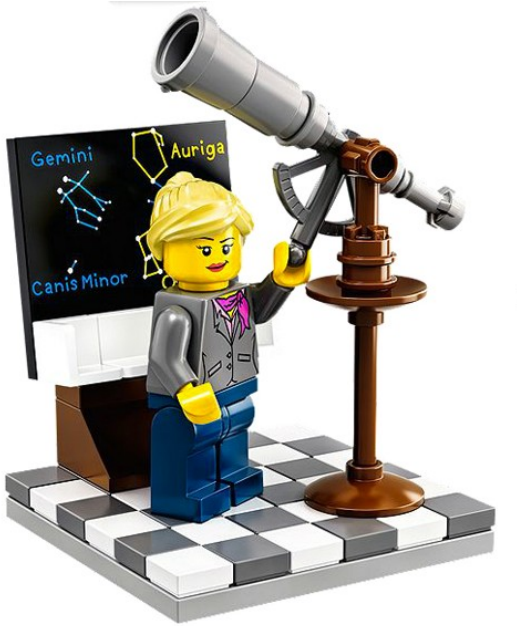
Quantitative Evaluation of Gender Bias in Astronomical Publications from Citation Counts

Neven Caplar

Sandro Tacchella, Simon Birrer

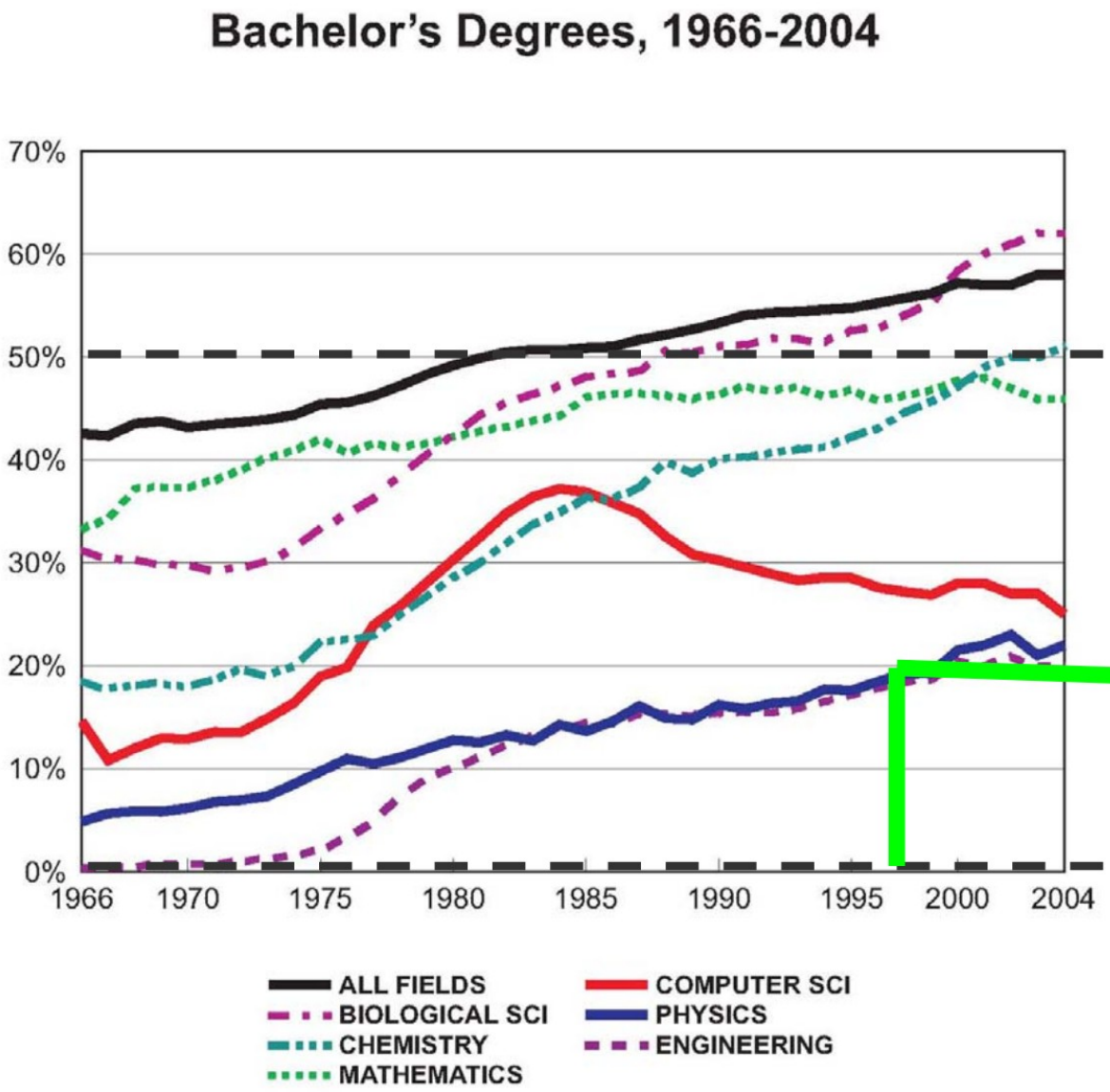


ETH

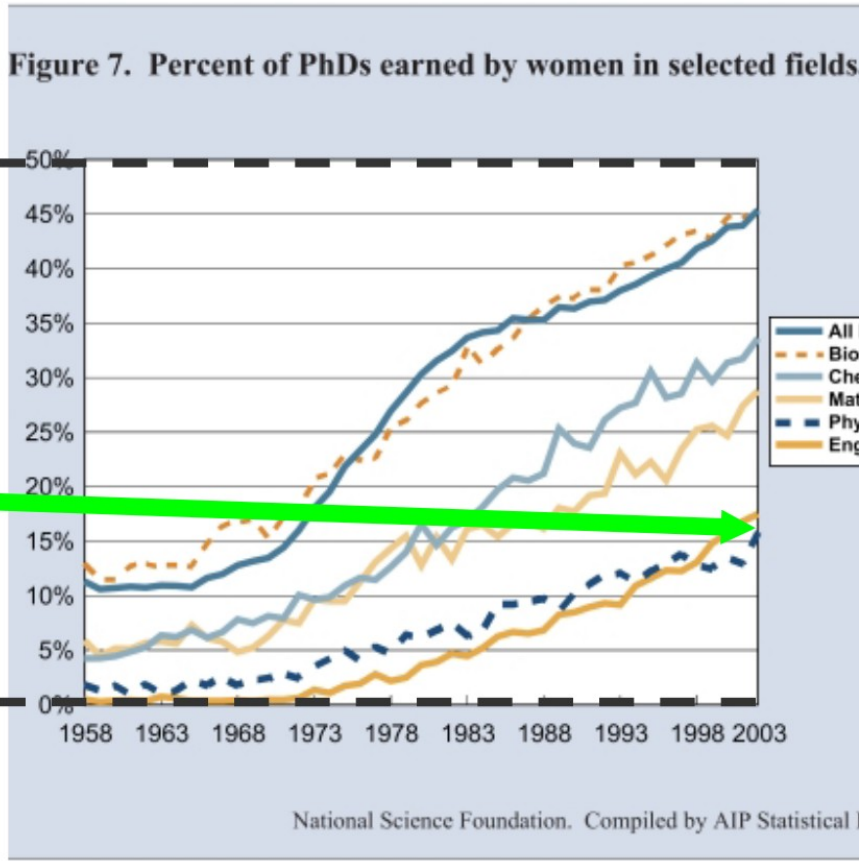
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich


ALL FIELDS
BIOLOGICAL SCI
CHEMISTRY
MATHEMATICS
COMPUTER SCI
PHYSICS
ENGINEERING

Attrition between B.S. and Ph.D. degrees



19% → 15% Physics



From talk by Meg Urry

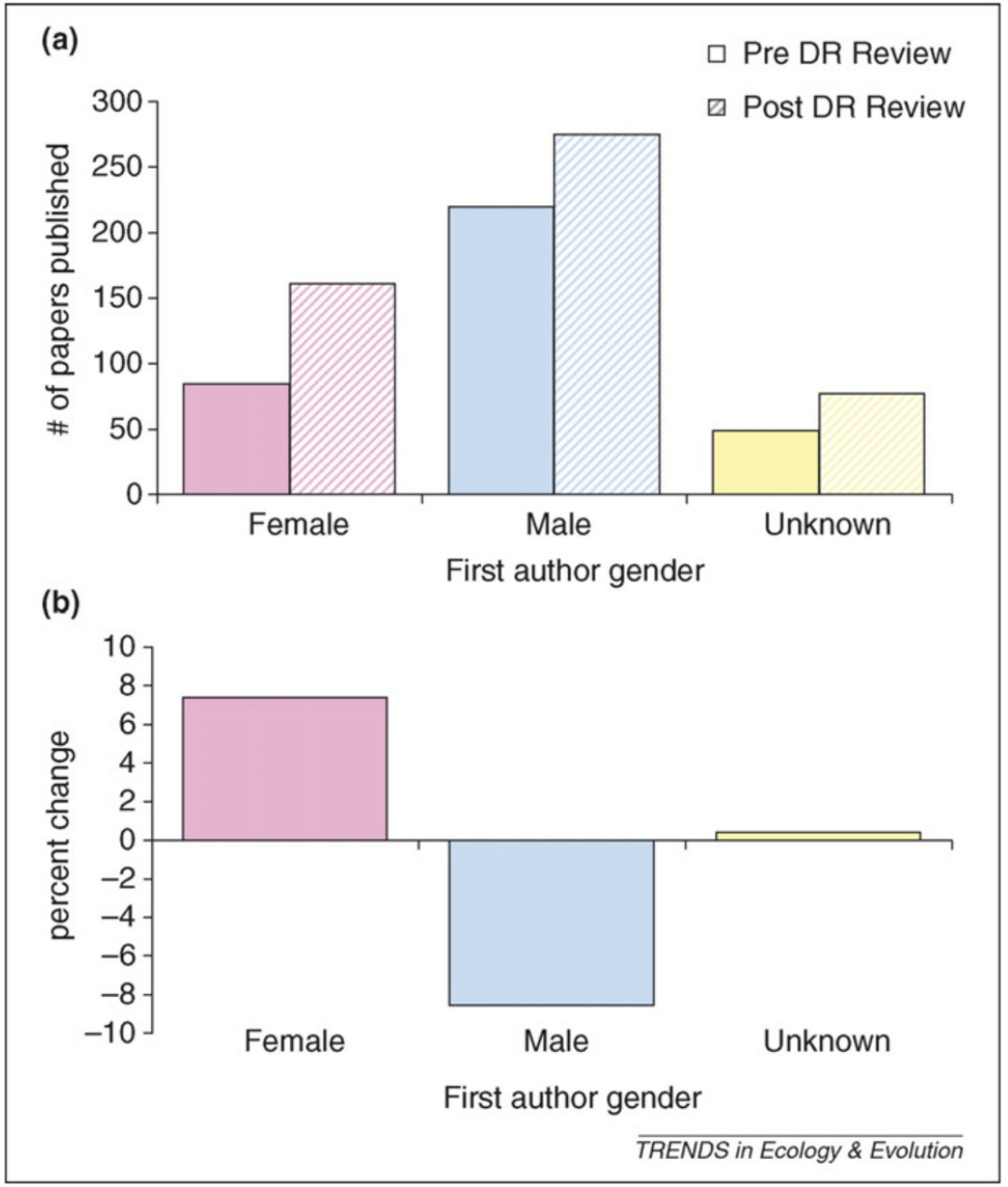
Gender difference in science

Table I. Mean Evaluation Scores of Men and Women

	Author of article			
	John T.	Joan T.	J. T.	Mean
Masculine article				
Men	1.9	2.9	2.5	2.4
Women	2.3	3.3	2.6	2.7
Mean	2.1	3.1	2.6	
Feminine article				
Men	1.8	3.7	2.9	2.8
Women	2.1	2.4	2.6	2.4
Mean	2.0	3.0	2.8	
Neutral article				
Men	2.0	2.4	2.7	2.4
Women	2.6	3.3	2.5	2.8
Mean	2.3	2.9	2.6	
Mean of combined articles				
Men	1.9	3.0	2.7	
Women	2.3	3.0	2.6	

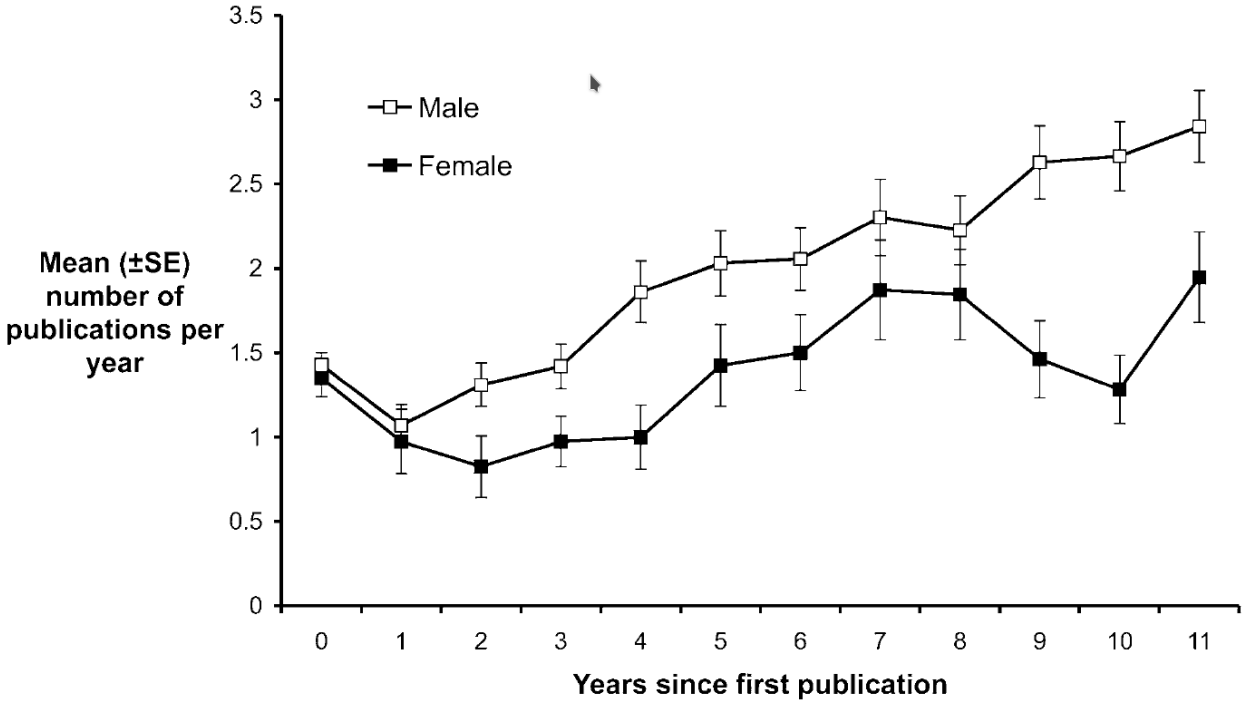
Paludi & Bauer, 1983

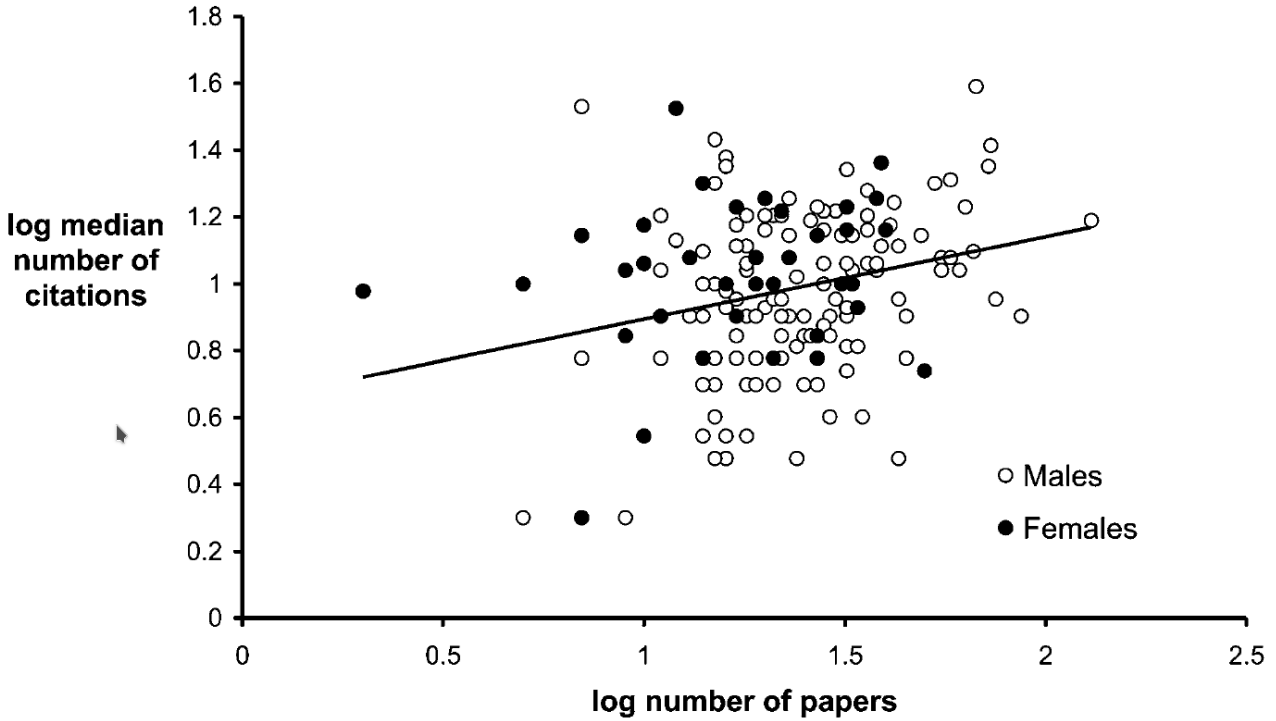
- Articles with the women listed as an author received the lower score than the same articles with a man listed an author
- Effect present for both women and men as referees



