

# Stand der Forschung zum Gender-Gap/Bias in der Wikipedia



Tilman Bayer (Benutzer:HaeB)  
[FemNetz 2022](#), 29.1.2022

# Vorschau

- \* Intro, Disclaimer
- \* Gender Gap: Lesende
- \* Gender Gap: Beitragende
- \* Was ist eigentlich Bias, und wie kann man ihn nachweisen?
- \* Gender Gap: Inhalte
- \* Interventionen
- \* Ausblick: Interessante offene Fragen
- \* Q&A

# Intro, Disclaimer

- \* Fokus auf **empirische Forschung** (aber Theorie ist ebenfalls wichtig)
- \* Fokus auf **quantitativen Resultaten** (aber qualitative sind ebenfalls wichtig)
- \* Nur eine kleine **Auswahl aus 2 Jahrzehnten** Forschung
- \* **Meine Perspektive:** Wikipedianer seit 2003, 2011-2019 bei der Foundation, Herausgeber des Wikimedia Research Newsletters seit 2011, männlich
- \* Sorry for my Auswandererdeutsch ;-)
- \* Q&A am Ende (Fragen per Chat oder Video)
- \* ...aber bitte gerne unterbrechen bei kurzen Verständnisfragen während des Talks

# Gender Gap: Lesende

\* Johnson et al.: "Global Gender Differences in Wikipedia Readership" (2021)

\* Multilinguale WMF-Umfrage(n) unter 65.031 Wikipedia-Leser/innen

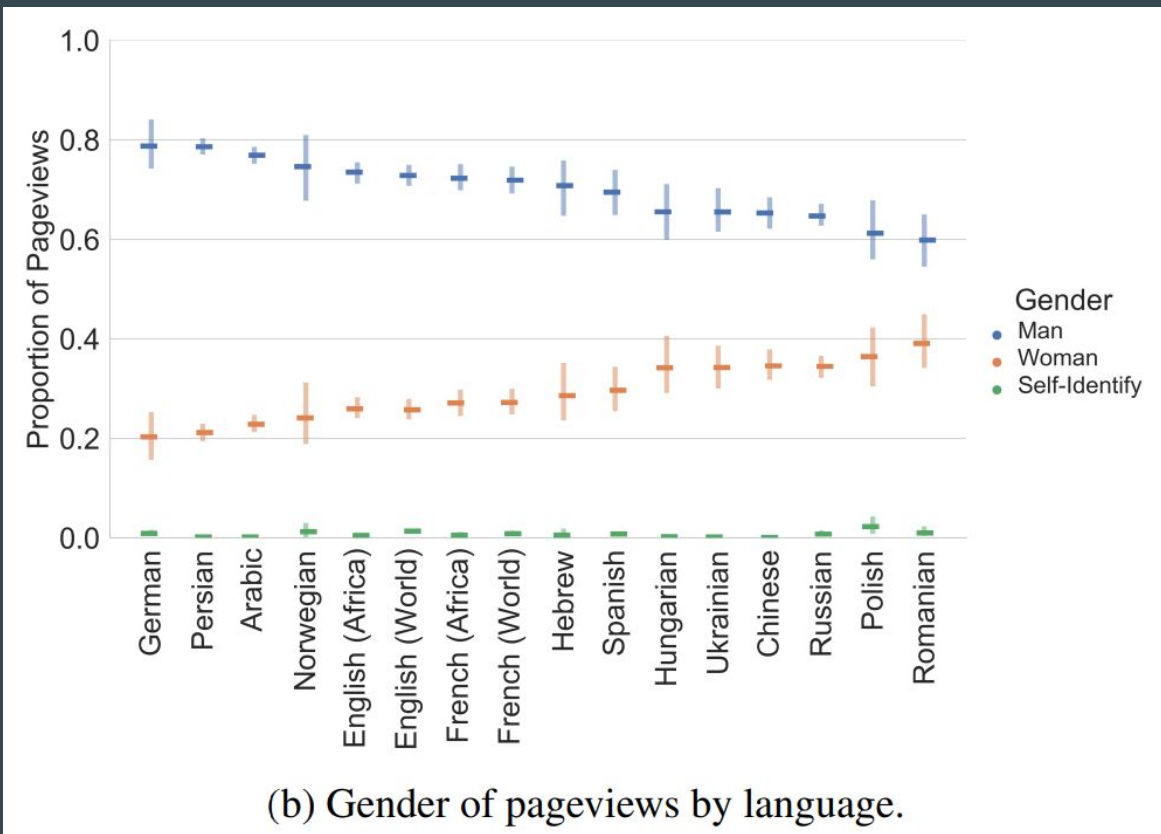
\* Kombiniert mit Webserver-Logs (welche Artikel wurden gelesen)

