authors to papers that received the number of citations on the horizontal axis. We assume that each author contributed equally as $1/N_{\text{aut}}$ to papers with $N_{\text{aut}} > 1$. **Right:** fractional contribution of all F authors to papers with N_{aut} authors.

we can only proceed indirectly. Fig. 9 indicates a gap that opens at an age roughly consistent with maternity (but also consistently with the transition to scientific independence), and that does not close at older ages. We searched for a possible gender difference in career gaps: for each author we computed the longest time gap between consecutive papers, using arXiv dates to have precise information about publication dates. Fig. 8 shows the distribution of longest gaps among M and F authors: it does not show significant gender differences. A similar null result is found restricting to hired authors. The M/F gap seems strong in Germany, UK, Italy; weaker in USA, France, null in Japan (maternity laws are different in different countries; but single-country statistics is poor).

As we restricted to active authors, the averages computed in this section are not affected by the mildly higher abandonment rate of F authors found in fig. 6 at any scientific age.

Stopping writing papers might however be the extremum of a tendency towards reduced productivity. A gender difference in career gaps or abandonment rates (possibly due to maternity issues) might reduce the cumulative number of papers and of received citations of authors. We thereby avoid summing over author careers, in order to see if a productivity gap persists.

Each year we determine the scientifically active authors that produced papers, and show in fig. 10 their average productivity, separately for M and F authors. We see that F authors produce on average roughly 30% less papers than M authors, and receive roughly half citations. We adopted fractional counting: without it hyper-authored publications would lead to a recent boom, roughly equal for M and F authors. Alternatively, we observe the productivity gap by restricting to theorists and avoiding fractional counting.

Furthermore, fig. 11 (left panel) shows the gender contribution to single papers, finding a smaller F percentage among authors of top-cited papers, especially restricting to single-author papers (see also [West et al. (2013), Jagsi et al. (2006)]). The right panel shows that F authors tend to work in larger collaborations.

3.4 Distribution of individual citations

The left panel of fig. 12 shows the distributions in the number of individual citations N_{cit} received by female and male authors in fundamental physics, considering the whole INSPIRE data-base and without applying any correction nor cut. The bell-shaped distributions spread through a few orders of magnitude in N_{cit} . The dotted curves show that each bell is well approximated, at least in its upper side, by a log-normal as function of N_{cit} (namely, by a Gaussian as function of $\log N_{\text{cit}}$, the variable used on the horizontal axis of fig. 12; \log is the logarithm in base 10). A log-normal, already observed in bibliometrics [Thelwall et al.(2014)], arises when many positive independent random variables contribute multiplicatively. We find that the number of citations received by authors in physics tends to grow linearly or quadratically rather than exponentially with their scientific age, and that papers written by elder authors receive, on average, less citations than papers written by younger authors. This suggests 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 et al.(2017)]).

The difference between the F and M distributions in fig. 12 is statistically significant.

The black curve in fig. 12 shows the ratio N_F/N_M of female vs male authors (left axis) as function of the number of received individual citations (horizontal axis).⁹ The gender ratio is not constant: the M fraction progressively grows when going from average to top authors in terms of individual citations. In the raw data, both averages and variances differ.

Since we deal with a composite sample of data, we consider whether such difference can be a byproduct of confounder factors. This issue is discussed in appendix A.3: we don't find any confounder that washes away the trend. One confounder gives a correction worth taking into account: senior authors had more time to receive citations and, for historical reasons, male authors are presently on average more senior than female authors. This confounder cannot remove the difference, given that fig. A.12 finds the same difference within sub-samples of authors with same scientific age, and that fig. 1 shows a productivity gap present within papers with same publication dates. Correcting for this confounder is however needed to precisely quantify the difference. We compensate for the different time evolution $N_{F,M}^{\text{start}}(t)$ (number of F and M authors that produced their first paper during year t) by assigning to each male author A a weight proportional to $N_F^{\text{start}}(t_A)/N_M^{\text{start}}(t_A)$, where t_A is the date of his/her first paper. This is equivalent to select every year a random sub-set of new male authors such that M and F authors are numerically equal, and averaging over the possible choices. The result is shown in the right panel of fig. 12. No significant gender differences are now seen in the lower sides of the distributions, controlled by social phenomena such as hiring thresholds (such sides would

⁹In precise mathematical notation this is dN_F/dN_M , and the bells are $N_G^{-1}dN_G/d\log N_{\text{cit}}$.

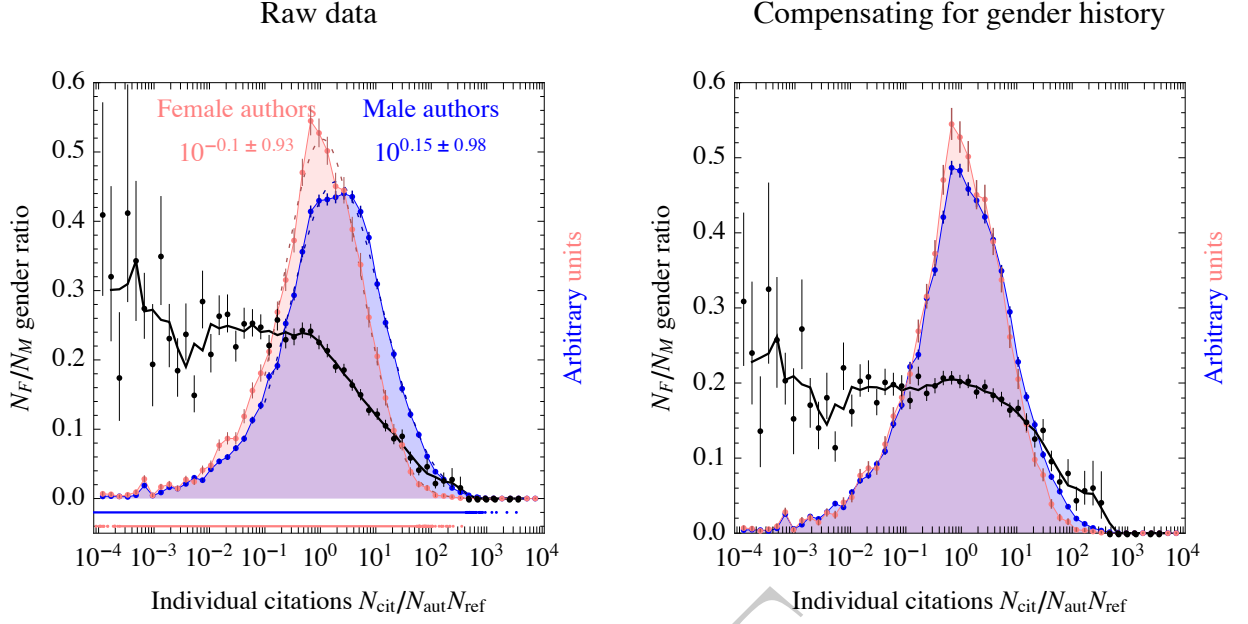


Figure 12: The bells show the probability distributions of authors in fundamental physics as function of their number of individual citations, separately for Male (blue) and Female (pink) authors and plotted with a common normalization (right axis). The numbers in the figure show the logarithmic average and standard deviation; the dots below the bells 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 smoothened 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:** raw data. **Right:** compensating for gender history.

be mostly removed restricting to hired authors). The difference in variances among the upper sides persists.

We can account for different ages using a complementary strategy: we 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. Since N_{icit} averaged over authors approximatively scales as Δt_A^p with $p \approx 1.8$ within all main topics, we choose this value of p . The left panel of fig. 13 shows that the trend persists, and that the distributions become narrower, having removed one source of their spread. The right panel of fig. 13 shows that applying both corrections has negligible extra effect.

The difference in variances in the upper side is seen in any case.

The M fraction is largest among top authors, as clear from the dots below the bells in the left panel of fig. 12, which show the individual authors. A physicist might focus on the sub-set of top authors, read their names (the top 50 authors are listed in table 7 of [Strumia et al.(2018)] and received 7% of the individual citations, like all women), and consider their

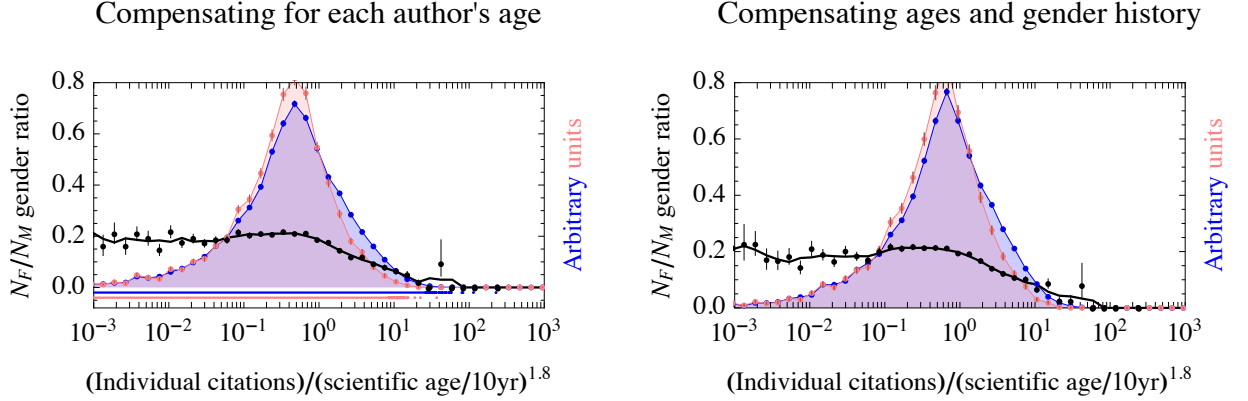


Figure 13: As in fig. 12, adopting a measure that, on average, does not depend on the scientific age of authors. The M distribution still has a longer upper tail.

