

# Gender Imbalance in Mathematics Citations

## Do female mathematicians get cited less than men?

Undergraduate project by Katharina Limbeck supervised by Janine Illian

### Introduction

The gender gap in STEM subjects has been widely studied and it is a known fact that women are under-represented in mathematics, especially in the most reputable mathematics journals. One aspect of this gender gap, which has not yet been analysed specifically for mathematics is gender imbalance in papers' reference lists. Because citations are often used to measure academic success, it is of great interest to see if evidence of gender bias can be observed in the citation behaviour of mathematicians.

Analysing gender imbalance in neuroscience reference lists a recently published study by Dworkin et al. (2020) found that:

- Neuroscience articles written by female last or first authors get cited less than expected.
- This imbalance is driven by the citation behaviour of men.
- Gender inequality in citations is increasing over time.
- Citation imbalance is growing faster in male authored reference lists.

Now, the main aim of this project is to successfully adapt the methods used by Dworkin et al. (2020) to analyse citation imbalance in mathematics.

### Data Collection

|    | 10 Top Mathematics Journals                       | 2019 MCQ | No. Articles |
|----|---|----------|--------------|
| 1  | Annals of Mathematics                             | 5.24     | 1,460        |
| 2  | Journal of the American Mathematical Society      | 5.14     | 759          |
| 3  | Publications Mathematiques                        | 4.67     | 243          |
| 4  | Acta Mathematica                                  | 3.97     | 409          |
| 5  | Inventiones Mathematicae                          | 3.56     | 2,269        |
| 6  | Communications in Pure and Applied Mathematics    | 3.43     | 1,397        |
| 7  | Memoirs of the American Mathematical Society      | 2.87     | 878          |
| 8  | Duke Mathematical Journal                         | 2.73     | 2,499        |
| 9  | Transactions of the American Mathematical Society | 2.36     | 7,832        |
| 10 | Journal of the European Mathematical Society      | 2.28     | 927          |

Select ten best entries from two journal rankings:

- pure mathematics journals rates as A\* by the 2010 ERA ranking
- with the highest Mathematical Citation Quotients in 2019.

Download data about all papers published in these journals between 1990 and 2019 from the Web of Science Core Collection. Authors' names, publication dates, publication journals, full reference lists and unique Digital Object Identifiers (DOI) are extracted for all 18,673 articles.

Full author names are checked online using "Crossref".

An author name disambiguation algorithm is implemented to further complete name data.

### Conclusions

#### Representation in 10 Top Journals

The overwhelming majority, 90%, of all identified authors of top mathematics papers are men. Only 65 papers written by multiple women but 6,025 papers written by multiple men were found.

#### Citation gaps after accounting for papers' specific characteristics

Papers written by multiple men are cited significantly more than expected.

Articles authored by multiple women or mixed teams are cited significantly less than expected.

Papers written by single authors of both genders are neither significantly under- nor overcited.

#### Recommendations

- Cite conscientiously!
- Actively consider how you want to reproduce your own discipline, for example by computing the gender diversity in your own reference lists.

#### Effect of Citing Authors' Genders

Citation imbalance does not differ significantly between man- or women-led reference lists.

The under-citation of female authors does not grow significantly faster in man-led reference lists.

#### Trends in citation imbalance

The over-citation of articles only authored by men and the corresponding under-citation of papers also written by women are increasing over time.

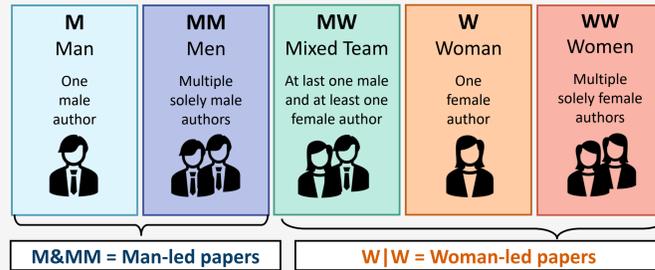
#### Further Questions

- What would be a fair distribution of citations?
- How do research sub-fields in mathematics influence citation imbalance?
- Can gender differences in citations also be observed in other journals?