**“men represent approximately two-thirds of Wikipedia readers on any given day”**

# Gender Gap: Lesende

Deutsche Wikipedia hat das größte Gender Gap in Pageviews: 79% männlich

Rumänische Wikipedia hat das kleinste Gender Gap, mit immer noch 60% männlich



# Gender Gap: Lesende

Frauen lesen weniger Artikel per Wikipedia-Session

Jedoch **keine Gender-Unterschiede** bei Häufigkeit von “**Information Needs**”:

1. Fakten nachschlagen
2. Überblick über ein Thema gewinnen
3. “In-Depth Information” (Artikel ganz lesen etc.)

# Gender Gap: Lesende

Artikelthemen mit relativ **mehr männlicher Leserschaft:**

- Sport
- Spiele/Games
- Mathematik

Artikelthemen mit relativ **mehr weiblicher Leserschaft:**

- “Broadcasting” (Fernsehsendungen etc.)
- Gesundheitsthemen
- “Entertainment”

# Gender Gap: Lesende

“self-focus bias”:

Männliche Leser lesen relativ mehr Artikel über Männer

Weibliche Leser lesen relativ mehr Artikel über Frauen

...**Ausnahme**: Keine Unterschiede auf der deutschsprachigen Wikipedia(!)



# Gender Gap: Lesende

NB: “closing content gaps is not a panacea [kein Allheilmittel]”

Gegenbeispiel Walisische Wikipedia:

- \* Mehrheit der Biografien ist über Frauen
- \* Frauen stellen die Mehrheit der Walisischsprachenden
- \* ... trotzdem sind auch dort männliche Wikipedia-Leser in der Überzahl

# Gender Gap: Lesende

## Mögliche Gründe (größtenteils spekulativ):

1. Frauen nehmen Wikipedia als **weniger nützlich** wahr (basierend auf einigen wenigen Umfragen, hauptsächlich unter Studierenden)
2. **Geringere Selbstwirksamkeitserwartung** (*self-efficacy*) von Frauen
  - d.h. etwa: Vertrauen in die eigene Fähigkeit, die Qualität von “anonymen” Wikipedia-Artikeln auch ohne traditionelle Autoritäten (Brockhaus etc.) einzuschätzen
3. **Content Gaps** (mit Einschränkungen, s.o.)
4. “Wikipedia content has a **STEM bias**” (MINT-Inhalte überwiegen) ...?

# Gender Gap: Beitragende

Kleiner **Geschichts-Exkurs** - das Editor Gender Gap war in Wikipedias erstem Jahrzehnt weitgehend **unbekannt oder ignoriert**. Konkrete Zahlen sind wichtig!

- 2001: Wikipedia gegründet
- Ca. 2004: Erste Real-Life-Treffen, subjektive Eindrücke...
- 2005: Erste Wikimania, erste Umfrage (Uni Würzburg)
- 2009-10: Ergebnisse der UNU-MERIT/WMF-Umfrage veröffentlicht: 13% weiblich
- 2011: Foundation (Sue Gardner) bringt das Thema proaktiv auf die Titelseite der New York Times

Bayer, Tilman: *How many women edit Wikipedia?* <https://wikimediafoundation.org/news/2015/04/30/how-many-women-edit-wikipedia/>

