Soziale Normen, Kooperation und Konflikte in der wissenschaftlichen Zusammenarbeit

Heiko Rauhut

Universität Zürich, Soziologisches Institut CHESS talk, 8.12.2016

Chancen und Risiken von Teamwork

Vorteile von Teamwork

- Arbeitsteilung
- Spezialisierung
- Weisheit der Vielen
- Networking

Nachteile von Teamwork

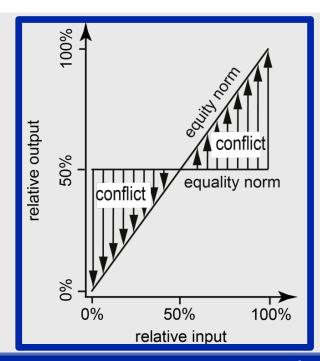
- Kollektivgutprobleme
- Normenkonflikte zwischen
 Gleichheits- und Leistungsnormen,
 insbesondere bei Autorenreihenfolge
 (alphabetisch vs. Erstautorenschaft)
- Wissenschaftliches Fehlverhalten (?)

SNSF Starting Grant: CONCISE



Theorie der Normen-konflikte

Winter, Rauhut, Helbing, 2011



Fokus dieses Vortrags

The increasing dominance of networking in the production of knowledge

Heiko Rauhuta,b,1,2 and Fabian Winterc,1,2

^aUniversity of Zurich, Institute of Sociology, Andreasstrasse 15, CH-8050 Zürich, Switzerland; ^bETH Zurich, Swiss Federal Institute of Technology; ^cMax Planck Institute for Research on Collective Goods, MPRG "Mechanisms of Normative Change", Kurt-Schumacher-Strasse 10, 53113 Bonn, Germany, orcid.org/0000-0002-4838-4504

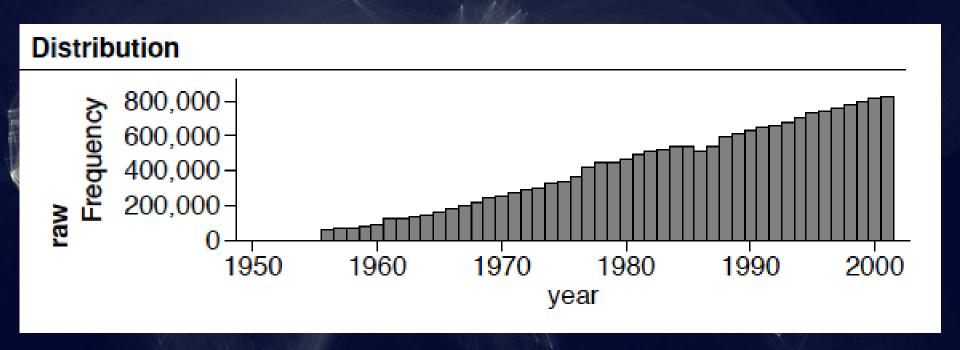
This manuscript was compiled on July 14, 2016







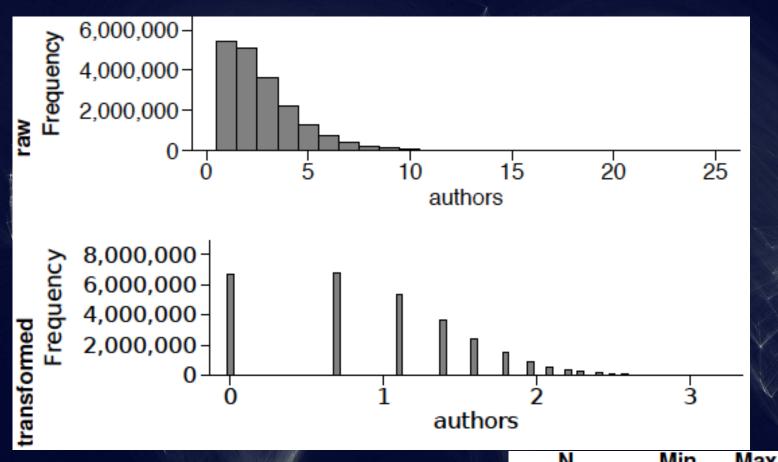
Publikationsjahr



N	Min	Max	Mean
19,473,631	1956	2001	1985.9



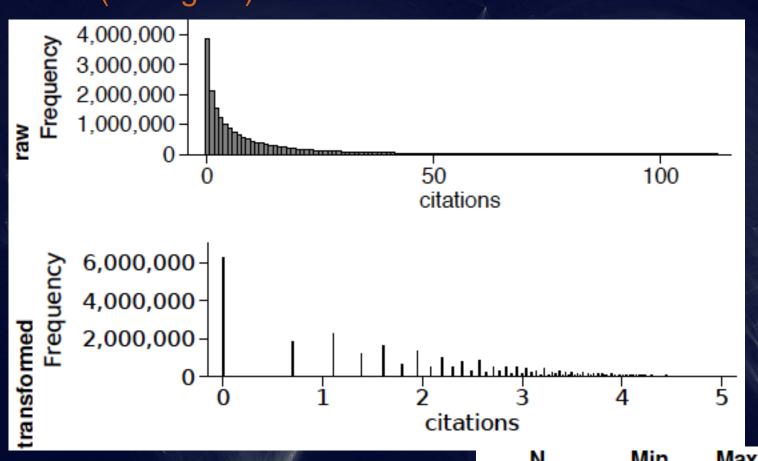
Autoren



N	Min	Max	Mean	SD
9,473,631	1	917	2.9	4.5
p 1	p 5	p10	p50	p90
1	1	1	2	5



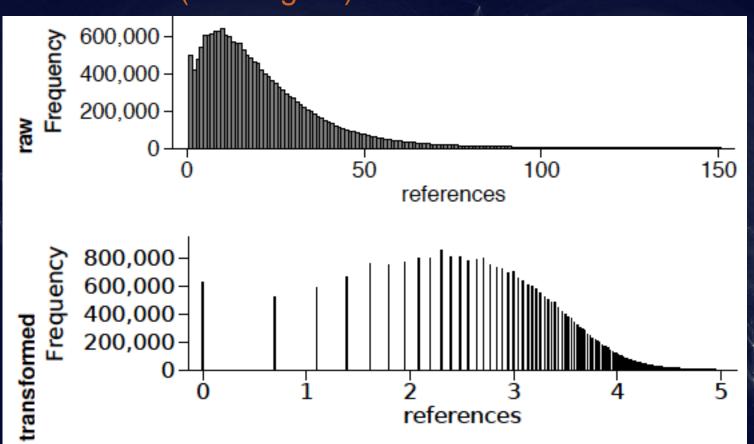
Zitate (in-degree)

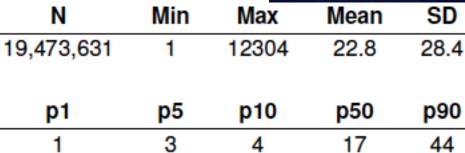


N	IVIIN	wax	wean	อบ
19,473,631	0	30343	13.3	39.1
р1	p5	p10	p50	p90
0	0	0	4	32



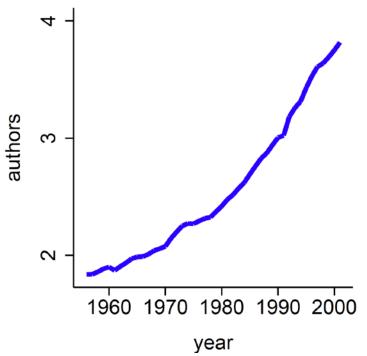
Referenzen (out-degree)





Wandel in Wissensproduktion: Von Einzelforschern zu Teams

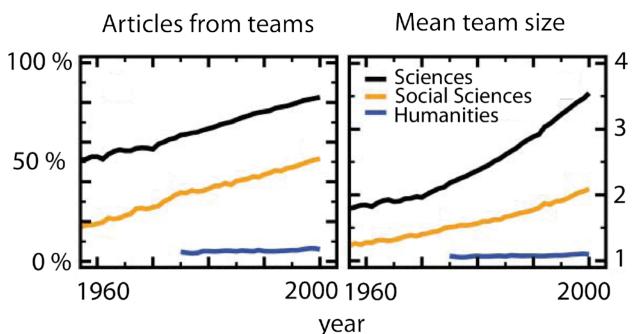
Mittlere Autorenanzahl, gesamter Web of Science (N > 20 Mio), Rauhut & Winter, 2016, R&R:

