Embracing Complexity in the Science of Science

Dissertation Defense for **Dakota Murray**

PhD in Informatics, Computing, Culture, and Society

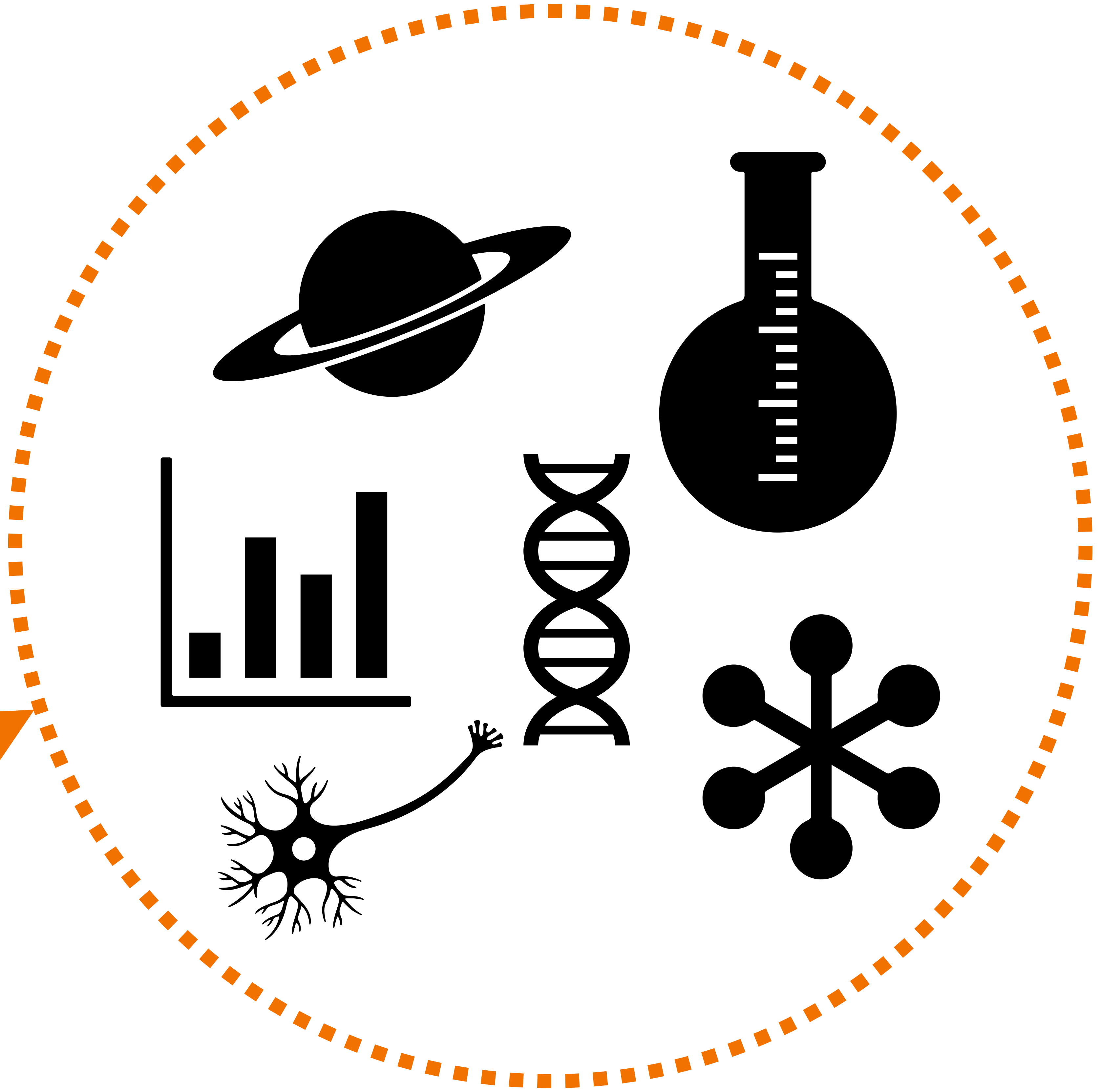
Minor in Statistics

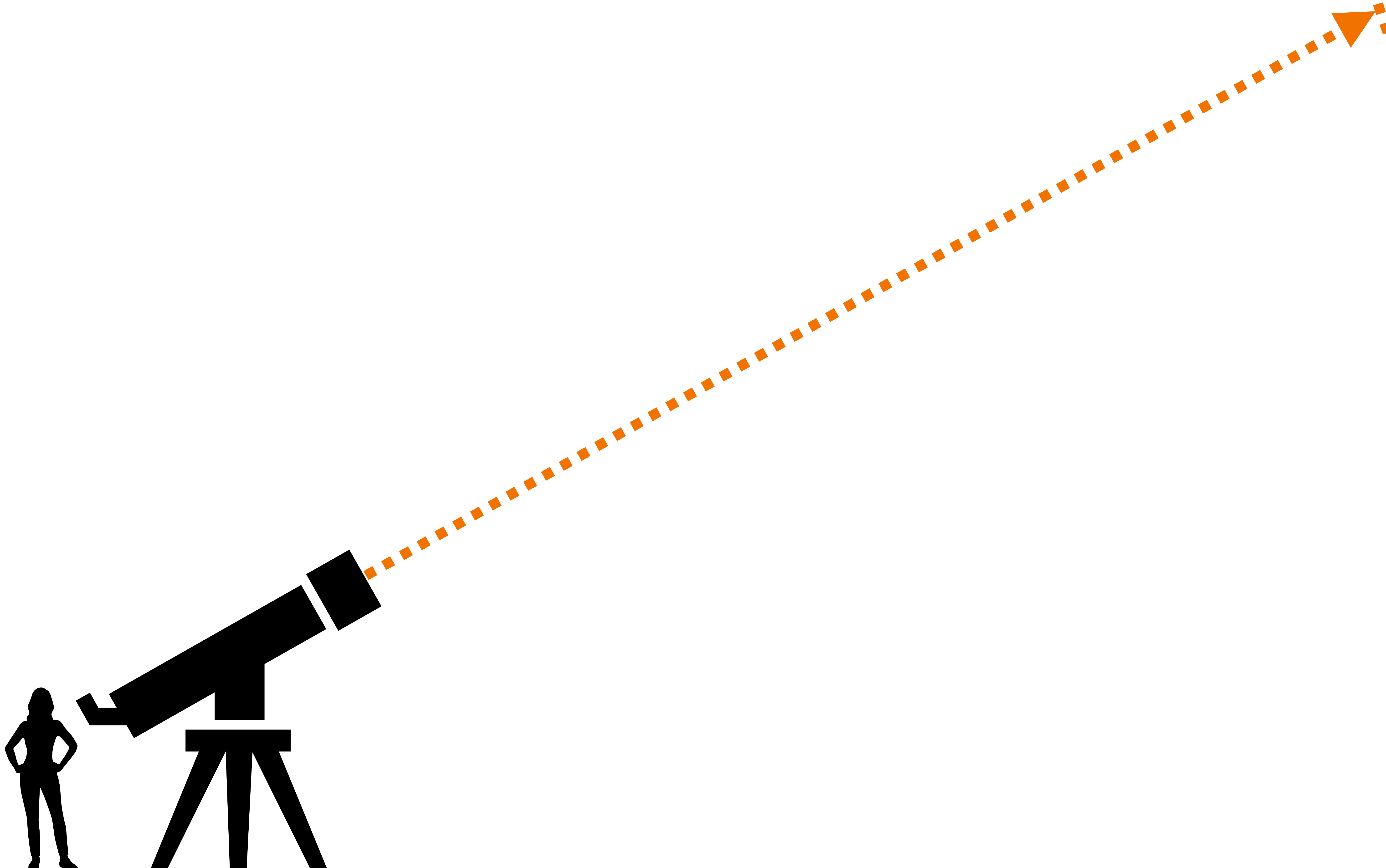
INDIANA UNIVERSITY
SCHOOL OF INFORMATICS, COMPUTING, AND ENGINEERING

23rd August, 2021

Science of Science

Turning the tools of science upon itself





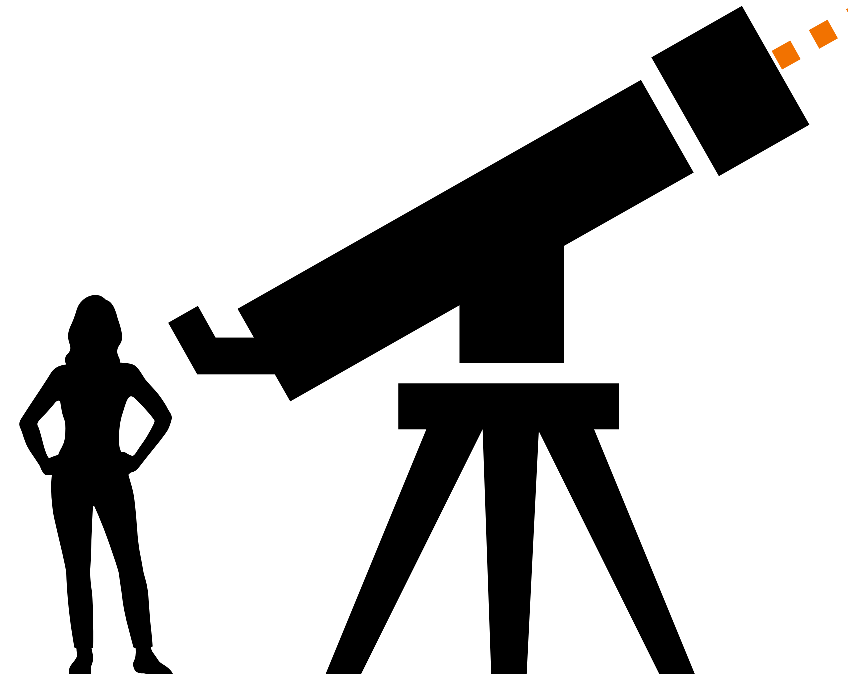
Fortunato, S., Bergstrom, C. T., Börner, K., Evans, J. A., Helbing, D., Milojević, S., Petersen, A. M., Radicchi, F., Sinatra, R., Uzzi, B., Vespignani, A., Waltman, L., Wang, D., & Barabási, A.-L. (2018). Science of science. *Science*, *359*(6379), eaao0185.