User:HaeB: *New York Times article sparks extended discussion of Wikipedia's "gender gap"*, Wikipedia Signpost, 2. Februar 2011  
[https://en.wikipedia.org/wiki/Wikipedia:Wikipedia\\_Signpost/2011-02-07/Gender\\_gap](https://en.wikipedia.org/wiki/Wikipedia:Wikipedia_Signpost/2011-02-07/Gender_gap)

# Gender Gap: Beitragende

Der **Anteil weiblicher Benutzerinnen** variiert je nach Land, Wikipedia-Sprachversion und Umfragemethode (teilweise einige % unterschätzt aufgrund von Participation-Bias: Männer nehmen eher an Umfragen teil)...

...aber liegt praktisch durchweg immer noch weit unter 50%.

Maung, Rebecca: *2021 Community Insights Report*

[https://meta.wikimedia.org/wiki/Community\\_Insights/Community\\_Insights\\_2021\\_Report/Thriving\\_Movement#Community\\_and\\_Newcomer\\_Diversity](https://meta.wikimedia.org/wiki/Community_Insights/Community_Insights_2021_Report/Thriving_Movement#Community_and_Newcomer_Diversity)

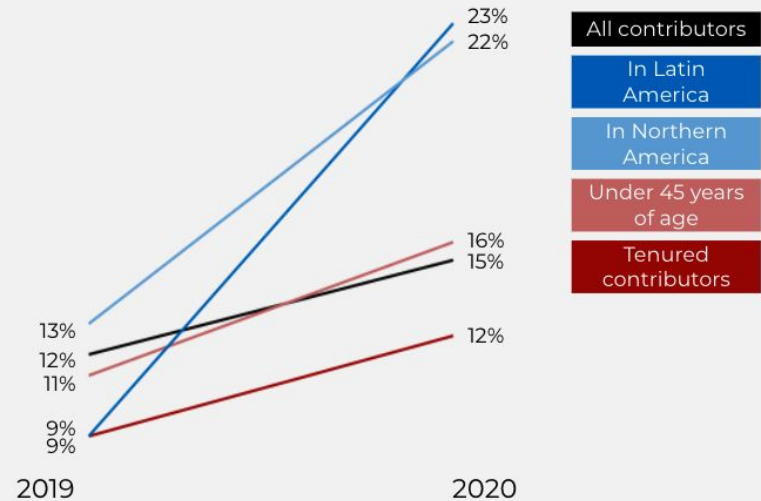
Hill, Benjamin Mako; Shaw, Aaron (2013): ["The Wikipedia Gender Gap Revisited: Characterizing Survey Response Bias with Propensity Score Estimation"](#) PLoS ONE Volume: 8, Issue: 6, DOI:10.1371/journal.pone.0065782.

Bayer, Tilman: *How many women edit Wikipedia?* <https://wikimediafoundation.org/news/2015/04/30/how-many-women-edit-wikipedia/>

# Gender Gap: Beitragende

Langfristige Trends sind nicht eindeutig, aber dann und wann werden positive Nachrichten vermeldet...

The overall **proportion of women increased** in the last year, **particularly among more tenured contributors, younger contributors, and those in certain geographies.**



Maung, Rebecca:

[https://meta.wikimedia.org/wiki/Community\\_Insights/Community\\_Insights\\_2021\\_Report/Thriving\\_Movement#Community\\_and\\_Newcomer\\_Diversity](https://meta.wikimedia.org/wiki/Community_Insights/Community_Insights_2021_Report/Thriving_Movement#Community_and_Newcomer_Diversity)