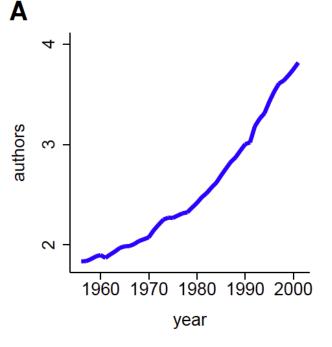


4 / 25

Trend gilt für alle Disziplinen

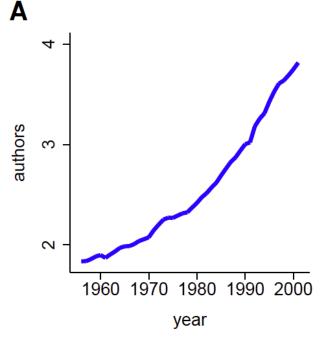
Wuchty et al. 2007, Science

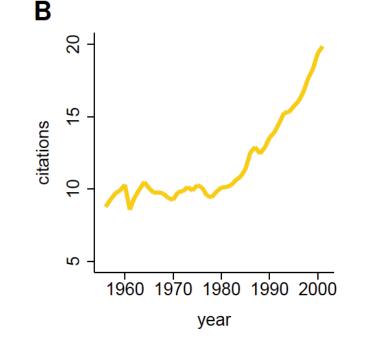




Teams mehr Resonanz als Einzelforscher

A: Mittlere Autorenzahl

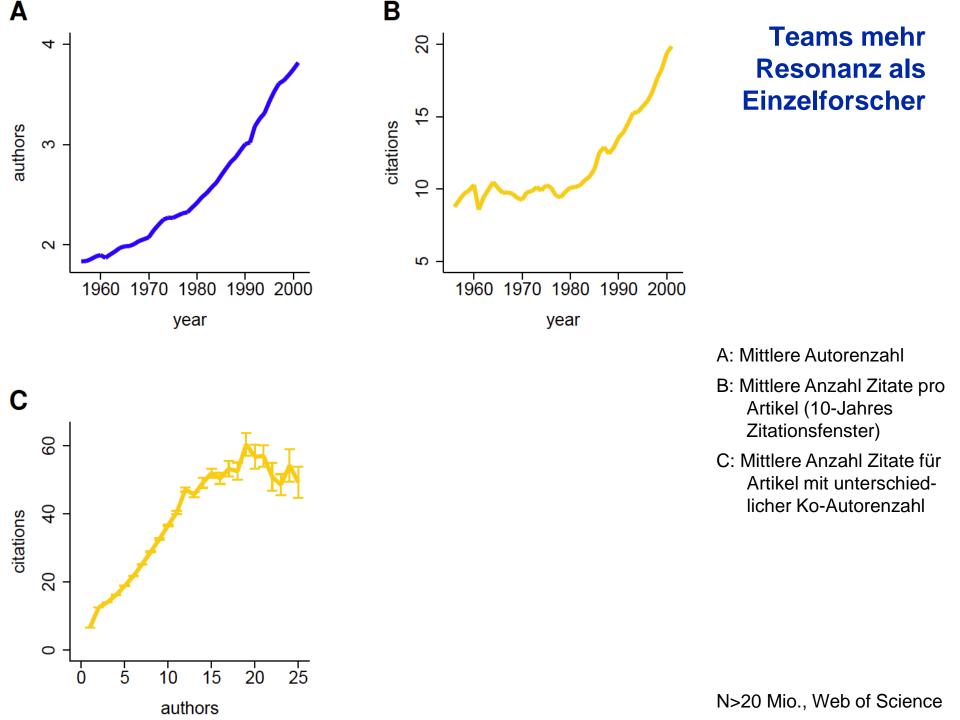




Teams mehr Resonanz als Einzelforscher

A: Mittlere Autorenzahl

B: Mittlere Anzahl Zitate pro Artikel (10-Jahres Zitationsfenster)



Teamwork Effekt

$$y'_{i,j} = b_1(x'_{i,j}) + v_j + \epsilon_i$$

 $y'_{i,j}$ Zitate (log) von Papier i in Journal j

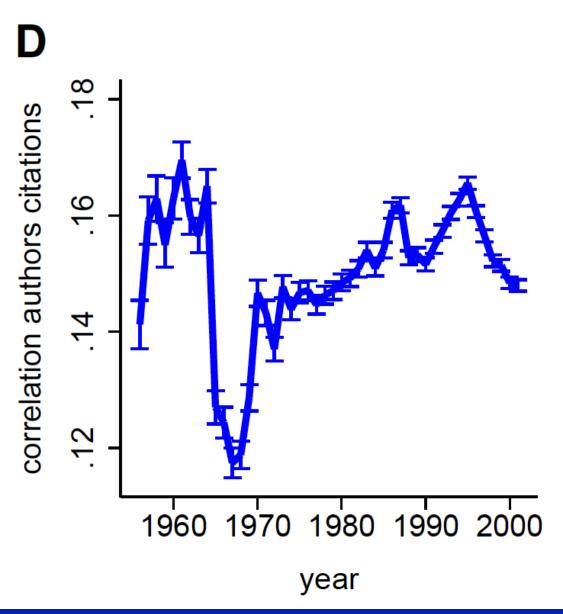
 $x'_{i,j}$ Teamgrössen(log) von Papier i in Journal j

 v_i Journal fixed effects

Für jedes Jahr eine getrennte Regression, um temporale und "Qualitätseffekte" zu kontrollieren

Journal fixed effects kontrolliert für: Disziplin, Fach, "Qualität" und "Substanz der Artikel"

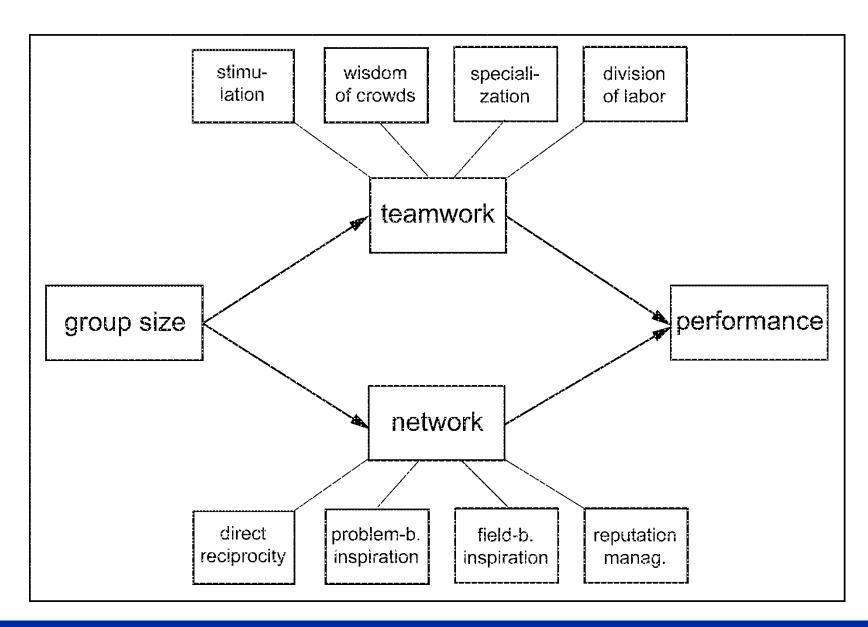
Teamwork Effekt



D: Standardisierte Journal-Fixed Effects Regressionskoeffizienten zwischen Autorenzahl und Zitate, separat pro Jahr.

N>20 Mio., Web of Science

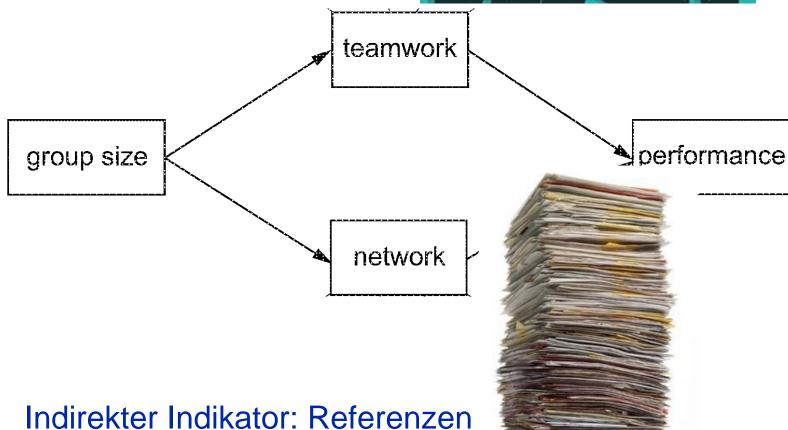
Welcher Mechanismus?

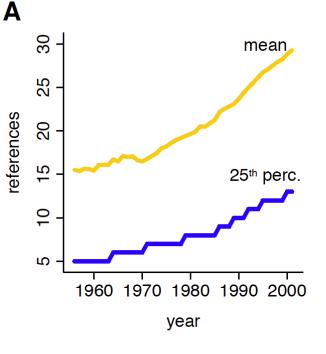


Welcher Mechanismus?