Science of Science

Turning the tools of science upon itself

Leverages massive data sets, mathematical modelling, and computational tools

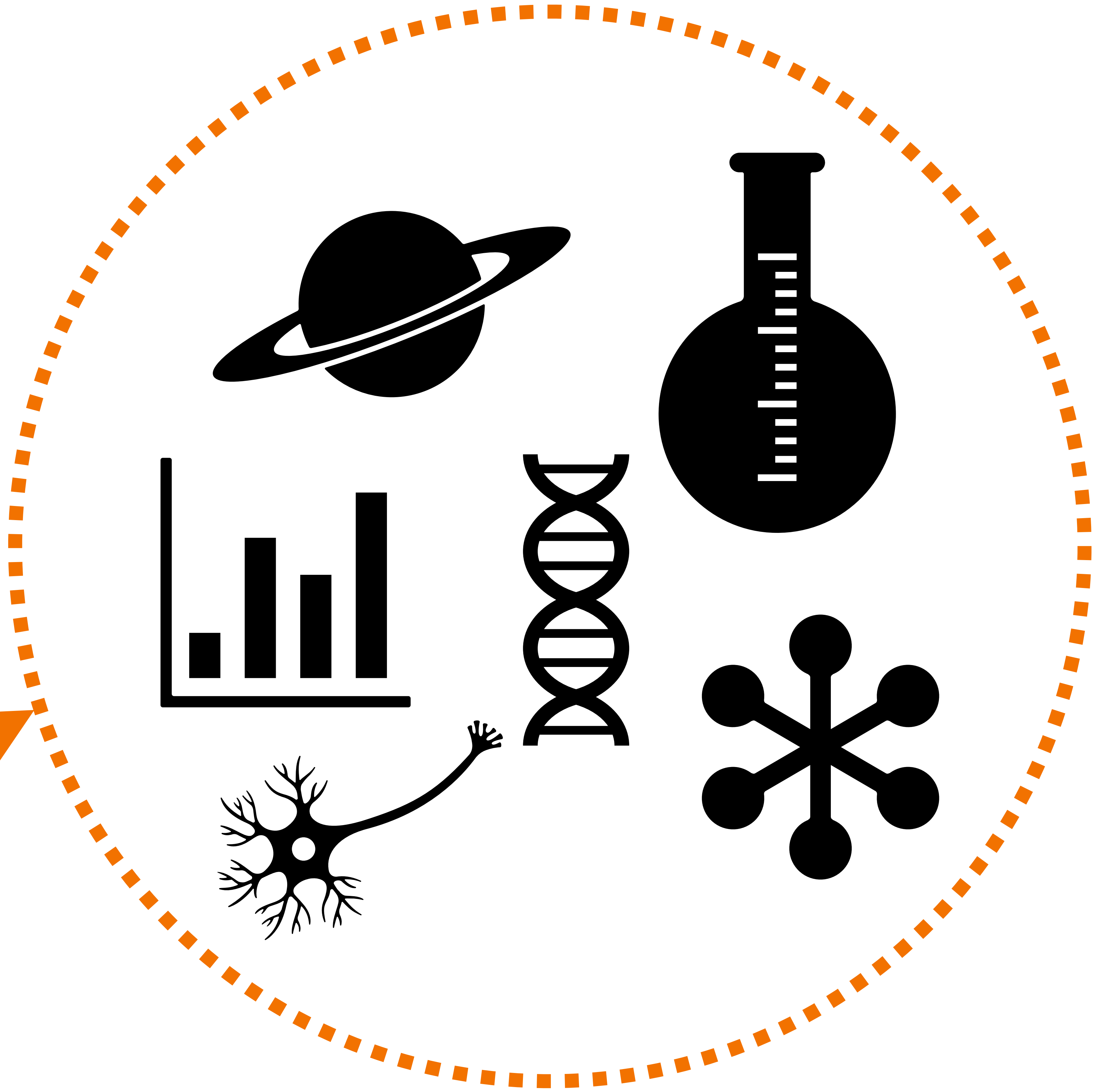
Understand the processes underpinning how science works



Fortunato, S., Bergstrom, C. T., Börner, K., Evans, J. A., Helbing, D., Milojević, S., Petersen, A. M., Radicchi, F., Sinatra, R., Uzzi, B., Vespignani, A., Waltman, L., Wang, D., & Barabási, A.-L. (2018). Science of science. *Science*, *359*(6379), eaao0185.

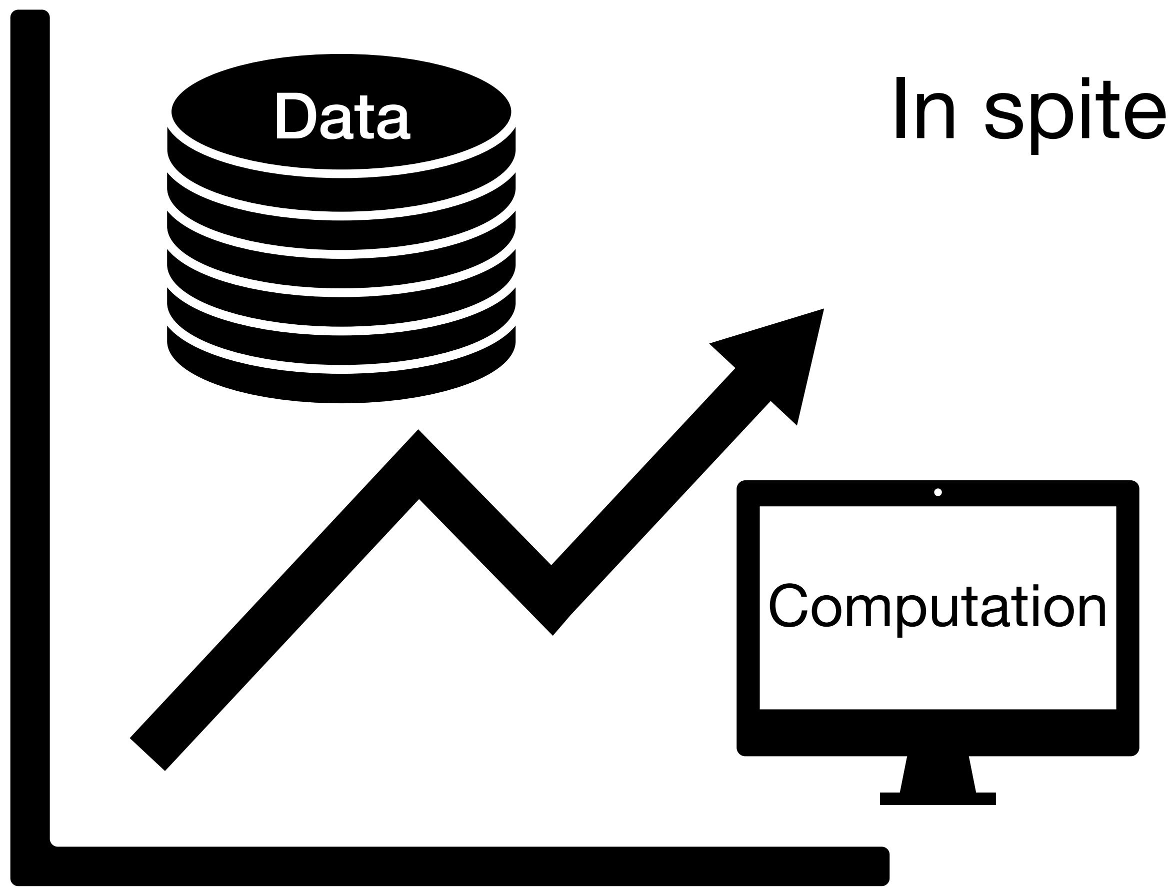
Science of Science

Turning the tools of science upon itself



Leverages massive data sets, mathematical modelling, and computational tools

Understand the processes underpinning how science works



In spite of technical advances, challenges remain

Fortunato, S., Bergstrom, C. T., Börner, K., Evans, J. A., Helbing, D., Milojević, S., Petersen, A. M., Radicchi, F., Sinatra, R., Uzzi, B., Vespignani, A., Waltman, L., Wang, D., & Barabási, A.-L. (2018). Science of science. *Science*, *359*(6379), eaao0185.