# Gender Gap: Beitragende

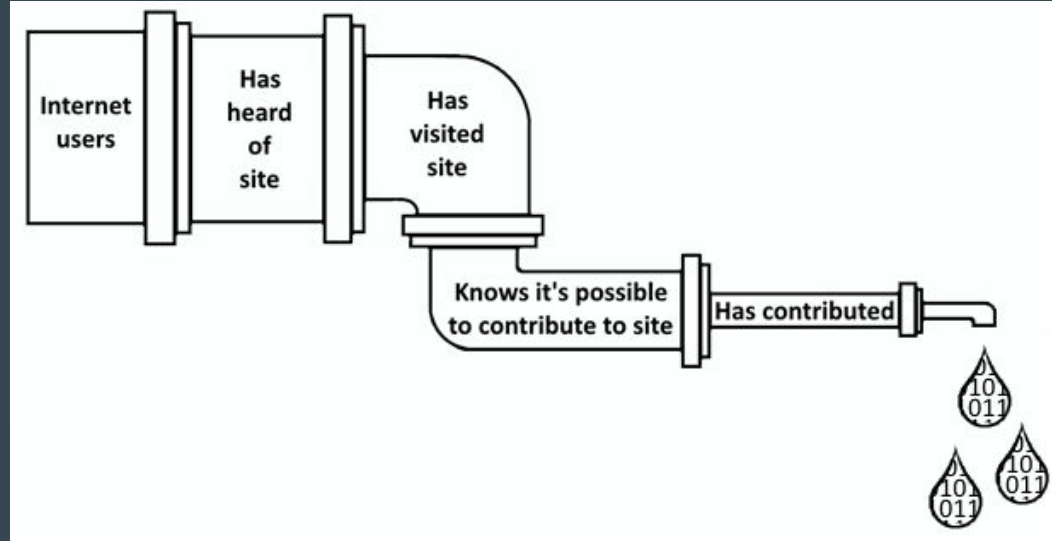
Shaw/Hargittai (2018):

“Pipeline of Online Participation Inequalities”

Umfragen unter US-Studierenden

**Pipeline-Lecks** (d.h. Gender Gaps) insbesondere bei “weiß, dass man Wikipedia bearbeiten kann”

→ **Empfehlung: Awareness Campaigns**



Shaw, Aaron; Hargittai, Eszter (2018-02-01). "The Pipeline of Online Participation Inequalities: The Case of Wikipedia Editing". Journal of Communication 68 (1): 143–168. ISSN 0021-9916. doi:10.1093/joc/jqx003 .

<https://web.archive.org/web/20180301134833/https://academic.oup.com/joc/article/68/1/143/4915319>

# Gender Gap: Beitragende

Hargittai/Shaw 2014 (ebenfalls Umfragen unter US-Studierenden):

“**Internet skills gap**” trägt zu Wikipedias Gender Gap bei

\* *“the most likely contributors are high-skilled males and ... among low-skilled Internet users no gender gap in Wikipedia contributions exists”*

\* NB: “skills” meint digitale Fähigkeiten **in der Eigenwahrnehmung**, d.h es scheint auch hier wieder um geringeres Selbstvertrauen zu gehen:

*“why women, on average, report lower level understanding of Internet-related terms remains a puzzle. Although studies with detailed data about actual skills based on performance tests suggest no gender differences in the observed skills, research that looks at self-rated know-how consistently finds gender variation with real consequences for online behavior”.*

# Gender Gap: Beitragende

Helfen Univorlesungen mit **Wikipedia-Assignments**? Vielleicht:

"Does a Wikipedia-based assignment increase self-efficacy among female students? A qualified maybe" (2021)

- *“an assignment in which students create original content for Wikipedia increases confidence in using the platform and information-seeking behavior among both female and male students, although the advantages continue to be greater for men”*

Kalaf-Hughes, Nicole; Cravens, R.G. (2021-12-22). ["Does a Wikipedia-based assignment increase self-efficacy among female students? A qualified maybe"](#). *Journal of Political Science Education*. 17 (sup1): 862–879. doi:[10.1080/15512169.2021.1921586](#). ISSN 1551-2169.



# Gender Gap: Beitragende

Shane-Simpson-Kristen (2016):

Labor-Experiment mit US-Studierenden:

Bearbeite eine Pseudo-Wiki-Seite mit sichtbarer “Versionsgeschichte” - entweder mit Gender-neutralen Benutzernamen ("Cheerios4Life", "AnonymousOne") oder mit weiblich/männlich konnotierten ("Ms Trouble", "Mr Football")

Frauen trugen weniger Text bei, wenn keine anderen weiblichen “Benutzerinnen” in der Versionsgeschichte sichtbar waren.