Indikator: Teamgrösse

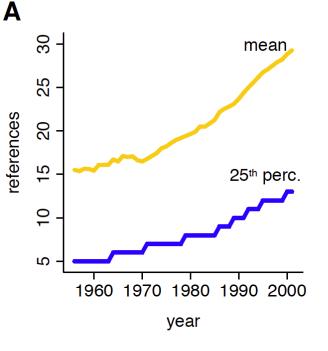


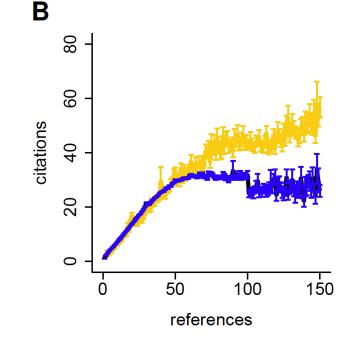




Netzwerkmechanismus: Mehr Referenzen führen zu mehr Zitaten

A: Mittlere Referenzenzahl

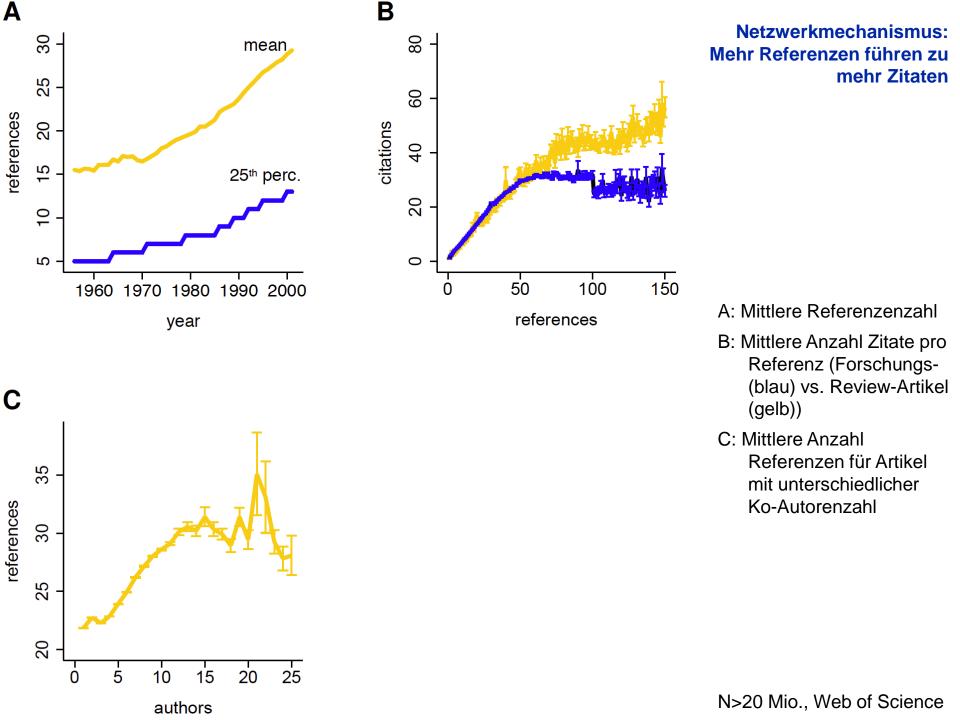




Netzwerkmechanismus: Mehr Referenzen führen zu mehr Zitaten

A: Mittlere Referenzenzahl

B: Mittlere Anzahl Zitate pro Referenz (Forschungs-(blau) vs. Review-Artikel (gelb))



Netzwerk Effekt

$$y'_{i,j} = b_1(x'_{i,j}) + v_j + \epsilon_i$$

 $y'_{i,j}$ Zitate (log) von Papier i in Journal j

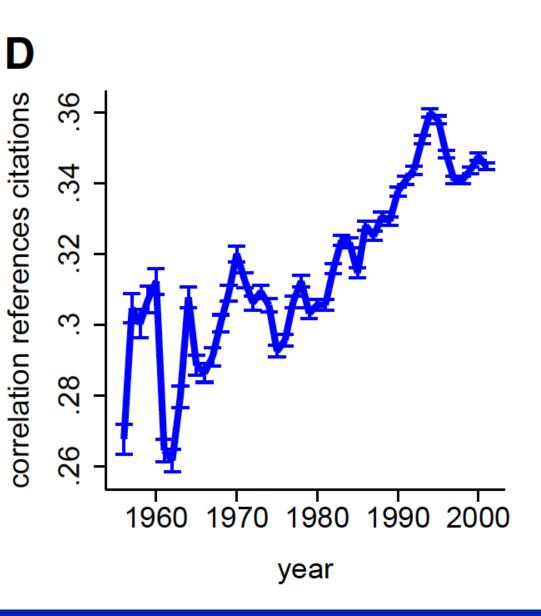
 $x'_{i,j}$ Referenzen (log) von Papier i in Journal j

 v_i Journal fixed effects

Für jedes Jahr eine getrennte Regression, um temporale und "Qualitätseffekte" zu kontrollieren

Journal fixed effects kontrolliert für: Disziplin, Fach, "Qualität" und "Substanz der Artikel"

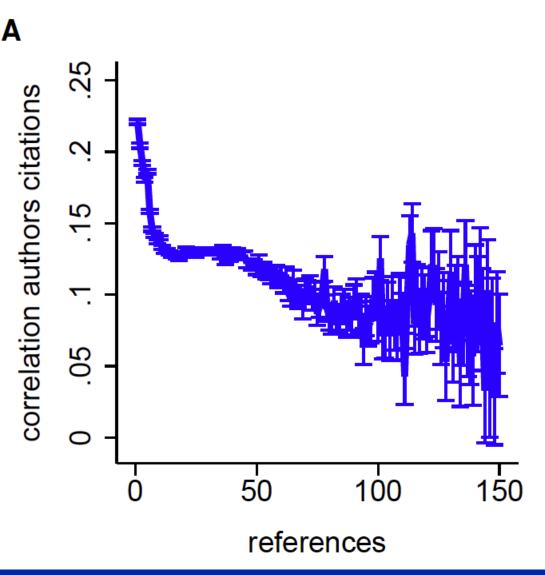
Netzwerk Effekt



D: Standardisierte Journal-Fixed Effects Regressionskoeffizienten zwischen Referenzen und Zitaten, separat pro Jahr.

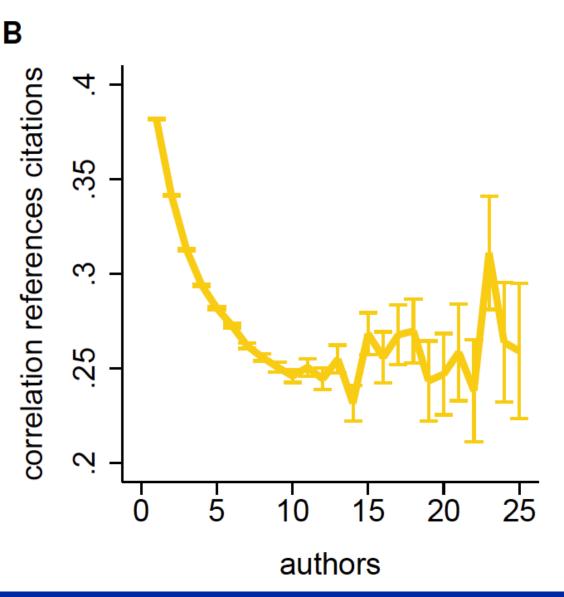
N>20 Mio., Web of Science

Substitution von Team- durch Netzwerkeffekt



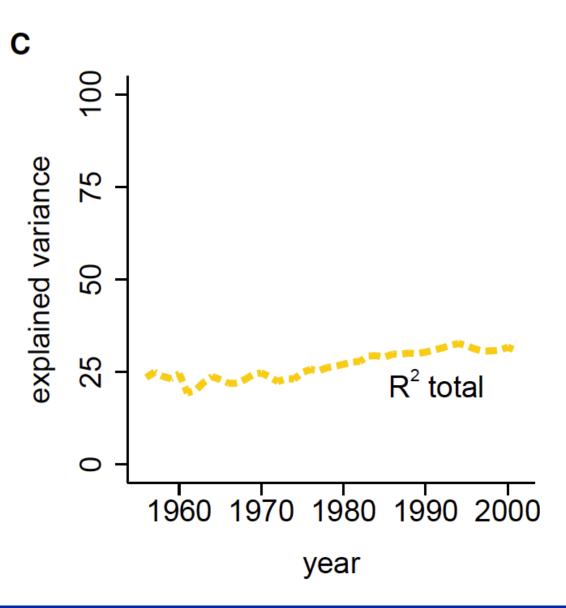
- Interaktion zwischen Teamgrössen, Zitaten und Referenzen
- Standardisierte
 Regressionskoeffizienten für Autoren
 (log) auf Zitate (log) getrennt für
 Papiere mit unterschiedlicher Anzahl
 Referenzen (Journal-Fixed Effects
 Regressionen)
- Fehlerbalken: Standardfehler

Substitution von Netzwerk- durch Teameffekt



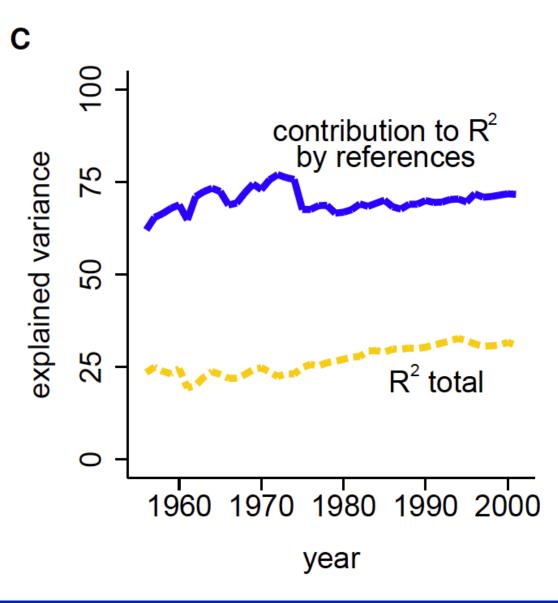
- Interaktion zwischen Referenzen, Zitaten und Teamgrössen
- Standardisierte
 Regressionskoeffizienten für
 Referenzen (log) auf Zitate (log)
 getrennt für Papiere mit
 unterschiedlicher Anzahl Ko-Autoren
 (Journal-Fixed Effects Regressionen)
- Fehlerbalken: Standardfehler