scientific results, concluding that no sociological confounder can wash away most of them. It is thereby interesting to show that the gender difference is statistically significant even restricting to top authors. In the raw data, the F author with most individual citations is in position $F_1 = 69$. Assuming that gender has no effect, the probability of being in this or lower position is $\varphi_1 = m^{F_1-1} \approx 3 \cdot 10^{-6}$, 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 k -th 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 $\varphi_2 \approx 3 \cdot 10^{-11}$), $F_3 = 191$ ($\varphi_3 \approx 2 \cdot 10^{-13}$), etc., roughly fitted by $F_k \approx 69k$. As discussed above, the time evolution of $f = N_F/(N_M + N_F)$ is a confounder: it is however not enough to remove the difference, since $f > 1/69$ at all relevant times. More precisely, performing the correction described above, the positions become $F_k = \{22, 61, 84, \dots\} \sim 27k$. Considering the age-corrected metric $N_{\text{cit}}/\Delta t_A^{1.8}$ the positions are $F_k = \{18, 89, 145, \dots\} \sim 37k$. An excess of male top-authors is found in any case.

A varying gender fraction that culminates in a small group of extremely productive, mostly male, ‘star authors’ has been observed in [Bordons et al. (2003), Abramo et al. (2009)b, Abramo et al. (2015)] (see also [Kwiek (2016)]). As we find that the non constance of N_F/N_M survives to confounders, it is interesting to investigate its quantitative shape.

A look at fig. 12 or 13 suggests that the M bell has a longer tail of top-authors (raw data show also a difference in averages, mostly due to confounders, that makes the difference in variances less easily visible). Appendix B shows that the difference in upper variances is statistically significant. This can be visually appreciated 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 function of $\ell = \ln N_{\text{cit}}$) in their upper side. A common standard deviation σ and different averages $\mu_M \neq \mu_F$ for M and F

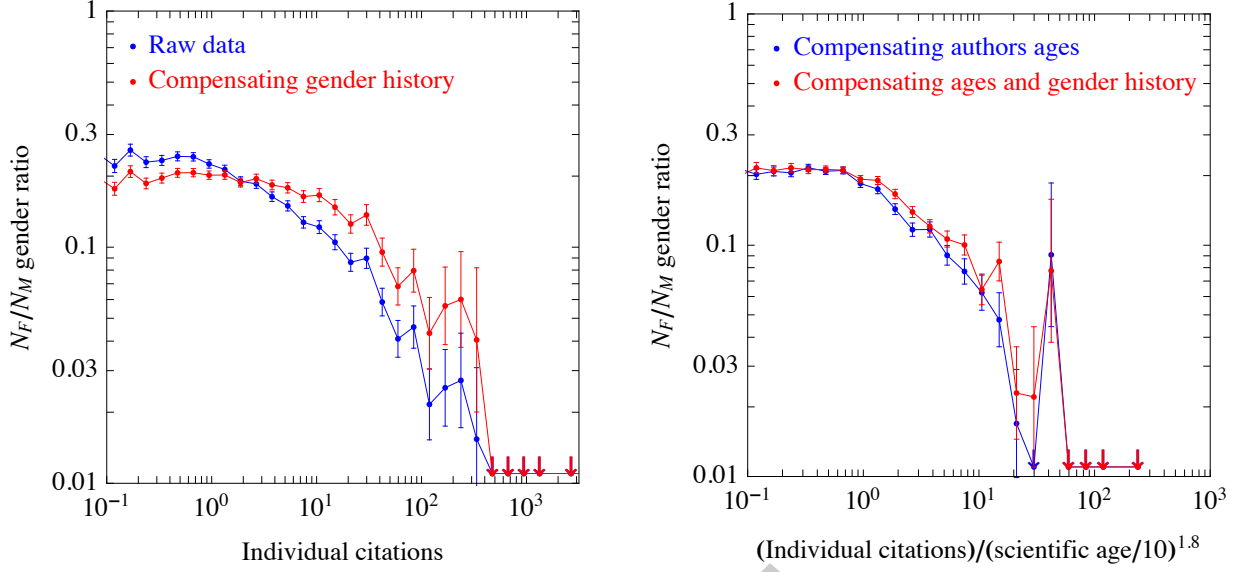


Figure 14: In the **left panel** the points with Gaussian 1σ statistical errors are data about the N_F/N_M female/male ratio as function of the number of individual citations received. Points with $N_M > N_F = 0$ are shown as down arrows. The blue points are raw data, the red points are corrected compensating for the different time evolution of the overall number of M and F authors. In the **right panel** we consider a different bibliometric index which approximatively compensates also for the scientific age of each author. Data are not well fitted by a linear function in log-log scale (which corresponds to $p = 1$ in eq. (6)) 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. (4), rather than as different averages, as in eq. (3).

authors would produce a N_F/N_M of the form

$$\frac{N_F}{N_M} \propto \frac{e^{-(\ell-\mu_F)^2/2\sigma_F^2}}{e^{-(\ell-\mu_M)^2/2\sigma_M^2}} = \exp\left[-\frac{R_\sigma}{2}\left(\ell - \frac{\mu_M + \mu_F}{2}\right)\right], \quad R_\sigma = 2\frac{\mu_M - \mu_F}{\sigma^2}. \quad (3)$$

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

$$\frac{N_F}{N_M} \propto \frac{e^{-(\ell-\mu)^2/2\sigma_F^2}}{e^{-(\ell-\mu)^2/2\sigma_M^2}} = \exp\left[-\frac{R_\sigma}{2}(\ell - \mu)^2\right], \quad R_\sigma = \frac{1}{\sigma_F^2} - \frac{1}{\sigma_M^2}. \quad (4)$$

Thereby a dominant difference in averages (standard deviations) would produce a line (a parabola) when N_F/N_M is plotted as function of N_{cit} in log-log scale. Such plot is shown in fig. 14, finding a parabolic shape along the upper sides of the bells, where the log-normal approximation is accurate enough (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, we repeat the analysis within the independent sub-samples of fig. A.12: plotted in

log-log scale they independently tend to show a parabolic (rather than linear) trend in N_F/N_M . The dotted curves in fig. A.12 show how well each sub-sample 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 sub-samples of fig. A.12 (where the top authors are plotted as points), consistently with [Bordons et al. (2003), Abramo et al.(2009)b].

A similar difference in upper variances is found using different bibliometric indicators, see appendix A.3.

4 Conclusions

We performed a bibliometric analysis of gender issues in fundamental physics world-wide from ~ 1970 to now, with the following results:

1. First, we see the well-known initial gender difference in representation: among new authors that appear at PhD-level, there are roughly 4 males for each female. The initial female fraction is not positively correlated with the Global Gender Gap Index of the countries,¹⁰ and evolves little in the subsequent career stages.
2. Citations exhibit no or small gender asymmetry and thereby encode the common opinion of male and female authors about what is interesting in fundamental physics (fig. 3). M and F authors give themselves a similar fraction of self-references (fig. A.1); the M fraction is higher in single-authored papers (fig. 11).
3. Female authors do not have, at hiring moments, higher average bibliometric indicators based on individual citations or fractionally-counted paper than male authors (fig. 4 and 5). Among authors identified by INSPIRE (that misses authors that write very few papers) we do not find a gender difference in hired percentages (fig. 7); F authors show a $\approx 30\%$ higher abandonment rate at all ages (fig. 6) and do not take longer breaks among papers than M authors (fig. 8).

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”. 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)]. We find:

4. A productivity gap both in the fractionally-counted number of publications and in their citational impact (fig. 9), which does not appear to be concentrated in specific countries, topics, periods, bibliometric indicators, journals (fig. A.2), etc. The gap is also found without integrating over careers (see fig. 10, 11).

¹⁰Similar studies about students (less mobile than researchers) found an anti-correlation known as “gender equity paradox” [Stoet et al.(2018)].

5. A gradually increasing male fraction when going from average to top authors in terms of individual citations (or other indices). The quantitative shape of this trend appears dominantly due to male and female distributions with different variances in their upper sides rather than to different averages (see fig. 14, eq. (3), eq. (4)).