Budden+, 2008

- Articles with the women listed as an author received the lower score than the same articles with a man listed an author
- Effect present for both women and men as referees
- Fraction of papers authored by women increased after switching to double-blind refereeing system





- Articles with the women listed as an author received the lower score than the same articles with a man listed an author
- Effect present for both women and men as referees
- Fraction of papers authored by women increased after switching to double-blind refereeing system
- Men tend to publish more

Symonds+, 2006

Gendered Language in Teacher Reviews

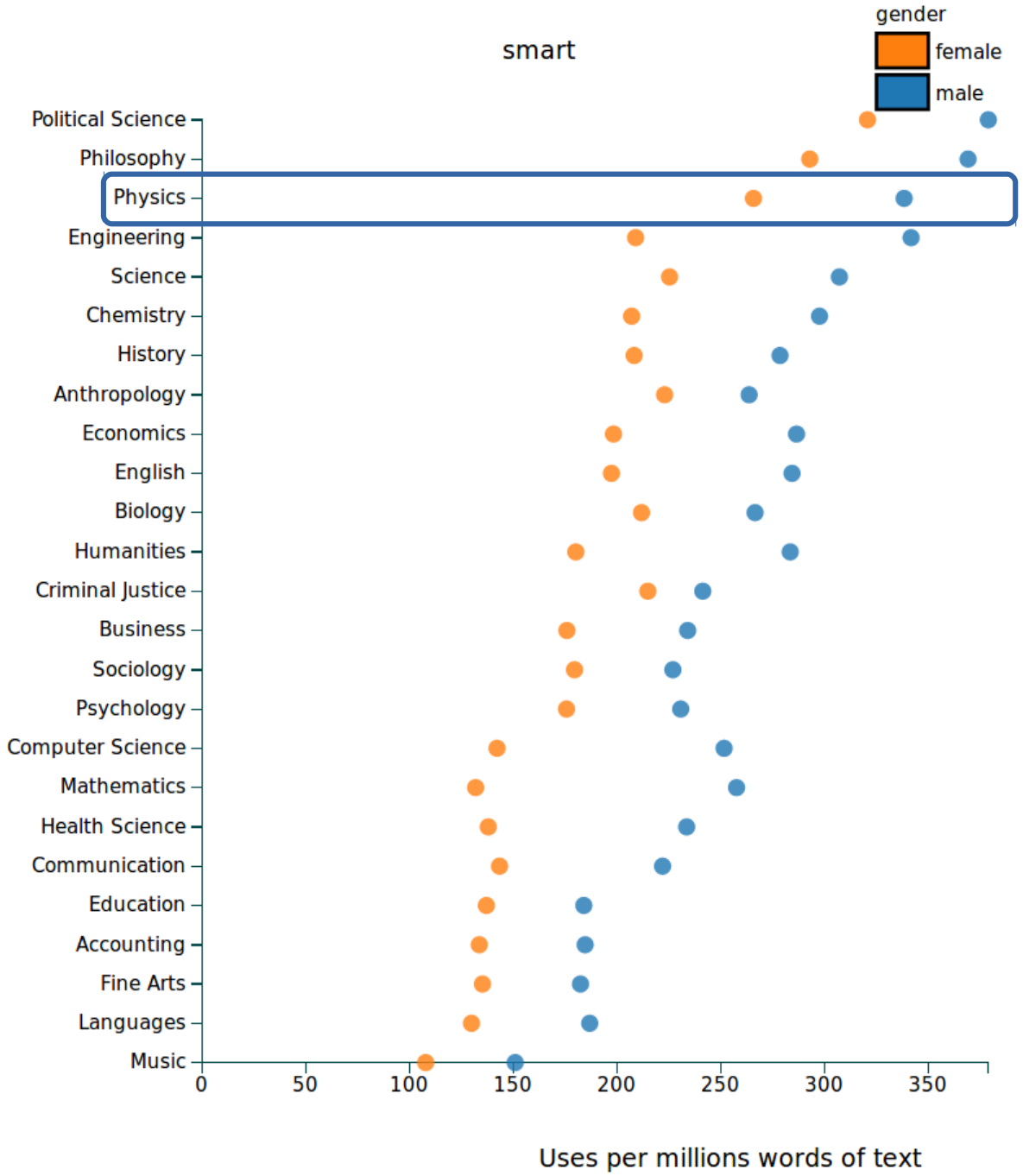
This interactive chart lets you explore the words used to describe male and female teachers in about 14 million reviews from RateMyProfessor.com.

You can enter any other word (or two-word phrase) into the box below to see how it is split across gender and discipline: the x-axis gives how many times your term is used per million words of text (normalized against gender and field). You can also limit to just negative or positive reviews (based on the numeric ratings on the site). For some more background, see here.

Not all words have gender splits, but a surprising number do. Even things like pronouns are used quite differently by gender.

Search term(s) (case-insensitive): use commas to aggregate multiple terms

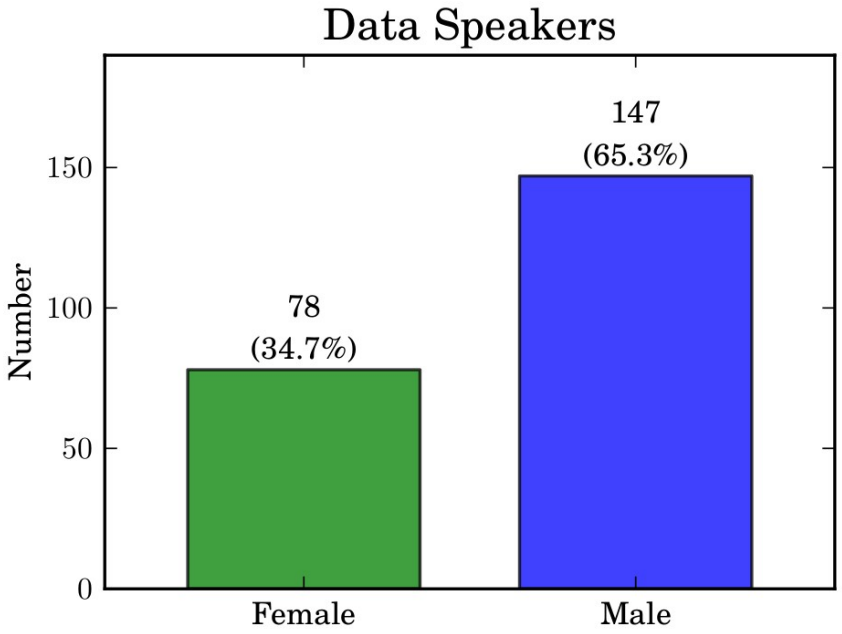
smart

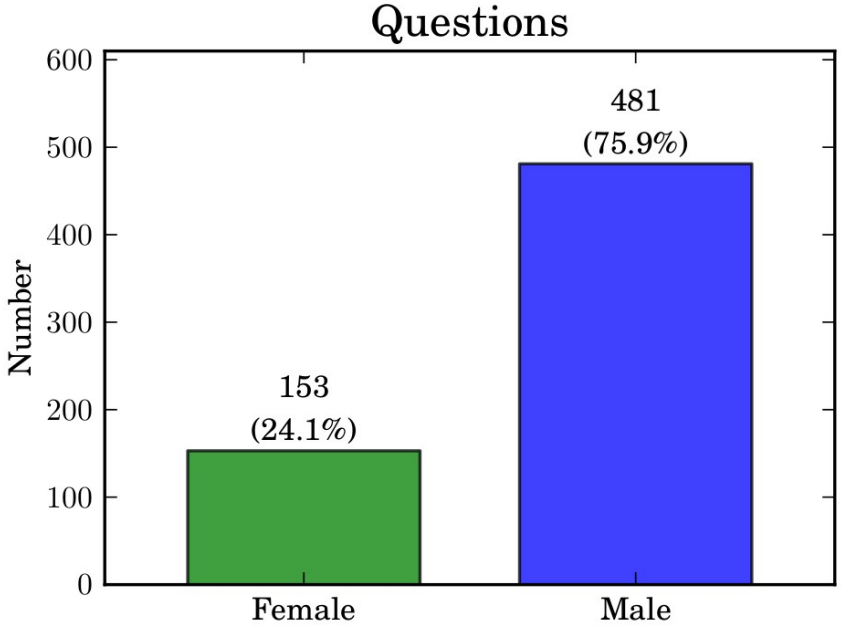


Overview

- Introduction
  - Gender difference in science
  - **Gender difference in astronomy**
- Method
  - Data gathering
  - Discussion of the sample
- Results
  - Gender difference in citation counts
  - Gender bias
  - Self citation and productivity
  - Discussion

Gender difference in astronomy





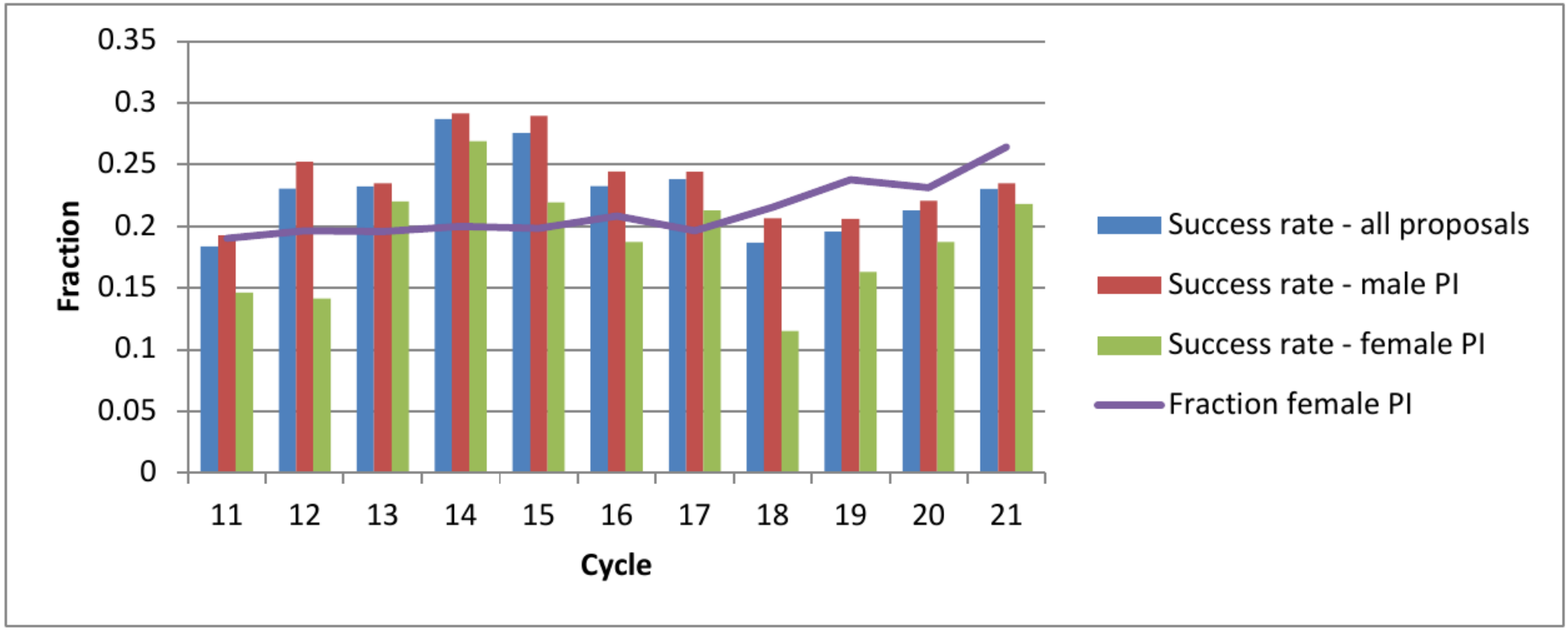
- Women ask less questions on conferences

Davenport+, 2014

Gender difference in astronomy

- Women ask less questions on conferences
- Women are less likely to get telescope time (seems even more so for older women)

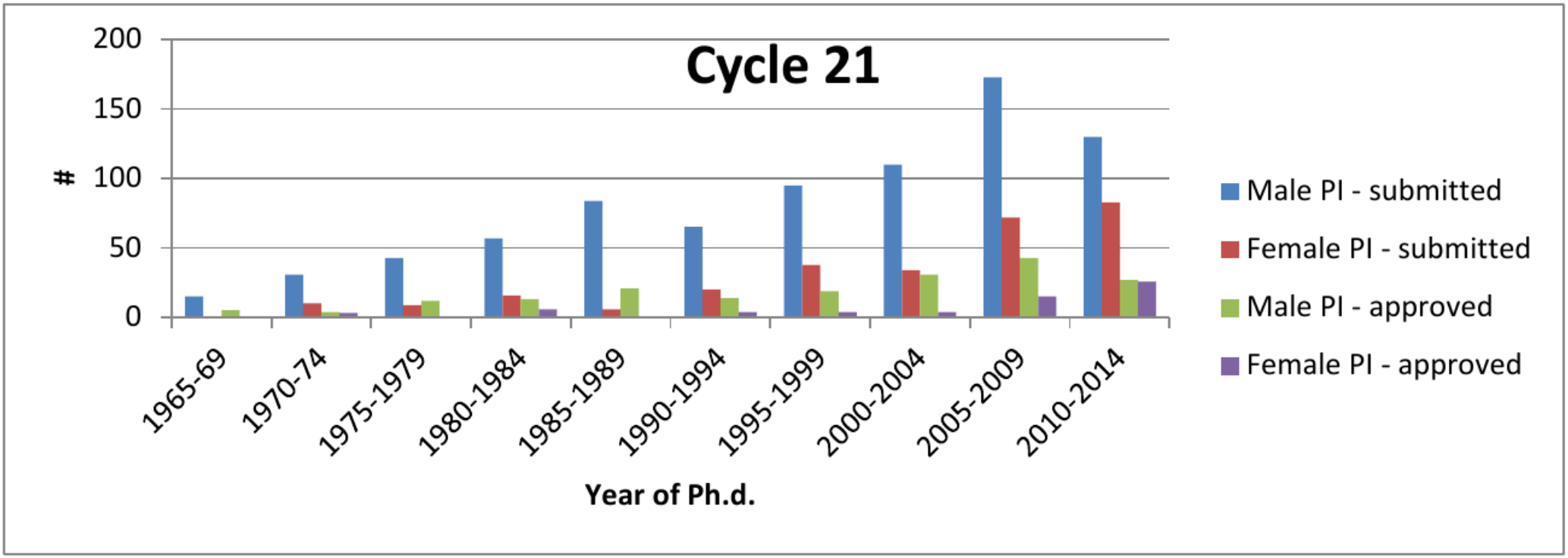
Reid+, 2014



Gender difference in astronomy

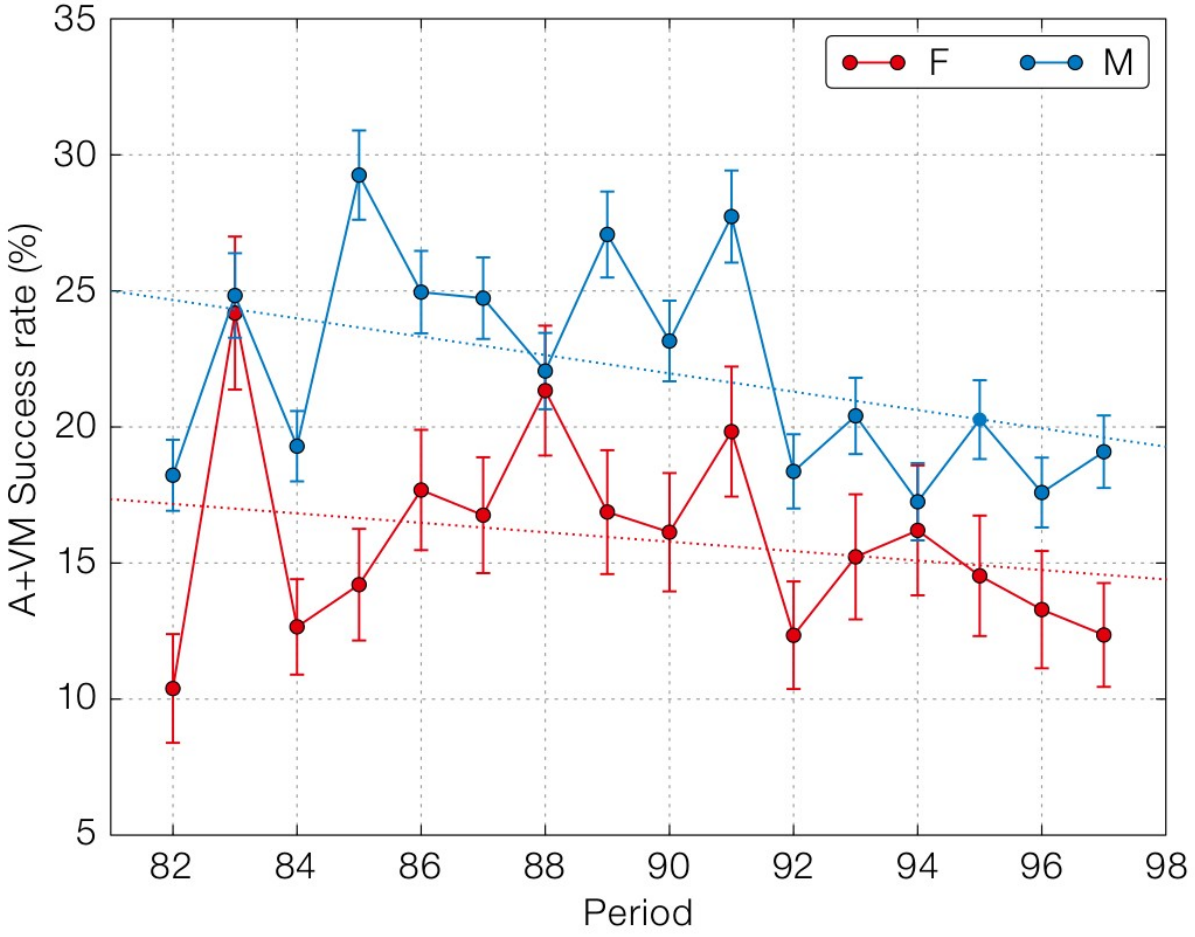
- Women ask less questions on conferences
- Women are less likely to get telescope time (seems even more so for older women)