Science is ***big***, and getting ***bigger***

There are more publications, and more people, than before

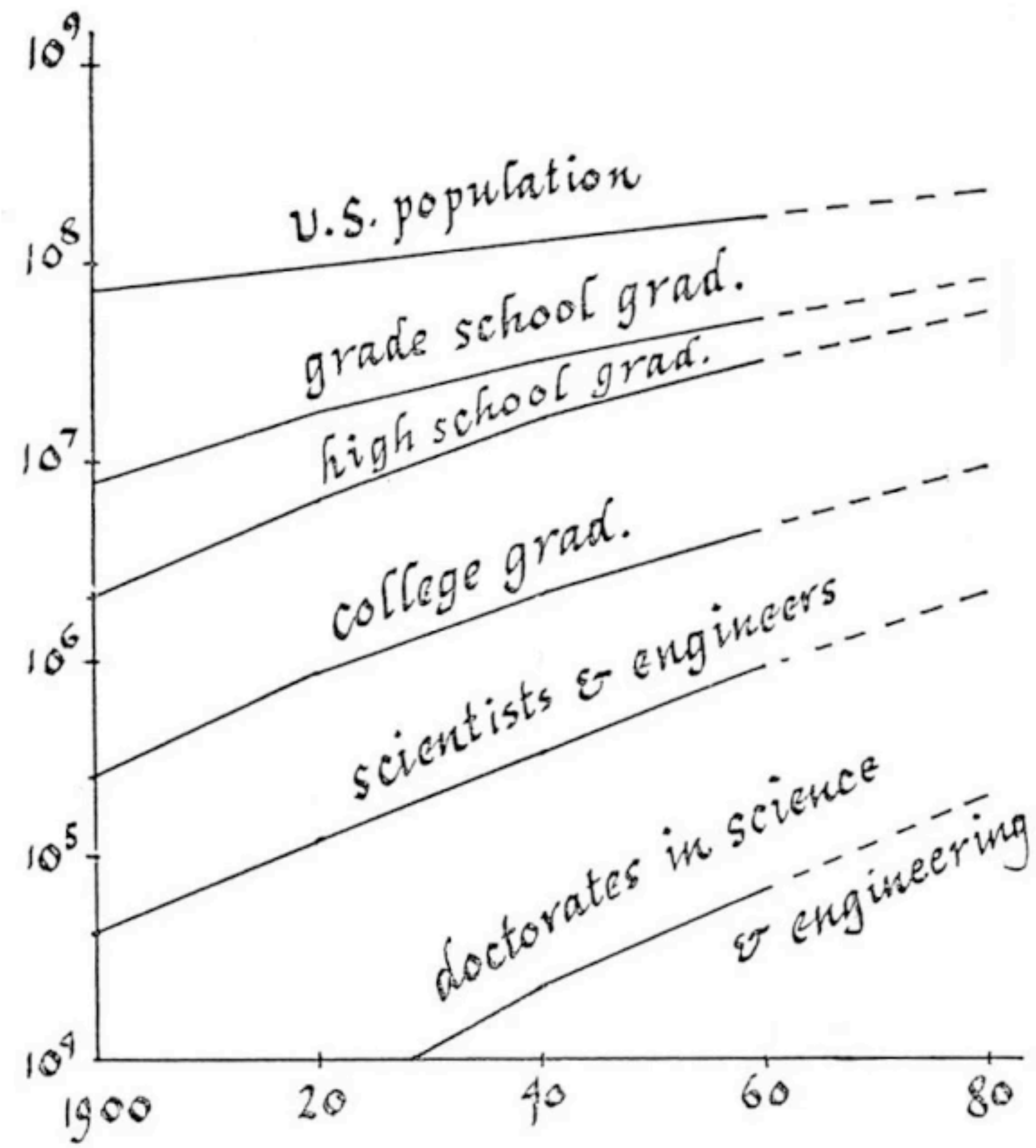
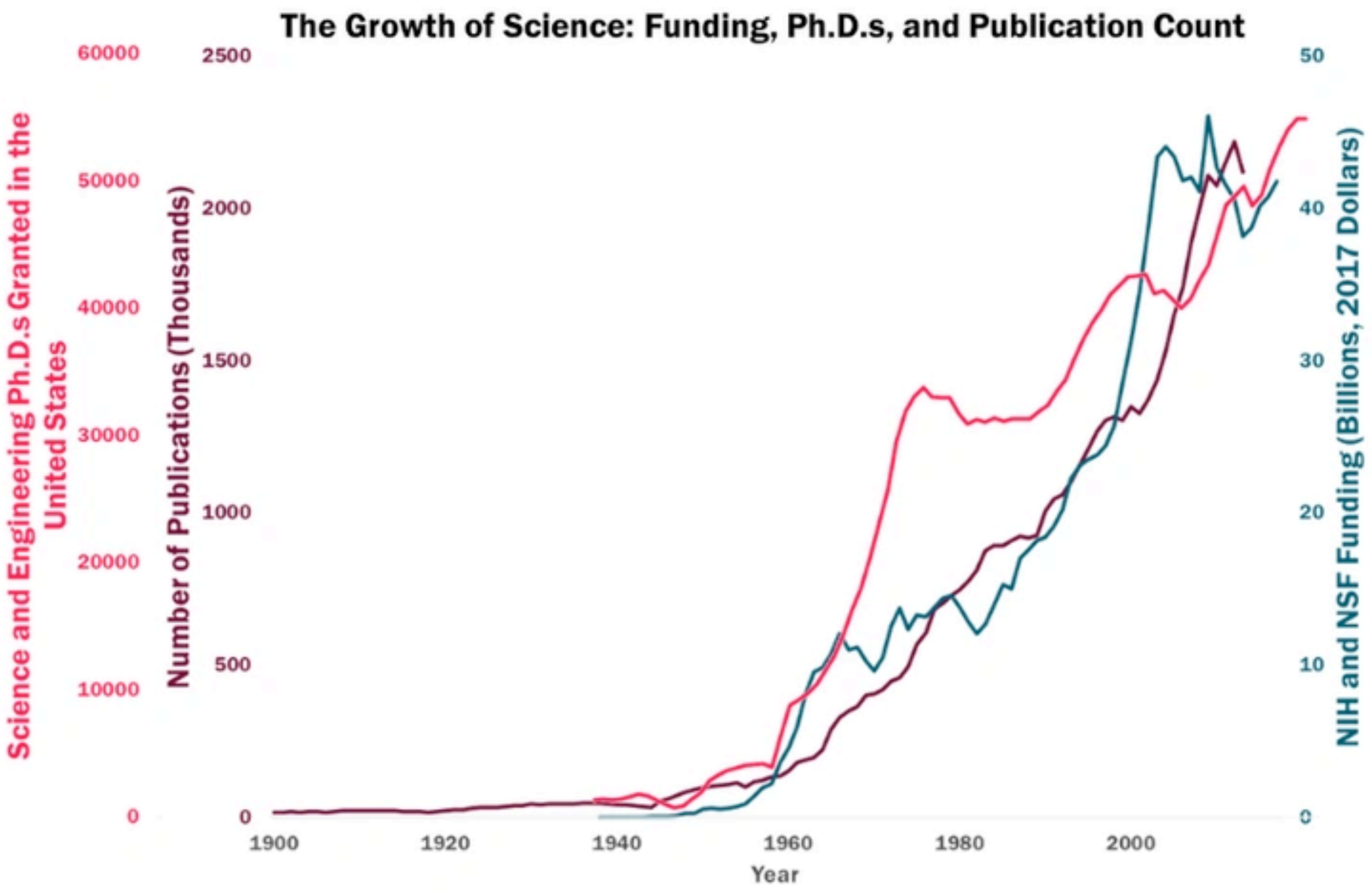


Fig. 3. GROWTH OF SCIENTIFIC MANPOWER AND OF GENERAL POPULATION IN THE UNITED STATES

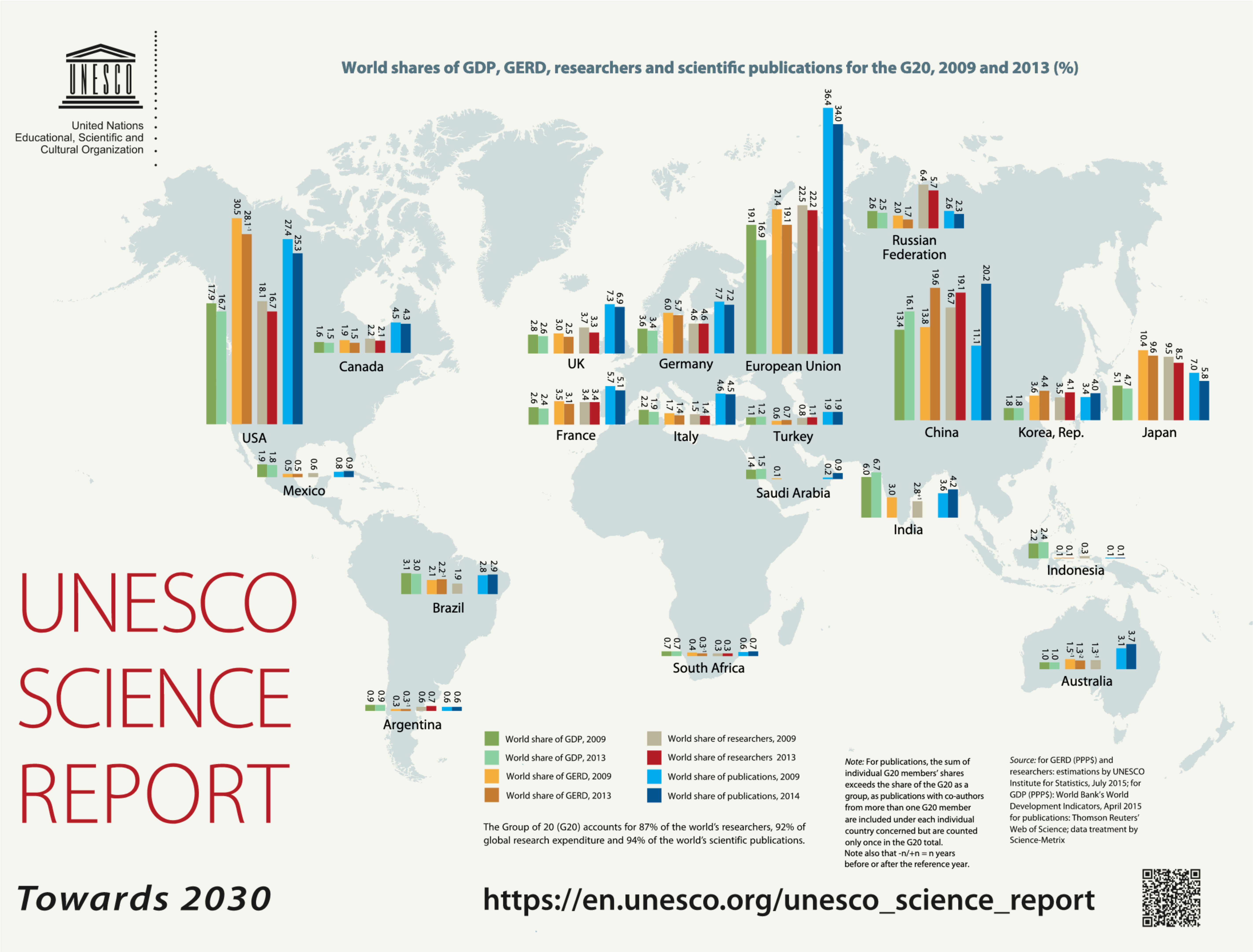


de Solla Price, D. (1963). *Little Science Big Science* (1st ed.). Columbia University Press.

Collison, P., & Nielsen, M. (2018, November 16). Science Is Getting Less Bang for Its Buck. *The Atlantic*.