—> **Existierende Benutzerinnen mehr sichtbar zu machen** könnte helfen.

Shane-Simpson, Christina; Gillespie-Lynch, Kristen (2016-10-06). "Examining potential mechanisms underlying the Wikipedia gender gap through a collaborative editing task". *Computers in Human Behavior* **66** (January 2017): 312–328. doi:10.1016/j.chb.2016.09.043.

Review/Zusammenfassung:

[https://meta.wikimedia.org/wiki/Research:Newsletter/2016/October#Why\\_women\\_edit\\_less:\\_a\\_controlled\\_experiment](https://meta.wikimedia.org/wiki/Research:Newsletter/2016/October#Why_women_edit_less:_a_controlled_experiment)

# Was ist eigentlich Bias, und wie kann man ihn nachweisen?

**Viele verschiedene Definitionen** (z.B. [[Gender Bias]] sagt “eine verzerrte Wahrnehmung durch sexistische Vorurteile und Stereotype”)

Hier Fokus auf **Kausalität** - d.h. sind ungleiche Ergebnisse tatsächlich ursächlich durch Geschlecht bedingt, oder durch andere Faktoren (die mit Geschlecht korrelieren)?

**Ideal** wären randomisierte Experimente:

- Beispiel medizinische Studien: Eine Hälfte der Patientengruppe bekommt Impfstoff, die andere ein Placebo. Messe Unterschied in Häufigkeit der Corona-Erkrankungen
- Problem: Das Geschlecht von Wikipedia-Biographien etc. kann nicht zufällig zugewiesen werden...
- Audit Studies: zB Bewerbungen mit zufällig weiblichen/männlichen Namen verschicken

# Gender Gap: Inhalte

Beytía/Wagner: “Visibility layers”

Versuch der **Systematisierung**

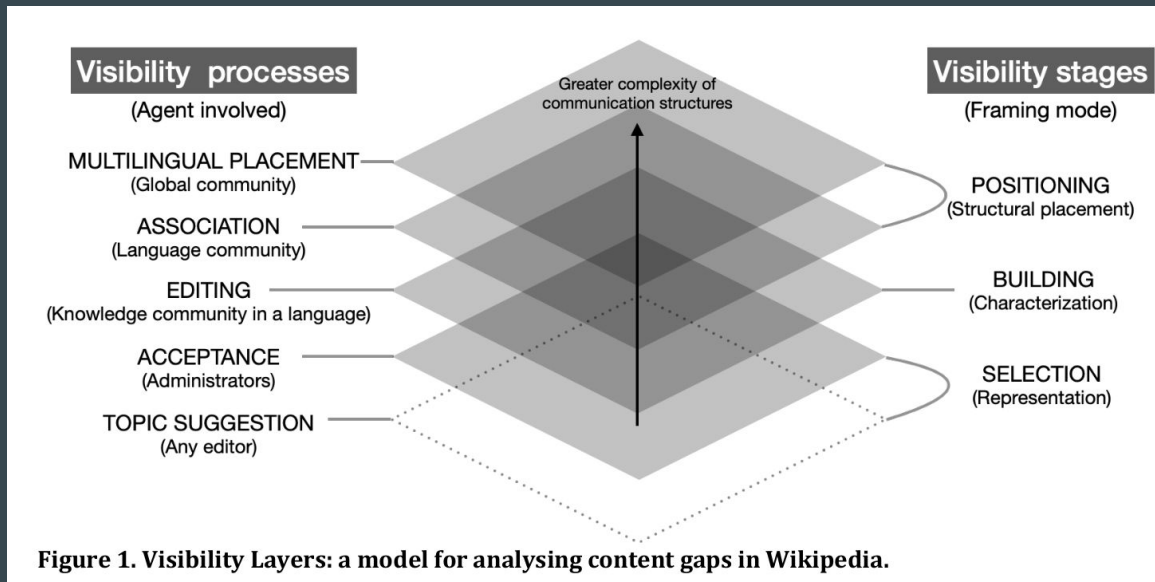
1. Artikelerstellung

2. Admins akzeptieren Artikel

3. Inhaltliche Verbesserungen

4. Verlinkungen etc. (d.h. Artikel nicht verwaist, ...)

5. “Multilingual placement”  
(interwikis etc.)