Reid+, 2014



Gender difference in astronomy

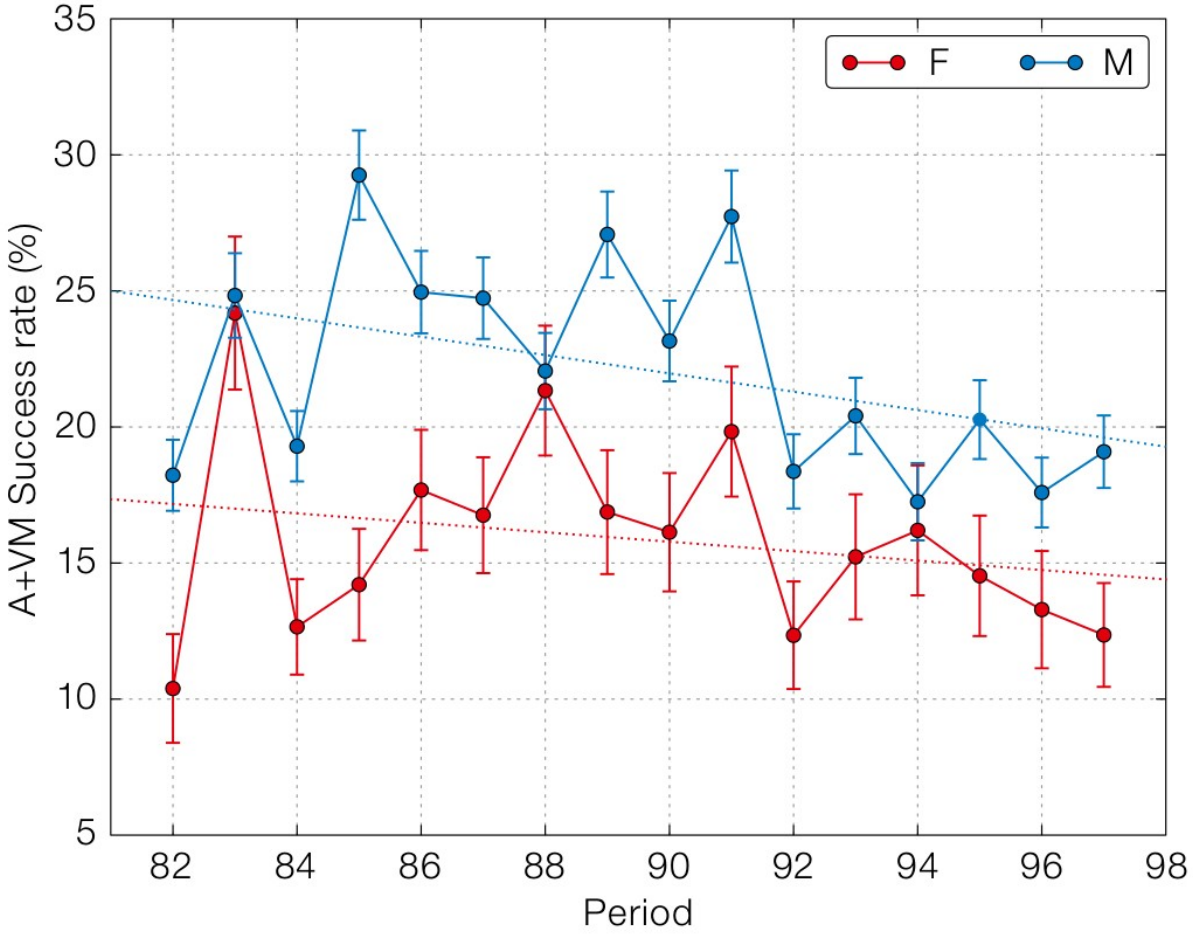
- Women ask less questions on conferences
- Women are less likely to get telescope time (seems even more so for older women)



Patat, 2016

Gender difference in astronomy

- Women ask less questions on conferences
- Women are less likely to get telescope time (seems even more so for older women)



Is there a difference between men and women in citations counts?

Patat, 2016

Overview

- Introduction
  - Gender difference in science
  - Gender difference in astronomy
- **Method**
  - Data gathering
  - Discussion of the sample
- Results
  - Gender difference in citation counts
  - Gender bias
  - Self citation and productivity
  - Discussion

Method

voxcharta

ArXiv discussions for 383 institutions including Adler Planetarium, UWC Cosmology, Galactic and Nongalactic, Cornell-Cosmo, Subaru JC, and IANCU.

ETH-Extragalactic

Home | About | Help | Tools | Keyword

Archive for March 3rd, 2017

ETH-Extragalactic's Discussion Agenda for Friday

Show votes from other institutions.

1 Reconciling Mass Functions with the Star-Forming Main Sequence Via Mergers
Neven Caplar View PDF Mark as not discussed

Yesterday's Postings

Sort: By everyone's votes in descending order. (Listings sorted in 0.651 seconds.)

Filters: ASTRO-PH CO EP GA HE IM SR GR-QC HEP-PH HEP-TH HEP-LAT HEP-EX NUCL-TH NUCL-EX

Show: Titles Only Conf. Proceedings Submitted

New papers (56) Collapse

[astro-ph #1] The Formation of the First Quasars in the Universe

0 votes @ETH-Extragalactic
(35 votes from 26 institutions)
Promote Demote
Comment
Suggest this paper! Suggest

Joseph Smidt[1†], Daniel J. Whalen, Jarrett L. Johnson[2†], Hui Li[3†]
[1]UC Irvine, [2]MPE, [3]NJU
[†]Listed affiliation is based on previous publications and was not specified in this preprint.
ArXiv #: 1703.00449 (PDF, PS, ADS, Papers, Other)
Comments: 6 pages, 5 figures. Submitted to ApJ
Originally posted 03/02/2017

GA

Supermassive black holes are the central engines of luminous quasars and are found in most massive galaxies today. But the recent discoveries of ULAS J1120+0641, a 2 × 10^9 $M_\odot$ black hole at $z \sim 7.1$, and SDSS J0100+2802, a 1.2 × 10^{10} $M_\odot$ black hole at $z = 6.3$, challenge current paradigms of cosmic structure formation because it is not known how quasars this massive appeared less than a billion years after the Big Bang. Here, we report new cosmological simulations of SMBHs with x-rays fully coupled to primordial chemistry and hydrodynamics that show that J1120+0641 and J0100+2802 can form from direct collapse black holes if their growth is fed by cold, dense accretion streams, like those thought to fuel the rapid growth of some galaxies at later epochs. Our models reproduce the mass, luminosity and ionized near zone of J1120+0641, as well as the star formation rate and metallicity in its host galaxy. They also match new observations of the dynamical mass of the central 1.5 kpc of its emission region just obtained with ALMA. We find that supernova feedback from star formation in the host galaxy regulates the growth of the quasar from early times.

[astro-ph #2] Reconciling Mass Functions with the Star-Forming Main Sequence Via Mergers

1 vote @ETH-Extragalactic
Charles L. Steinhardt, Dominic Yurk, Peter Capak

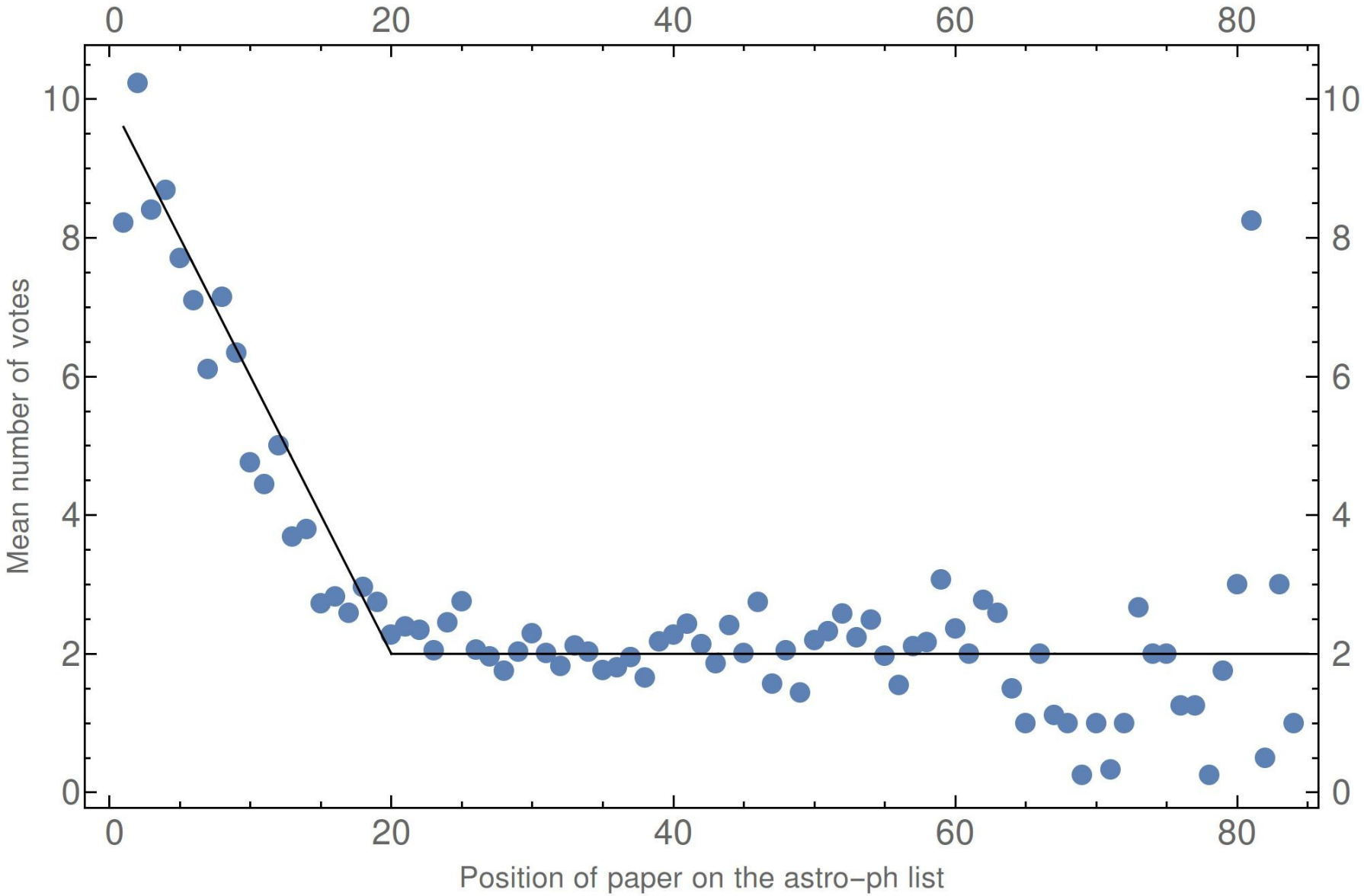
ASTRO-PH
Cosmology and Nongalactic
Earth and Planetary
Galaxies
High Energy
Instrumentation and Methods
Solar and Stellar
GR-QC
HEP-EX
HEP-LAT
HEP-PH
HEP-TH
NUCL-EX
NUCL-TH
ANNOUNCEMENTS
SPECIAL TOPICS
VOX CHARTA BLOG

S	M	T	W	
			1	
5	6	7	8	
12	13	14	15	
19	20	21	22	
26	27	28	29	

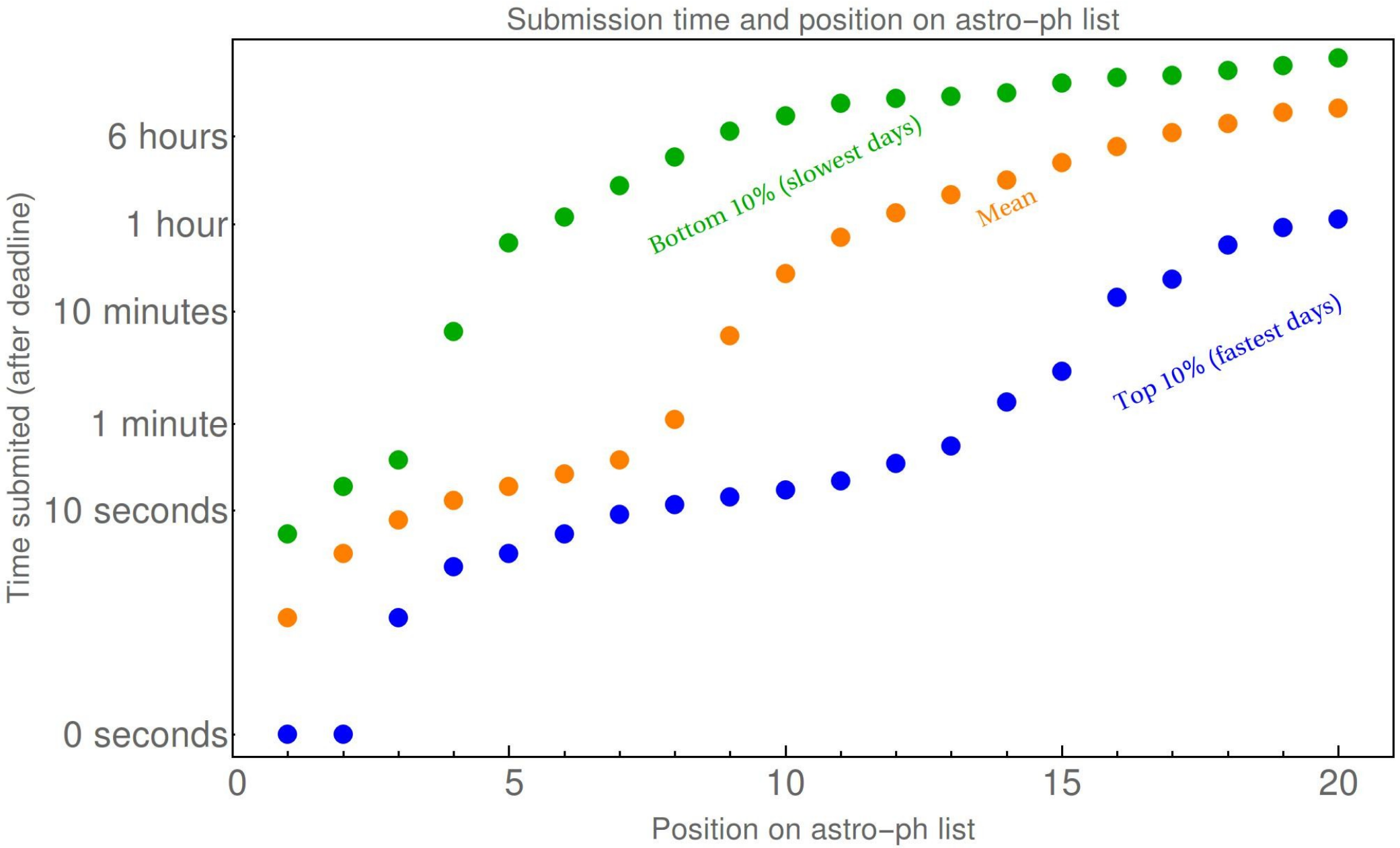
« Feb

CATEGORY RSS FEEDS

Method



- Number of “upvotes” correlated with the position on the arXiv list



- Top 5 on ArXiv papers are usually submitted within 10 seconds of deadline

Method

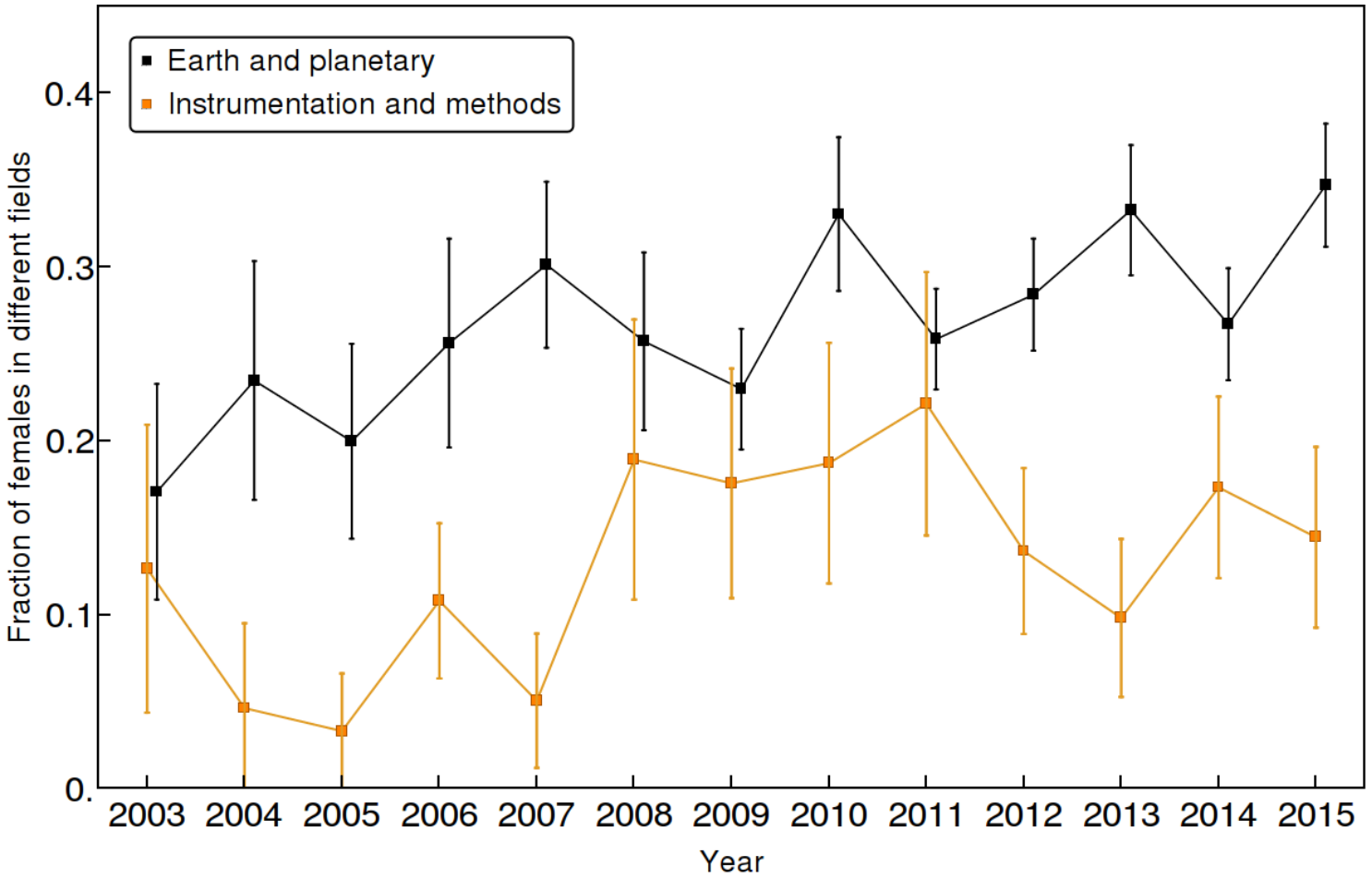
- Gathering data
  - Every paper in ADS database “astronomy” and published in Science, Nature, APJ, A&A, MNRAS from 1950 to 2015
  - All the information gathered in single effort in June 2016
  - If paper is available on arXiv, also record the subfield of the paper and download the source *.tex file
    - ArXiv data via querying available for papers after 2002
    - *.tex file (via S3 Amazon server) available for papers after 2007