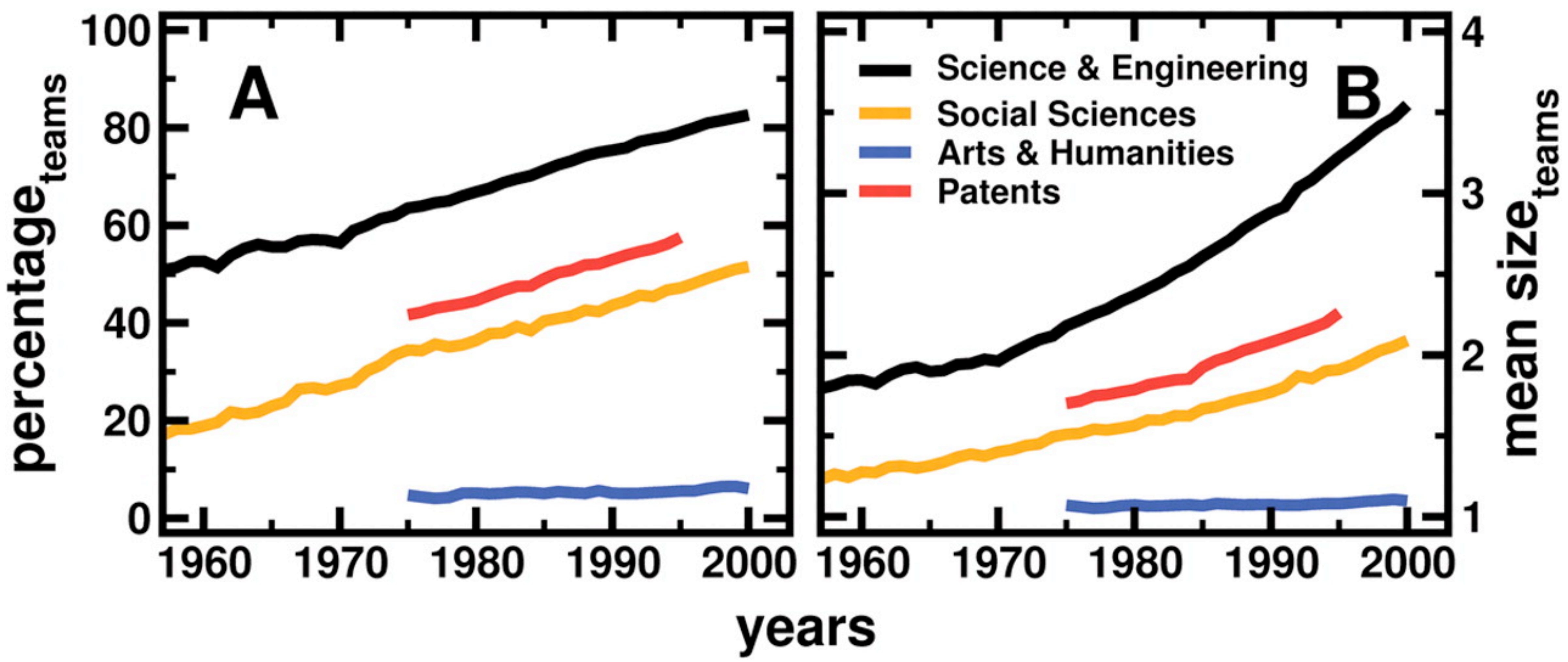
Science is more ***heterogeneous***

More different people, from different backgrounds, contributing to science

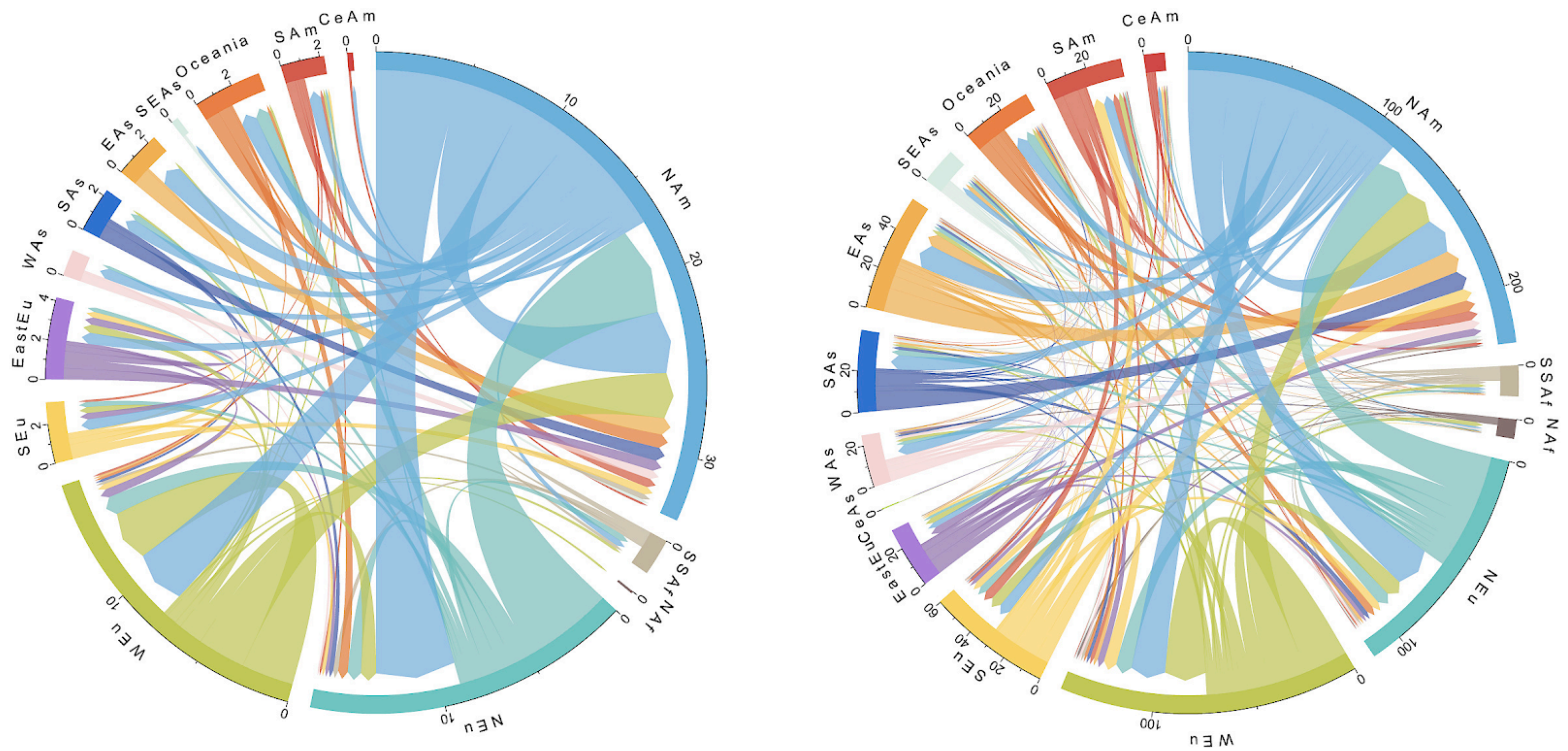


Science is more ***interconnected***

Individuals connected through collaboration, mobility



Wuchty, S., Jones, B. F., & Uzzi, B. (2007). The Increasing Dominance of Teams in Production of Knowledge. *Science*, *316*(5827), 1036–1039.



Czaika, M., & Orazbayev, S. (2018). The globalisation of scientific mobility, 1970–2014. *Applied Geography*, *96*, 1–10.

Science is *complex*

Science is *complex*

Challenges traditional approaches to its study

Science is *complex*

Challenges traditional approaches to its study

Opportunity for new perspectives for making sense of science

Complexity Perspective

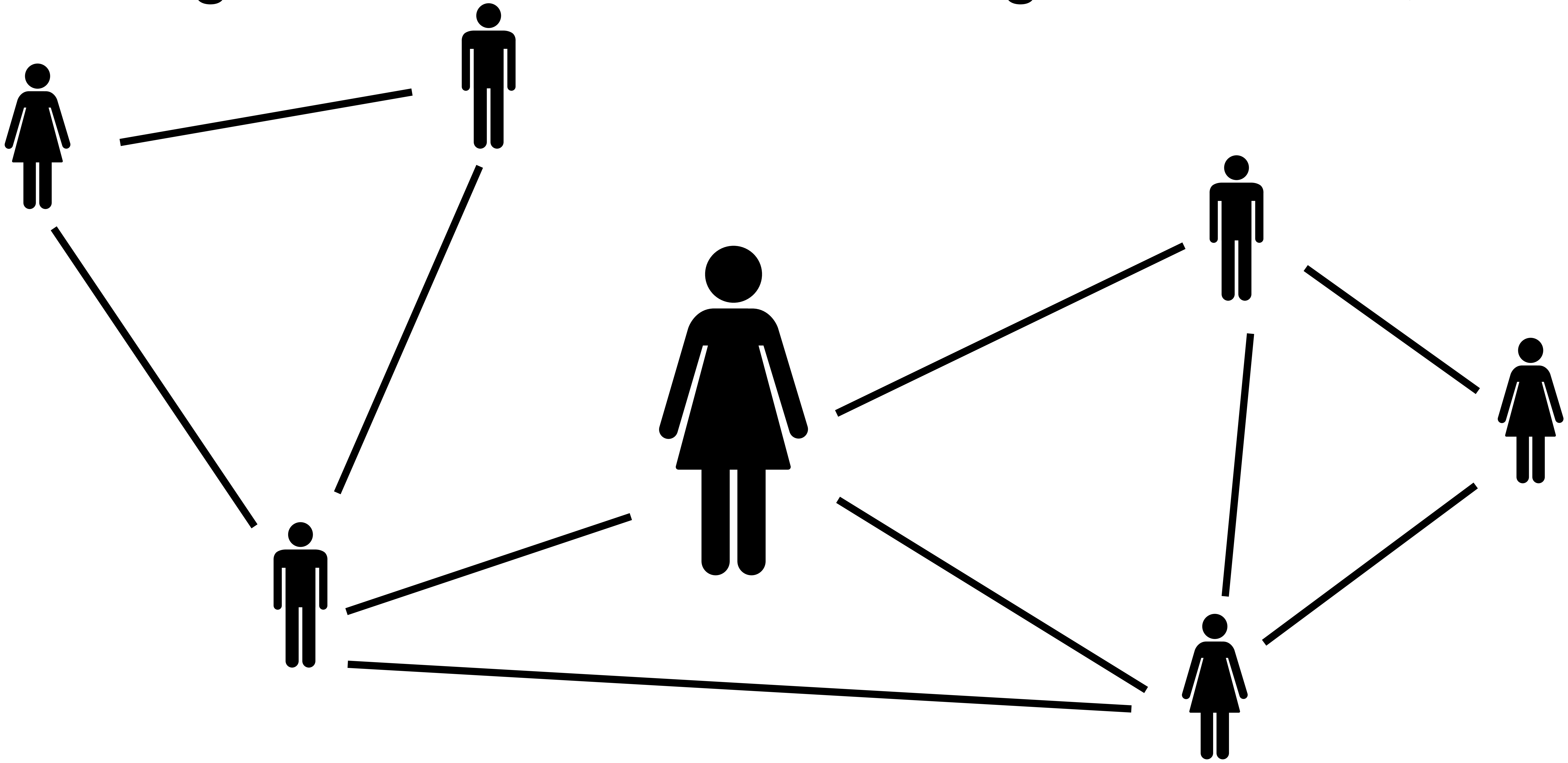
A sense-making framework for understanding science

- A complex network of individual scientists, connected via their various interactions
- Shifts focus towards the *everyday* actions of individual scientists
- Flexible, easily fit onto existing Science of Science ideas
- A new and powerful perspective for conceiving of and understanding science

Past approaches view scientists as isolated, singular, static

Each is embedded in a wider and dynamic social system

Connected through their different and shifting interactions,



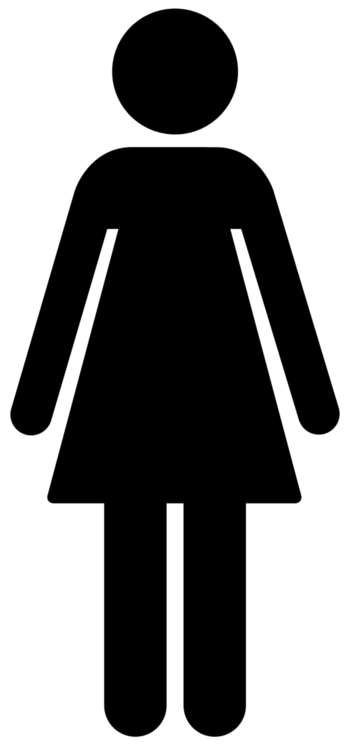
At they same time, they are complex individuals