While many social phenomena could produce different averages, producing different variances would need something that specifically disadvantages top female authors. Thereby, 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 [Su et al.(2009), Lippa (2010), Su et al.(2015), Thelwall (2018)b] and in variability [Halpern et al.(2007), Hyde (2014), Stevens et al.(2017), Wang et al.(2013)]. Difference in interests dominantly accounts for the initial difference in representation. Difference in variability accounts for the difference in productivity.¹¹ The small amount of higher male variance suggested by bibliometric data in fundamental physics is roughly consistent with independent observations of presumably relevant traits, assuming that top-cited physicists correspond to a $\sim 5\sigma$ upward fluctuation, as expected from a pool of $\sim 10^9$ persons.

Needless to say, “for every complex natural phenomenon there is a simple, elegant, compelling, wrong explanation”. 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 hypothesis.

We conclude addressing ethic 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 proportions in this comparison). The interpretation in terms of different variances implies that we should keep giving gender-neutral equal opportunities to everybody by considering each person based on his/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 is a politically-charged choice that can lead to discriminations (allegedly aimed at establishing equal outcomes) and to intolerance against diversity in ideas.

Acknowledgements I thank Guy Madison and many others for discussions and suggestions; Riccardo Torre for (among many things) having implemented the WGND name-gender association; Sabine Hossenfelder for having independently replicated the results in fig. 9b and section 3.1 using arXiv data [Hossenfelder et al.(2018)].

¹¹This is consistent with [O’Dea et al.(2018)], that confirms the difference in variabilities looking at grades, and observes that this difference alone cannot reproduce the representation gap.

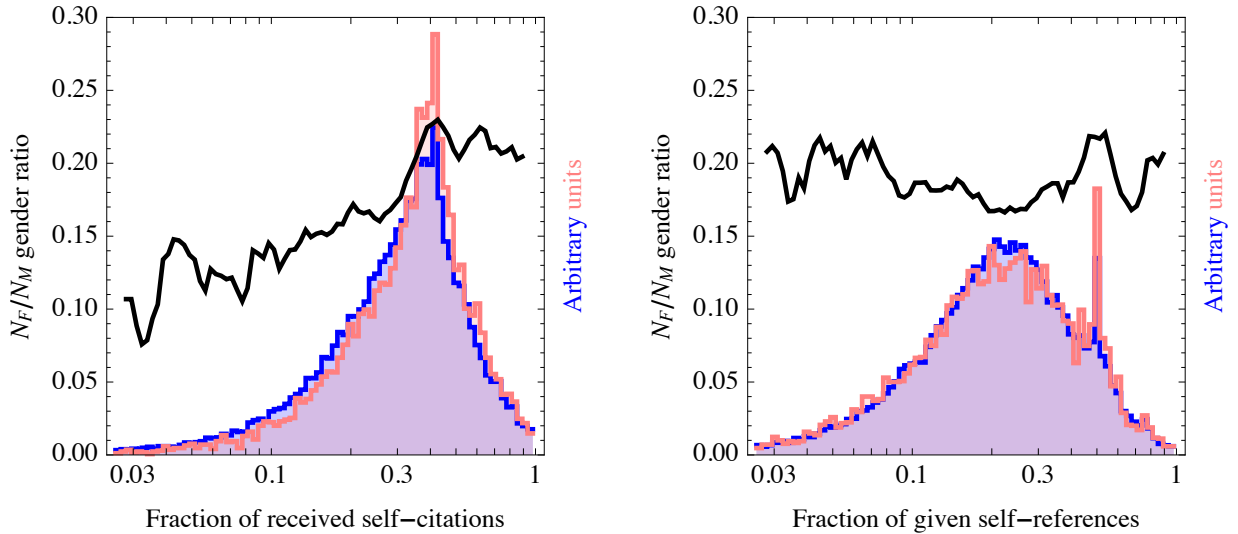


Figure A.1: **Left:** *distribution of authors as function of their self-citation rate.* **Right:** *distribution of authors as function of their self-reference rate.*

A Checking possible confounders

We here show that the gender differences discussed in the main text remain after accounting for possible confounder ‘backgrounds’. To start we consider generic confounders, and in the following sections we consider confounders specific to the various analyses.

A first generic worry is the effect of self-references. We count a reference as self-reference whenever the citing and cited paper have at least one author in common. The fraction of self-references is 20 – 30% higher in the sub-sample of papers written by solo M authors,¹² who tend to have more past papers. We define the self-citation (self-reference) rate of each author as the fraction of self-citations among all individual citations he/she received (gave). Considering all papers, fig. A.1 shows the distribution of M and F authors as function of their self-citation and self-reference rates. The narrow peak in the self-reference distribution is due to large experimental collaborations. Gender differences are small. We see an excess of M authors (mostly top authors) among those with a small fraction of self-citations. Similar results are found restricting within the main topics.

Next, we consider a possible gender bias in the publishing system. Studies generically find gender-fair refereeing processes [Borsuk et al.(2009), Ceci et al.(2014), Edwards et al.(2018)]. Fig. A.2 shows roughly the same gender percentage among authors of main journals, up to small variations related to the gender percentage of their topic (e.g. more F authors in astrophysics). Furthermore, our analyses are not restricted to published papers, as the current main publishing tool in fundamental physics is the pre-print bulletin arXiv.

¹²This agrees with [Hossenfelder et al.(2018)] that considered single-authored papers.

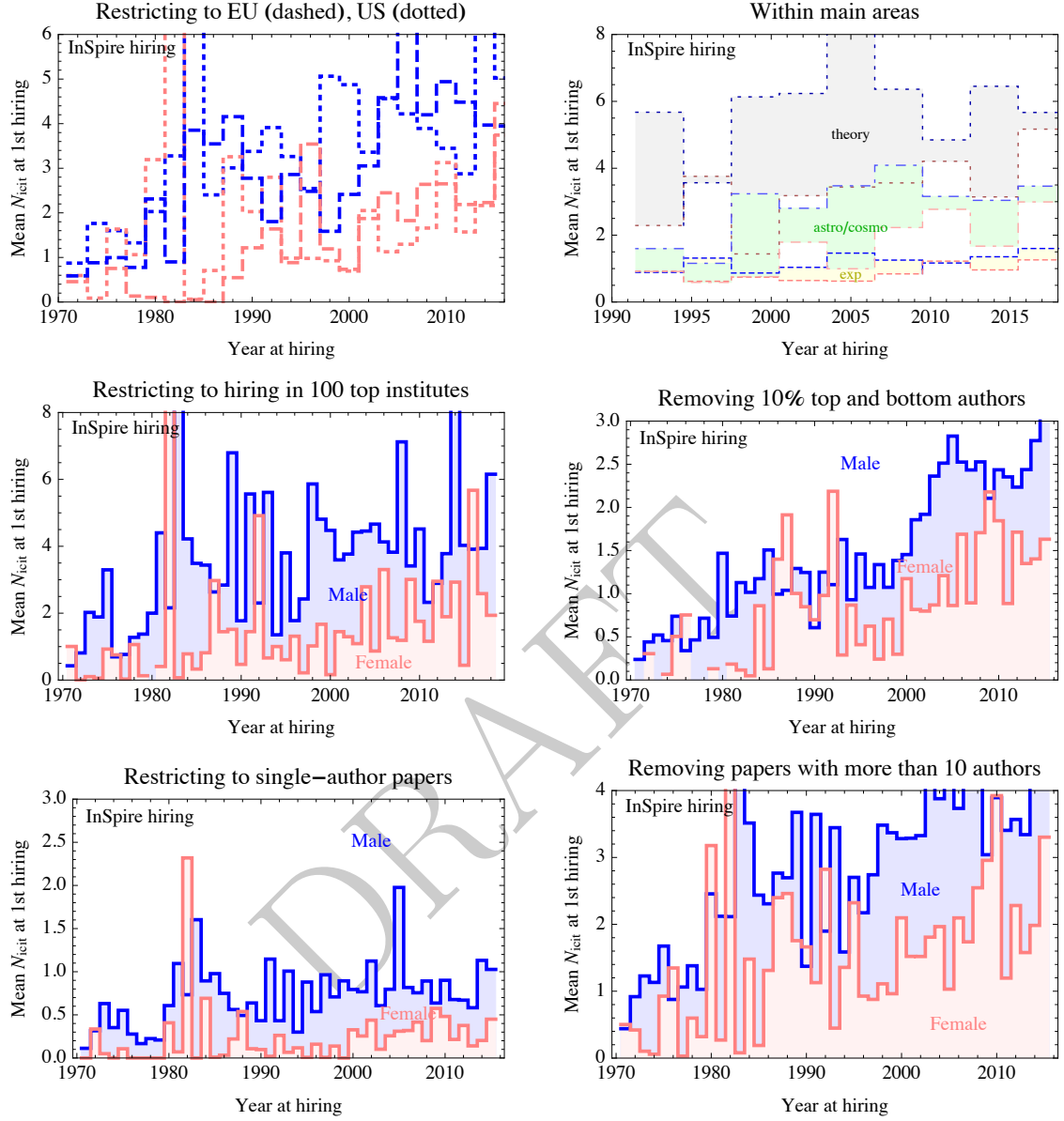


Figure A.3: **Top-left:** as in the left panel of fig. 4, restricting to hiring in EU and US institutions. **Top-right:** computing means separately among authors with most papers in theory (dotted), astro/cosmo (dot-dashed), experiment (dashed). **Middle-left:** considering only authors hired by 100 top institutes. **Middle-right:** recomputing the mean after dropping every year the 10% of hired authors with most and with least individual citations at hiring. **Bottom-left:** considering only single-author papers. **Bottom-right:** considering only papers with less than 10 authors.