### Gender Assignment

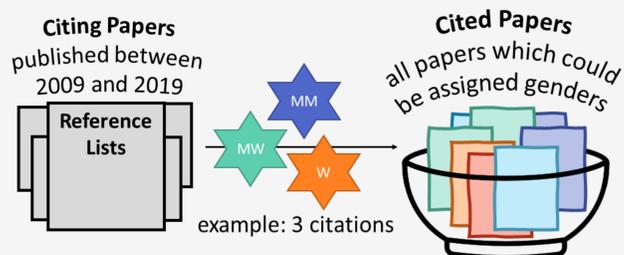
Gender assignment services are used to assign probabilities of belonging to a specific gender to all unique first names. Names are checked with U.S. baby name data, gender-guesser, genderze.io and NamSor. An author is labelled man or woman if the corresponding probability is bigger than 70%.

Articles are categorised according to all authors' genders:



This assignment of gender categories is approximately 92% accurate and can assign 11,870 papers. All unassigned papers are excluded from further analysis.

### Observed Citations



Analyse citations given by citing papers (published between 2009 and 2019) to cited papers (all articles which could be assigned to gender categories).

Identify citations by matching DOIs of reference list entries to other papers in the data. If an entry does not include DOI other details such as publication year and author name are matched. This process finds 11,870 observed citations by 4,148 citing papers after excluding 6,540 self-citations i.e. instances of authors citing their own work.

Observed citations are compared to expected citation counts under two different models of fair citation behaviour independent of gender.

### Measuring Citation Imbalance

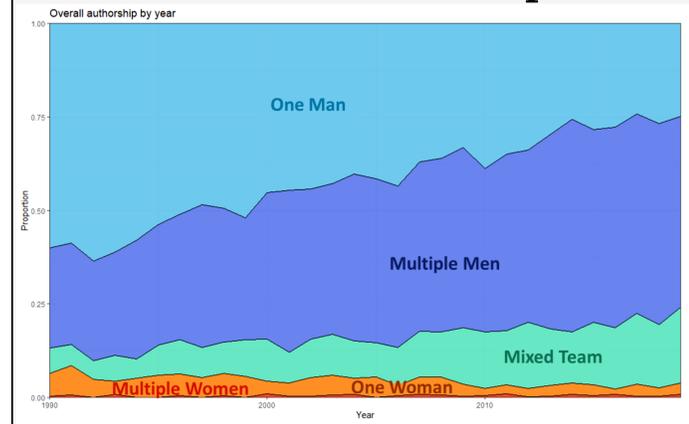
Citation imbalance is computed as the percent difference between the numbers of observed and the expected citations given to each gender group.

For example the over- or under-citation of single female authors is

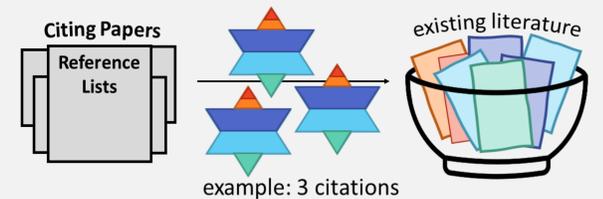
$$\Delta W = \frac{obs(W) - exp(W)}{exp(W)}$$

- obs(W) is the number of all observed citations given to W papers.
- exp(W) is the number of all expected citations given to W papers.

### Trends in Authorship

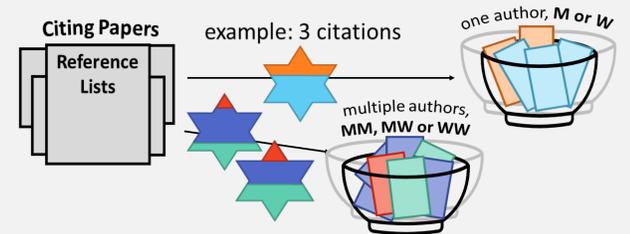


### 1 Random Draw Model



Assume citations are random draws from existing literature. Citations are split up according to the gender-make up of all articles published before the citing paper. This gives unconditional expected citation counts.

