Do male researchers disregard the work of female researchers? The role of gender in citation decisions

Analytical Sociology: Theory and Empirical Applications
Venice International University, Nov 2019

Alexander Tekles, Katrin Auspurg & Lutz Bornmann
LMU Munich / MPG
Motivation: The Gender Gap in Citations

• Male’s work is more frequently cited than female’s work

Based on 5,483,841 research papers in the Web of Science database 2008-2013 (Sugimoto 2018)

• Caused by a “gender homophily bias”? 
Gender Homophily Bias in Citations

• Disproportionately citing references of own gender

<table>
<thead>
<tr>
<th>Male authors</th>
<th>% references of male authors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female authors</td>
<td>% references of female authors</td>
</tr>
</tbody>
</table>

• Example: articles in sociology journals 1985-1994
  (Davenport 1995; random sample out of SSCI-journals)

Gender homophily rate (GHR):
88% - 67% = 21ppts

Cited references
- Male (first) author
- Female (first) author
Gender Homophily Bias in Citations

- Authors disproportionately cite references of own gender

Example: articles in sociology journals 1985-1994
(Davenport 1995, random sample out of SSCI-journals)

- So far all studies report evidence for gender homophily
  - \( N = 12 \) studies, covering many disciplines
  - Mean GHR (also including mixed gender teams): 12ppts
Partly Caused by Self-Citations

• In particular men tend to cite themselves

• But even without self-citations studies find evidence for gender homophily bias (e.g. Ghiasi et al. 2018; Pothoff & Zimmermann 2017)
Further Explanations

• “Matilda effects”: Less recognition of the work of females
  (Ferber et al. 1986; Kanter 1970)

• But females might be more likely be aware of and cite females’ work (e.g. because of same-gender networks)

• “Implicit biases” (but how, and why??)

• So far there is not any clear and consistent explanation

• But there are lots of policy recommendations
Already Some Actions Taken

Row over 40% gender quota for reading lists at Lund
Jan Petter Myklebust 24 November 2017

114TH 2019 ANNUAL MEETING | AUGUST 10-13, 2019 | NEW YORK, NY | UNITED STATES
3156. Presidential Panel. White Supremacy and Sociology
Sheraton New York, Metropolitan Ballroom
West, Second Floor, 8:30-10:00 AM
Session Organizer: Wendy Leo Middelkoop, University of Texas
President: Glenn Edward Bracey, Sociology: An Organ of White Supremacy
Racism - Johnny E. Williams
White Ignorance, Foreclosed Knowledge
Damning Case of Sociology - Jennifer C. Mueller, Skidmore College
Revisiting White Sociology through the Publication Process - David G. Embrick, University of Connecticut; David L. Brunsma, Virginia Tech
The Citation Inflation of White Masculinity - Saida Grundy, Boston University

Ruth Pearce @NotRightRuth · 12 Aug.
Grundy: Segregated citation networks emerge because white men are not reading black feminism, queer theory etc. But also because of what we are (and are not) taught. Notes that black feminist scholars in sociology often taught only in feminist theory classes. #ASA19
Already Some Actions Taken

- The content of the entry needs to be accessible - Could an upper level UG student or PG student new to the topic understand this flagship?
  - Yes  ☐  No

Please briefly explain your reasoning, especially if your answer is no:

- Does it appropriately cite male and female authors?
  - Yes  ☐  No

If no, please make suggestions for citations to add/delete to the author to ensure a balance:
Citation Patterns Could Emerge from Unbiased Science?

- Homophily bias would exist with a direct gender effect
  
  ![Diagram: Male authors → % references of male authors](http://example.com/diagram.png)

- Unbiased selection of references would exist with
  - Maximum substantive fit to research question
  - Maximum quality (rigor, impact)
  - Selection based on the whole population of existing references

- These factors could lead to indirect effects (mediators)
  
  ![Diagram: Male authors → % references of male authors](http://example.com/diagram.png)
Horizontal Segregation: Research Field as Mediator

![Gender Distribution of Authorships](image)

- **Benign Prostatic Hyperplasia & Prostatitis (N = 1,487)**
- **Control of Gene Expression (N = 67,319)**
- **Cell Signaling (N = 72,995)**
- **Medical Genetics (N = 81,831)**
- **Fertility Regulation (N = 1,604)**