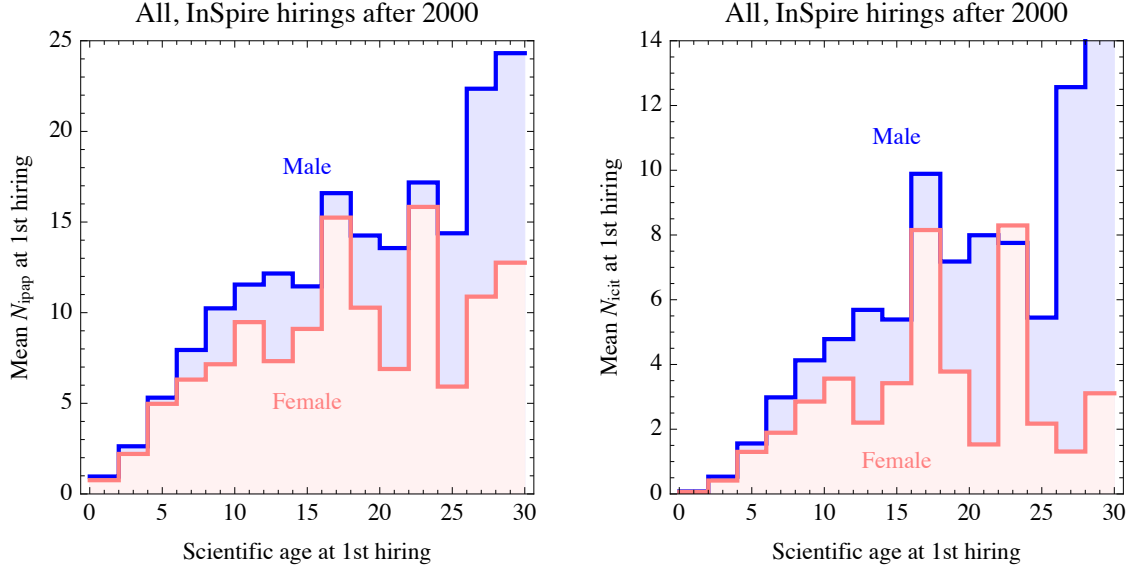


Figure A.4: As in fig. 4, as function of scientific age at hiring (years passed since earliest paper).

Finally, fig. A.4 shows the bibliometric indicators at hiring as a function of “scientific age” at hiring (defined as the time passed since the earliest paper of each author): the gender difference shows no notable dependence on this variable.

A.2 Productivity

We here study whether the productivity gap found in fig.s 9 could be due to confounders.

Fig. A.5 shows checks similar to what performed in fig. A.3: we restrict to institutions in the EU and US; to main topics; to 100 top institutes; to papers with few authors; we drop top and bottom authors. We find a productivity gap among theorists and astrophysicists but not among experimentalists, who work in large experimental collaborations inside which bibliometric indicators cannot recognise individual merit (indices of experimentalists are significantly correlated). The gap is present even using non-fractional citation counts (namely, assigning a paper with N_{aut} authors fully to each author) provided that the analysis is restricted to theorists and/or to papers with less than $\mathcal{O}(10)$ authors; otherwise average citation counts are dominated by large collaborations.¹³ Furthermore, a gap in the same direction is observed computing the median, restricting to authors with more than 5 papers; removing single-author papers (possibly affected by a hypothetical collective gender bias); replacing individual citations with fractional counting or PageRank-based metrics.

A.3 Distribution of individual citations

We here study whether the trend found in fig. 12 of section 3.4 (a higher M fraction among top-authors) is faked by confounders. The higher average seniority of M authors has been already considered in the main text.

¹³We thank Sabine Hossenfelder for discussions.

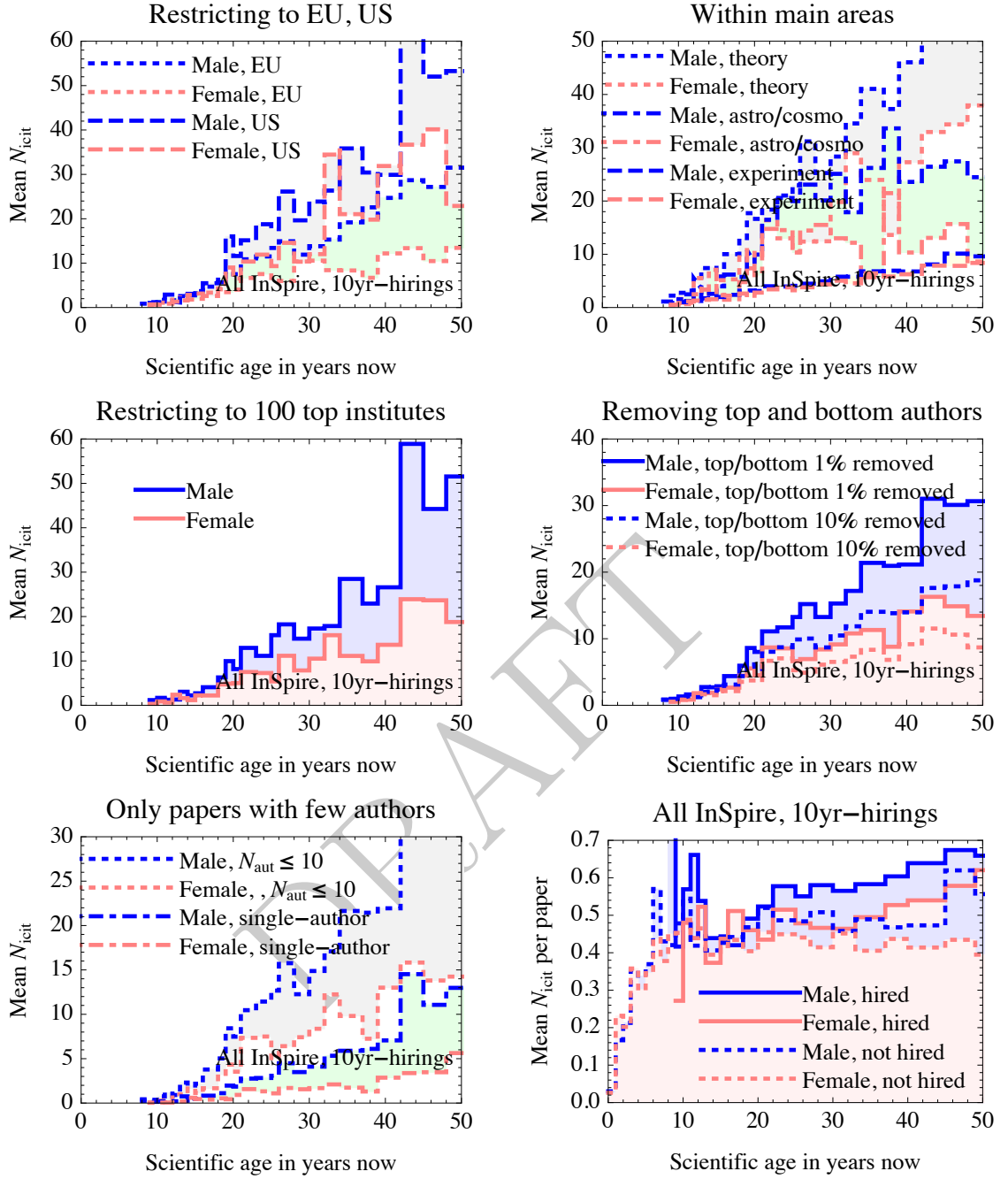


Figure A.5: **Top-left:** as in fig. 9, restricting to authors in EU and US institutions. **Top-right:** computing means separately among authors with most papers in theory (dotted), astro/cosmo (dot-dashed). **Middle-left:** considering only authors mainly affiliated to 100 top institutes. **Middle-right:** recomputing the mean after dropping the 1% or 10% fraction of hired authors with most and least individual citations. **Bottom-left:** considering only papers with few authors. **Bottom-right:** considering the average number of fractionally counted citations per fractionally-counted paper.