Method

- Adding paper information
  - *.tex file used to establish length of papers
  - Subfield determined from abstract for papers where subfield is not recorded
- Adding information about authors
  - Country of origin from affiliation
  - Seniority = time since the first paper in our database
  - Gender
    - We run the name through 3 different databases
      - SexMachine (40,000 names, done by native speakers)
      - Data from USA Social Security Administration and UK Office of National Statistics (highly complete but geographically limited)
      - Gender API (commercial service)
    - Agreement between databases around 98.5%

Method

- Adding paper information
  - *.tex file used to establish length of papers
  - Subfield determined from abstract for papers where subfield is not recorded



- Total: 208,577 entries
- Final dataset: 149,741 entries

- Cleaning data
  - entries with zero citations or zero references (4,417 ADS entries);
  - authors that have only published in Science and/or Nature (5,484 ADS entries);
  - entries with no authors specified (491 ADS entries);
  - entries with no first name for the first author (e.g. collaboration articles; 7,713 ADS entries);
  - entries for which first author only used initials for all publications available in the dataset (42,448 ADS entries)
  - entries for which the gender of the first name of first author could not be determined (2,260 ADS entries)

Table 1A
Example of the data available (first 8 columns)

Bibcode	First Author[1]	First name	Gender	first publication year[2]	# citations	# references	# authors
1978ApJ...222..745C	Condon, J. J.	James	male	1973	19	22	2
1988ApJ...333..611W	Wilson, Christine D.	Christine	female	-99	18	14	5
1990MNRAS.246..565A	Aspin, C.	Colin	male	1981	19	26	4
1990Natur.345...49T	Torbett, Michael V.	Michael	male	1980	48	11	2
1992ApJ...392..760B	Burrows, Christopher J.	Christopher	male	1991	37	7	3
1993A&A...277..677M	Meier, R.	Roland	male	1993	97	77	4
1996A&A...309..171S	Shibanov, Y. A.	Yurii	male	1992	42	18	2
1997A&A...324L...5C	Cambresy, L.	Laurent	male	1997	58	12	8
2002A&A...381L..25M	Meynet, G.	Georges	male	1985	82	31	2
2002MNRAS.329L..67B	Ballantyne, D. R.	David	male	2000	31	29	3
2010ApJ...711.1310K	Khatri, Rishi	Rishi	male	2010	3	37	2
2014ApJ...780..111H	Heitmann, Katrin	Katrin	female	2006	63	57	5
...							

[1] Name of the first author as specified in the paper

[2] Year in which the leading author of the paper in question published their first paper

Table 1B
Example of the data available (continued, last 9 columns)

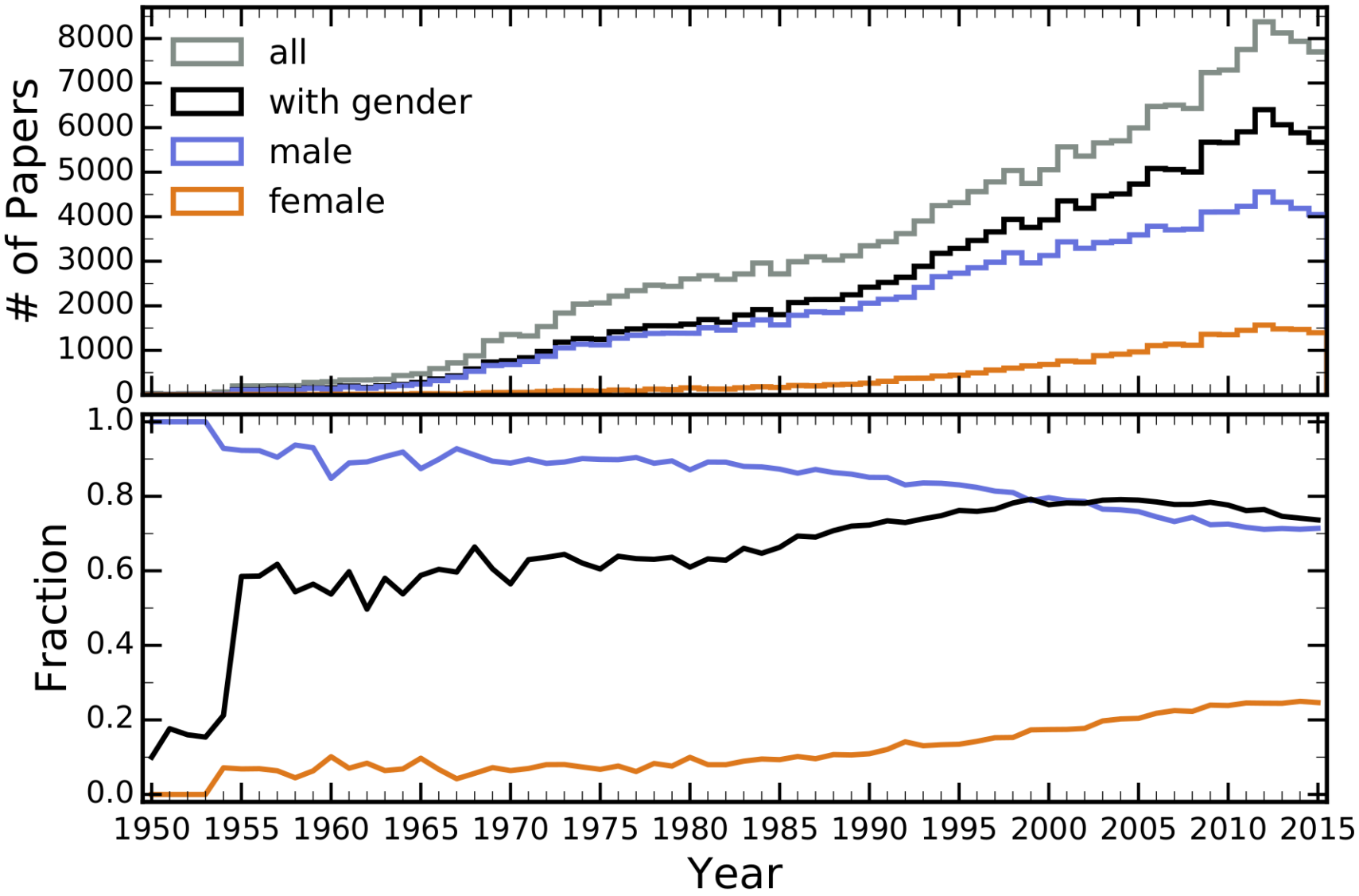
Region	Year	Journal	# field[3]	# floats[4,5]	# equations	# math inline	# words	# Bibcode of first publication
NAMERICA	1978	APJ	3	-99	-99	-99	-99	1973ApJ...183.1075C
NAMERICA	1988	APJ	4	-99	-99	-99	-99	-99
OTHER	1990	MNRAS	4	-99	-99	-99	-99	1981MNRAS.194..283A
NAMERICA	1990	NAT	1	-99	-99	-99	-99	1980Natur.286..237T
NAMERICA	1992	APJ	6	-99	-99	-99	-99	1991ApJ...369L..21B
OTHER	1993	AA	4	-99	-99	-99	-99	1993A&A...277..677M
OTHER	1996	AA	2	-99	-99	-99	-99	1992A&A...266..313S
OTHER	1997	AA	4	-99	-99	-99	-99	1997A&A...324L...5C
EUROPE	2002	AA	2	-99	-99	-99	-99	1985A&A...150..163M
EUROPE	2002	MNRAS	5	-99	-99	-99	-99	2000ApJ...536..773B
NAMERICA	2010	APJ	3	8	10	160	2709	2010ApJ...711.1310K
NAMERICA	2014	APJ	3	17	14	502	11456	2006ApJ...642L..85H
...								

[3] 1="Earth and Planetary Astrophysics", 2="Solar and Stellar Astrophysics", 3="Astrophysics of galaxies", 4="Cosmology and Extragalactic Astrophysics", 5="High Energy Astrophysical Phenomena", 6="Instrumentation and Method for Astrophysics"

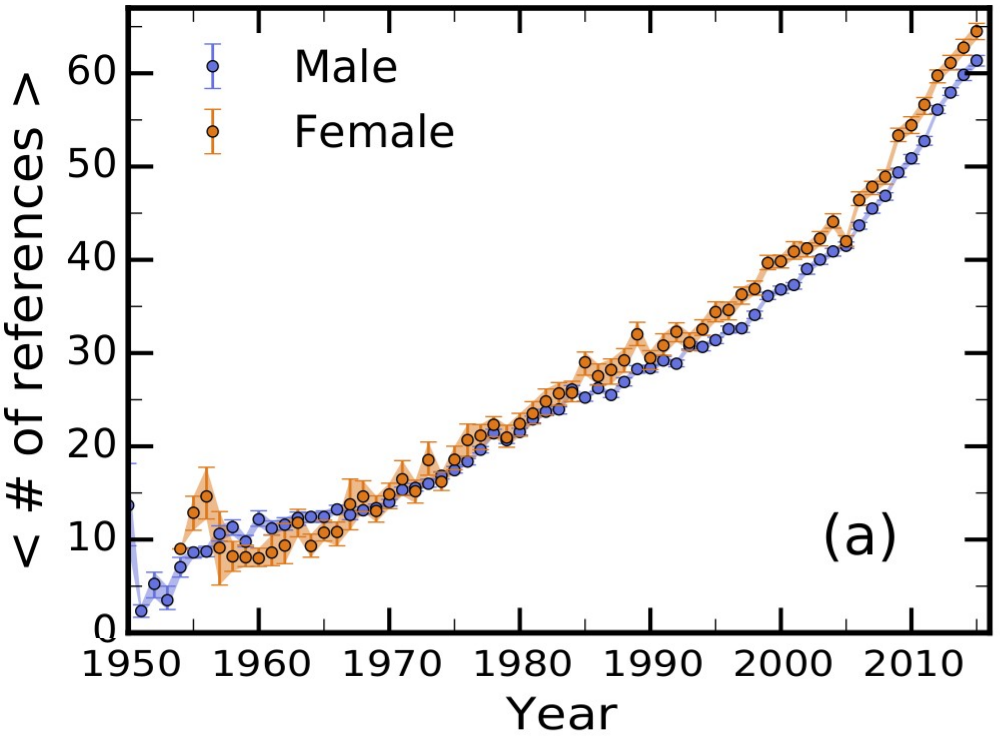
[4] floats include both figures and tables

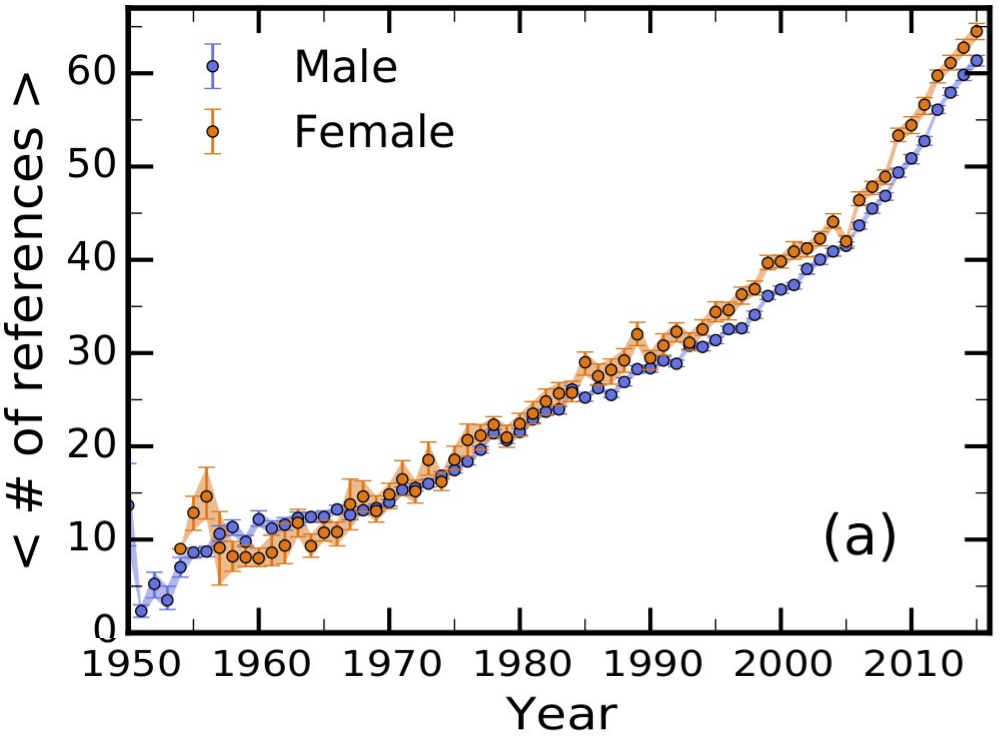
[5] with -99 we denote that there is no data available for this quantity