### 2 Conditional Citation Model

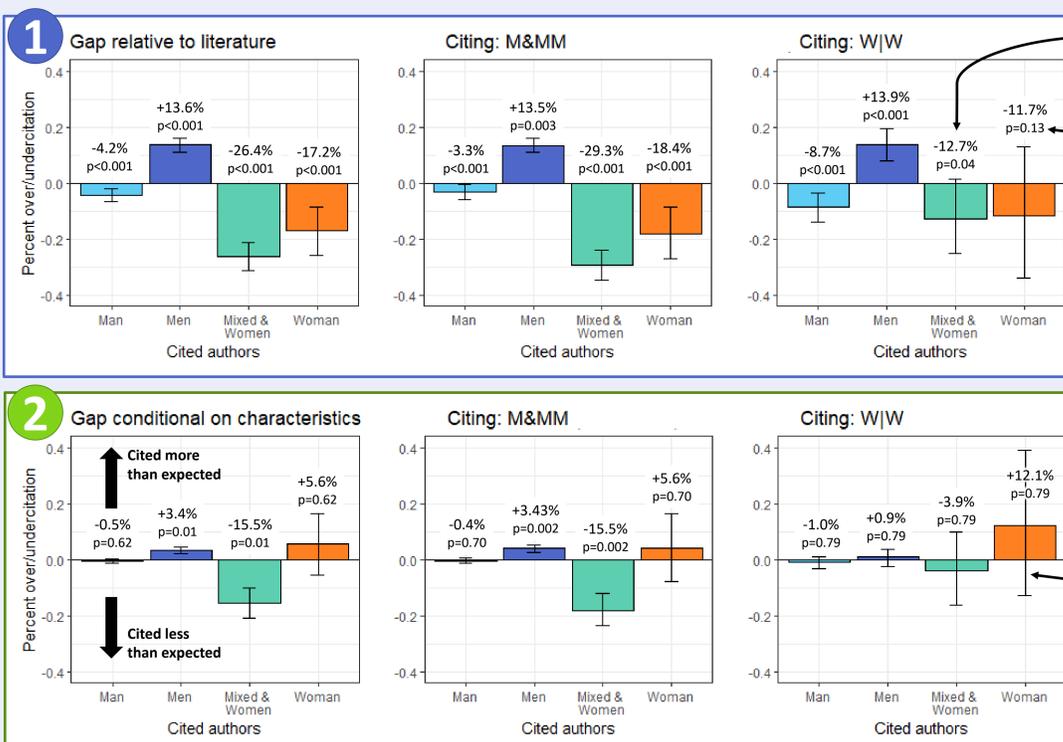


- Assume citations are draws from papers similar to the cited paper with
- the same number of authors (one or multiple)
  - similar publication dates, similar publication journals
  - similar senior authors and similar team size.

Fit two multinomial generalized additive models conditional on these characteristics. One model assigns M or W to papers with one author, the other assigns MM, MW or WW to papers with multiple authors.

These models give probabilities of belonging to specific gender categories for each cited papers. Conditional expected citation counts are then split up according to these probabilities.

### Results - Gender Imbalance in Citations



#### Mixed & Women

Analyse MW and WW papers together because very few WW articles were found.

#### Statistical Significance:

Compute "mock" observed citation counts, 10,000 times as random draws or 5,000 times as draws from similar papers. Re-calculate percent difference to get null estimates.

The p-values then indicate the proportion of null estimates, which are more extreme than the observed citation imbalance corrected by the Holm-Bonferroni method

#### 95% confidence intervals

Re-calculating citation imbalance across 1,000 bootstrap samples of citing papers. Take the 2.5 and 97.5 percentiles of these bootstrap estimates