With their own characteristics, making choices about their career

Gender

Location

Ethnicity



Interests

Economic Status

Personality

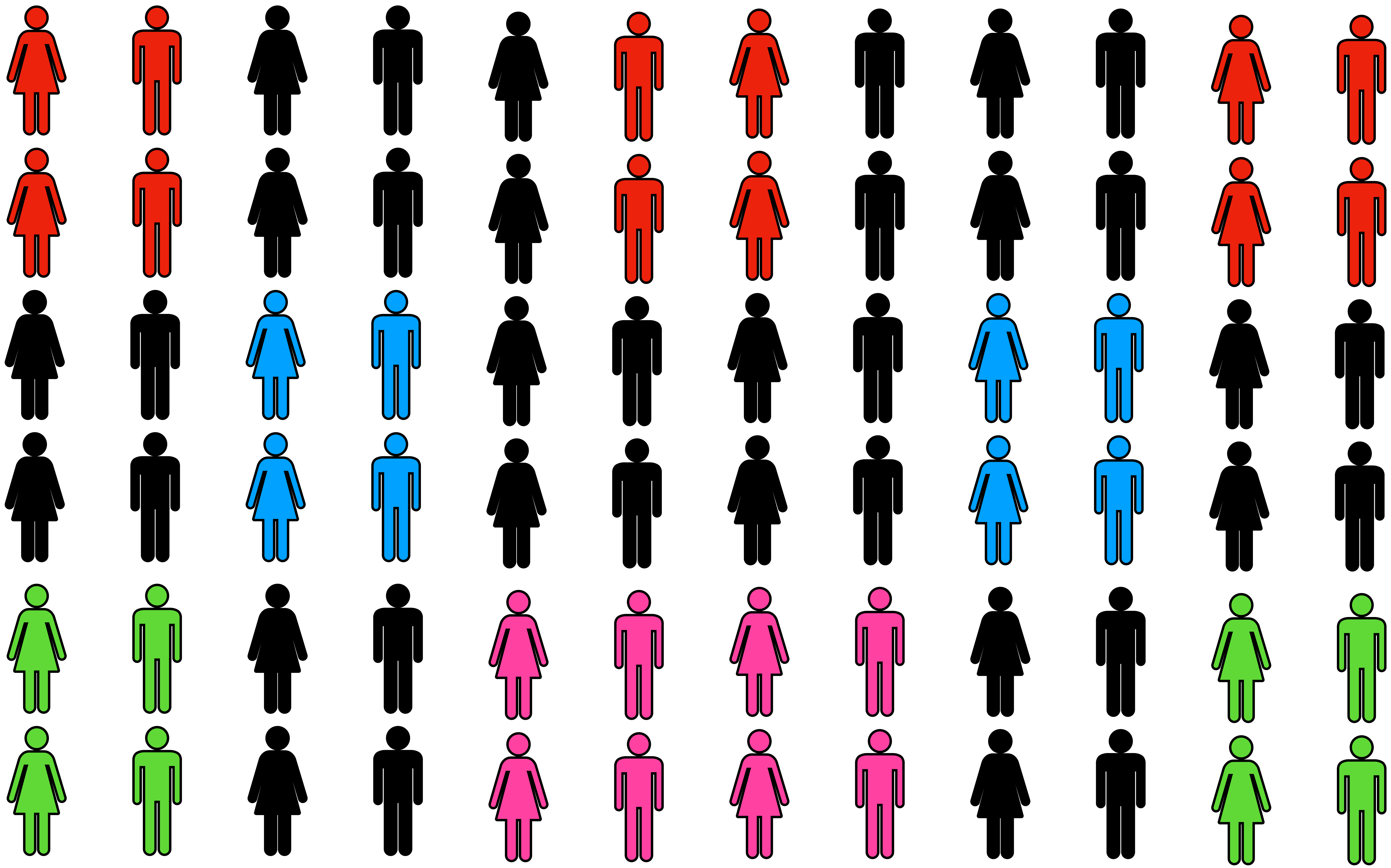
Nationality

Agency

Human Capital

Reputation

Bottom-up forces allow for self-organization



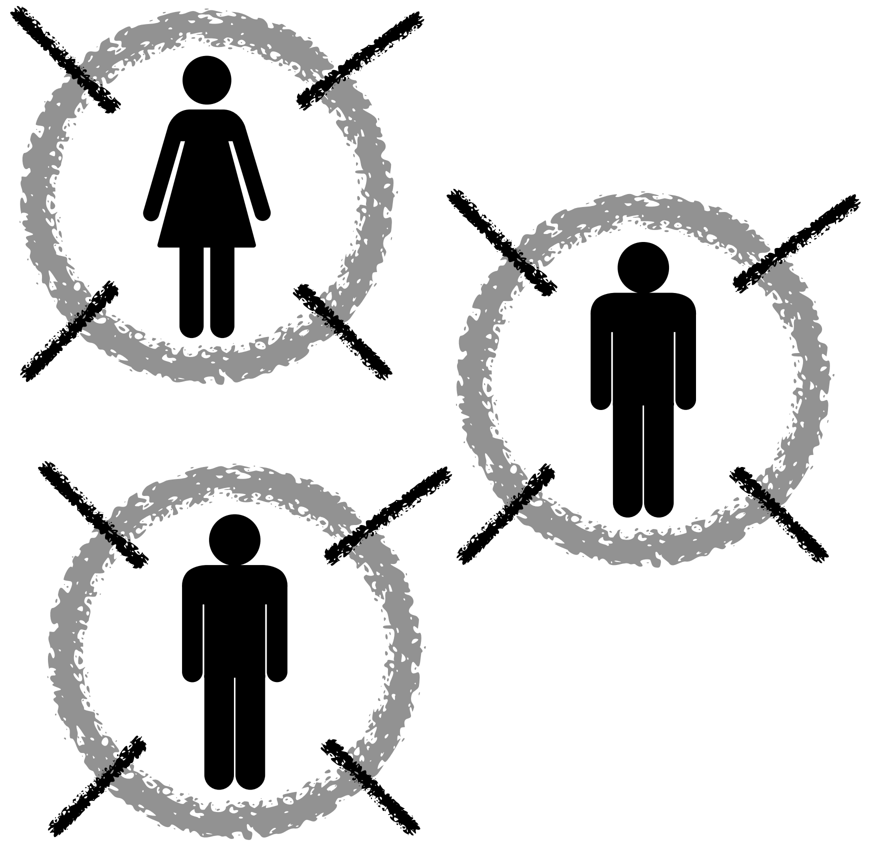
Bottom-up forces allow for self-organization