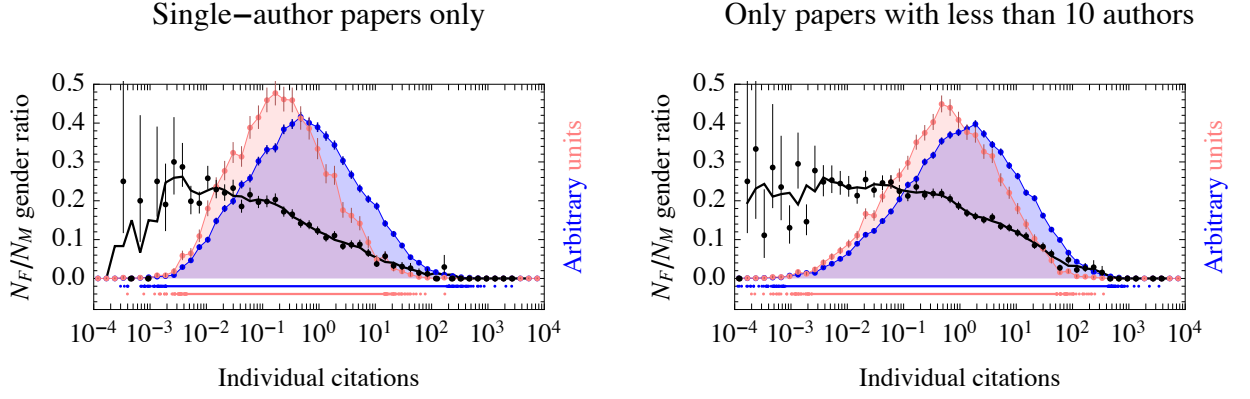


Figure A.6: As in fig. 12, but restricting to single-author papers (left) or to papers with less than 10 authors (right).

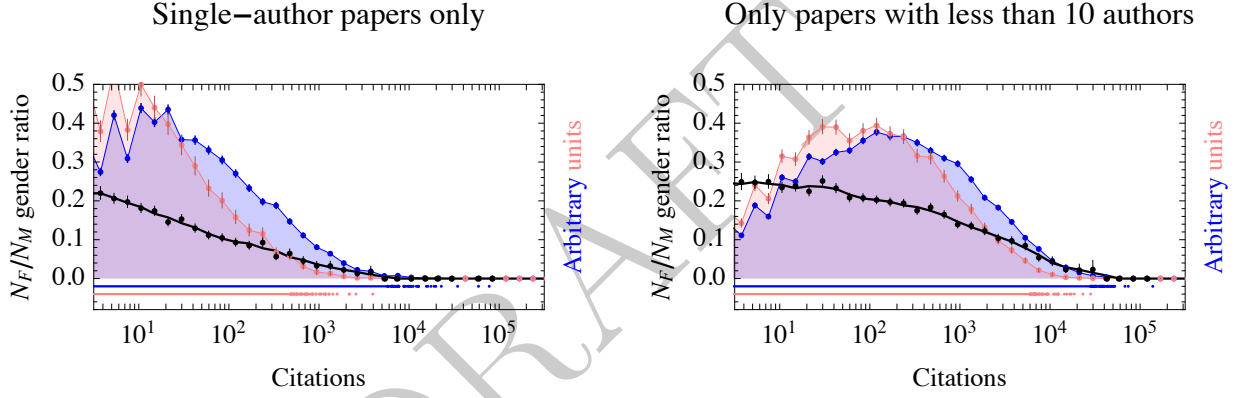


Figure A.7: As in fig. A.6, but considering citations.

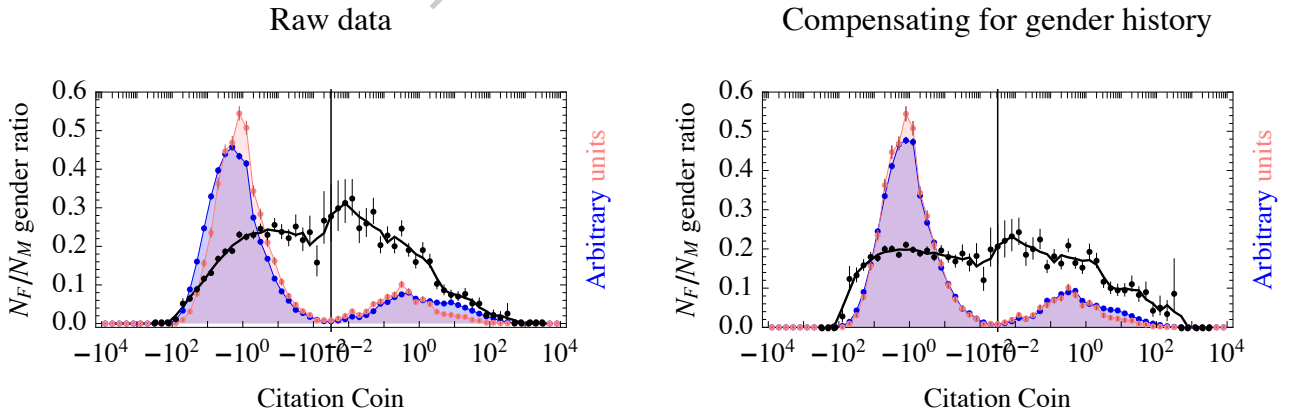


Figure A.8: As in fig. 12, but considering the Citation Coin.

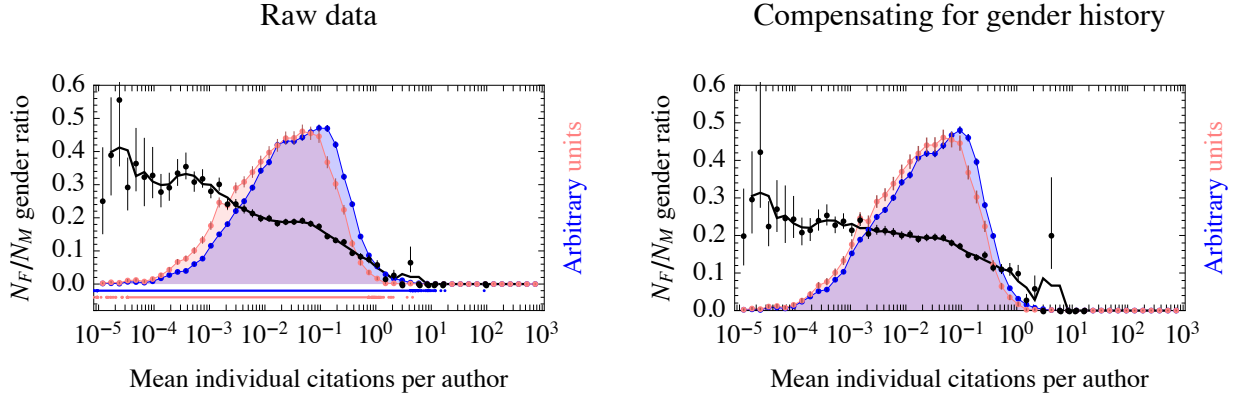


Figure A.9: As in fig. 12, but considering the average number of individual citations received by each author, $N_{\text{icit}}/N_{\text{pap}}$.

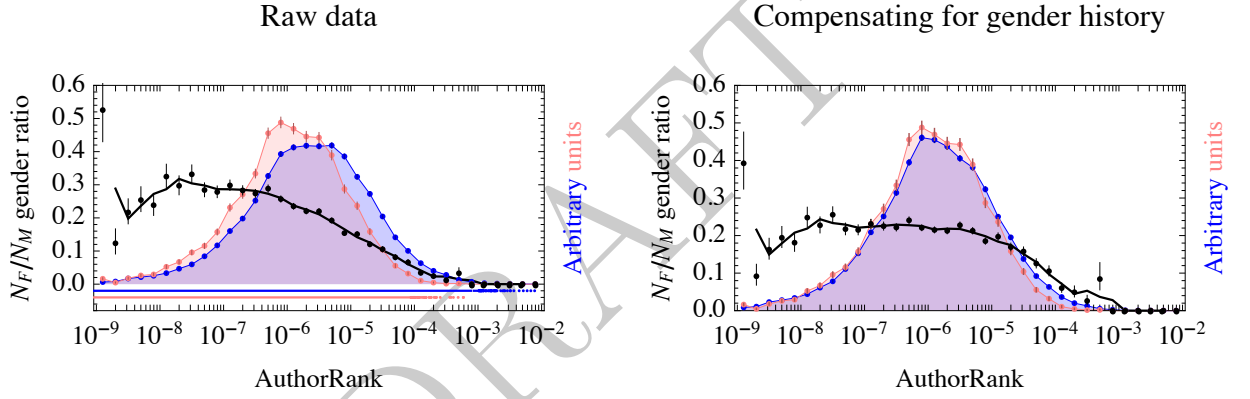


Figure A.10: As in fig. 12, but considering the Author Rank.

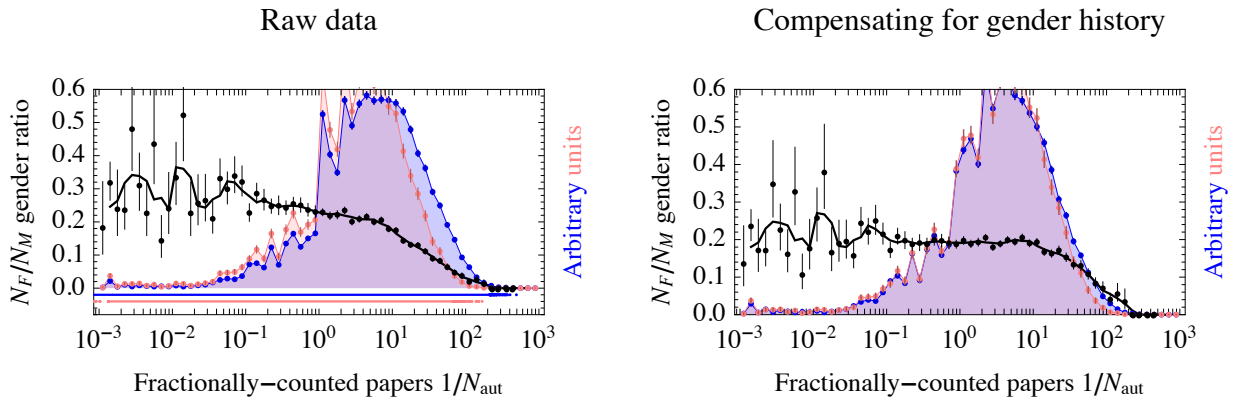


Figure A.11: As in fig. 12, but considering the number of fractionally-counted papers.