# Gender Gap: Inhalte

**Table 1. Visibility stages and gender asymmetries in Wikipedia**

Visibility stage	Visibility process(es)	Agent(s) involved	Asymmetry	Framing mode
Selection	Topic suggestion / Acceptance	Any editor / Administrators	Article coverage	Representation
			Deletion*	
Building	Editing	Knowledge community in a language	Writing length*	Characterisation
			Lexical	
			Topical	
			Visual	
			Source*	
Positioning	Association	Language community	Classification	Structural placement
	Multilingual placement	Global community	Network position	
			Multilingual notability*	

\* = These asymmetries have not shown a systematic preference for male content in the empirical tests.

# Gender Gap: Inhalte

Wikipedias **Gender Gaps in der Anzahl existierender/behaltener Artikel** sind eindeutig und wohlbekannt (Beispiel: 16,8 % weibliche Biografien auf dewiki derzeit)

Aber wie weit reflektieren sie **Wikipedias eigenen Gender Bias**?

Extrembeispiel Fields-Medaille (“Nobelpreis für Mathematik”): 60 Mal verliehen seit 1936, bisher lediglich eine weibliche Preisträgerin - nicht wirklich Wikipedias Schuld

...aber ab wann wird das zur Ausrede?

# Gender Gap: Inhalte

Adams/Brückner/Naslund: "Who Counts as a Notable Sociologist on Wikipedia? (2019)

Sample von 3000 US-Soziologinnen und Soziologen an Top-Universitäten

Ergebnis: **Frauen hatten weniger oft einen enwiki-Artikel** als männliche Kollegen mit gleichen Anzahl der Veröffentlichungen, h-index etc. (**Kontrollvariablen**)

...aber **keine Unterschiede bei Löschhäufigkeit**, d.h. Gap schon bei Artikelerstellung

Die Autorinnen machen die **Relevanzkriterien der enwiki verantwortlich**:

*"The Wikipedia criteria for notability of academics include references to making a 'significant impact on the field' and being an 'elected member of a highly selective and prestigious scholarly society or association' or the 'highest-level elected or appointed administrative post at a major academic institution,' [...] There are gendered and racialized patterns and criteria already embedded in these judgments. ..."*

Adams, Julia; Brückner, Hannah; Naslund, Cambria (2019-01-01). "Who Counts as a Notable Sociologist on Wikipedia? Gender, Race, and the 'Professor Test;". *Socius* 5: 2378023118823946. ISSN 2378-0231. doi:10.1177/2378023118823946

# Gender Gap: Inhalte

Zhang, Terveen: "Quantifying the Gap: A Case Study of Wikidata Gender Disparities". (2021)

“in Wikidata, ... less than 22% of items that represent people are about women. We asked: what is the source of this bias? Specifically, does it originate from the actions of Wikidata editors or from external factors; that is, does it simply reflect existing real world gender bias? We conducted a quantitative case study that found: (i) the most popular categories of people included in Wikidata represent **male-dominant professions, such as American football**; (ii) within a selected set of professions where we could obtain gender distribution data, **Wikidata is no more biased than the real world**: men and women are included at similar percentages, and the quality of items representing men and women also is similar.”

Zhang, Charles Chuankai; Terveen, Loren (2021-09-15). "[Quantifying the Gap: A Case Study of Wikidata Gender Disparities](#)". *17th International Symposium on Open Collaboration*. New York, NY, USA: Association for Computing Machinery. pp. 1–12. ISBN [9781450385008](#).