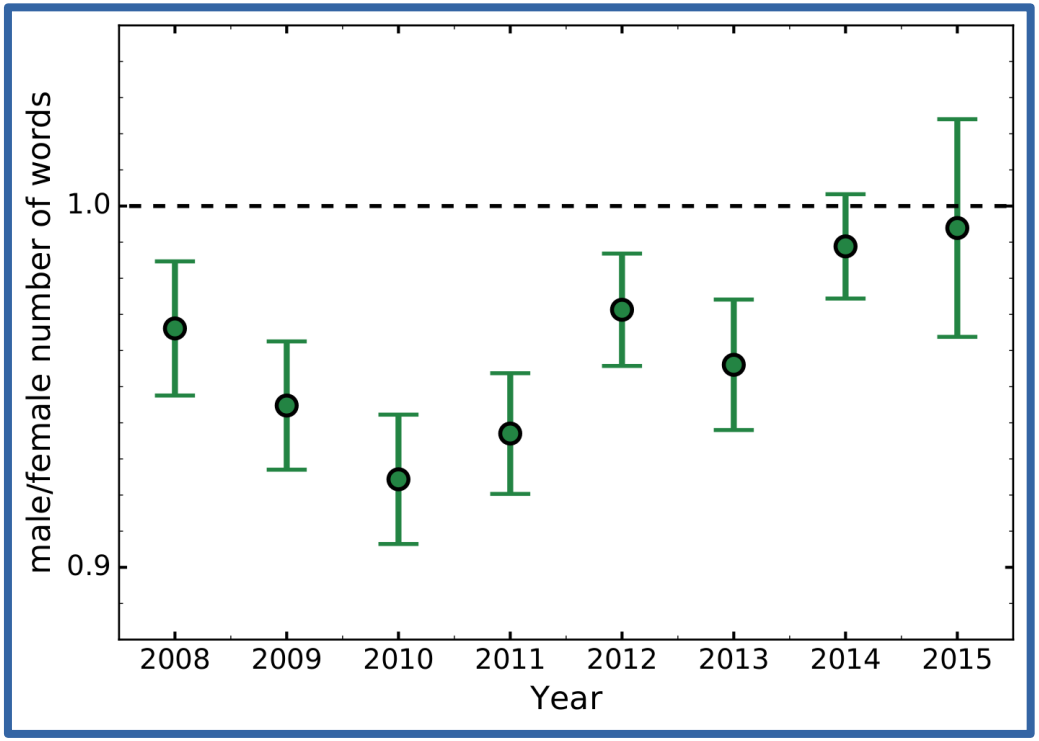
Properties of the sample

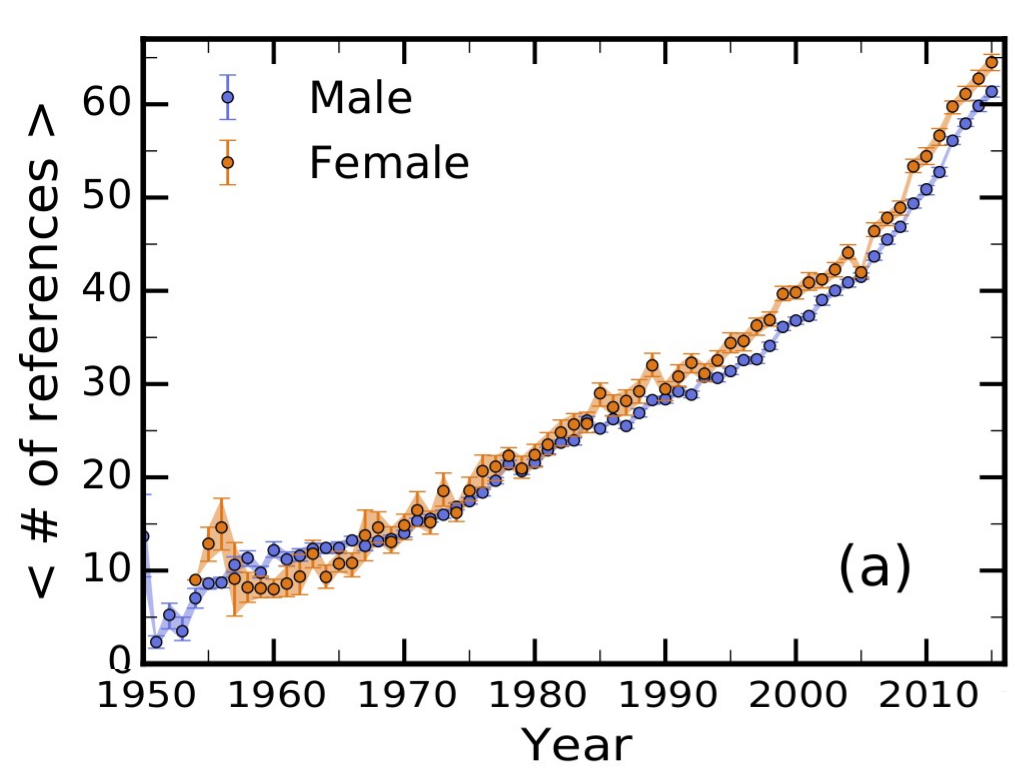


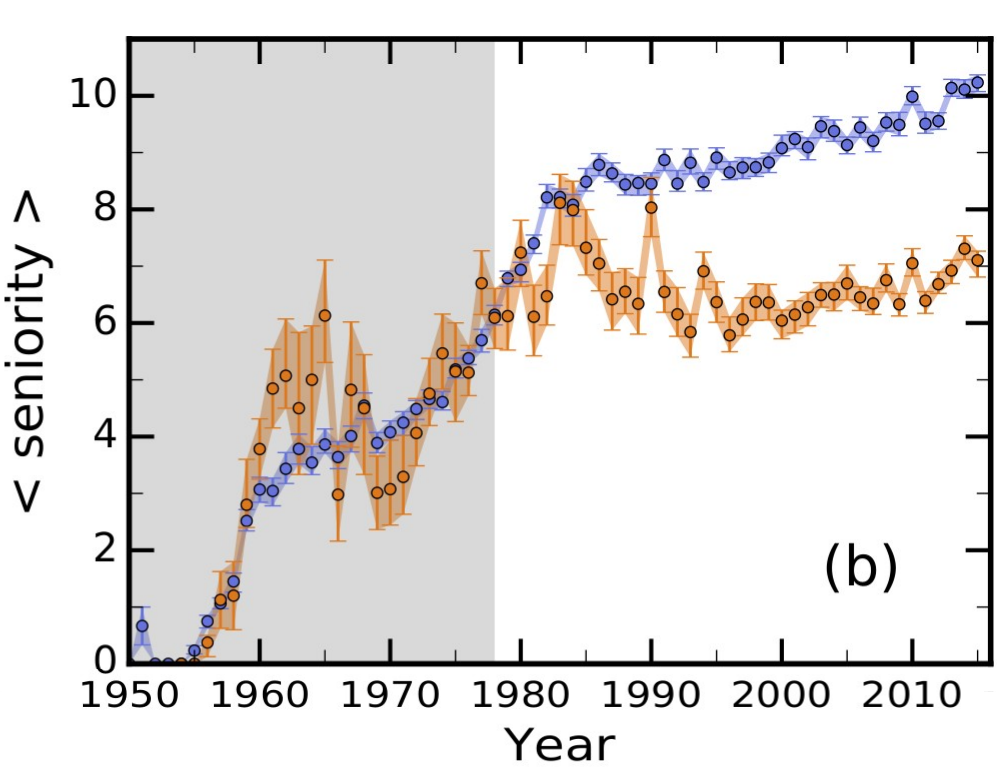
- Slow increase of the fraction of the papers written by women

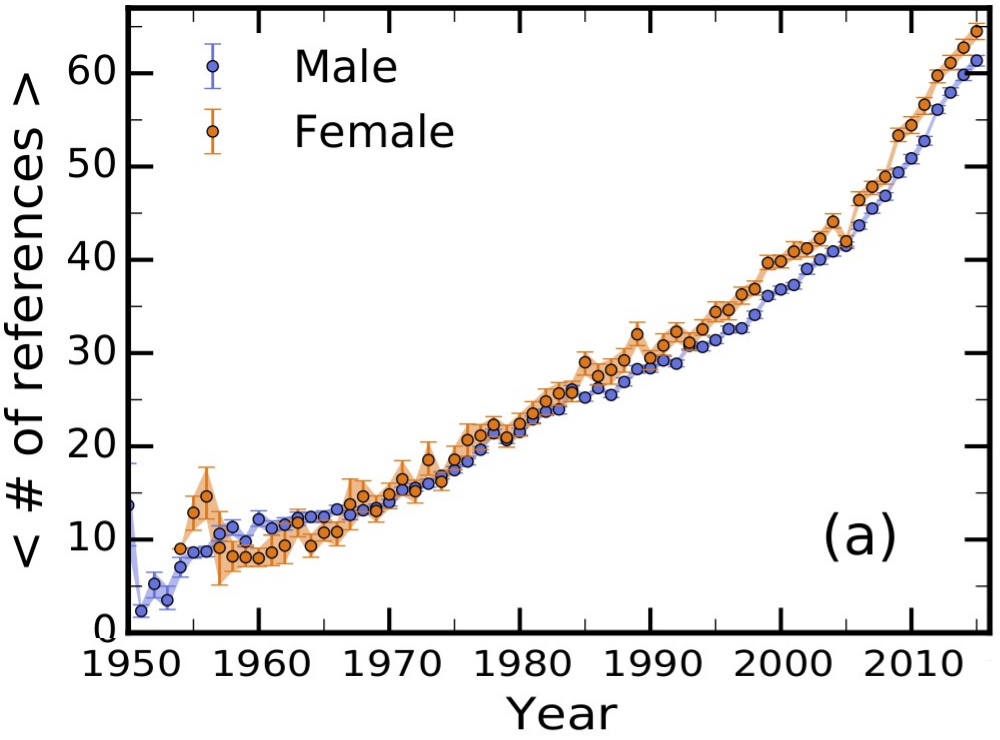

Male
Female
(a)
< # of references >
Year
0
10
20
30
40
50
60
1950
1960
1970
1980
1990
2000
2010

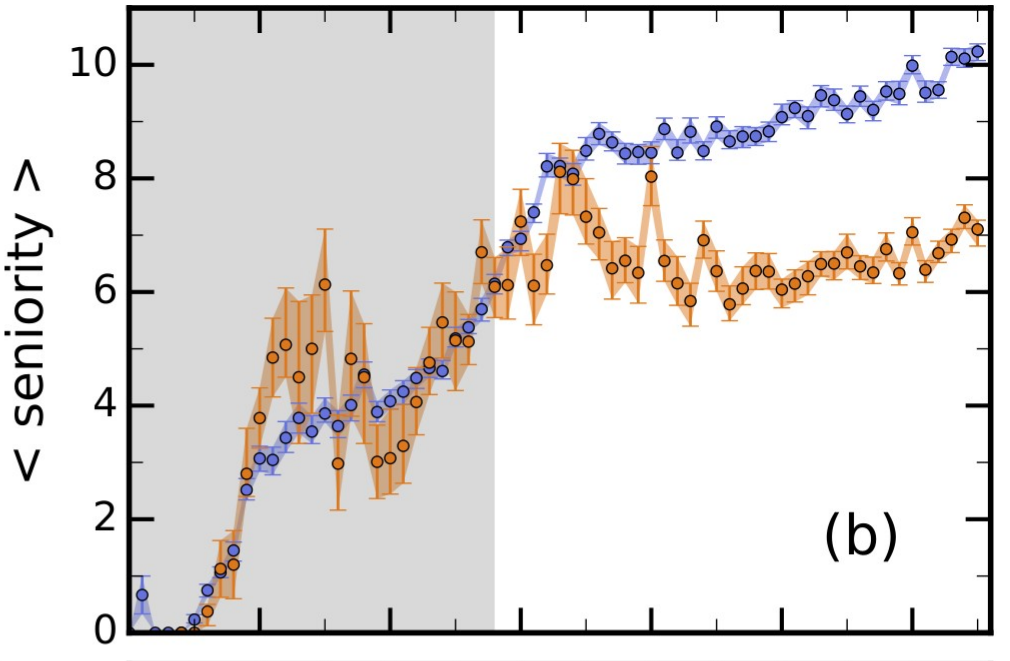

Male
Female
(a)
< # of references >
Year
0
10
20
30
40
50
60
1950
1960
1970
1980
1990
2000
2010

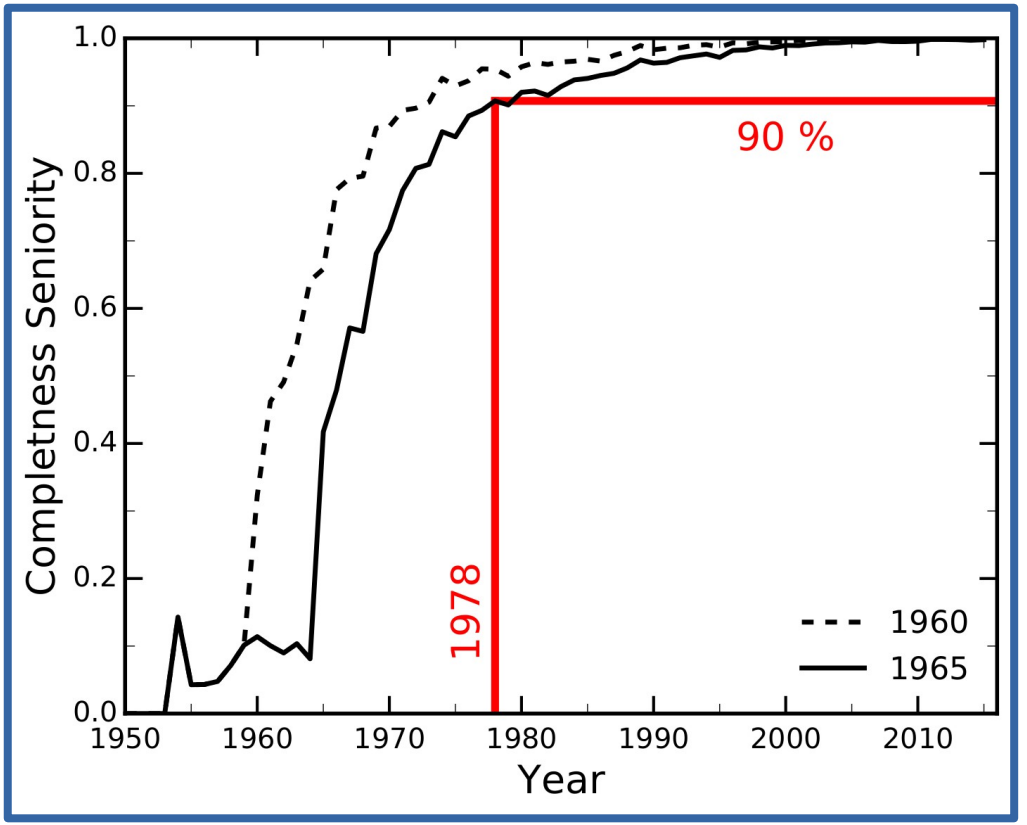

male/female number of words
1.0
0.9
2008
2009
2010
2011
2012
2013
2014
2015
Year

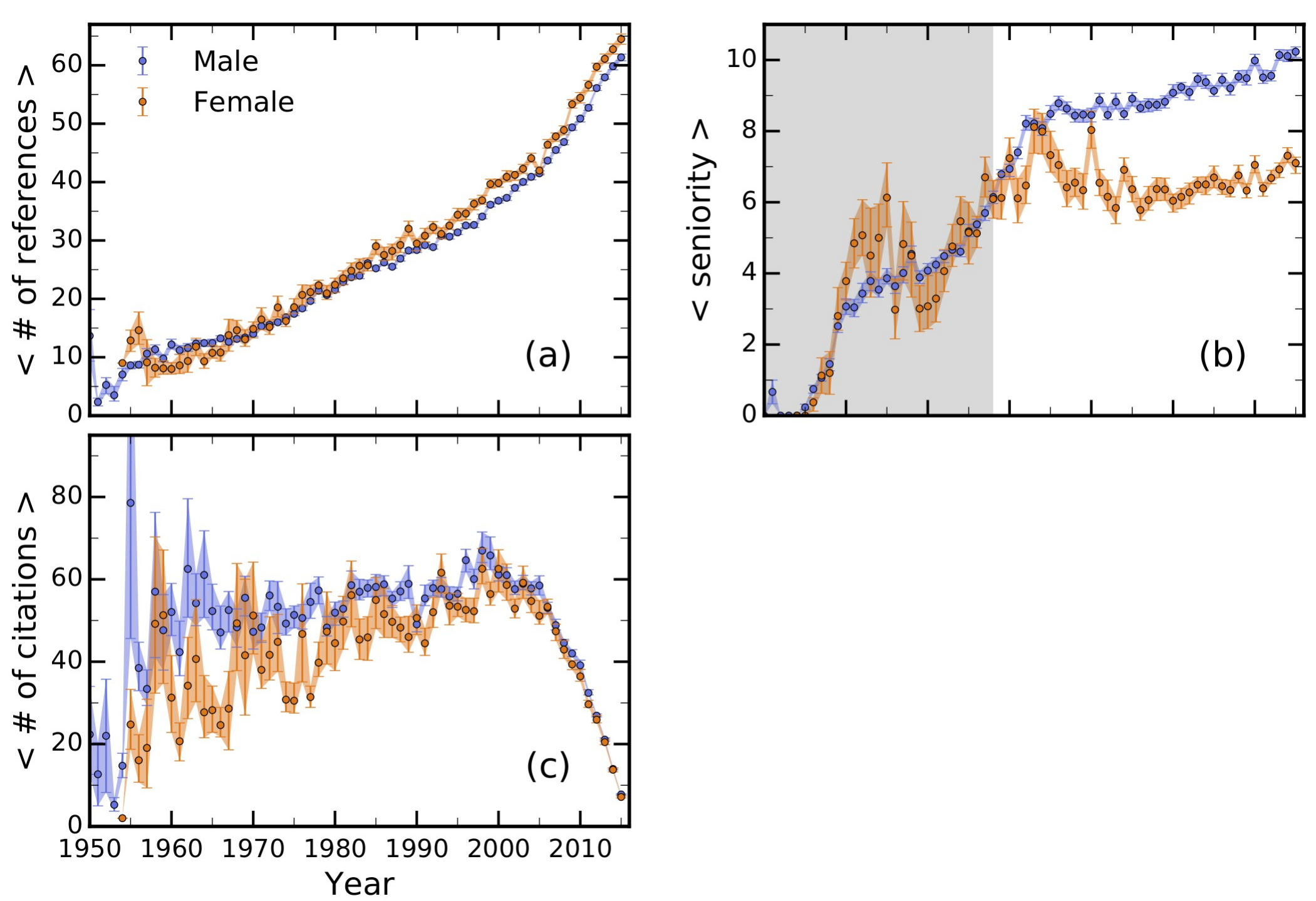

Male
Female
< # of references >
Year
(a)