One can worry that papers with many authors might contain gift authorships, or that the trend might be a byproduct of gender differences in co-authorship. The trend becomes stronger restricting to papers with less than 10 authors, as well as restricting to single-author papers, see fig. A.6. With these cuts, the trend is visible also considering simple citation counting, see fig. A.7. Furthermore, fig. 11 shows the percentage female contribution to papers as function of the number of citations received by papers: we see that F authors contributed to high-impact collaboration papers, but not yet to high-impact single-author papers.

One can worry that the trend is due to our specific choice of bibliometric index, individual citations. We thereby repeat the same analysis using different bibliometric indicators:

- Fig. A.8 show the result using the CitationCoin [Strumia et al.(2018)]: we again find a gender difference among top authors, that persists after compensating for the different gender history as at point 1.
- Fig. A.9 considers the mean individual citations per paper received by each author, $N_{\text{icit}}/N_{\text{pap}}$. Like the CitationCoin, this metric combines the information on the number of citations and of papers. However, it is a less effective figure of merit, as it leads to lists of top authors that contain authors who abandoned the field after writing one or few well-cited papers.¹⁴
- Fig. A.10 show the result using the AuthorRank computed as in [Strumia et al.(2018)] and shifted such that its lowest value is 0. The trend is again visible.
- Fig. A.11 shows that a similar conclusion applies considering the fractionally-counted number of papers written by each author, even dividing by the scientific age. The number of papers (published or not) written by each author is a bibliometric indicator less correlated with scientific merit than the number of received individual citations, but more objective (for example, one might think that citations are affected by a gender bias).

We next check that not only the trend, but also its specific quantitative shape discussed around eq. (6) is not due to confounders. Fig. 14 shows that this is the case when correcting for different author ages. Fig. A.12 shows that the trend is present in independent slices of the data-base selected according to possible confounders:

- The top-left panel of fig. A.12 shows that the trend is independently present when the data-base is splitted into the main topics of fundamental physics (experiment, theory and astrophysics, as deduced from arXiv categories available after ~ 1995). A hint of the trend is present among experimentalists, despite their large recent collaborations.
- The top-left panel of fig. A.12 also shows that the trend is independently present among hired (dashed) and non-hired (dotted) authors. As expected, this distinction only affects the ‘lower’ side of the $\log N_{\text{icit}}$ distribution. We here used 10yr-hires. Furthermore the trend is independently present restricting to authors hired (or not hired) by top institutions, defined in various ways along the lines of [Strumia et al.(2018)].

¹⁴This metric has 0.36 correlation with N_{icit} . The related metric $N_{\text{icit}}/N_{\text{ipap}}$ (mean individual citations per fractionally-counted paper, 0.04 correlation with N_{icit}) magnifies the problem up to the point that top authors are unknown authors who wrote one or few well-cited papers in collaborations [Strumia et al.(2018)].

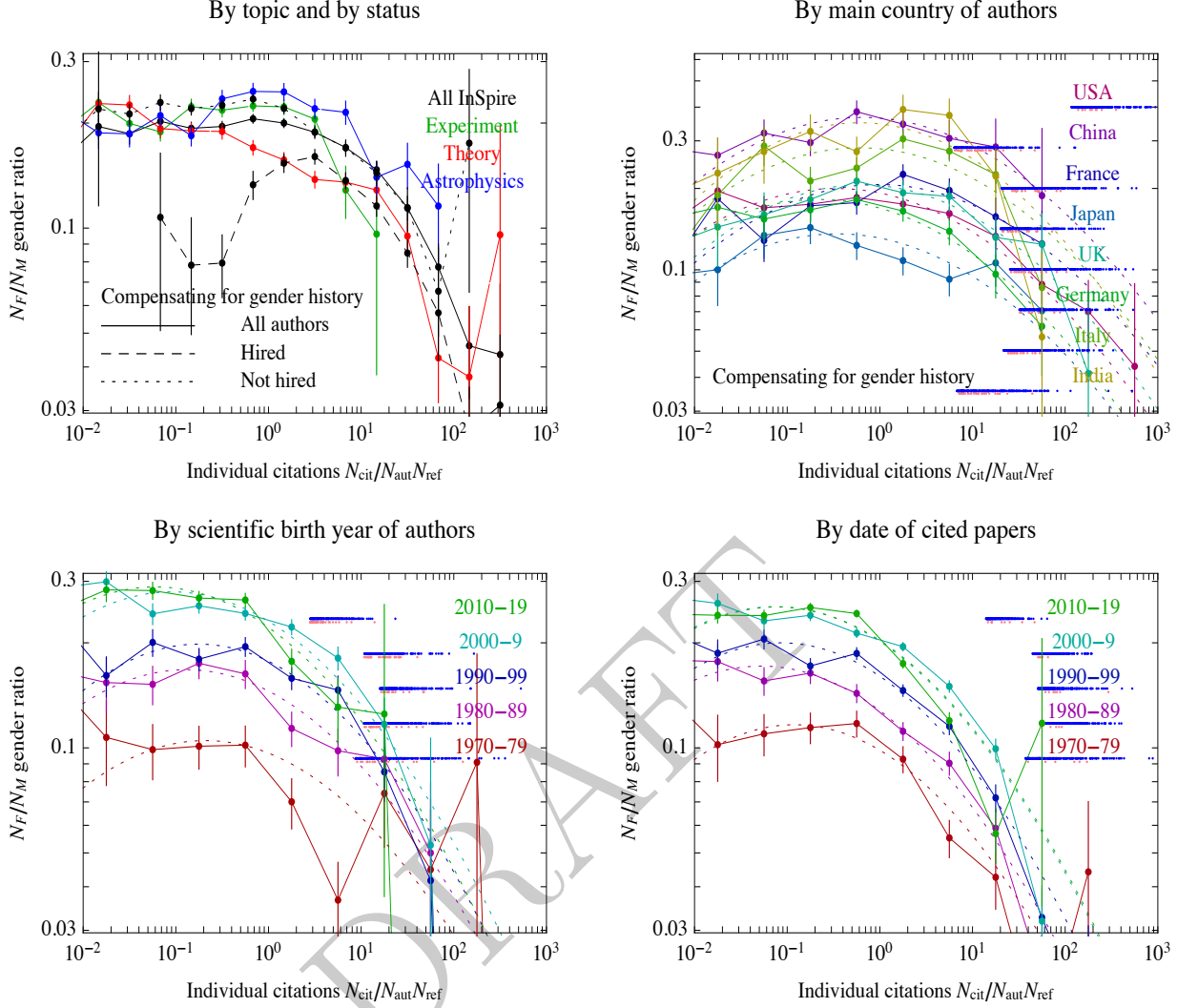


Figure A.12: **Top-left:** We compute the numbers N_F and N_M of female and male authors in fundamental physics who received a given amount N_{cit} of individual citations, and plot the ratio N_F/N_M (vertical axis) as function of $\log N_{cit}$ (horizontal axis). The papers are here spitted according to their arXiv topic (all, experiment, theory, astrophysics) and the authors as all, hired, not hired (using 10yr-hires). **Top-right:** same quantity splitting authors according to the most frequent country in their affiliations. **Bottom-left:** same quantity splitting authors according to their scientific birth (publication year of the earliest paper), restricting to scientifically active authors up to the present decade. **Bottom-right:** same quantity splitting papers according to their publication date. Data points with zero female or male authors are not shown; however in the latter 3 panels the points below the labels show the positions of the 300 top authors (M are blue, F are pink). The dotted curves show how eq. (6) with $p \approx 2$ and $R_\sigma \approx 2$ can fit the observed N_F/N_M .

- The fraction of M and F authors is approximatively constant within the countries that mostly contributed to fundamental physics: see the right panel of fig. 1. Thereby we do not expect that national differences are a confounder. Indeed, the top-right panel of fig. A.12 shows that the trend is independently present dividing authors in non-overlapping samples according to the most frequent country in their affiliations (at least for countries with enough statistics).

The plots above have been corrected for the generation effect as described in section 3.4; the plots below are restricted to time periods so that such correction is not necessary:

- The bottom-left panel of fig. A.12 shows that the trend is independently present dividing authors in non-overlapping samples according to their scientific birth date (year of their first paper). We restricted to authors still scientifically active in the present decade (this restriction makes a little difference). This plot means that the trend is not due to physicists preferentially citing famous older authors who built their scientific reputation in years around 1970 when progress was faster and when the female fraction was lower. On the contrary, the trend is present also restricting to recent authors.
- The bottom-right panel of fig. A.12 shows that the trend is independently present dividing papers in non-overlapping samples according to their publication date, since 1970 (when the INSPIRE data-base started being complete) up to now (recent papers are still accumulating citations).