# Gender Gap: Inhalte

Field, Park, Tsvetkov: "Controlled Analyses of Social Biases in Wikipedia Bios" (2021):

\* *“While prior research has examined man/woman gender bias in biography articles, possible influences of **confounding variables limit conclusions**. In this work, we present a methodology for reducing the effects of confounding variables in analyses of Wikipedia biography pages. Given a target corpus for analysis (e.g. biography pages about women), we present a method for constructing a **comparison corpus** that matches the target corpus in as many attributes as possible, except the target attribute (e.g. the gender of the subject).”*

D.h. finde möglichst “ähnliche” Artikel, die sich nur im Geschlecht (“target attribute”) unterscheiden

Beispiel: [[Tim Cook]] → [[Marissa Mayer]]

Die Foundation arbeitet derzeit an einem Tool dafür.

Field, Anjalie; Park, Chan Young; Tsvetkov, Yulia (2021-10-22). ["Controlled Analyses of Social Biases in Wikipedia Bios"](#) (PDF). *arXiv:2101.00078 [cs]*. (v2) (preprint)

<https://phabricator.wikimedia.org/T290447>



# Gender Gap: Inhalte

Wagner et al.: "It's a Man's Wikipedia? Assessing Gender Inequality in an Online Encyclopedia" (2015)

Gute Nachrichten:

\* Vergleich mit drei externen Personen-Listen (Freebase, MITs "Pantheon", "Human Accomplishment"): Frauen auf enwiki "**slightly overrepresented**" (Erwartung war unterrepräsentiert).

Wagner, Claudia; Garcia, David; Jadidi, Mohsen; Strohmaier, Markus (2015-04-21). ["It's a Man's Wikipedia? Assessing Gender Inequality in an Online Encyclopedia"](#). *Ninth International AAAI Conference on Web and Social Media*.

# Gender Gap: Inhalte

Wagner et al.: "It's a Man's Wikipedia? Assessing Gender Inequality in an Online Encyclopedia" (2015) (*Fortsetzung*)

Schlechte Nachrichten:

\* Wikipedia-Artikel über Wissenschaftlerinnen fielen oft durch den **“Finkbeiner-Test”** (Betonung auf Familie, Kochkünsten, “erste Frau in..”)

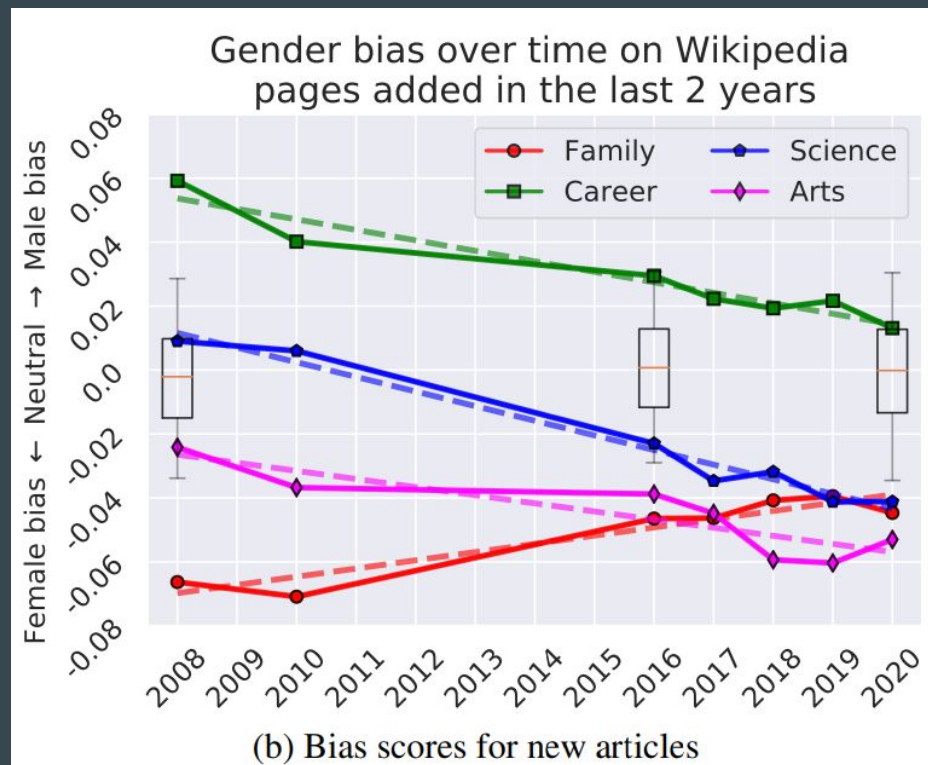
\* **“Schlumpfine-Prinzip”** (“wenn in einer Gruppe von männlichen Hauptcharakteren nur eine weibliche Hauptfigur vorkommt”) zeigt sich in der Struktur von Wikilinks