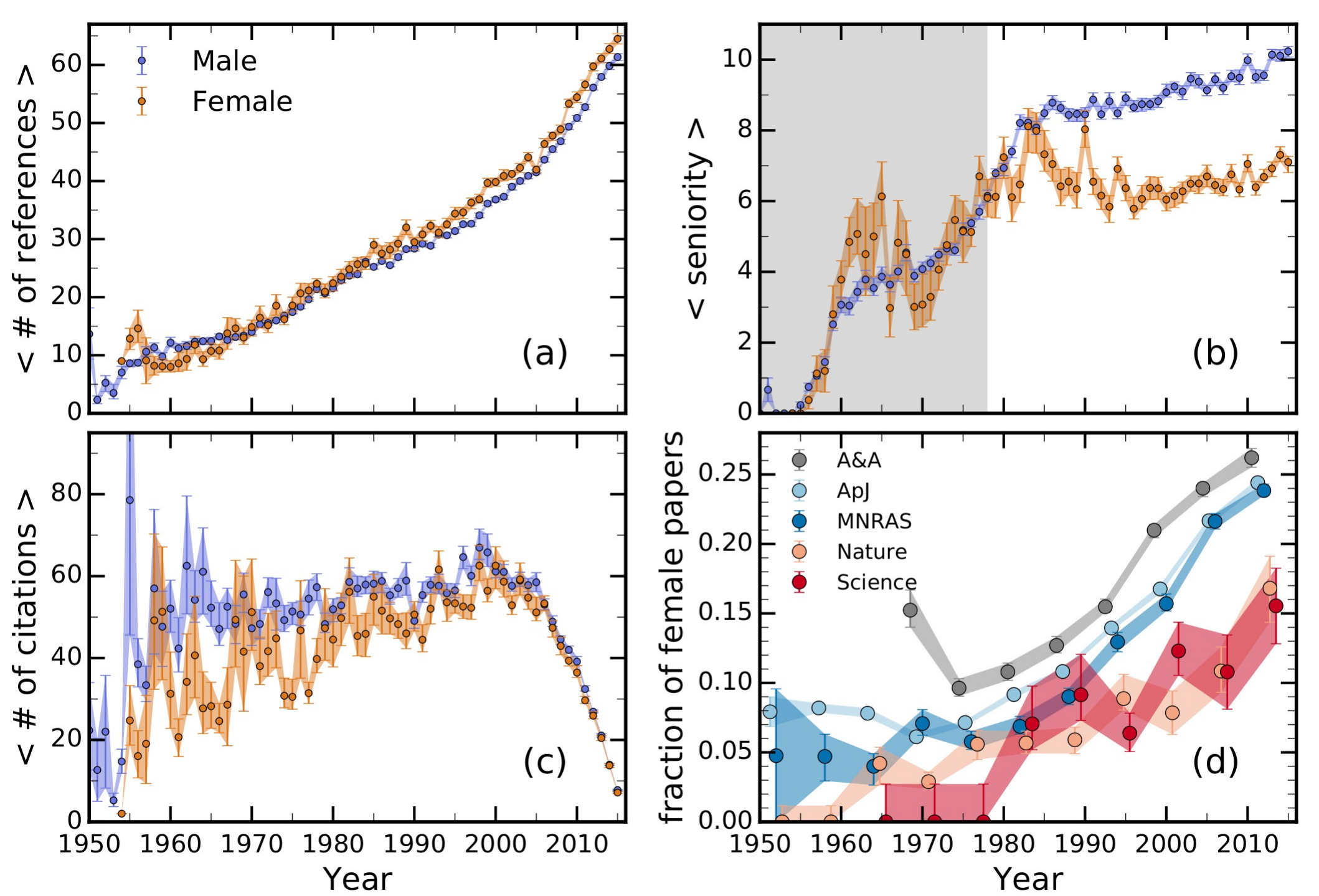

< seniority >
Year
(b)


Male
Female
< # of references >
Year
(a)


< seniority >
(b)

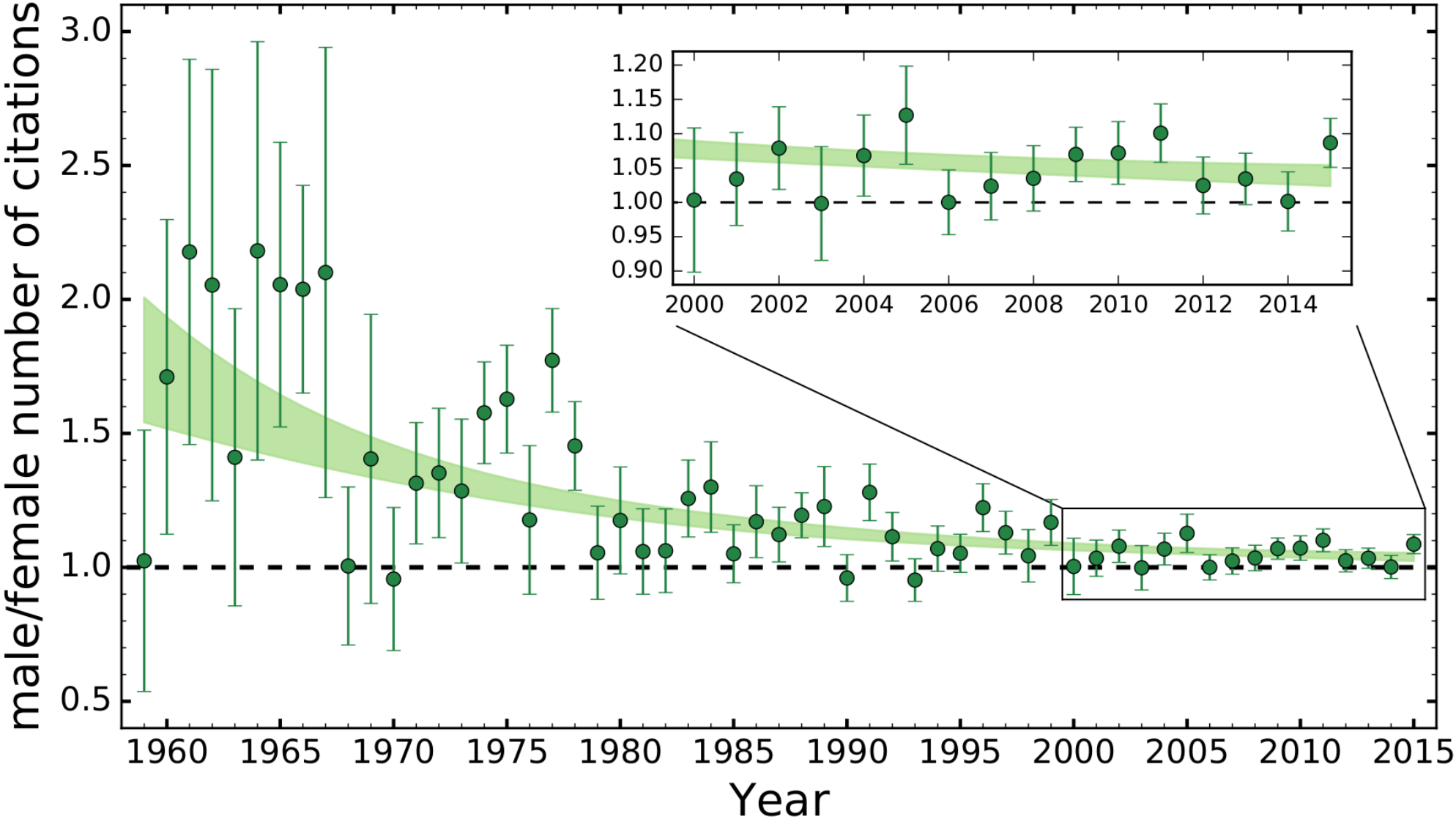

Completness Seniority
Year
90 %
1978
1960
1965


Male
Female
< # of references >
(a)
< seniority >
(b)
< # of citations >
(c)
Year
0
10
20
30
40
50
60
2
4
6
8
80
1950
1960
1970
1980
1990
2000
2010


Male
Female
< # of references >
(a)
< seniority >
(b)
< # of citations >
(c)
Year
A&A
ApJ
MNRAS
Nature
Science
fraction of female papers
(d)
Year

Overview

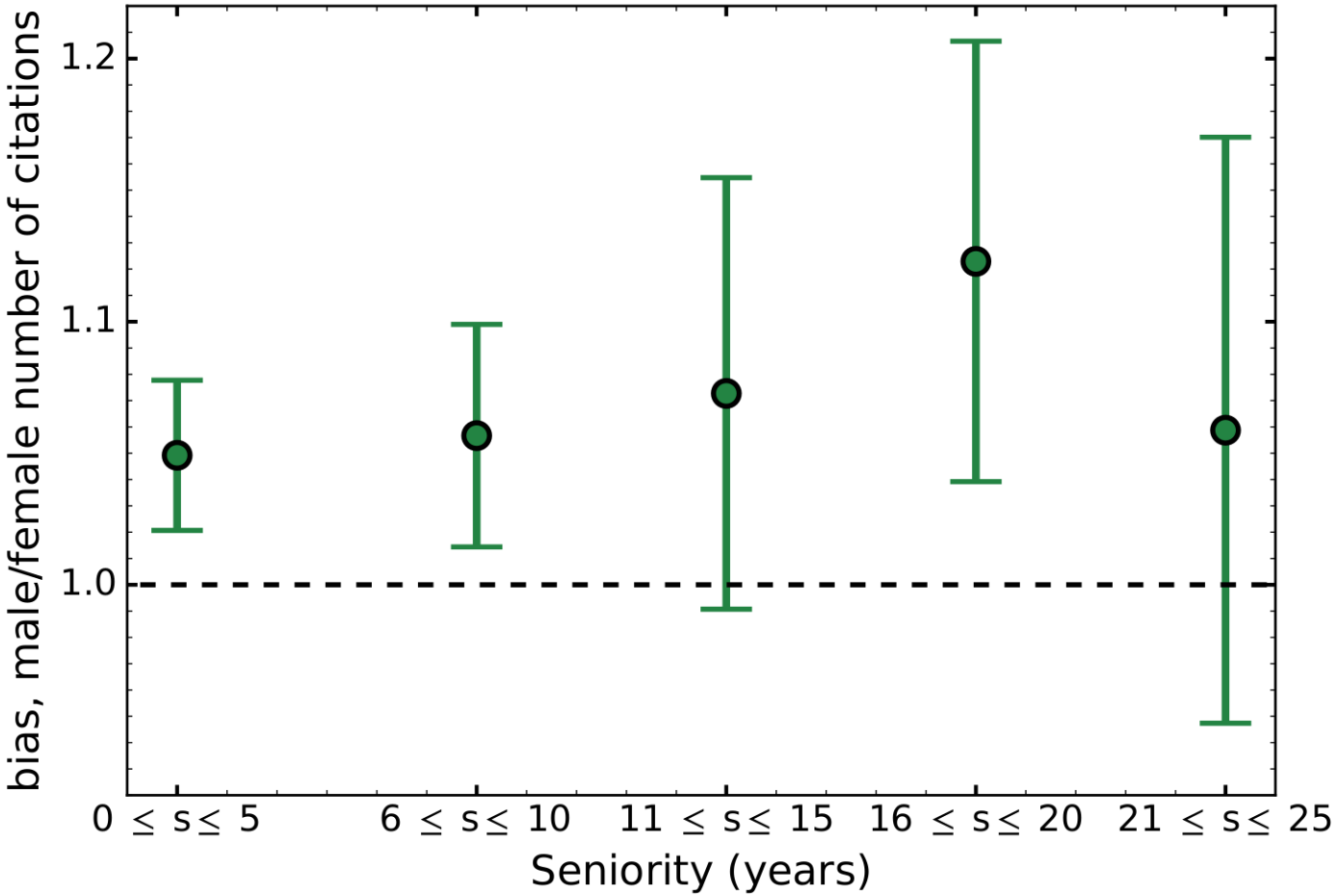
- Introduction
  - Gender difference in science
  - Gender difference in astronomy
- Method
  - Data gathering
  - Sample discussion
- **Results**
  - Gender difference in citation counts
  - Gender bias
  - Self citation and productivity
  - Discussion



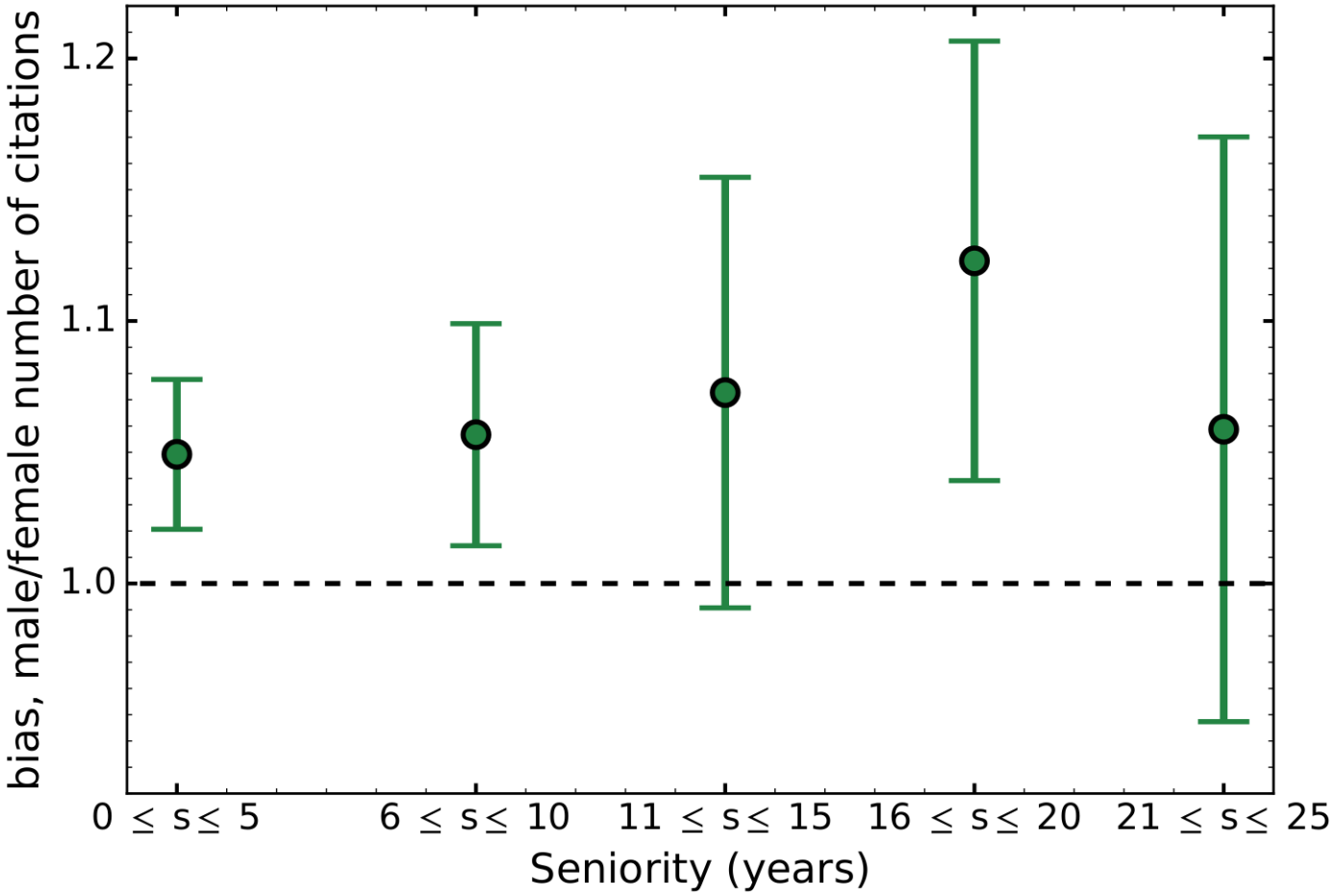
Gender difference: ratio of mean number of citation for papers written by men over mean number of citations for papers written by women

Constant fit to data since 1985: Men receive ~6% more citations

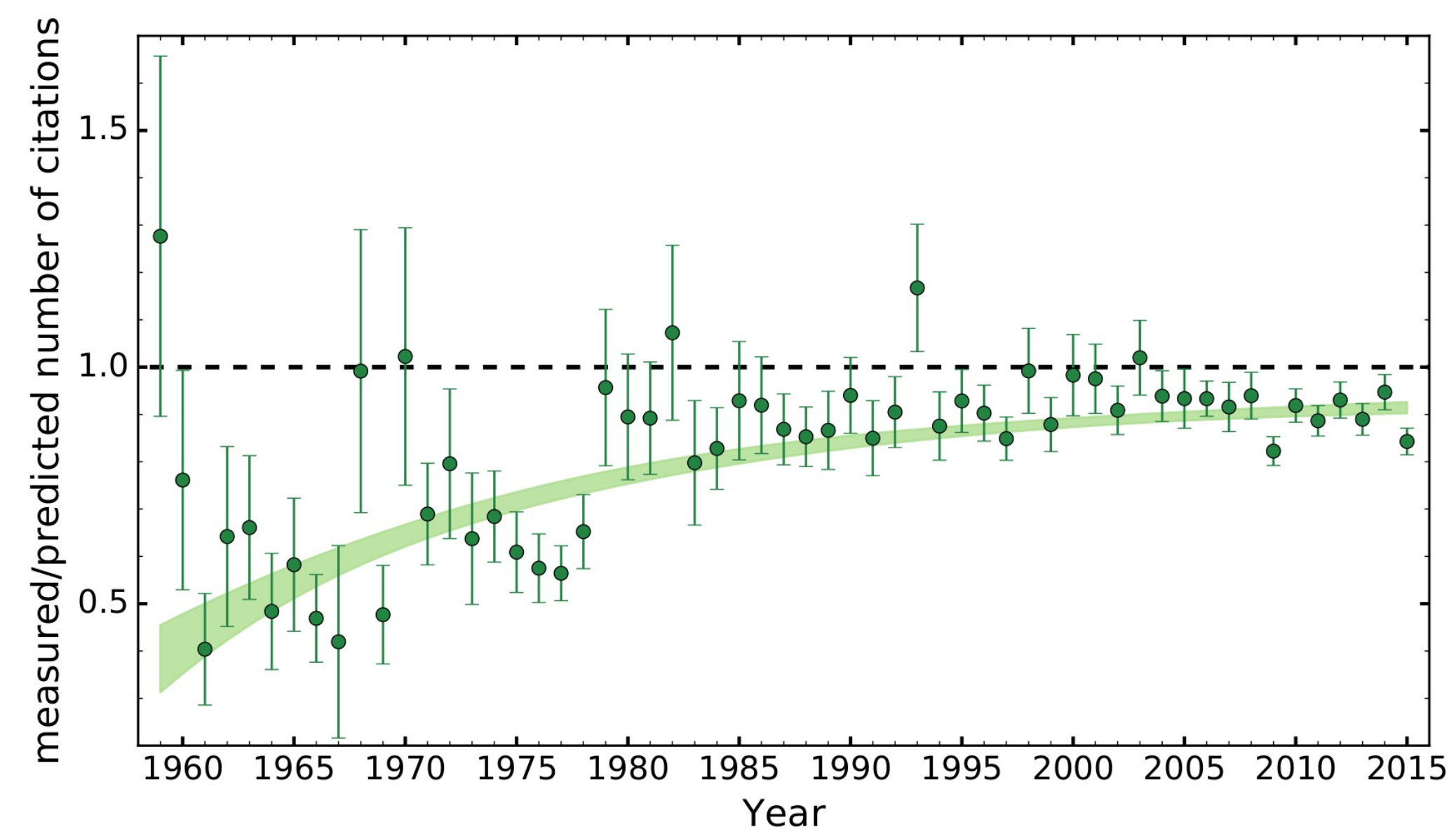
- How to control for difference in the properties of the sample?
  - Match the samples… match all of the parameters?