Even though each may lack knowledge of the broader whole

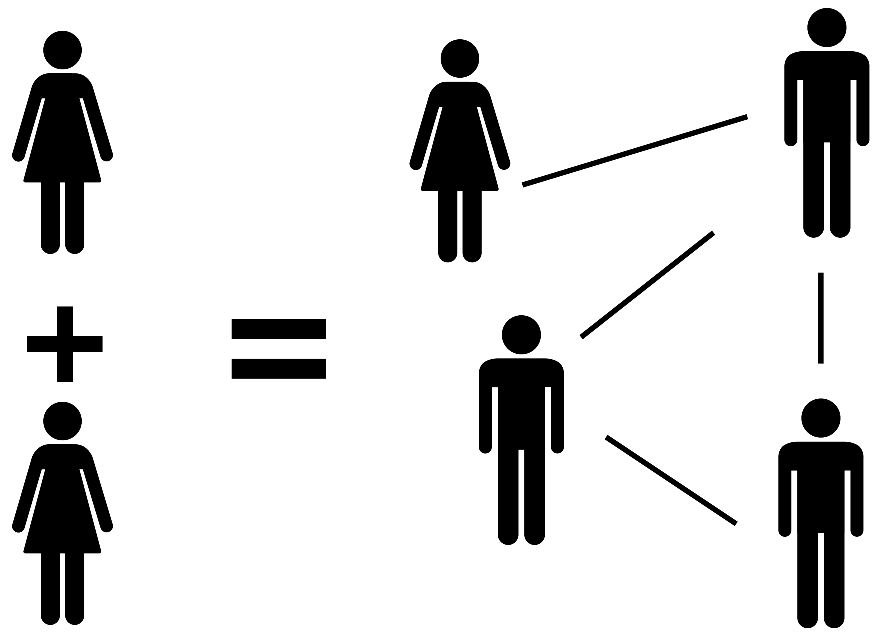
Concepts from complexity science

Can be employed as sense-making tools for understanding science

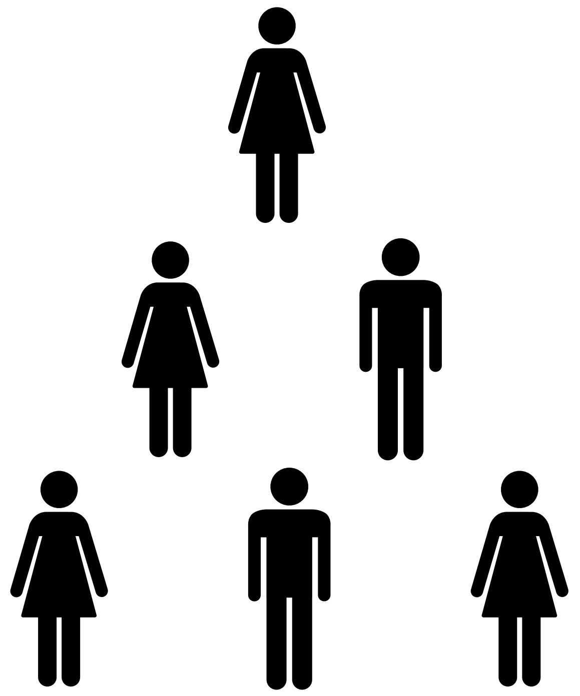
Autonomy



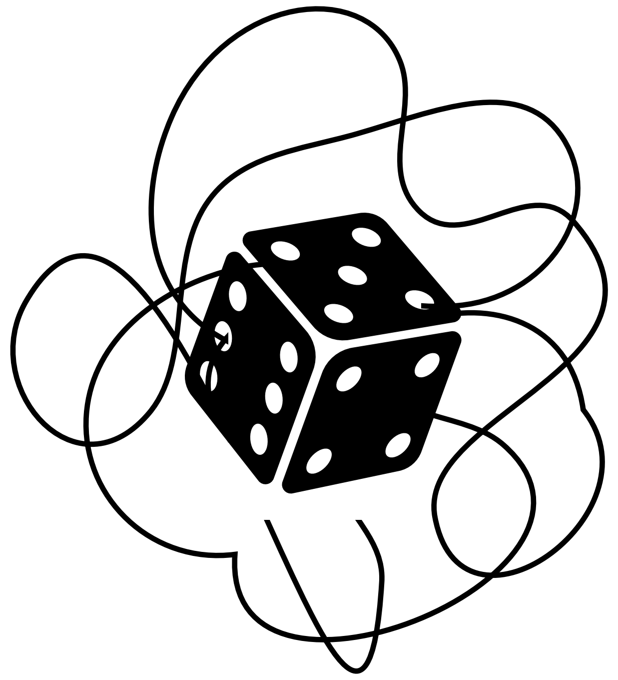
Non-linearity



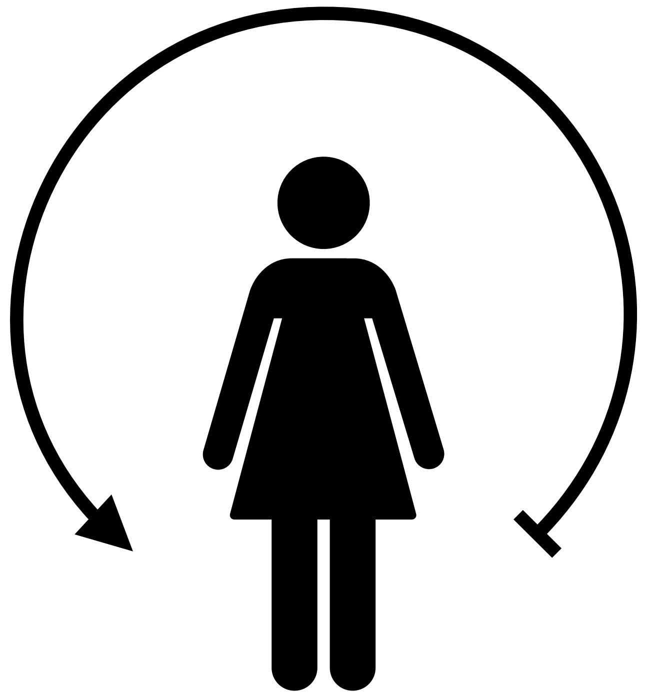
Hierarchical structure



Chaotic dynamics



Feedback

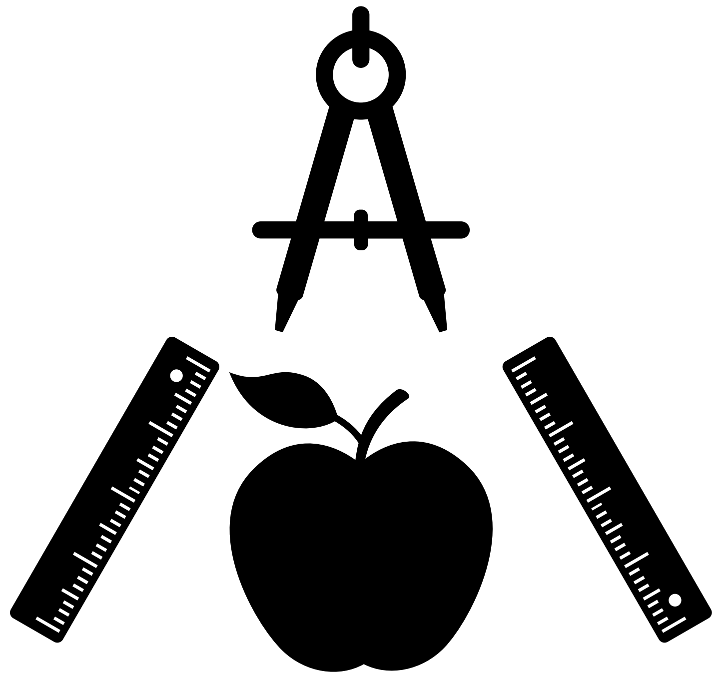


Structure of dissertation

Four studies examining topics relevant to the Science of Science

Peer review at *eLife*

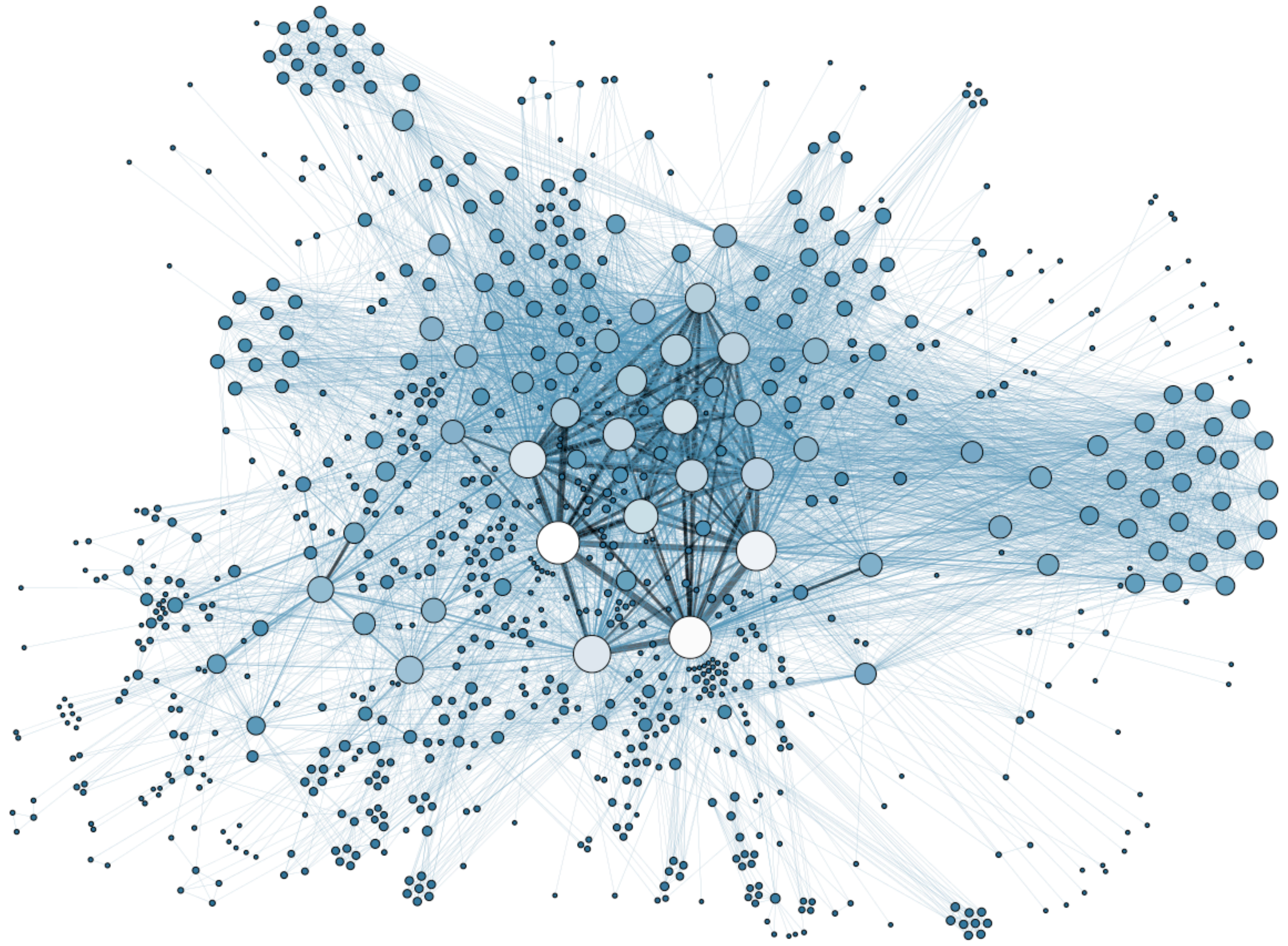
Disagreement in science



Student-teacher evaluations

Global scientific mobility

Interpret using the *<u>Complexity Perspective</u>*



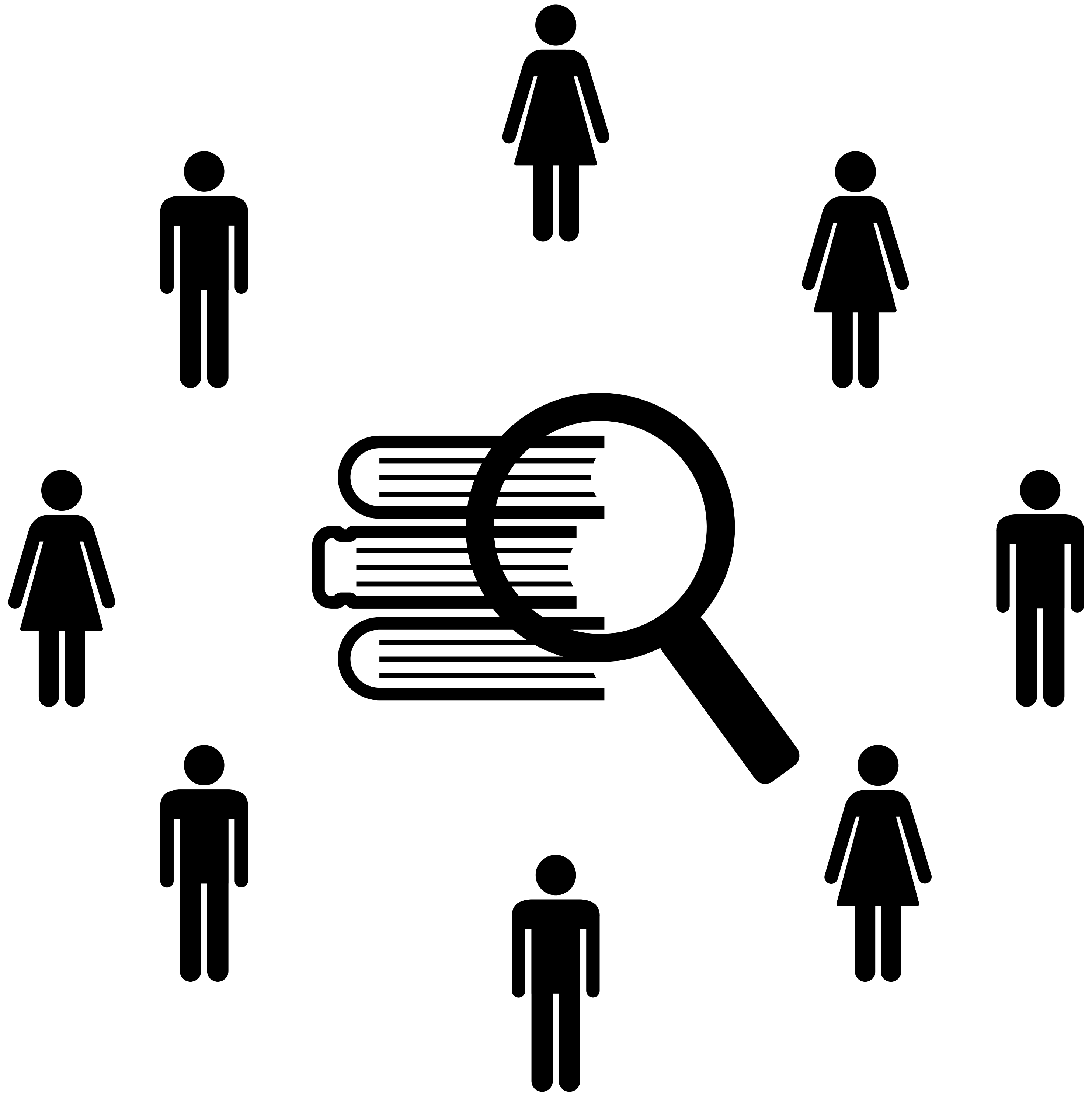
Study 1:

Peer review at eLife

Murray, D., Siler, K., Larivière, V., Chan, W. M., Collings, A. M., Raymond, J., & Sugimoto, C. R. (2019). Author-Reviewer Homophily in Peer Review. *BioRxiv*, 400515.