B Statistical significance of the difference in variances

We here quantify the statistical significance of the difference between the M and F binned distributions in fig. 12 or 13 through

$$\chi^2 = -2 \sum_i \ln P(\mu_{Mi}, N_{Mi}) P(\mu_{Fi}, N_{Fi}), \quad \mu_{Gi} = N_G^{\text{tot}} \frac{N_{Mi} + N_{Fi}}{N_M^{\text{tot}} + N_F^{\text{tot}}} \quad (5)$$

where $P(\mu, N) = \mu^N e^{-\mu} / N!$ is the Poissonian probability of observing N events when μ are expected. We use Poissonian probabilities because some bins have zero or few authors.

We compute the χ^2 as a function of two fit parameters that artificially shift the mean and rescale the upper width of one of the two bells. We sum over bins i in the middle and upper parts of the bells, ignoring the lower tails. The middle part of the bells is included to reduce the uncertainties on the means, allowing to better see the difference in variances. We marginalise over the shift in means doing statistical inference in the common Gaussian limit of Bayesian and frequentistic techniques: in this limit marginalisation over parameters reduces to minimisation of the χ^2 , and the final $\Delta\chi^2$ is approximatively distributed as a χ^2 with one degree of freedom. We find that the χ^2 decreases by a statistically significant amount ($\Delta\chi^2 \approx 50 - 100$ depending on how much of the middle region is included) when the width of the M distribution is artificially reduced by 10 – 15%.

We can also quantify the statistical significance through the approximated analysis that relies on eq. (3) and eq. (4) by fitting the N_F/N_M distributions plotted in fig. 14 to the following analytic form

$$R = R_N \exp \left(- \frac{R_\sigma}{2} \ln^p \frac{N_{\text{icit}}}{N_{\text{icit}}^0} \right) \quad (6)$$

where \ln is the natural logarithm and R_N , R_σ , p , N_{icit}^0 are free parameters. We can then fit the observed N_{Fi} , N_{Mi} to the theoretical $R = N_F/N_M$ by defining a $\chi^2 = -2 \sum_i \ln P(R_i, N_{Fi}, N_{Mi})$ statistics, where the sum runs over the bins i , excluding the lower tails of the N_{icit} distributions. Furthermore

$$P(R, N_F, N_M) \propto R^{2N_F} (1 + R)^{-2(N_F + N_M)} \quad (7)$$

(written up to an R -independent multiplicative factor) is the maximal probability of observing N_G authors with gender G assuming a given expected gender ratio R . We find that the fit parameter R_N depends on the overall number of M and F authors in the given sample; the fit parameter N_{icit}^0 depends on the overall amount of citing literature relevant for the given sample. Different sub-samples are fitted reasonably well by common values of the extra fit parameters: $p \approx 2$ and $R_\sigma \approx 2 - 3$. Compensating for gender history the preference for $p = 2$ is again $\chi_{p=1}^2 - \chi_{p=2}^2 \approx 50 - 100$, again depending how much of the middle region is included.

References

- [Abramo et al.(2009)] G. Abramo, C.A. D’Angelo, A. Caprasecca, “*Gender differences in research productivity: A bibliometric analysis of the Italian academic system*”, *Scientometrics* 79 (2009) 517.
- [Abramo et al.(2009)b] G. Abramo, C.A. D’Angelo, A. Caprasecca, “*The contribution of star scientists to sex differences in research productivity*”, *Scientometrics* 81 (2009) 137.
- [Abramo et al.(2015)] G. Abramo, T. Cicero, C.A. D’Angelo, “*Should the research performance of scientists be distinguished by gender?*”, *Journal of Informetrics* 9 (2015) 25 [arXiv:1810.12737].
- [Abramo et al. (2018)] G. Abramo, C.A. D’Angelo, F. Rosati, “*Selection committees for academic recruitment: does gender matter?*”, *Research Evaluation* 24 (2015) 392 [arXiv:1810.13236].
- [Allen-Hermanson (2017)] S. Allen-Hermanson, “*Leaky Pipeline Myths: In Search of Gender Effects on the Job Market and Early Career Publishing in Philosophy*”, *Front. Psychol.* 8 (2017) 953.
- [Birnholtz et al.(2006)] J.P. Birnholtz, “*What does it mean to be an author? The intersection of credit, contribution, and collaboration in science*”, *J. Ass. Inf. Sci. Tec.* 57 (2006) 1758.
- [Bordons et al. (2003)] M. Bordons, F. Morillo, M.T. Fernandez, I. Gomez, “*One step further in the production of bibliometric indicators at the micro level: Differences by gender and professional category of scientists*”, *Scientometrics* 57 (2003) 159.
- [Borsuk et al.(2009)] R. Borsuk et al., “*To name or not to name: The effect of changing author gender on peer review*”, *BioScience* 59 (2009) 985.
- [Caplar et al.(2017)] N. Caplar, S. Tacchella, S. Birrer, “*Quantitative evaluation of gender bias in astronomical publications from citation counts*”, *Nature Astronomy* 1 (2017) 0141.
- [Ceci et al.(2011)] S.J. Ceci, W.M. Williams, “*Understanding current causes of women’s underrepresentation in science*”, *PNAS* 108 (2011) 3157.
- [Ceci et al.(2014)] S.J. Ceci, D.K. Ginther, S. Kahn, W.M. Williams, “*Women in academic science: A changing landscape*”, *Psychological Science in the Public Interest* 15 (2014) 75.
- [Ceci et al.(2015)] S.J. Ceci, W.M. Williams, “*Women have substantial advantage in STEM faculty hiring, except when competing against more-accomplished men*”, *Front. Psychol.* 6 (2015) 1532.
- [Chen et al.(2007)] P. Chen, H. Xie, S. Maslov, and S. Redner, “*Finding scientific gems with google*”, *J. Informet.* 1 (2007) 8 [arXiv:physics/0604130].
- [Cole et al.(1987)] J.R. Cole, H. Zuckerman, “*Marriage, motherhood, and research performance in science*”, *Scientific American* 256 (1987) 119.
- [Edwards et al.(2018)] H.A. Edwards, J. Schroeder, H.L. Dugdale, “*Gender differences in authorships are not associated with publication bias in an evolutionary journal*”, *PloS One* 29 (2018) e0201725.

- [Flaherty (2018)] K. Flaherty, “*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:1810.01511].
- [Fox (1995)] M.F. Fox, “*Women in scientific careers*” (1995), in S. Jasanoff et al. editors, “*Handbook of science and technology studies*” (pag. 205), Thousand Oaks, Sage.
- [Fox (2005)] M.F. Fox, “*Gender, family characteristics, and publication productivity among scientists*”, *Social Studies of Science* 35 (2005) 131.
- [Ginther et al.(2009)] D.K. Ginther, S. Kahn, “*Does science promote women? Evidence from Academia 1973-2001*” in R.B. Freeman et al. (editors), “*Science and engineering careers in the United States (National Bureau of Economic Research Conference Report)*” (pag. 163), University of Chicago Press.
- [Ginther et al.(2015)] D.K. Ginther, S. Kahn, “*Comment on “Expectations of brilliance underlie gender distributions across academic disciplines”*”, *Science* 24 (2015) 391.
- [Glass et al.(2010)] C. Glass, K. Minnotte, “*Recruiting and hiring women in STEM fields*”, *J. Divers. High. Educ.* 3 (2010) 218.
- [Halpern et al.(2007)] D.F. Halpern, C.P. Benbow, D.C. Geary, R.C. Gur, J.S., Hyde, M.A. Gernsbacher, “*The science of sex differences in science and mathematics*”, *Psychological Science in the Public Interest* 8 (2007) 1.
- [Hargens et al.(1978)] L.L. Hargens, J.C. McCann, B.F. Reskin, “*Productivity and reproductivity: Fertility and professional achievement among research scientists*”, *Social Forces* 57 (1978) 154.
- [Hill (2017)] T.P. Hill, “*An Evolutionary Theory for the Variability Hypothesis*” [arXiv:1703.04184], accepted for publication on the Mathematical Intelligencer but never published see “*Academic Activists Send a Published Paper Down the Memory Hole*”.
- [Holman et al.(2018)] L. Holman, D. Stuart-Fox, C.E. Hauser, C. Sugimoto, “*The gender gap in science: How long until women are equally represented?*”, *PloS Biol.* 16 (2018) e2004956.
- [Hooydonk (1997)] G. V. Hooydonk, “*Fractional counting of multiauthored publications: Consequences for the impact of authors*”, *JASIS* 48 (1997) 10.
- [Hossenfelder et al.(2018)] S. Hossenfelder, “*Do women get fewer citations than men?*”, talk at Chapman University, 2018/11/29.
- [Hyde (2014)] J.S. Hyde, “*Gender Similarities and Differences*”, *Ann. Rev. Psy.* 65 (2014) 3.1.
- [InSPIRE(2010)] *High-Energy Physics Literature IN- SPIRE Database*. A. Holtkamp, S. Mele, T. Simko, and T. Smith (INSPIRE collaboration), “*Inspire: Realizing the dream of a global digital library in high-energy physics*”, 2010.
- [Jagsi et al. (2006)] R. Jagsi, E.A. Guancial, C.C. Worobey, L.E. Henault, Y. Chang et al., “*The gender gap in authorship of academic medical literature — a 35-year perspective*”, *N. Engl. J. Med.* 355 (2006) 281.
- [Knapp (2010)] M. Knapp, “*Are participation rates sufficient to explain gender differences in chess performance?*”, *Proc. Biol. Sci.* 277 (2010) 2269.
- [Kwiek (2016)] M. Kwiek, “*The European research elite: a cross-national study of highly productive academics in 11 countries*”, *Higher Education* 71 (2016) 379.
- [Lariviere et al.(2013)] V. Lariviere, C. Ni, Y. Gingras, B. Cronin, C.R. Sugimoto, “*Bibliometrics: Global gender disparities in science*”, *Nature* 504 (2013) 211.
- [Levin et al.(1998)] S.G. Levin, P.E. Stephan, “*Gender differences in the rewards to publishing in Academe: Science in the 1970s.*”, *Sex Roles* 38 (1998) 1049
- [Leydesdorff et al.(2016)] L. Leydesdorff and H. W. Park, “*Full and fractional counting in bibliometric networks*”, *CoRR* abs/1611.06943 (2016) [arXiv:1611.06943].
- [Ley et al.(2008)] T. Ley, B. Hamilton, “*Gender gap in NIH grant applications*”, *Science* 322 (2008) 1472.
- [Lippa (2010)] R.A. Lippa, “*Gender Differences in Personality and Interests: When, Where, and Why?*”, *Social and Personality Psych. Compass* 4 (2010) 1098.
- [Ma et al.(2008)] N. Ma, J. Guan, and Y. Zhao, “*Bringing pagerank to the citation analysis*”, *Inf. Process. Manage.* 44 (2008) 800.