- How to control for difference in the properties of the sample?
  - Match the samples… match all of the parameters?

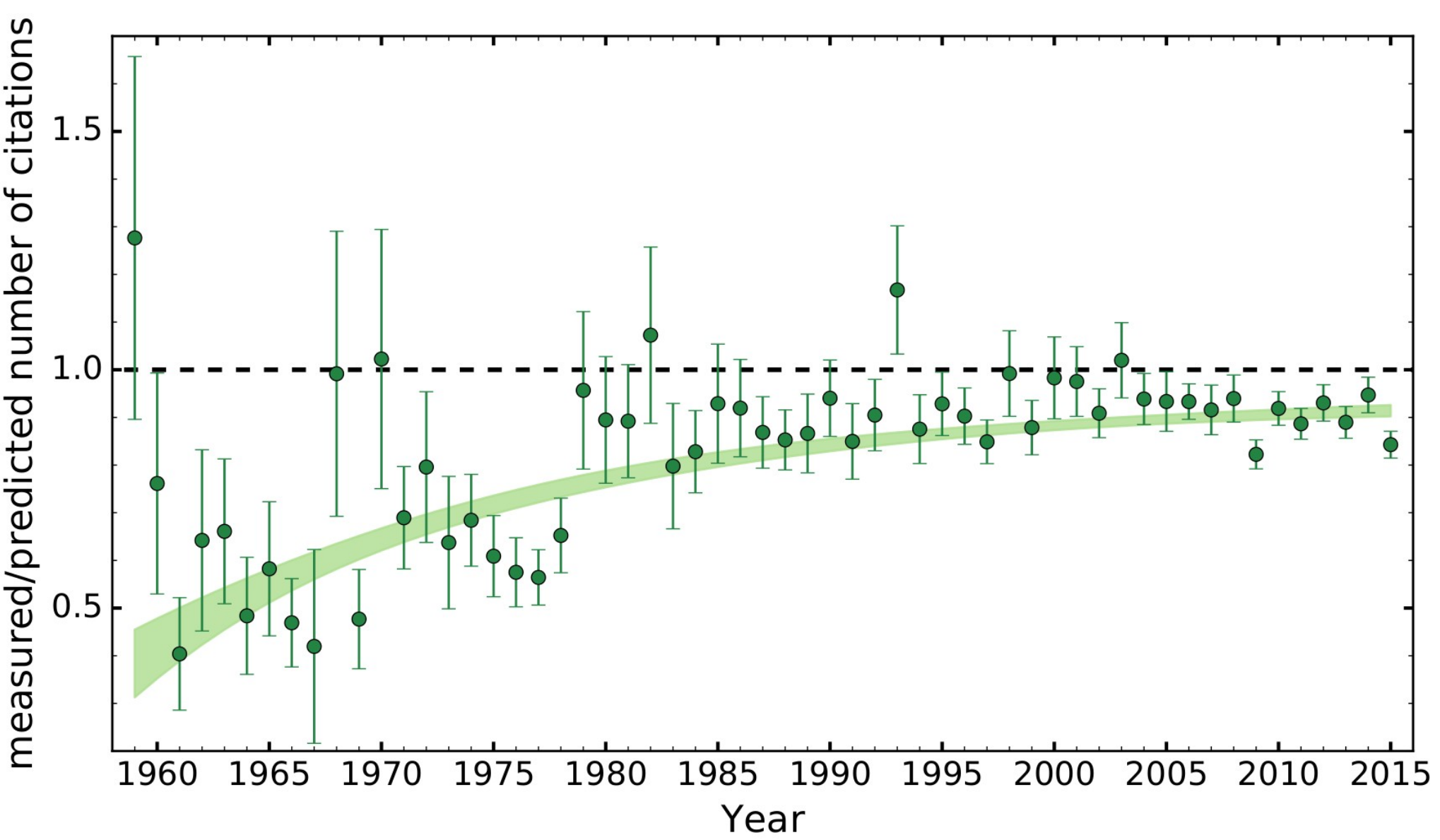


Alternative idea: Train random forest algorithm on the sample of papers written by men and use it on the sample of papers written by women

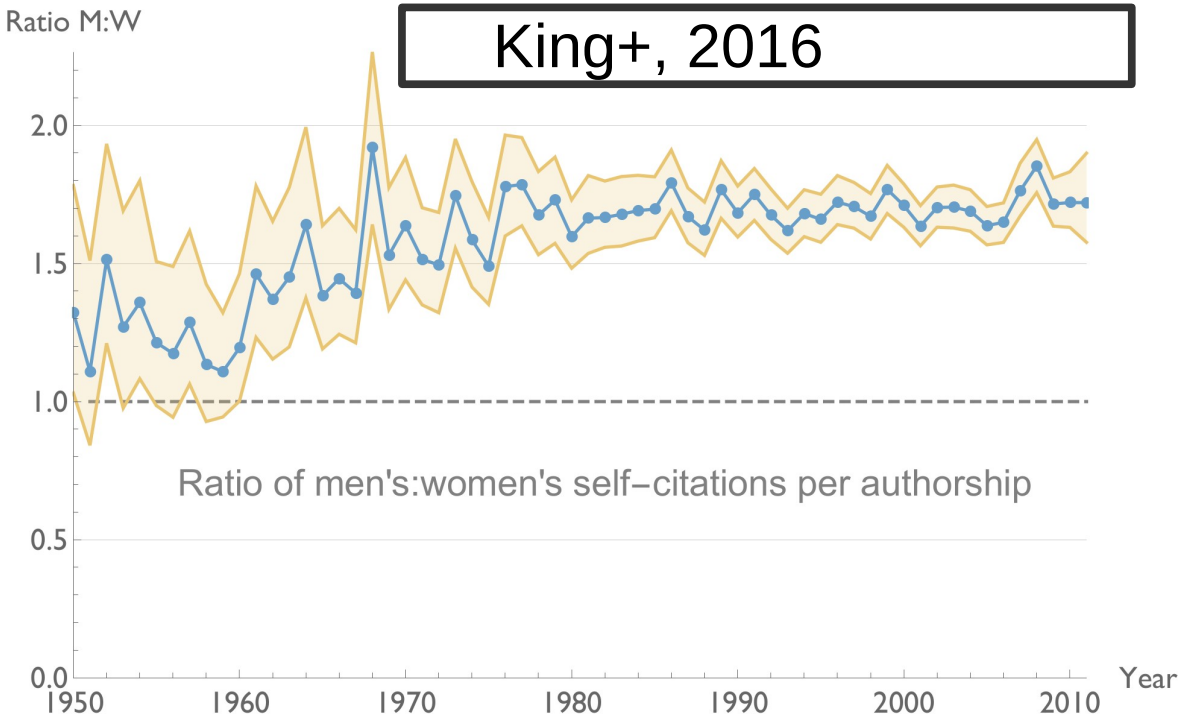


- Gender bias: measured over predicted number of citations for papers authored by women

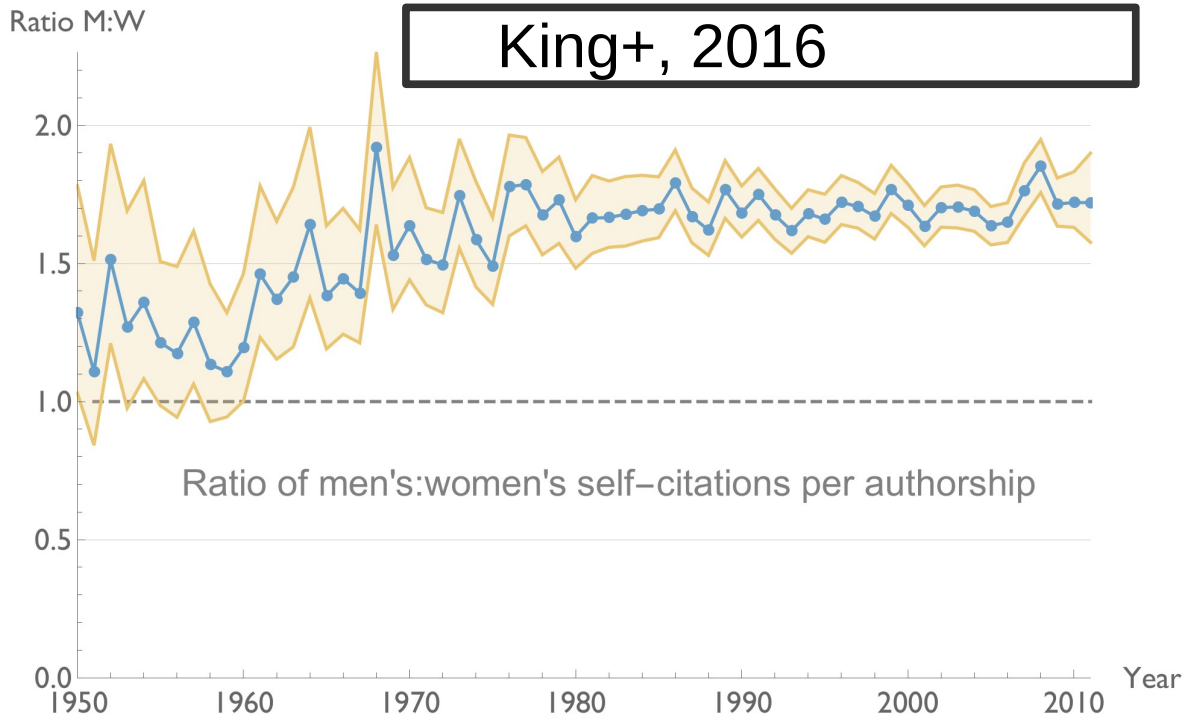
- Constant fit to data since 1985: Women receive 10.4±0.9% less citations

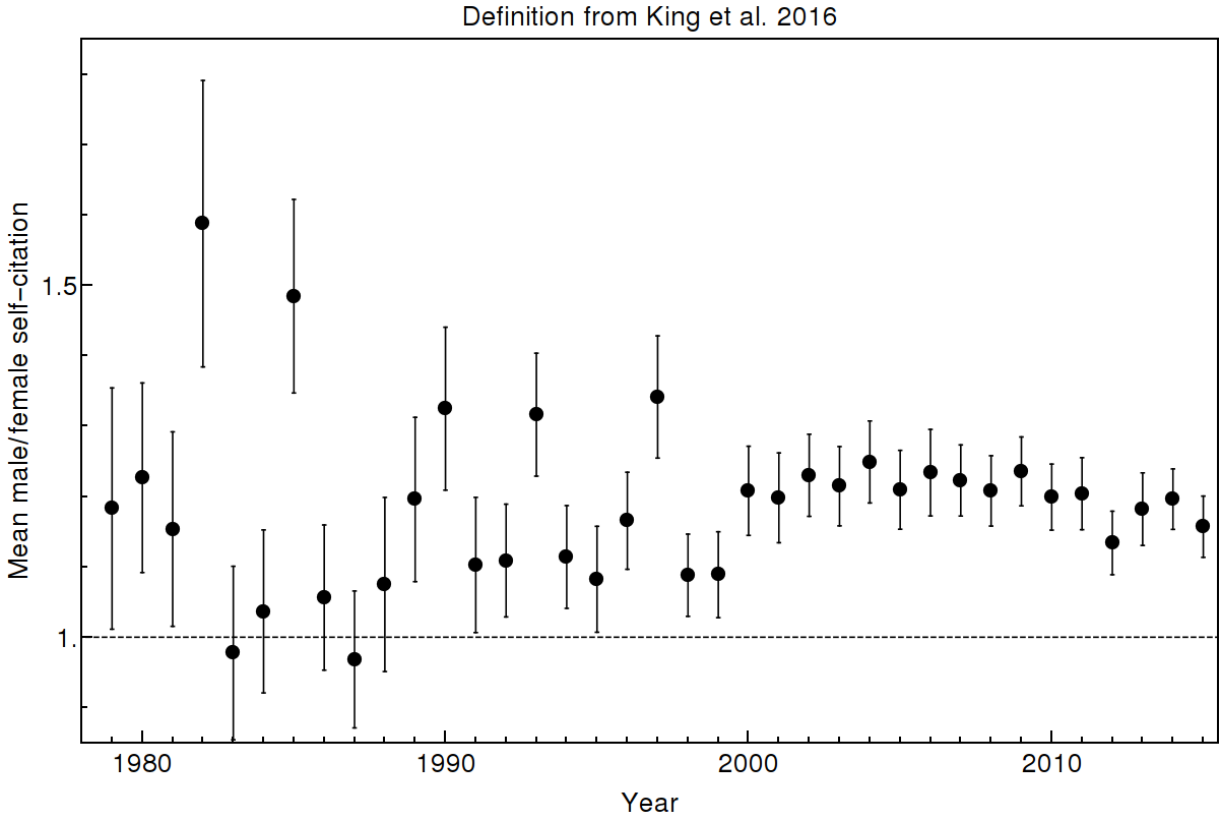


- Bias~10%, difference~6%, we expect that if there was no bias men should receive 4% fewer citations in the sample (also seen in the dedicated analysis)
- Most important parameters (Gini importance): 1. number of references, 2. year of publication, 3. journal

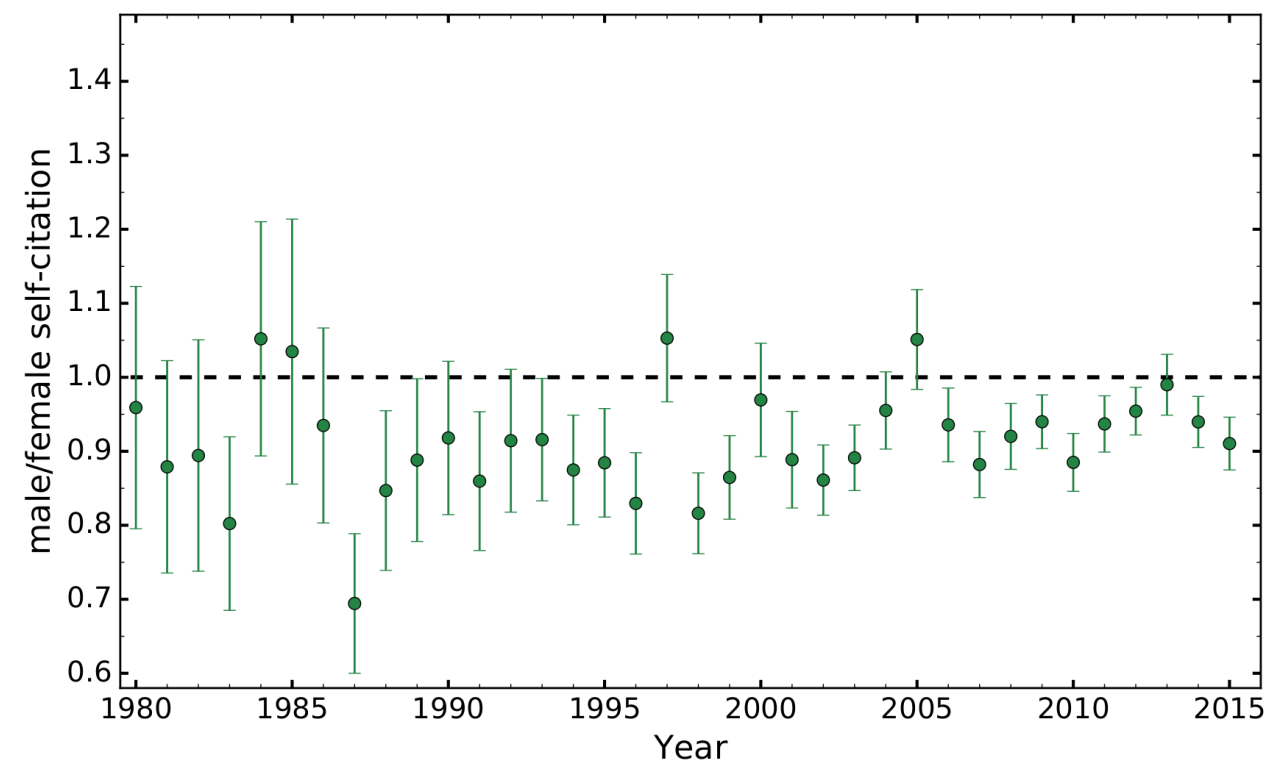


- Men self-cite 70% more?
- How to define self-citations?
- King definition:
  (Number of self citations)/
  (Number of authorships)