Wagner, Claudia; Garcia, David; Jadidi, Mohsen; Strohmaier, Markus (2015-04-21). ["It's a Man's Wikipedia? Assessing Gender Inequality in an Online Encyclopedia"](#). *Ninth International AAAI Conference on Web and Social Media*.

# Interventionen

Schmahl et al.: "Is Wikipedia succeeding in reducing gender bias? Assessing changes in gender bias in Wikipedia using word embeddings" (2020)

Beispiel: "mother" kommt öfter als "father" vor im Themenbereich "family"



# Interventionen

Langrock et al.: “The Gender Divide in Wikipedia: Quantifying and Assessing the Impact of Two Feminist Interventions.” (2020)

Auswirkungen von “Art+Feminism” und “500 Women Scientists” auf 11.000 enwiki-Artikel

“We find that the interventions are **successful at adding content about women that would otherwise be missing**, but they are **less successful at addressing structural biases that limit the visibility of that content.**” - insbesondere Verlinkungen (Wikilinks von, aber nicht zu den bearbeiteten Frauen-Biografien)

Langrock, Isabelle; González-Bailón, Sandra (2020-11-28). [\*The Gender Divide in Wikipedia: Quantifying and Assessing the Impact of Two Feminist Interventions\*](#). Rochester, NY: Social Science Research Network. (preprint)

[https://figshare.com/articles/presentation/Bridging\\_Wikipedia\\_s\\_Gender\\_Gap\\_Quantifying\\_and\\_Assessing\\_the\\_Impact\\_of\\_Two\\_Feminist\\_Interventions\\_-\\_Wikimedia\\_Research\\_Showcase\\_-\\_July\\_2021/15059532](https://figshare.com/articles/presentation/Bridging_Wikipedia_s_Gender_Gap_Quantifying_and_Assessing_the_Impact_of_Two_Feminist_Interventions_-_Wikimedia_Research_Showcase_-_July_2021/15059532)

# Ausblick: Interessante offene Fragen

- Offensive Kommunikation von Wikipedias Gender Gap kann motivieren, aber auch abschrecken
- Community-Ethnographie: Frauen in Führungspositionen
  - Beispiel: Wie sind Community-Policies entstanden?
  - Verifiability auf enwiki: User:SlimVirgin
  - Kurier, Checkuser auf dewiki: Elian, Bdk...
- Freizeit-Hypothese (Second Shift)
  - Wahrscheinlich kein Faktor für Gender Gap in z.B. USA
  - ...aber in anderen Ländern?

Hinnosaar, Marit (2015): "[Gender Inequality in New Media: Evidence from Wikipedia](#)". No 411, Carlo Alberto Notebooks from Collegio Carlo Alberto, ISSN 2279-9362.

U.S. Bureau of Labor Statistics: [American Time Use Survey - Charts by Topic: Volunteer activities](#),

OECD.Stat: [Employment : Time spent in paid and unpaid work, by sex](#)

# Vielen Dank! Was sind eure Fragen?

[@WikiResearch](#) auf Twitter

[\[\[meta:Research:Newsletter\]\]](#) :

- Archivsuche (über 1700 Publikationen seit 2011)
- Beiträge/Reviews stets willkommen
- Special Issues zu bestimmten Themen, u.a. Gender Gap/Bias

HaeBwiki@gmail.com