Dominanz des Netzwerkeffekts über den Team-Effekt



 Kombiniertes Modell mit Autorenanzahl (log) und Referenzen (log) auf Zitate (log), journal fixedeffects, getrennt nach Jahren

Dominanz des Netzwerkeffekts über den Team-Effekt



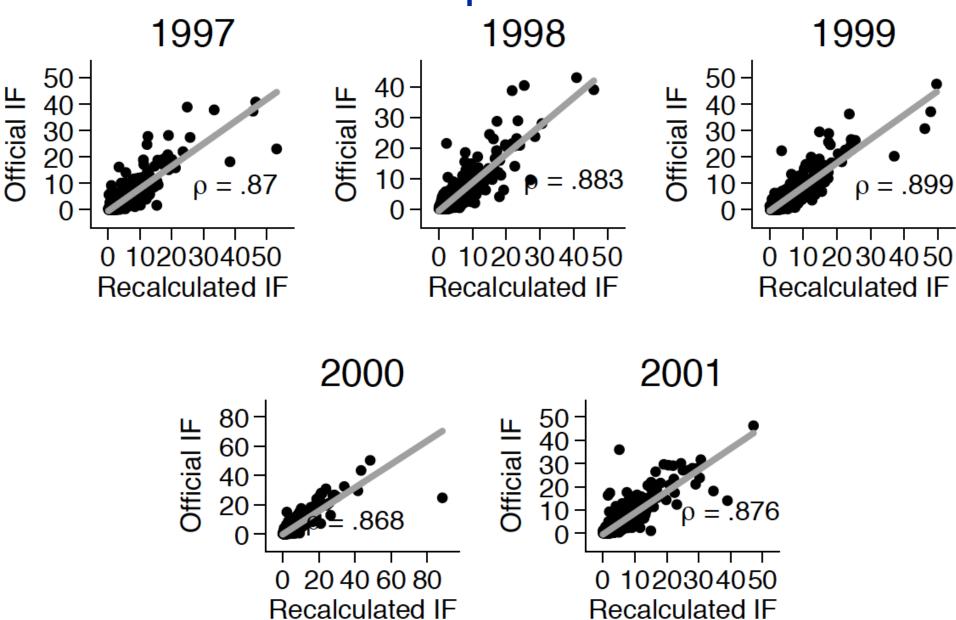
- Kombiniertes Modell mit Autorenanzahl (log) und Referenzen (log) und deren nicht-linearen Effekte (quadratisch) auf Zitate (log), journal fixed-effects, getrennt nach Jahren
- Shapley-Owen Decomposition der erklärten Varianz: Anteil an erklärter Varianz von Referenzen im Vergleich zu Teamgrössse (addiert sich auf 100%)

Robustheitsanalysen I: Journal Impact Factors

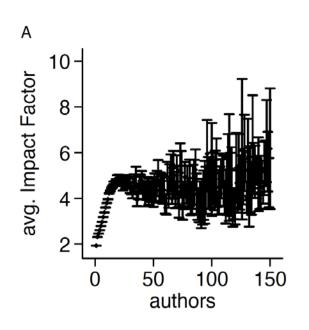
Robustheitsanalysen I: Journal Impact Factors

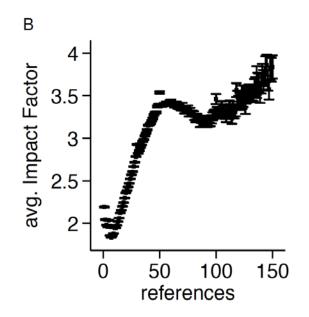
sum of citations to items in i in y-1 and y-2 by indexed publications in y the total number of "citable items" published in i in y-1 and y-2.

Robustheitsanalysen I: Journal Impact Factors



Robustheitsanalysen I: Journal Impact Factors





С			
	Impact Factor		
authors	0.171		
	(317.36)		
references	0.298		
	(781.87)		
year	0.0201		
	(854.07)		
constant	-38.61		
	(-831.69)		
Fields	264		
N	44,449,016		
t statistics in parentheses			

Fig. S2. Average impact factor of articles published by a given team size (A) and length of reference list (B). The table in (C) reports the correlation between impact factor and authors and references, respectively. The regression controls for field fixed effects.

Impact Factor_i = $b_0 + b_1 \ln (\text{authors}_i) + b_2 \ln (\text{references}_i) + b_3 \text{year}_i + u_i + \epsilon$,

Robustheitsanalysen II:

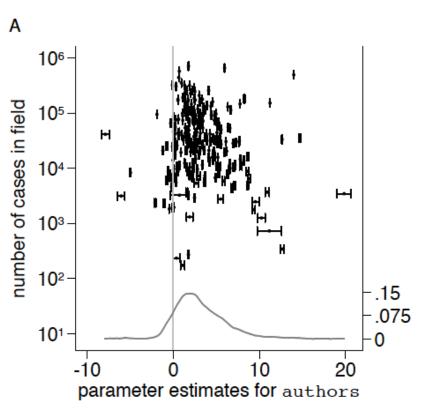
Jedes Fach Teams mehr Zitate und unterschiedliche Optima?

citations =
$$\beta_0 + \beta_1$$
 authors + β_2 authors² + ϵ

Robustheitsanalysen II:

Jedes Fach Teams mehr Zitate und unterschiedliche Optima?

citations =
$$\beta_0 + \beta_1$$
 authors + β_2 authors² + ϵ



Robustheitsanalysen II:

Jedes Fach Teams mehr Zitate und unterschiedliche Optima?

citations =
$$\beta_0 + \beta_1$$
 authors + β_2 authors² + ϵ

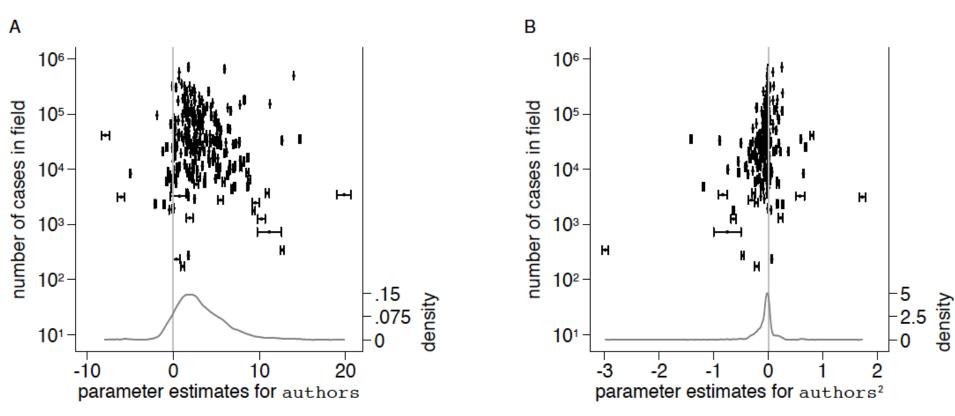
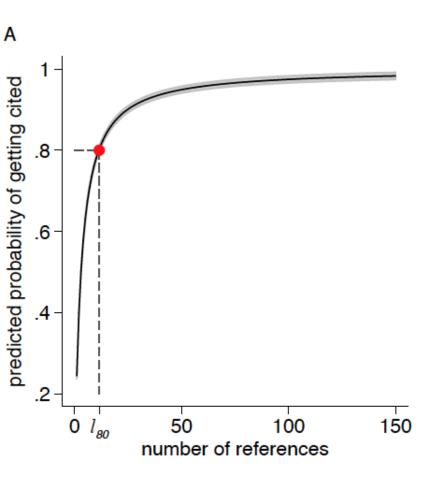


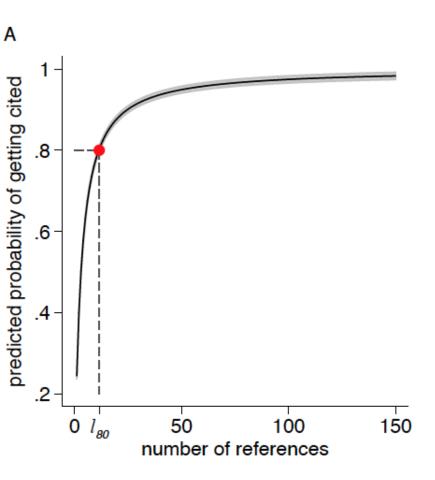
Fig. S3. Subfigure A plots the parameter estimated influence of one additional author on citations (β_1 in equation 2, black dots) and the corresponding 95% confidence intervals (whiskers) against the \log_{10} -number of articles in a field. The kernel density plot on the bottom of the figure with the corresponding axis on the right displays the distribution of estimates. Subfigure B plots the parameter estimated influence of the square of one additional author on citations (β_2 in equation 2, black dots) and the corresponding 95% confidence intervals (whiskers) against the \log_{10} -number of articles in a field. The kernel density plot on the bottom of the figure with the corresponding axis on the right displays the distribution of estimates. The estimates in A are mostly positive, while the estimates in B are mostly negative, suggesting a coherent pattern in different academic fields with variable "optimal" team sizes.