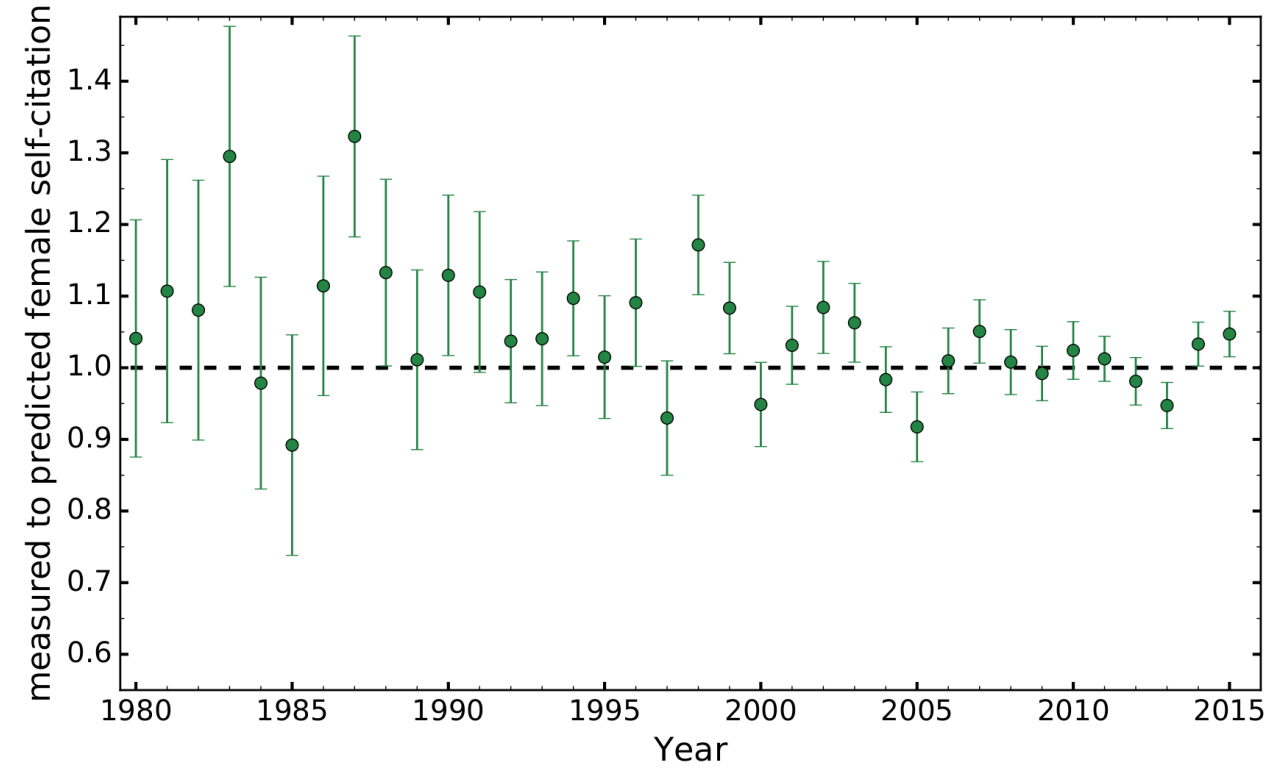




- Men self-cite 70% more?
- How to define self-citations?
- King definition:
  (Number of self citations)/
  (Number of authorships)

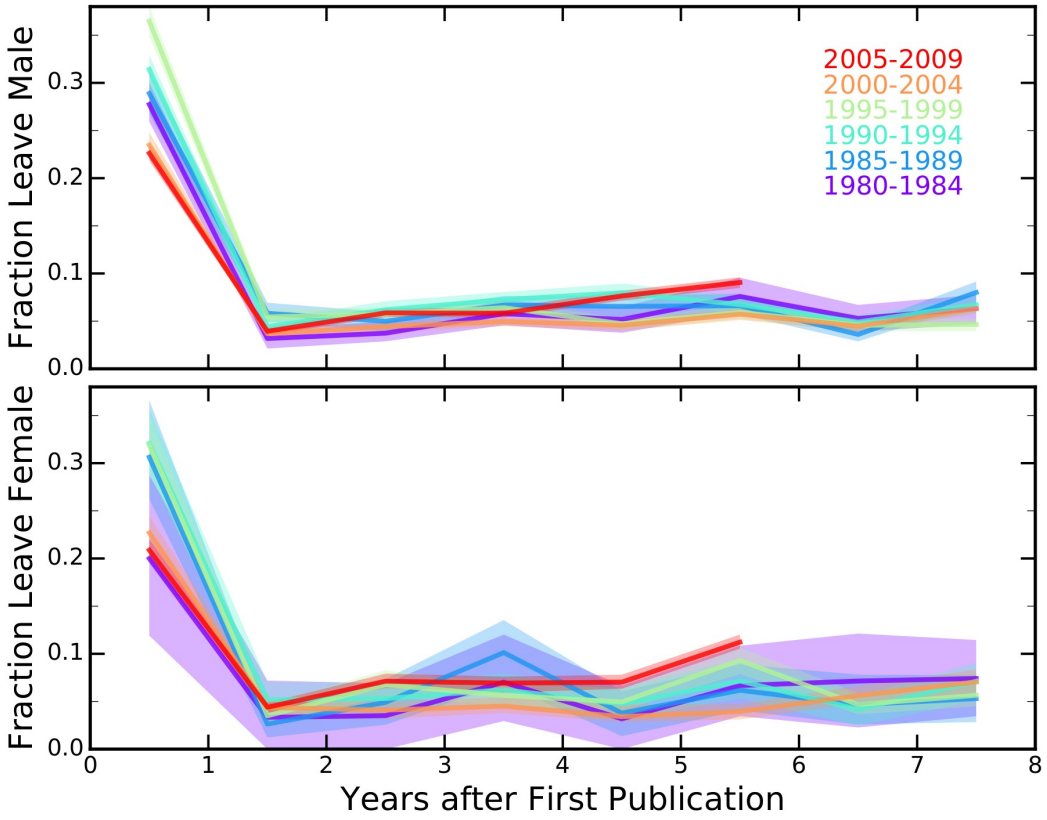


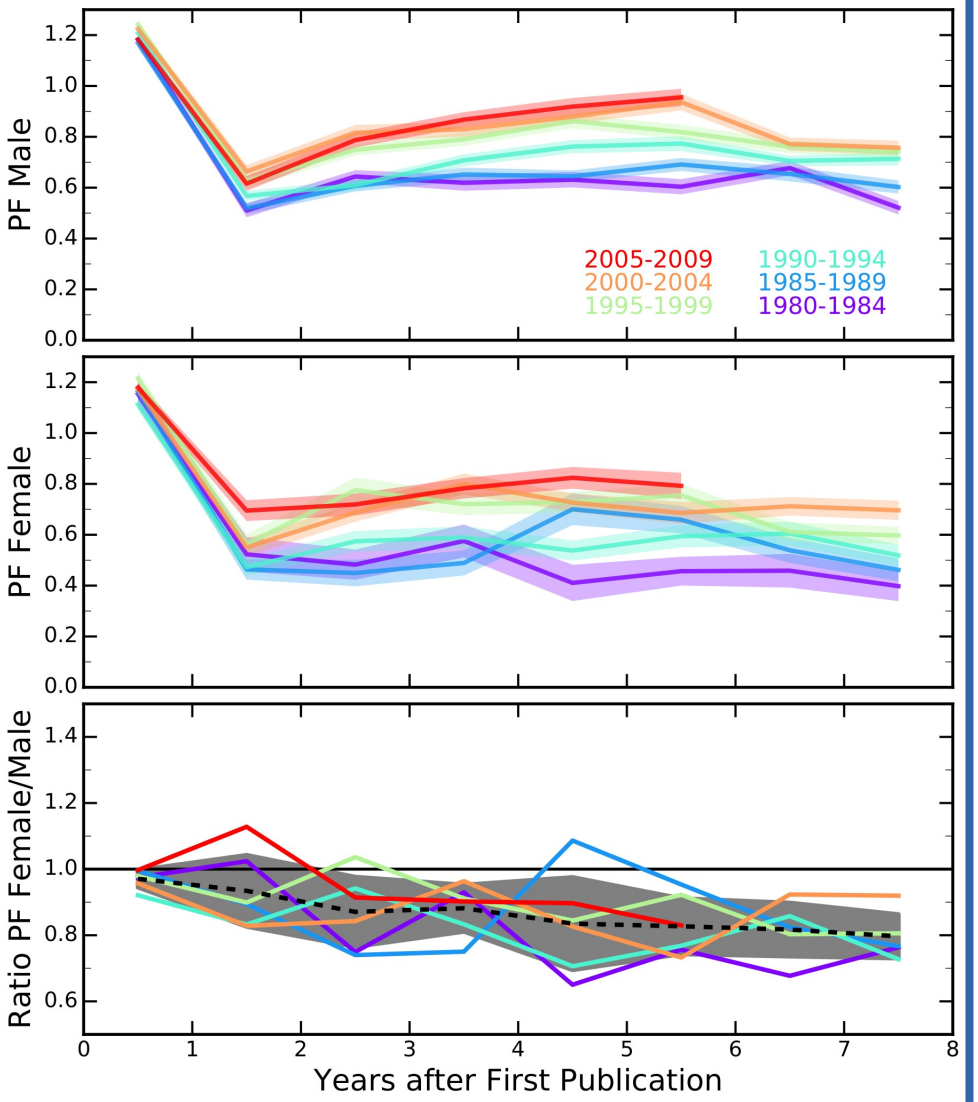
- Men self-cite 70% more?
- How to define self-citations?
- King definition:
(Number of self citations)/
(Number of authorships)



- We use as a measure self-citation of the last previous paper
- No difference is detected after controlled for parameters of the papers

- Do women leave astronomy more often than men?
- We find no difference in the fraction of authors who have left the field





- Women publish less than men in the sample

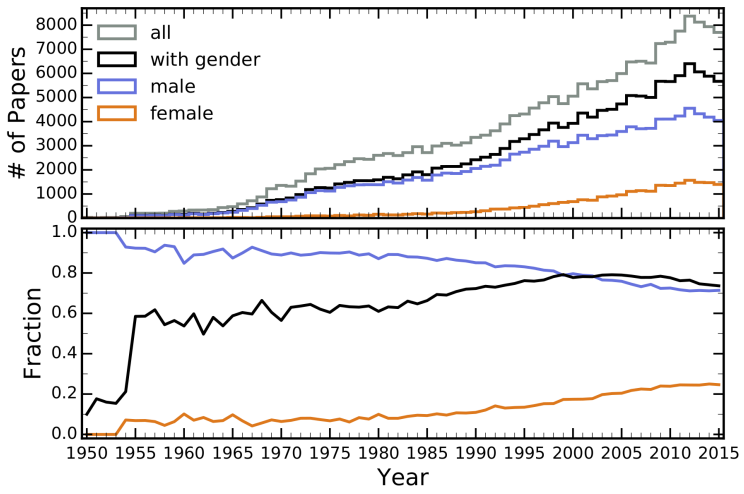
Discussion

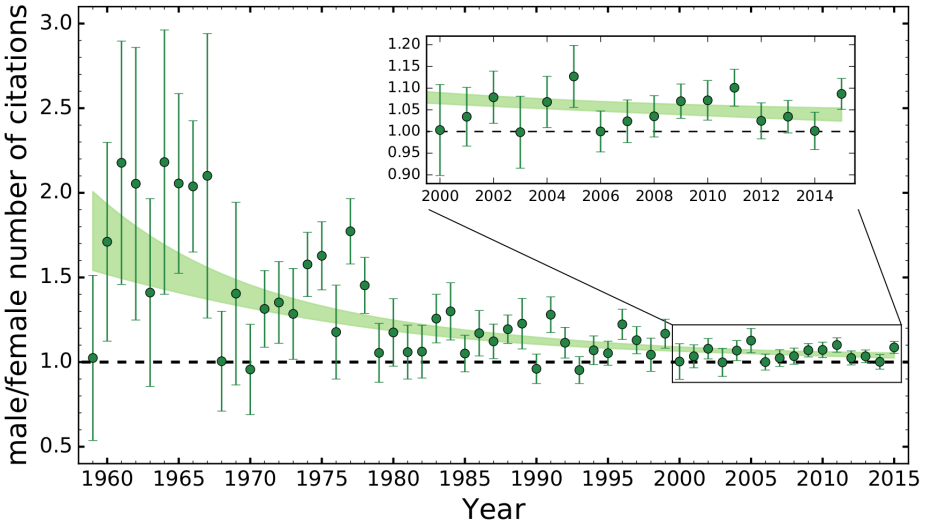
- Caveats of analysis
  - Is there bias in gender recognition?
    - Are we equally likely to recognize both men and women from their names?
  - Effect of changing surnames?
  - Additional parameters not considered?

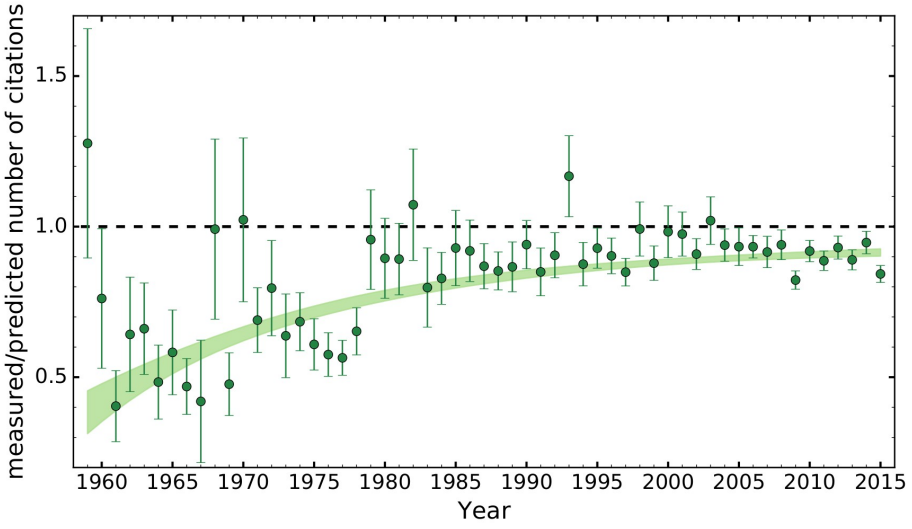
- Future?
  - “better” analysis, matching exactly every citation
  - “expensive” & time constraints
  - https://github.com/nevencaplar/Gender_Bias

Summary

- Analysis of over 200,000 publications from astronomy
- Gender difference of 6%
- But samples differ in their properties
  - We find that women receive 10.4±0.9% less citations than expected given the parameters of their papers
- No difference in self-citation







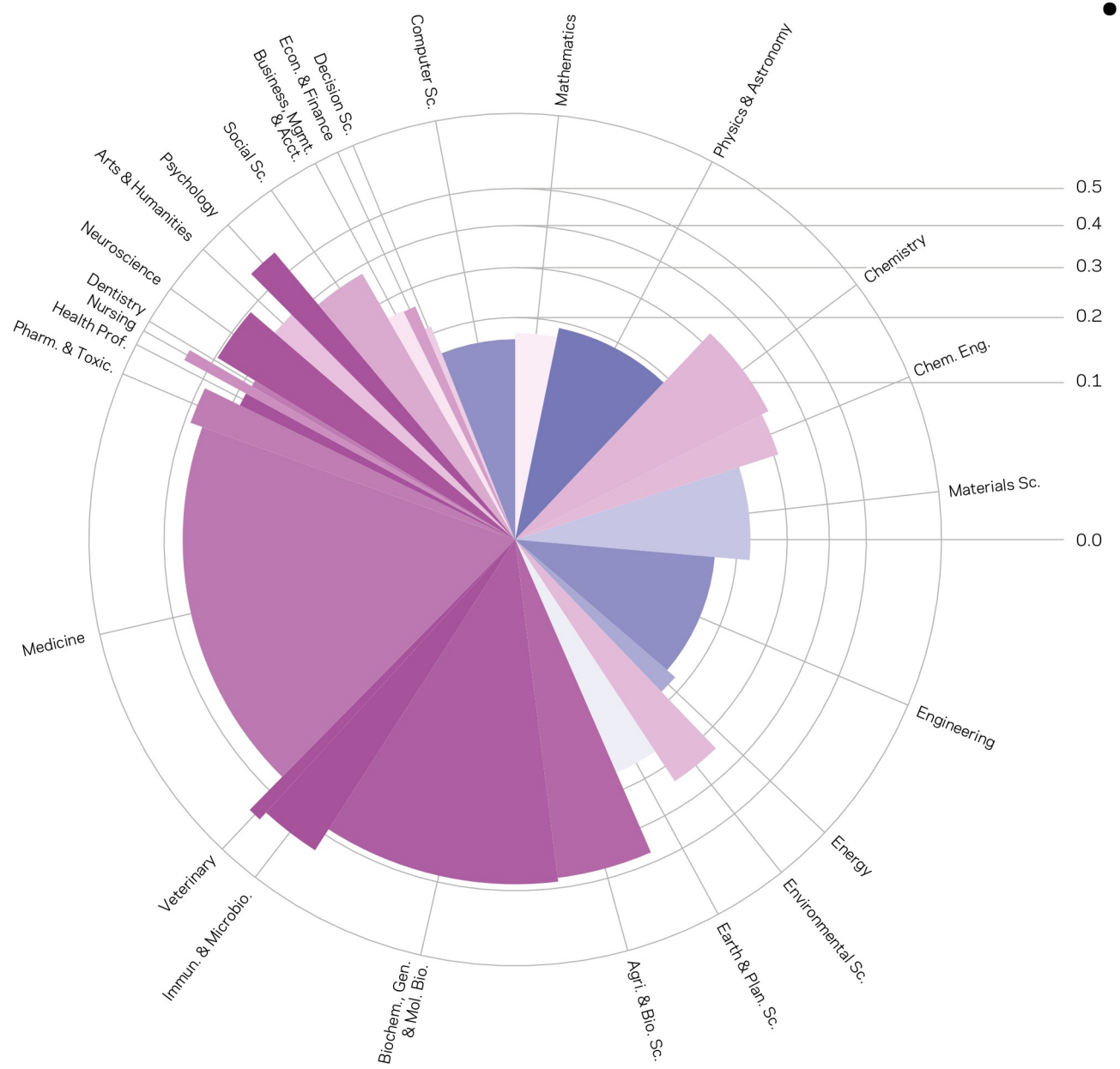


Figure 4 *— The number of researchers (denoted by the size of pie slices), the share of female researchers out of all researchers who published in each subject area (denoted by the length of pie slices), and the ratio between the productivity of female and male researchers (denoted by the colour of pie slices; the ratio between the productivity of female and male researchers increases when the colour changes from pink to blue); per subject; for Germany; 2010-2014.*

- In Germany
  - Women around 10% of researchers in “Physics and astronomy”
  - Women are more “productive” than men

From report “Mapping Gender in the German Research Arena”