Peer review is a ubiquitous tool for evaluation in science

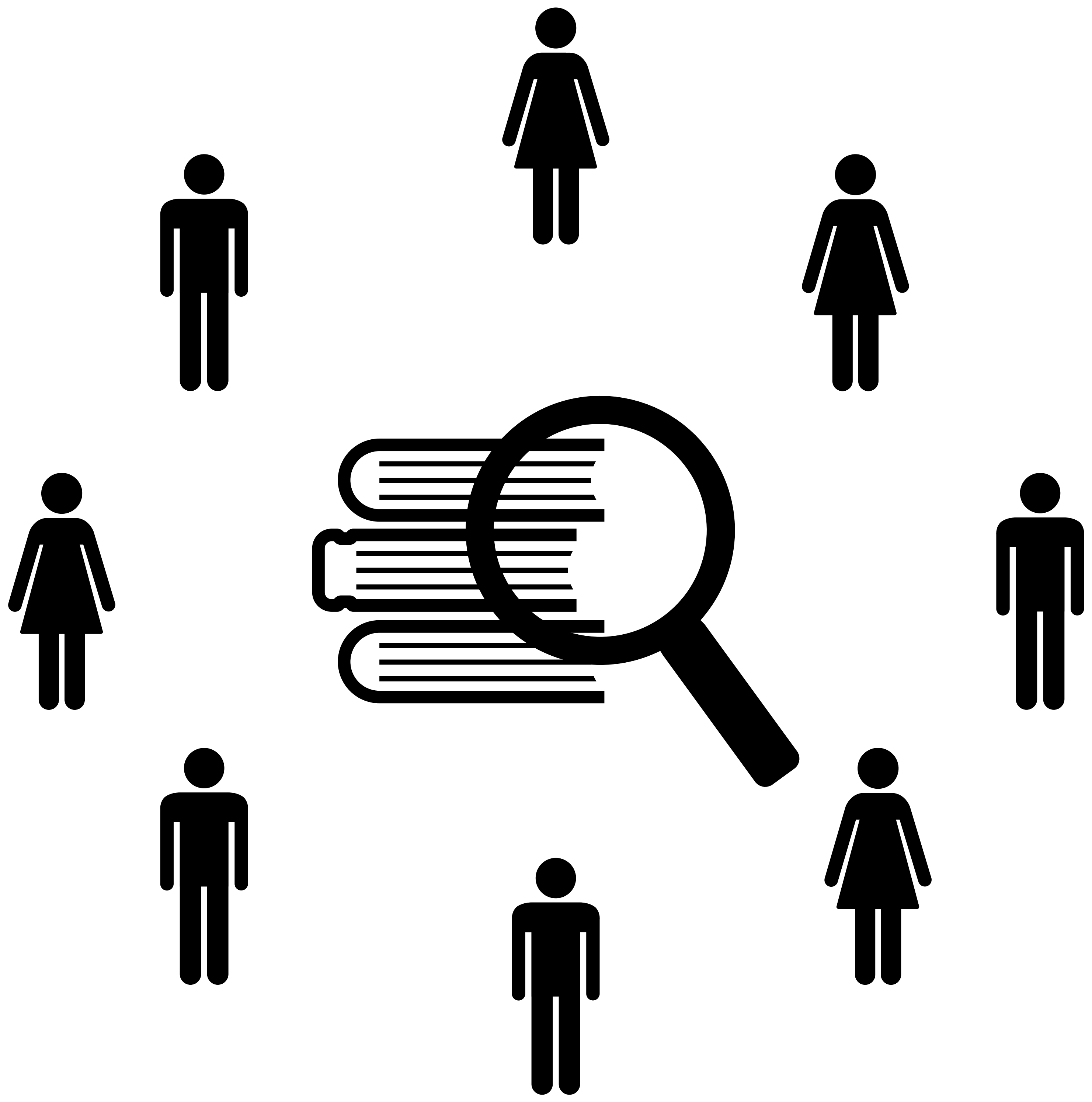
Supports hiring, promotion, editorial, and resource allocation decisions

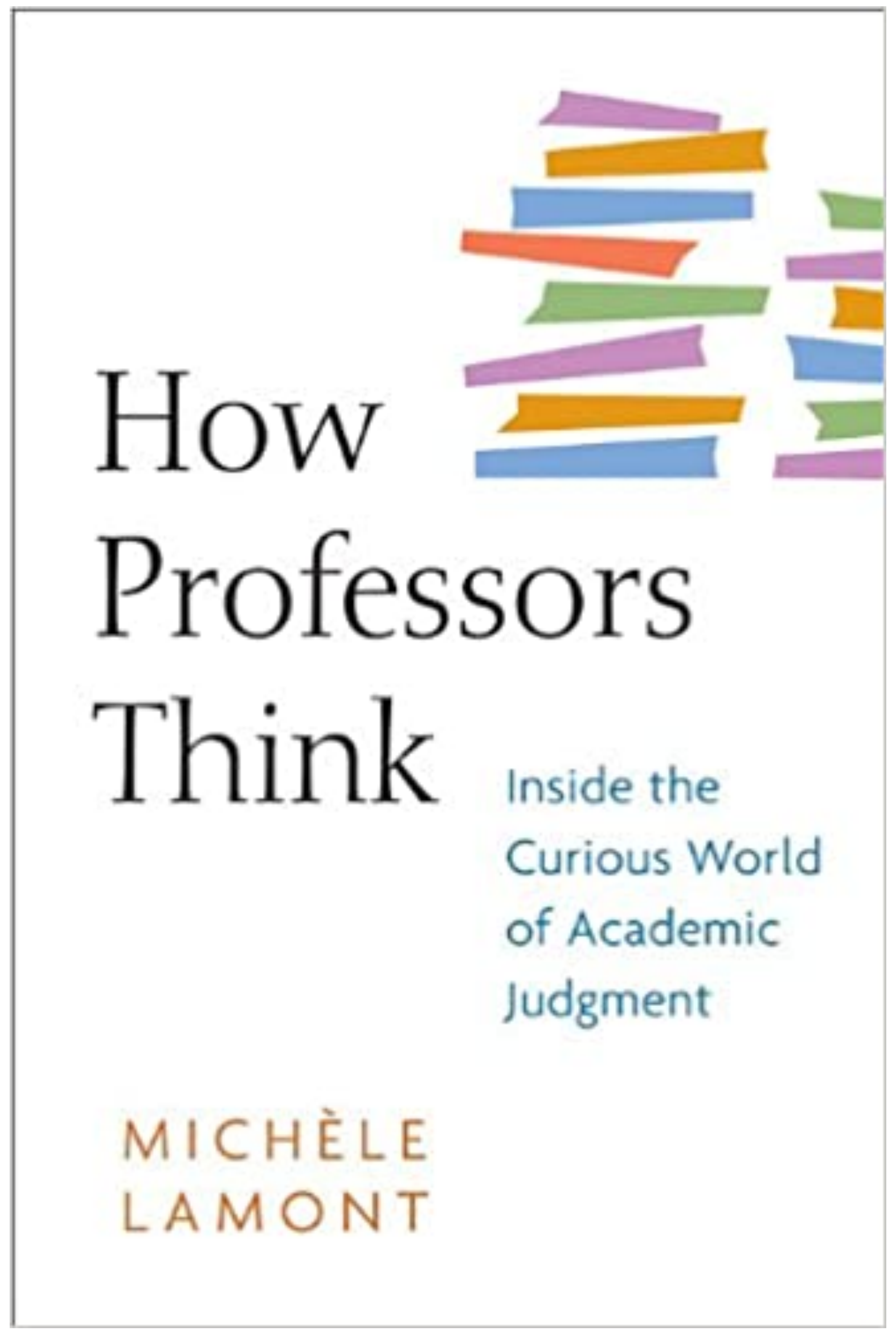


Peer review is a ubiquitous tool for evaluation in science

Supports hiring, promotion, editorial, and resource allocation decisions

But has also been challenged by claims of bias and subjectivity

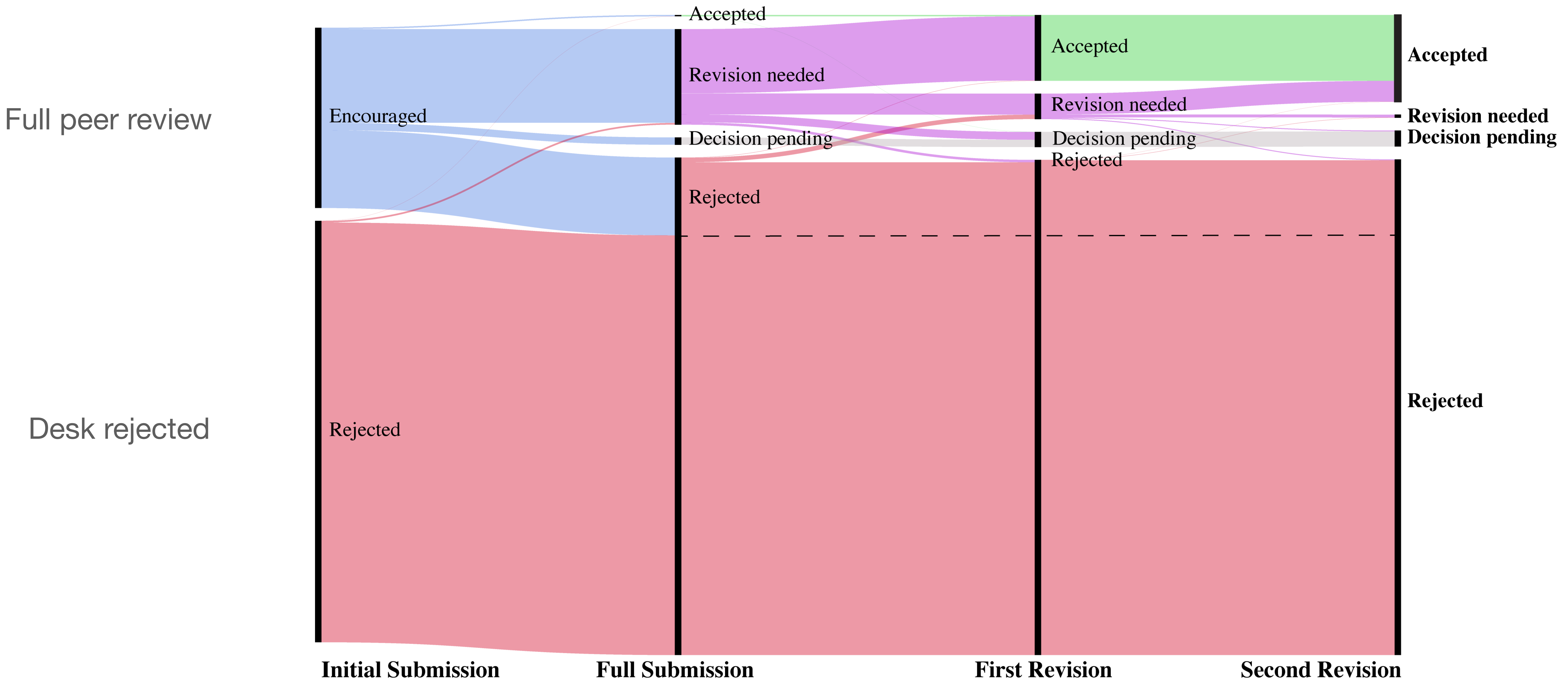




Lamont, M. (2009). *How Professors Think: Inside the Curious World of Academic Judgment*. Harvard University Press.