Robustheitsanalysen III: Wandelnde Normen "zitierbarer" Artikel

Robustheitsanalysen III: Wandelnde Normen "zitierbarer" Artikel



Robustheitsanalysen III: Wandelnde Normen "zitierbarer" Artikel



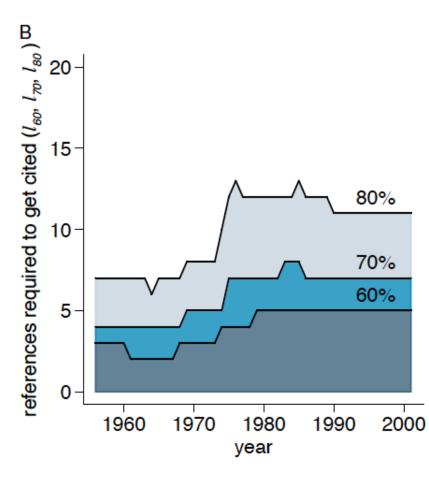
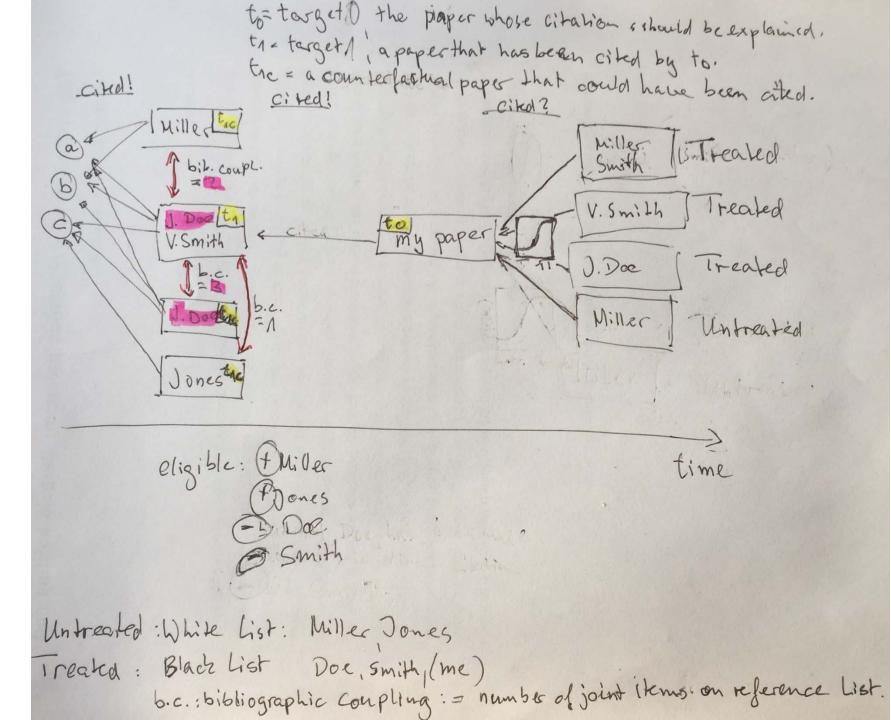


Fig. S4. Predicted probability of being cited at least once (A). Predictions are derived from a logistic regression of the form P(Being cited | references) = $\beta_0 + \beta_1$ References + ϵ . The point l_{80} on the x-axis indicates the length of the literature list l_{80} , where the probability of being cited is at least 80%. The grey area underlying the black line indicates the 95% confidence interval of the estimate. (B) Number of references required to have a predicted probability of being cited of at least 60%, 70%, and 80% over time. The red marker in A for instance corresponds to the l_{80} averaged over all years.

Aktuelle Zusatzanalysen: "Kausalanalyse" Netzwerkeffekt

- Nearest neighbor Kausalanalyse
- Identifikation von Artikeln, die in Referenzenliste zitiert wurden (treated)
- Identifikation von möglichst ähnlichen Artikeln, die hätten zitiert werden können, aber nicht zitiert wurden (untreated counterfactural control group)
- Ähnlichkeit anhand "bibliographic coupling": Möglichst ähnliche
 Referenzenlisten zwischen nicht-zitierten Papieren und zitierten Papieren
- Bibliographic coupling innerhalb eines Feldes des Web of Science
- Identifikation von Autoren auf diesen ähnlichen nicht-zitierten Artikeln
- Streichen von Autoren, die vom Ursprungspapier bereits bei anderen Artikeln zitiert wurden ("black list")
- Vergleich der Wahrscheinlichkeit eines Zitats von Autoren, die in der Referenzenliste des Ursprungsartikels zitiert wurden (treated) mit Autoren von "untreated counterfactual non-cited papers"
- Logistische Regression treated vs. non-treated auf Zitate
- Wenn zitierte Autoren eher zurückzitieren als unzitierte Autoren, die genausogut hätten genannt werden können -> Kausaleffekt des Netzwerkeffekts



Zusammenfassung

Teamwork wird dominierendes Arbeitsmodell in allen Fächern

 In allen Fächern nimmt Teamwork zu und Teams erzielen mehr Aufmerksamkeit (Zitate) als Einzelforscher

Vernetzung und "Marketing" zunehmend wichtiger

- Hinweise, dass Teams sowohl "bessere" Resultate erzielen ("Teamwork") als auch stärkere Verbreitung ihrer Forschung ("Netzwerk")
- Empirische Hinweise, dass Netzwerkmechanismus erklärungskräftiger anhand starker Effekte der Anzahl Referenzen ("out-degree") auf Zitate ("indegree")

Zusammenfassung

Individuelle Perspektive: Strategisches Zitieren und wissenschaftliche Karrieren

- Vermutlich häufig von Einzelforschern unterschätzt, dass viele Referenzen strategisch für Zitationserfolg eingesetzt werden können
- Optimale Anzahl um 50 Referenzen
- Zu viele Referenzen k\u00f6nnte "instrumentell" wirken und Glaubw\u00fcrdigkeit unterminieren

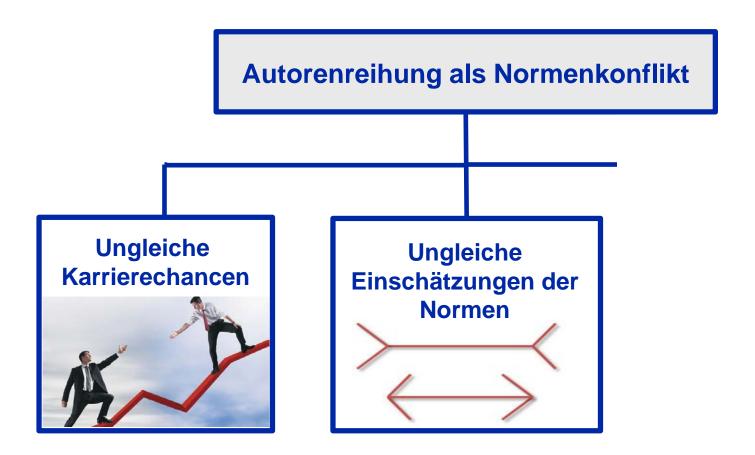
Forschungspolitische Perspektive

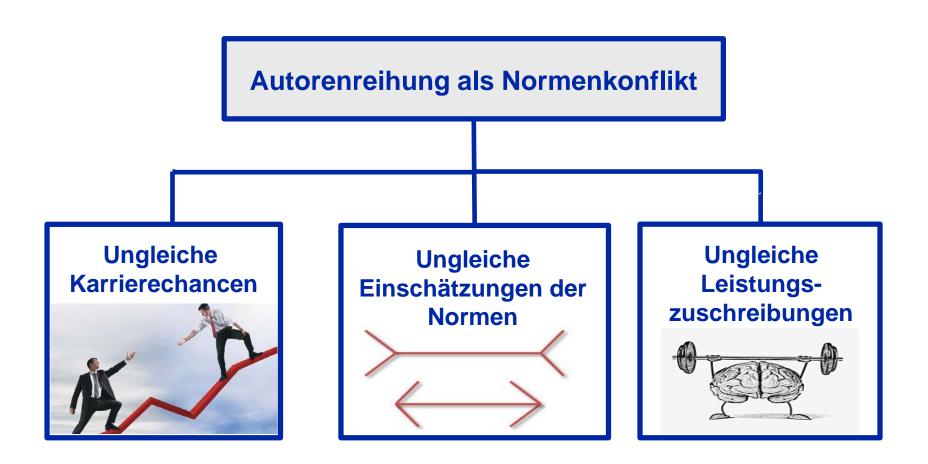
- Instrumentelle Zitate eher unerwünscht
- Referenzlimits wie bei general audience journals wie PNAS, Science, Nature mit etwa 20 Referenzen könnte sinnvolles, praktisches Modell für Fachzeitschriften sein

Autorenreihung als Normenkonflikt

Autorenreihung als Normenkonflikt







Diskussion

Verträge oder Vertrauen?

 "Prenuptials" und Verträge über Datenaustausch, Arbeitsteilung und Autorenschaften oder Vertrauen, informelle Normen und unvollständige Verträge?



Trennung zwischen Betreuer und Gutachter?
 Zunehmende Spezialisierung mit wenigen Experten vs. Bewertung eigener Arbeiten durch Betreuer

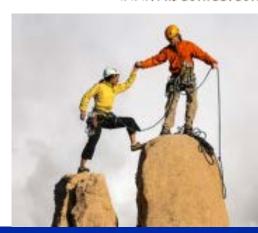
Werden prosoziale Präferenzen zunehmend wichtiger für wissenschaftliche Karrieren?