Authors

Pool of References

---

**correlation**

% (fe)male

---

Tekles/Auspurg/Bornmann

Gender in Citation Decisions
Time Trends: Age of Research Topic as Mediator

- Higher age of research topic → more references by males
- Males might more strongly focus on old, „classical“ topics (e.g. because of their higher academic age)

Source: Jagsi et al. 2006
Our Contribution

• Does gender homophily exist when controlling mediators?
  – Sophisticated measurement of fields/research topics
  – Indicator for quality of cited papers
Data

Web of Science

F1000Prime

Quality Ratings
- “good” vs. “very good” vs. “excellent”
- Ratings by different experts

Field/topic (keywords)
- Undefined no. of keywords per paper
- Assignments of different experts

Focal Papers
- Biology & medicine
- Selected by experts
- N = 162,071

Citing Papers
Without self-citations

Author & paper metadata
- Names & affiliations \(\rightarrow\) gender
- Publication year
Identification Strategy

• For focal papers, estimate the effect of focal paper gender on the share of male-authored papers among citing papers

• Linear regression
  – Units: focal papers
  – Dep. variable: share of male-authored papers among citing papers
  – Homophily: effect of focal paper gender
  – Control for keywords, quality rating, age of paper, team size

• Expectation
  – Without including control variables: positive effect of focal paper being authored by males
  – After adding control variables: smaller (no) effect of focal paper being authored by males
Preliminary Results (I)

DV: Share of male-authored papers among citing papers (%)

• M1: No controls (N = 42,718)
Preliminary Results (I)

DV: Share of male-authored papers among citing papers (%)

- M1: No controls (N = 42,718)
- M2: Controlling for field/topic keywords (N = 42,676)
Preliminary Results (I)

DV: Share of male-authored papers among citing papers (%)

- **M1**: No controls (N = 42,718)
- **M2**: Controlling for field/topic keywords (N = 42,676)
- **M3**: Full model (N = 42,676)
Preliminary Results (II)

- Considering binary variables for field/topic only controls effect of each keyword, independently of other keywords.
- But: Fields/topics may be better represented by certain combinations of keywords.
- Idea: for pairs of focal papers, use the number of shared keywords as indicator for topical similarity.
- For all pairs of focal papers (one female-authored, one male-authored) with at least x shared keywords: plot histogram of the difference in the share of male-authored papers in the citing papers.
Differences in share of male-authored papers among citing papers

Pairs of focal papers with at least 0 shared sections

Distribution w/o homophily
Preliminary Results (II)

Differences in share of male-authored papers among citing papers
Pairs of focal papers with at least 0 shared sections

Distribution w/o homophily
Average difference: 12.57

Difference in share of male-authored papers (percentage points):
male-authored focal paper - female-authored focal paper
Preliminary Results (II)

Differences in share of male-authored papers among citing papers
Pairs of focal papers with at least 1 shared sections

Average difference: 6.90
Preliminary Results (II)

Differences in share of male-authored papers among citing papers
Pairs of focal papers with at least 2 shared sections

Distribution w/o homophily

Average difference: 5.08

Difference in share of male-authored papers (percentage points):
male-authored focal paper - female-authored focal paper
Preliminary Results (II)

Differences in share of male-authored papers among citing papers
Pairs of focal papers with at least 3 shared sections

Distribution w/o homophily
Average difference: 3.54

Difference in share of male-authored papers (percentage points):
male-authored focal paper - female-authored focal paper
Preliminary Results (II)

Differences in share of male-authored papers among citing papers
Pairs of focal papers with at least 4 shared sections

Distribution w/o homophily

Average difference: 1.90

Difference in share of male-authored papers (percentage points):
male-authored focal paper - female-authored focal paper

Count

0

1000

500

0

-100

-50

0

50

100
Preliminary Results (II)

Differences in share of male-authored papers among citing papers
Pairs of focal papers with at least 5 shared sections

Distribution w/o homophily

Average difference:
-0.65
Preliminary Results (II)

Average difference: 12.57

Average difference: -0.65
Conclusions & Outlook

- Granularity of topological classification matters
- After thoroughly controlling for field/topic, evidence for gender homophily is completely gone
- Other variables have small effects (in our selective sample)
- General take home-message: Comparing citations (e.g. for evaluations) require thorough standardizations for fields/topics

- But only first results, we still work on robustness checks
  - Different operationalizations of author team’s gender
  - Analyses for non-F1000 Prime papers, e.g. in social sciences
  - Different approaches to control field/topic
    (e.g. similarity based on titles/abstracts)