- [Marsh et al.(2011)] H.W. Marsh, U.W. Jayasinghe, N.W. Bond, “*Gender differences in peer reviews of grant applications: A substantive-methodological synergy in support of the null hypothesis model*”, Journal of Informetrics 5 (2011) 167.
- [Miller et al.(2015)] D.I. Miller, J. Wai, “*The bachelor’s to Ph.D. STEM pipeline no longer leaks more women than men: a 30-year analysis*”, Front. Psychol. 6 (2015) 37.
- [Moldwin et al. (2018)] M.B. Moldwin, M.W. Liemohn, “*High-Citation Papers in Space Physics: Examination of Gender, Country, and Paper Characteristics*”, J. of Geophysical Research 123 (2018) 2557.
- [Moss-Racusin et al.(2012)] C.A. Moss-Racusin, J.F. Dovidio, V.L. Brescoll, M.J. Graham, J. Handelsman, “*Science faculty subtle gender biases favor male students*”, PNAS 109 (2012) 41 16474.
- [Mutz et al.(2012)] R. Mutz, L. Bornmann, H.D. Daniel, “*Does gender matter in grant peer review? An empirical investigation using the example of the Austrian Science Fund*”, Zeitschrift fur Psychologie 220 (2012) 121.
- [National Research Council] National Research Council, “*Gender Differences at Critical Transitions in the Careers of Science, Engineering and Mathematics Faculty*”, National Academies Press (2009).
- [O’Dea et al.(2018)] R.E. O’Dea, M. Lagisz, M. D. Jennions, S. Nakagawa, “*Gender differences in individual variation in academic grades fail to fit expected patterns for STEM*”, Nature Communications 9 (2018) 3777.
- [Perianes-Rodriguez et al.(2016)] Perianes-Rodriguez, L. Waltman, N.J. van Eck, “*Constructing bibliometric networks: A comparison between full and fractional counting*”, J. Informetrics 10 (2016) 1178 [arXiv:1607.02452].
- [Pinski et al.(1976)] G. Pinski, F. Narin, “*Citation influence for journal aggregates of scientific publications: Theory, with application to the literature of physics*”, Inf. Process. Manag. 12 (1976) 297.
- [Radicchi et al.(2009)] F. Radicchi, S. Fortunato, B. Markines, A. Vespignani, “*Diffusion of scientific credits and the ranking of scientists*”, Physical Review E80 (2009) 056103 [arXiv:0907.1050].
- [Rossi et al.(2019)] P. Rossi, A. Strumia, A. Torre, “*Bibliometrics for collaboration works*” [arXiv:1902.01693].
- [Ruocco et al.(2017)] G. Ruocco, C. Daraio, V. Folli, M. Leonetti, “*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 (2017) 17064.
- [Sax et al.(2002)] L.J. Sax, L.S. Hagedorn, M. Arredondo, M., F.A. Dicrisi, “*Faculty research productivity: Exploring the role of gender and family-related factors*”, Research in Higher Education 43 (2002) 423.
- [Sinatra et al.(2005)] R. Sinatra, P. Deville, M. Szell, D. Wang, A.L. Barabasi, “*A century of physics*”, Nature Physics 11 (2005) 791.
- [Stack (2004)] S. Stack, “*Gender, Children and Research Productivity*”, Research in Higher Education 45 (2004) 891.
- [Stevens et al.(2017)] S. Stevens, J. Haidt, “*The Google Memo: What Does the Research Say About Gender Differences?*”, Heterodox Academy (2017).
- [Stoet et al.(2018)] G. Stoet, D.C. Geary, “*The Gender-Equality Paradox in Science, Technology, Engineering, and Mathematics Education*”, Psychological Science 29 (2018) 581.
- [Strumia et al.(2018)] A. Strumia, R. Torre, “*Biblioranking fundamental physics*” [arXiv:1803.10713].
- [Su et al.(2009)] R. Su, J. Rounds, P.I. Armstrong, “*Men and things, women and people: A meta-analysis of sex differences in interests*”, Psychol. Bull. 135 (2009) 859.
- [Su et al.(2015)] R. Su, J. Rounds, “*All STEM fields are not created equal: People and things interests explain gender disparities across STEM fields*”, Front Psychol. 6 (2015) 189.
- [Thelwall (2018)] M. Thelwall, “*Do females create higher impact research? Scopus citations and Mendeley readers for articles from five countries*”, Journal of Informetrics 12 (2018) 1031 [arXiv:1808.03296].
- [Thelwall (2018)b] M. Thelwall, “*She’s Reddit: A source of statistically significant gendered interest information?*” [arXiv:1810.08091].
- [Thelwall et al.(2014)] M. Thelwall, P. Wilson, “*Regression for citation data: an evaluation of different methods*”, Journal of Infometrics 8 (2014) 963.

- [Torvik et al.(2016)] V.I. Torvik, S. Agarwal, "*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, 2016. *Ethnea web site*.
- [Wang et al.(2013)] M.T. Wang, J. Eccles, S. Kenney, "*Not lack of ability but more choice: Individual and gender differences in choice of careers in science, technology, engineering, and mathematics*", Psychological Science 24 (2013) 770.
- [Way et al.(2016)] S.F. Way, D.B. Larremore, A. Clauset, "*Gender, Productivity, and Prestige in Computer Science Faculty Hiring Networks*", Proc. 2016 World Wide Web Conference (WWW) (2016) 1169 [arXiv:1602.00795].
- [Wenneras et al.(1997)] C. Wenneras, A. Wold, "*Nepotism and sexism in peer-review*", Nature 387 (1997) 341.
- [West et al. (2013)] J.D. West, J. Jacquet, M.M. King, S.J. Correll, C.T. Bergstrom, "*The role of gender in scholarly authorship*", PloS One 8 (2013) e66212.
- [West et al.(2014)] J.D. West, M.C. Jensen, R.J. Dandrea, G.J. Gordon, C.T. Bergstrom, "*Author-level eigenfactor metrics: Evaluating the influence of authors, institutions, and countries within the social science research network community*", JASIST 64 (2014) 787.
- [WGND] Worldwide Gender-Name Dictionary.
- [Williams et al.(2017)] W.M. Williams, S.J. Ceci, "*2:1 faculty preference for women on STEM tenure track*", Proceedings of the National Academy of Sciences 112 (2017) 5360.
- [Wolfinger et al.(2008)] N.H. Wolfinger, M.A. Mason, M. Goulden, "*Problems in the pipeline: Gender, marriage, and fertility in the ivory tower.*", J. Higher Educ. 79 (2008) 388
- [Wolfram et al.] Wolfram Mathematica, Determine the Gender of Given Name.
- [World Economic Forum] World Economic Forum, "*The Global Gender Gap Report 2016*".
- [Xie et al.(1998)] Y. Xie, K.A. Shauman, "*Sex differences in research productivity: New evidence about an old puzzle*", American Sociological Review 63 (1998) 847.
- [Xie et al.(2003)] Y. Xie, Y. K. Shauman, "*Women in science: Career processes and outcomes*", Cambridge, Harvard University Press.