 Befragungen und Verhaltensexperimente mit Forschern: In welchen Fächern (Solo vs. Team-Disziplinen) Prosozialität und in welchen egoistisches Karrierestreben erfolgreicher?





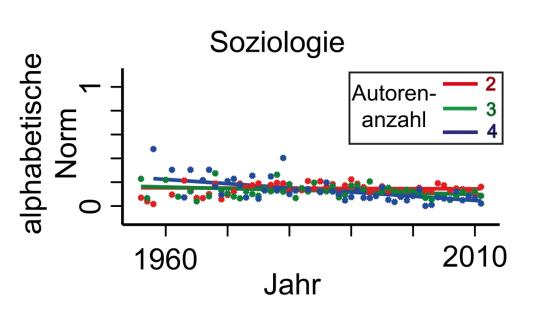
WWW.PHDCOMICS.COM

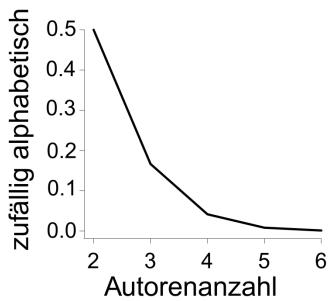


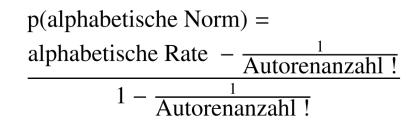
Appendix zu Forschungsresultaten

Autorenreihenfolge als Normenkonflikt?

Schätzung der Normen

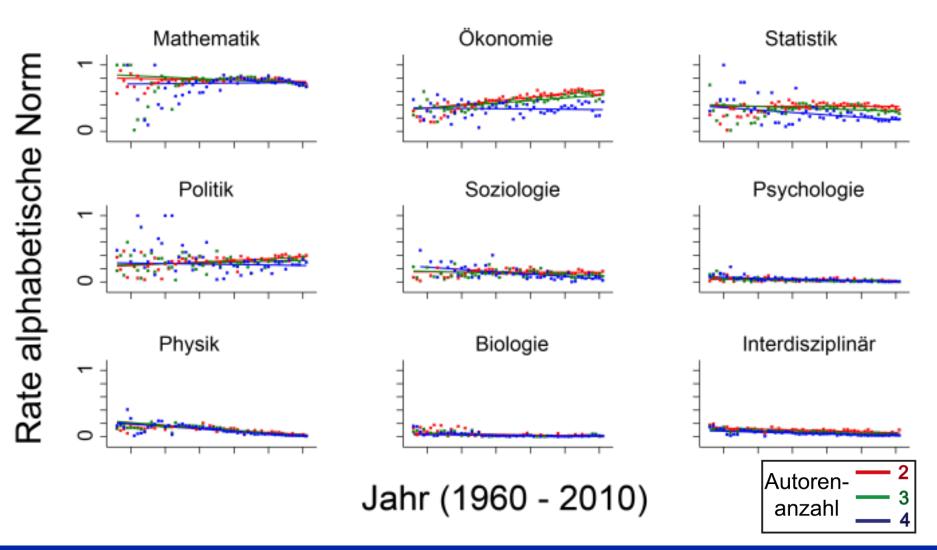






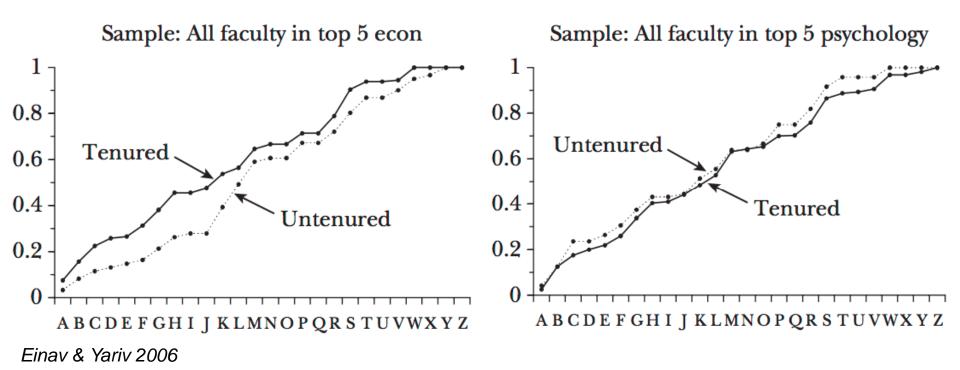
Autorenreihenfolge als Normenkonflikt?

Web of Science, 1'673'210 Artikel



Ungleiche Karrierechancen von "Z-Autoren"

Z-Autoren: seltener Tenure an Top-Econ-Unis, Fellows (Ec. Soc) und Nobelpreis

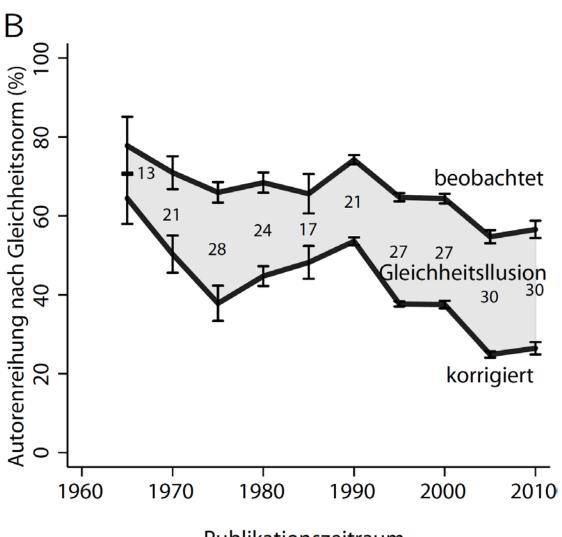


Work in progress:

Arbeiten Autoren mit späten Nachnamen häufiger alleine?

Unsere Analysen zeigen, dass Autoren in alphabetisch reihenden Fächern häufiger von anderen Ko-Autoren angefragt werden

Ungleiche Einschätzungen der Normen: "Gleichheitsillusion"



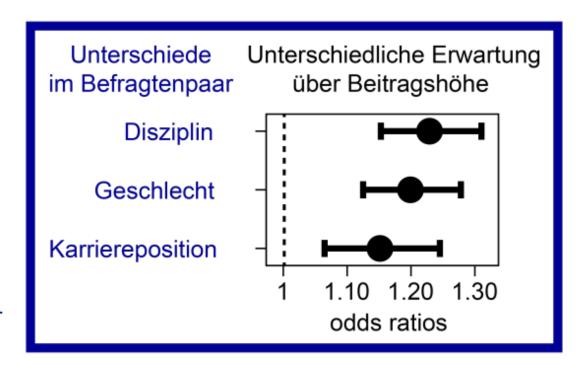
Obere Linie: unkorrigierte Rate alphabetisch gereihter Artikel; untere Linie: nach zufälliger Reihung korrigierte Rate. Differenz: Gleichheitsillusion.

Datenbasis: Deutsche Soziologie-Journals (KZfSS, ZfS, BJ, SW). N=3'378 Publikationen, 1'919 Autoren, 12'170 Zitationen.

Ungleiche Leistungszuschreibungen

- Umfragedesign:
 US Forscher aus Ökonomie,
 Management, Psychologie
 Maciejovsky, Budescu, Ariely, 2008
- Vignettenmethode:

 Zuschreibung von relativem Beitrag von Autoren in unterschiedlichen Positionen in Autorenliste
- Simulation Verhandlungsspiel: Kombination aller vergleichbaren Befragtenpaare: Konfliktwahrscheinlichkeit über Beitragsleistungen (N = 52'809)



10% mehr Konflikte bei Autorenpaaren aus unterschiedlicher Disziplin, mit unterschiedlichem Geschlecht oder in unterschiedlicher Karriereposition

Zusammenfassung

Teamwork wird dominierendes Arbeitsmodell in allen Fächern

 In allen Fächern nimmt Teamwork zu und Teams erzielen mehr Aufmerksamkeit (Zitate) als Einzelforscher

Implikation 1: Vernetzung und "Marketing" zunehmend wichtiger

- Hinweise, dass Teams sowohl "bessere" Resultate erzielen ("Teamwork") als auch stärkere Verbreitung ihrer Forschung ("Netzwerk")
- Empirische Hinweise, dass Netzwerkmechanismus erklärungskräftiger anhand starker Effekte der Anzahl Referenzen ("out-degree") auf Zitate ("in-degree")

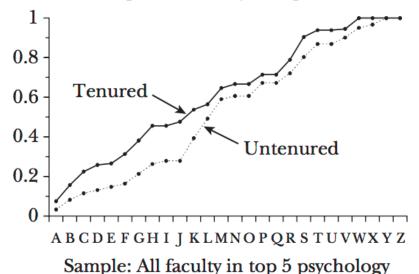
Implikation 2: Normen der Zusammenarbeit und Konfliktlösungen werden immer wichtiger

- Ungleiche Karrierechancen von Autoren mit spätem Nachnamen
- Fehleinschätzungen der Normen der Autorenreihung kreieren Konfliktpotential
- Unterschiedliche Kriterien für Leistung bei heterogenen Autorengruppen hinsichtlich Geschlecht, Disziplin und Karriereposition generieren Konfliktpotential

Unintendierte kollektive Konsequenzen

Z-Autoren schlechtere Karrierechancen?

Sample: All faculty in top 5 econ



O.6

O.4

O.2

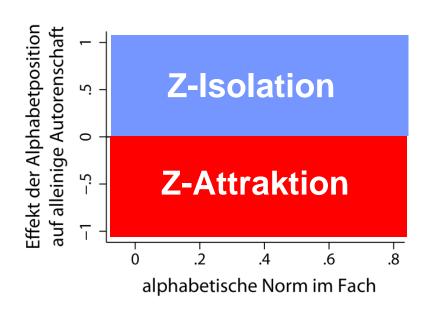
A B C D E F GH I J K L MN O P QR S T U V W X Y Z

Einav & Yariv 2006

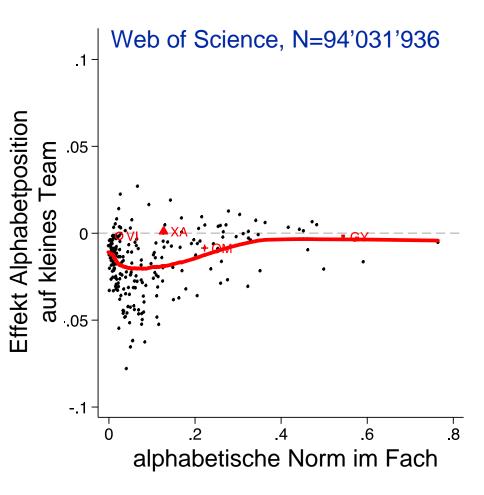
Z-Hypothesen:

da sie weniger Ko-Autoren finden, ohne in «Et.-Al-Falle» zu tappen

Attraktion: Z-Autoren werden häufiger als Ko-Autoren angefragt



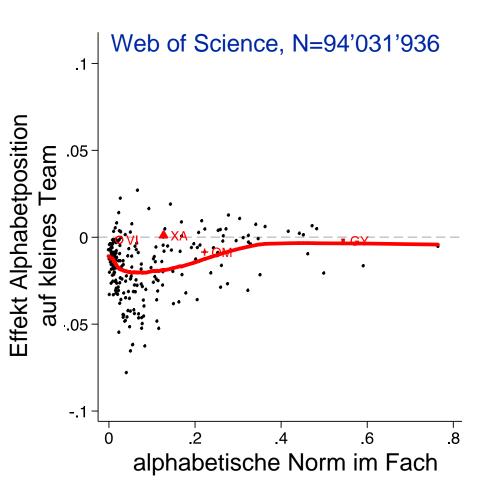
A-Autoren häufiger alleine!



241 logistische Regressionen (für jedes Fach)

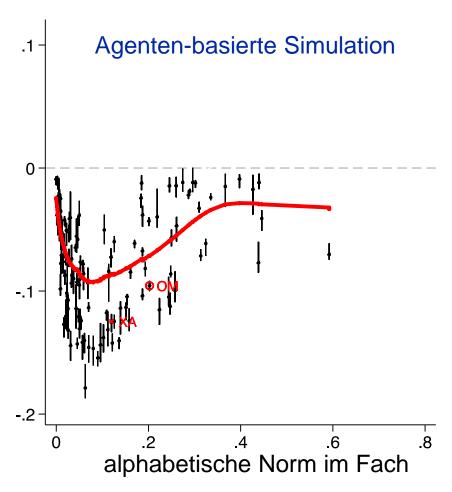
Rauhut & Winter, work in progress

A-Autoren häufiger alleine!



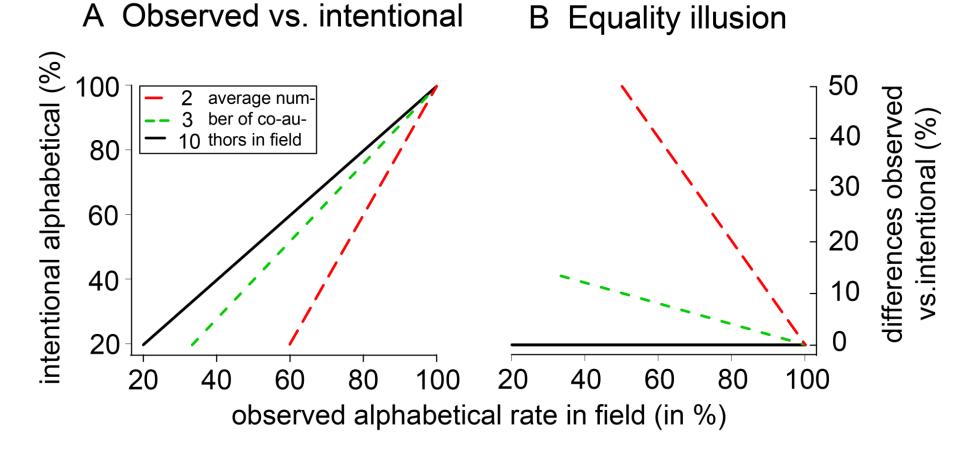
241 logistische Regressionen (für jedes Fach)

Rauhut & Winter, work in progress



- Ziehe zufällig Autor vom Web of Science
- Bestimme Teamgrösse von emp. Verteilung
- Bestimme Autorennorm mit p= Norm im Fach
- Alpha-Norm: Suche unbeschäftigten Z
- Nicht-Alpha: Suche irgendeinen n
- Wenn keiner übrig, suche kleineres Team
- Wiederhole bis jeder arbeitet
- Wiederhole Simulation 50 Mal

Equality illusion for 2, 3 and 10 average co-authors



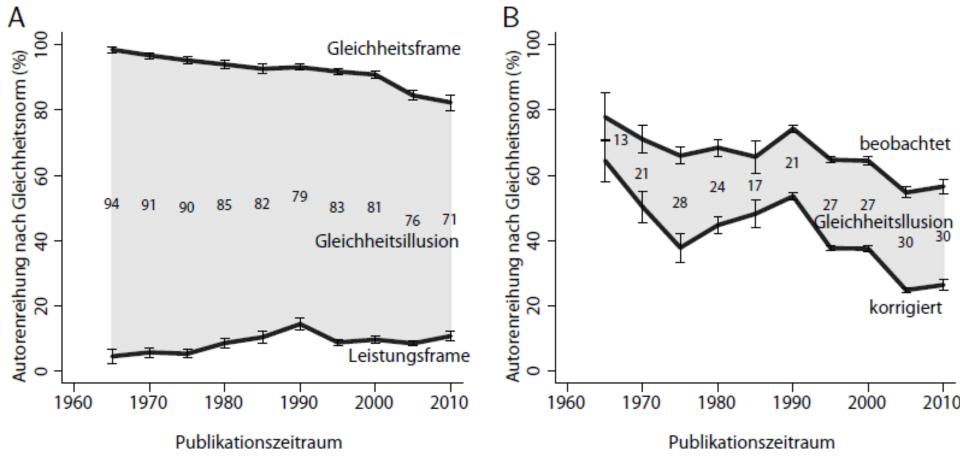
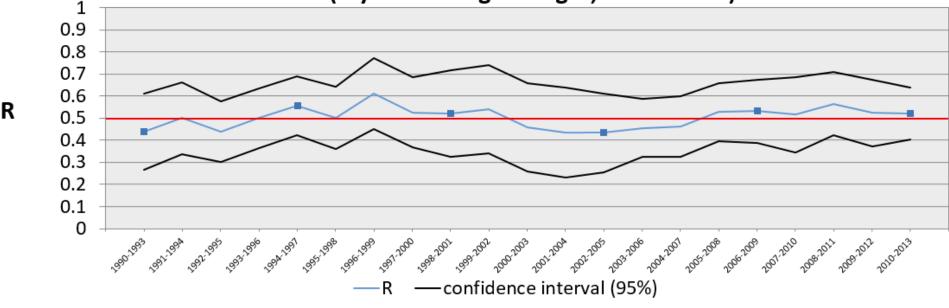


Abbildung 3. Gleichheitsillusion bei der Autorenreihenfolge mit Mechanismen "Framing" (A) und statistischer Fehlschluss (B). Alle Fehlerbalken sind 95% Bootstrap Konfidenzintervalle (bei Nicht-Überlappung entsprechen die Unterschiede einem Signifikanztest auf 5% Niveau). Prozentzahlen in grauen Flächen geben jeweils das prozentuale Ausmaß der Gleichheitsillusion im Fünfjahresblock an (als Differenz zwischen oberer und unterer Linie). (A) Empirisch möglicher Interpretationsspielraum zwischen Vertretern von Gleichheits- vs. Leistungsnorm bei gleicher Datenlage (Framing Mechanismus). Die obere Linie zeigt die mit der Gleichheitsnorm konsistente Rate von Artikeln (Einzelautorenschaften und alle alphabetisch gereihten Artikel). Die untere Linie zeigt die mit der Leistungsnorm konsistente Rate (Einzelautorenschaften und nach zufällig auftretender alphabetischer Reihung korrigierte Rate). (B) Ausmaß der Gleichheitsillusion unter statistischem Fehlschluss. Die obere Linie repräsentiert die unkorrigierte Rate alphabetisch gereihter Artikel, die untere Linie die nach zufälliger Reihung korrigierte Rate.

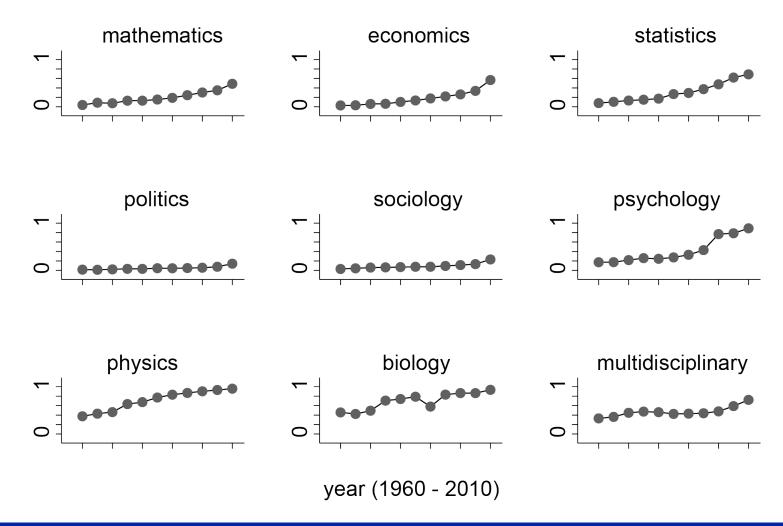
Substitution von Team and Netzwerkeffekten (A&B) and Dominanz des Netzwerkeffekts (C)

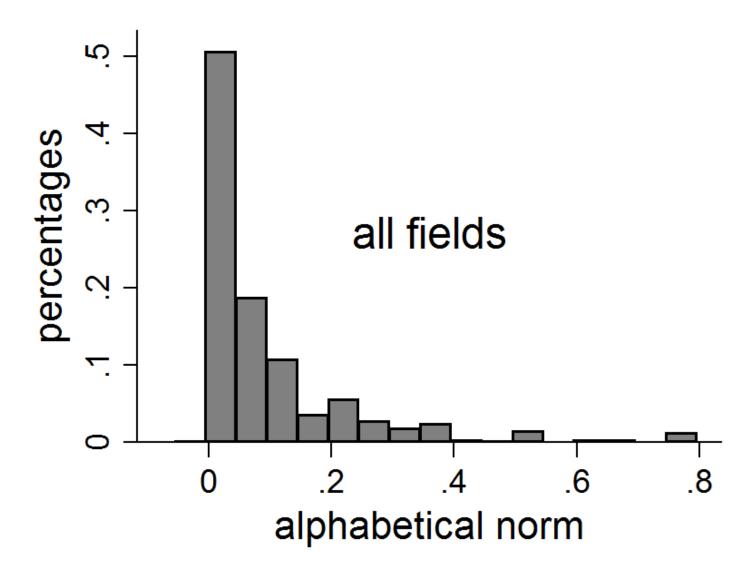
R over time for the 0.05 significance level, 10%-caliper (4-year moving averages, 1990 - 2013)



rate of multi-authored papers

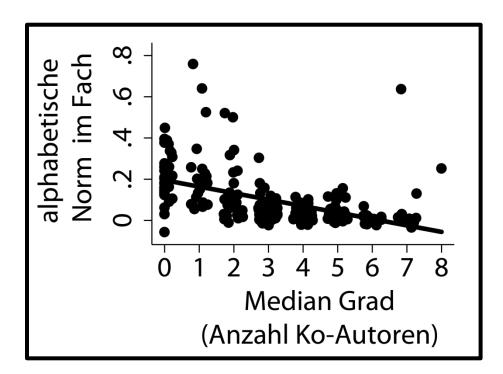
(Web of Science, 4,922,445 articles, weighted by number of authors)



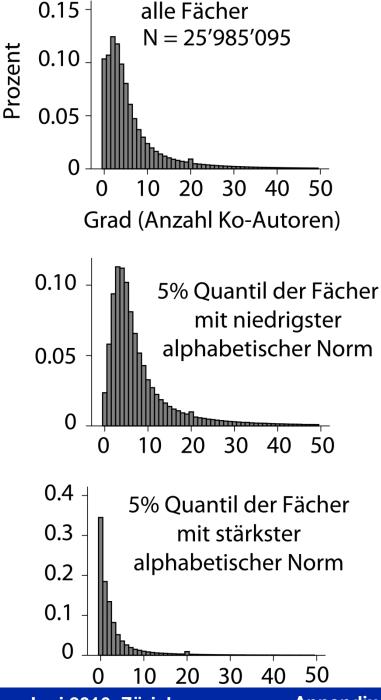


N = 241 fields



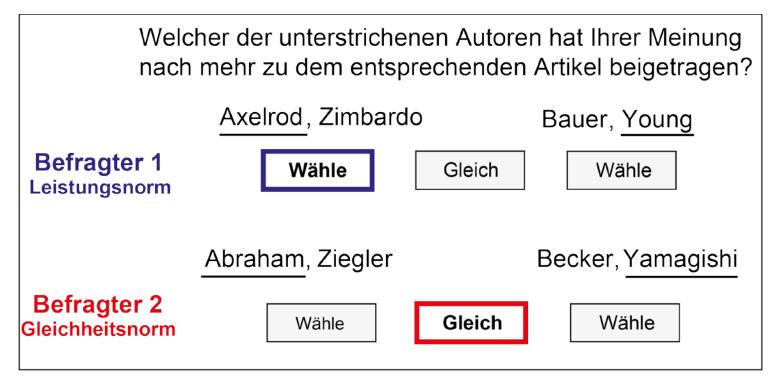


Vernetzungsgrad in Fächern erklärt Entstehung von Normheterogenität





- Sekundäranalyse der Umfragedaten von Maciejovsky, Budescu & Ariely (2008)
- Befragte: Doktoranden bis Professoren aus Ökonomie, Management und Psychologie



- Simulation von 52'809 vergleichbaren Autorenpaaren anhand 351 Befragter mit je etwa 30 Vignetten
- Normenkonflikt: Unterschiedliche Leistungszuschreibung bei gleicher Autorenkonstellation

Tabelle 1: Logistische Mehrebenenmodelle (odds ratios)

	(1)	(2)	(3)
	Konflikt	Konflikt	Konflikt
Fach unterschiedlich	1.292***	1.247***	1.229***
	(8.00)	(6.79)	(6.31)
Geschlecht unterschiedlich		1.210***	1.199***
		(5.86)	(5.58)
Karriereposition unterschiedlich			1.151***
			(3.53)
Konstante	0.744***	0.680***	0.616***
	(-12.24)	(-13.50)	(-12.08)
Standardabweichung der Konstanten	1.164***	1.157^{***}	1.154***
	(8.76)	(8.39)	(8.25)
\overline{N}	52809	52809	52809

Odds ratios; t-Werte in Klammern, * p < 0.05, ** p < 0.01, *** p < 0.001



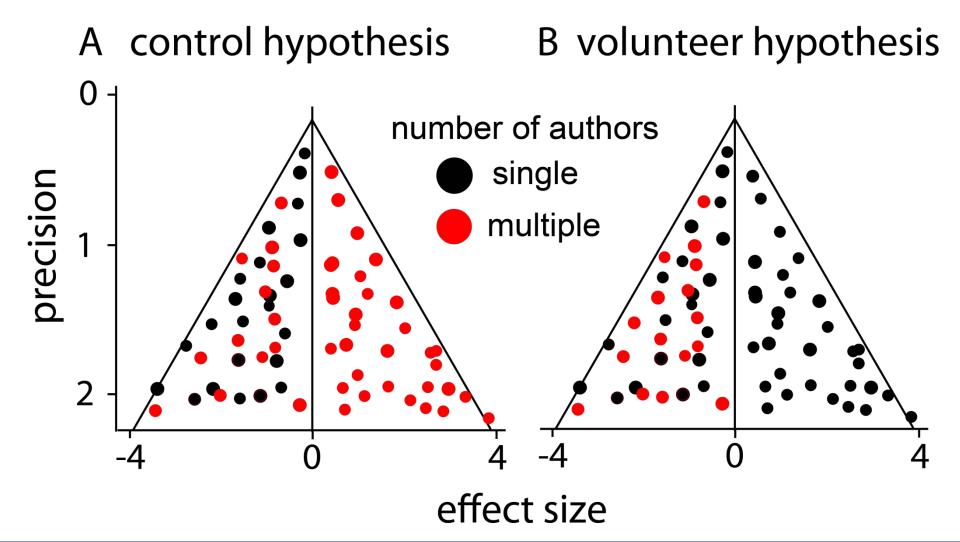
Social Value Orientation and Network Formation in Swiss Schools

More norm violations (misconduct) in teams?

problem: desirability, detection method: meta-analyses

volunteer vs. control hypothesis	larger teams → more or less misconduct?
data collection	regression coefficients and tests from published articles by team size
fraud detection	compare empirical distribution of coefficients with theoretical distribution
confirmation	larger deviations in team than solo articles
operationalization	fake data (e.g. deviation from Benford distribution)

Schematic funnel plots of hypothetical data



Benford distribution of first digits