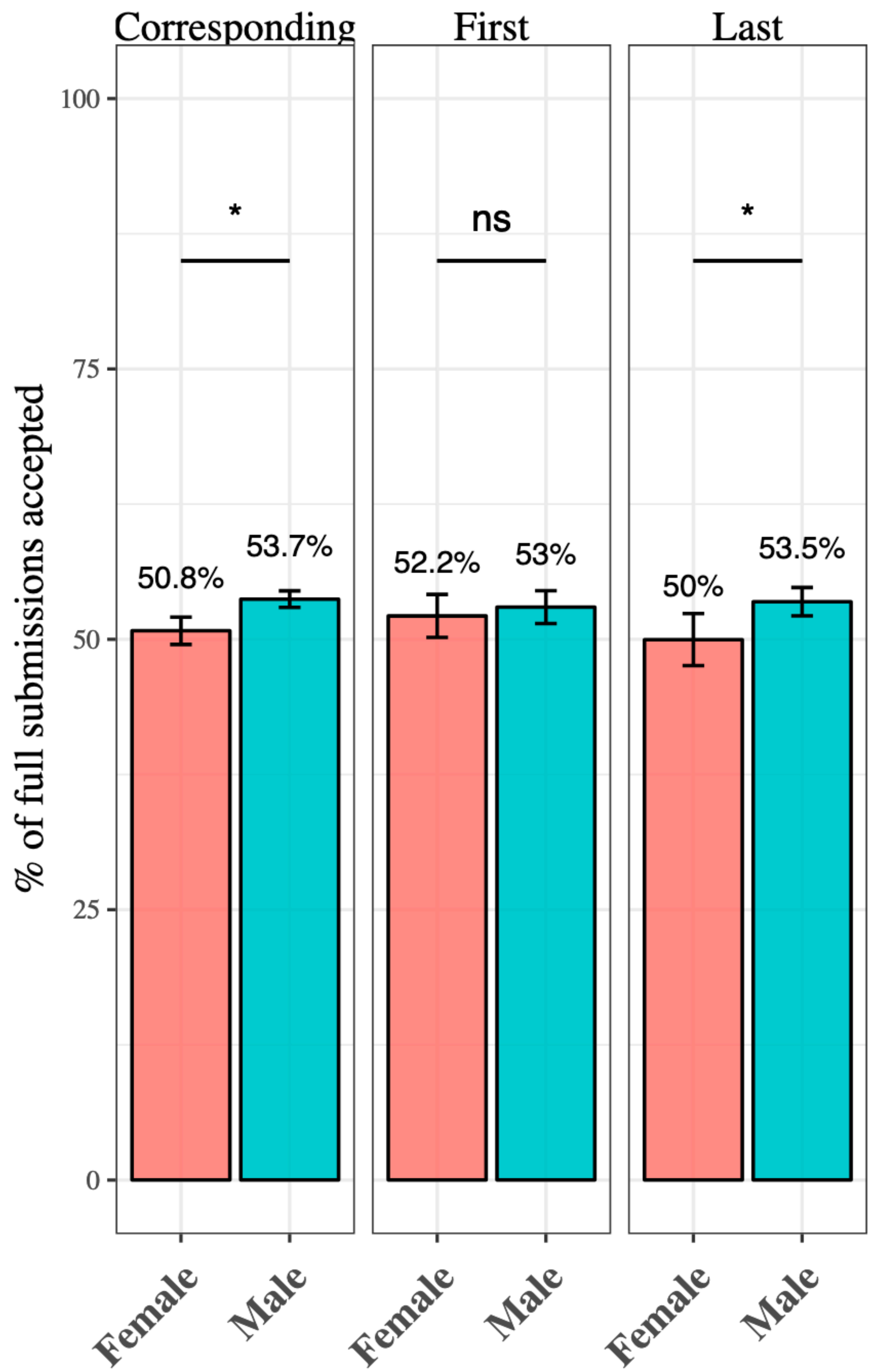
Peer review at *eLife*

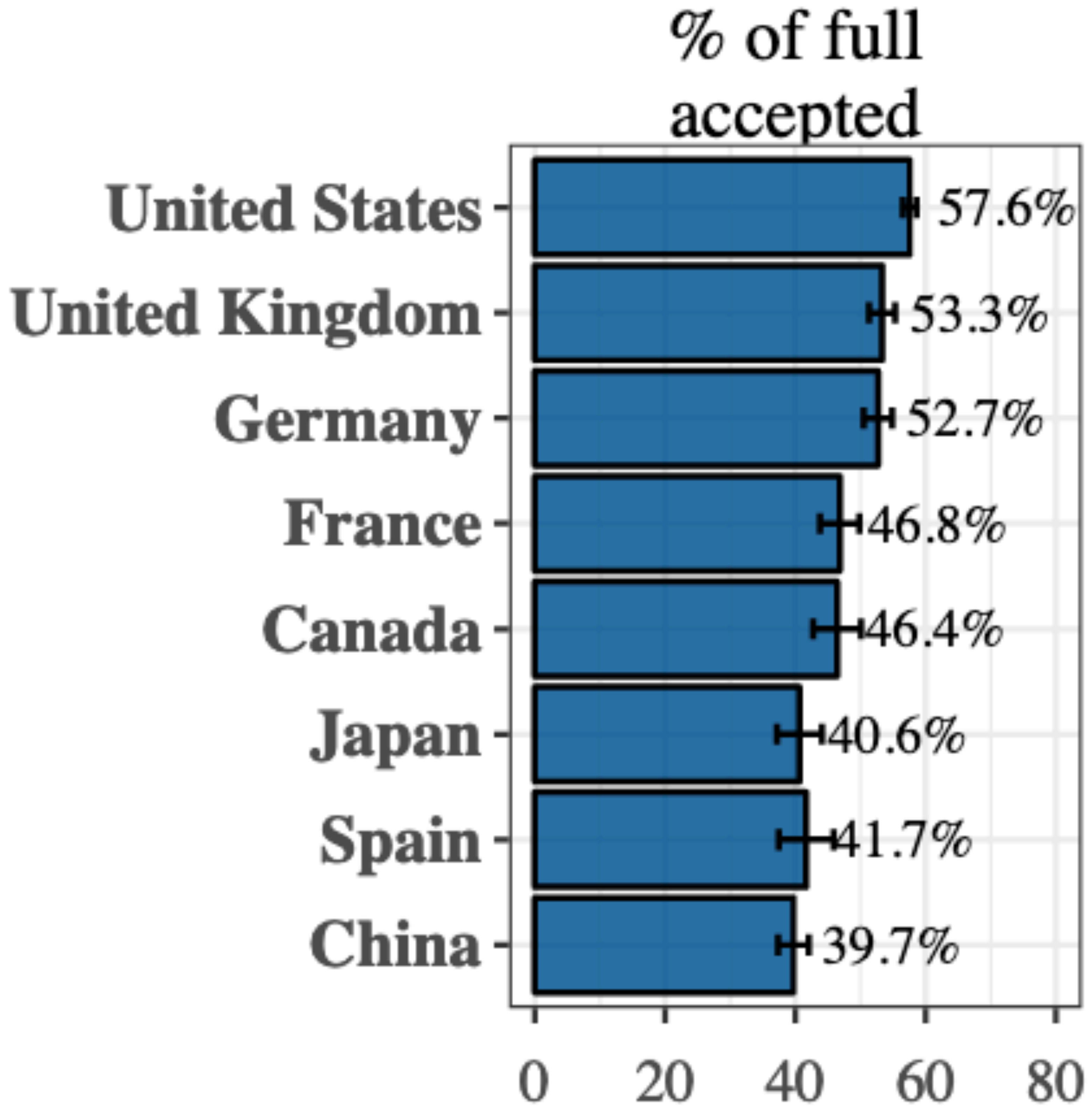
6,509 submissions reviewed and a decision made



Evidence of *disparity*

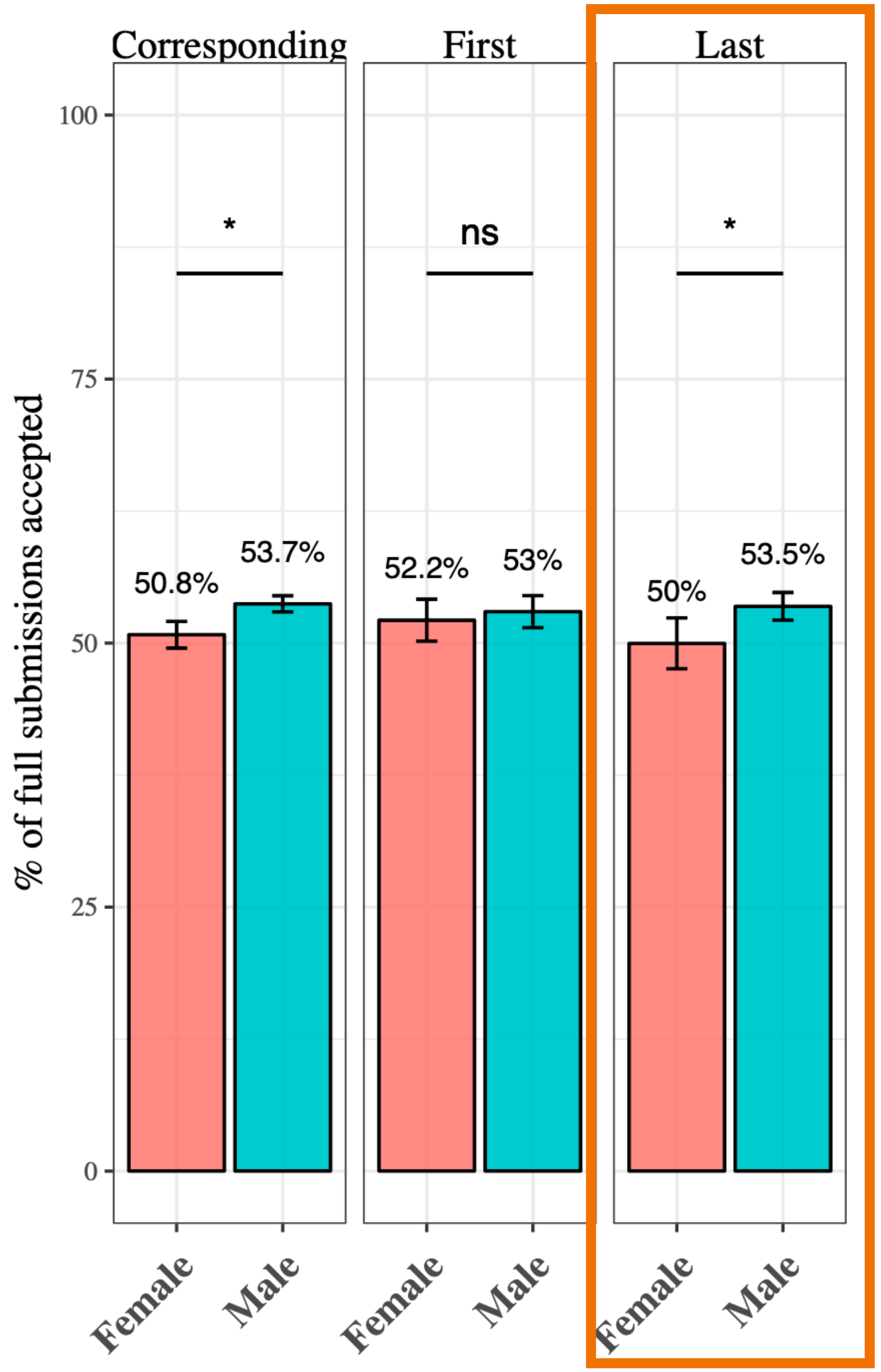
Women authors and those from peripheral nations accepted less



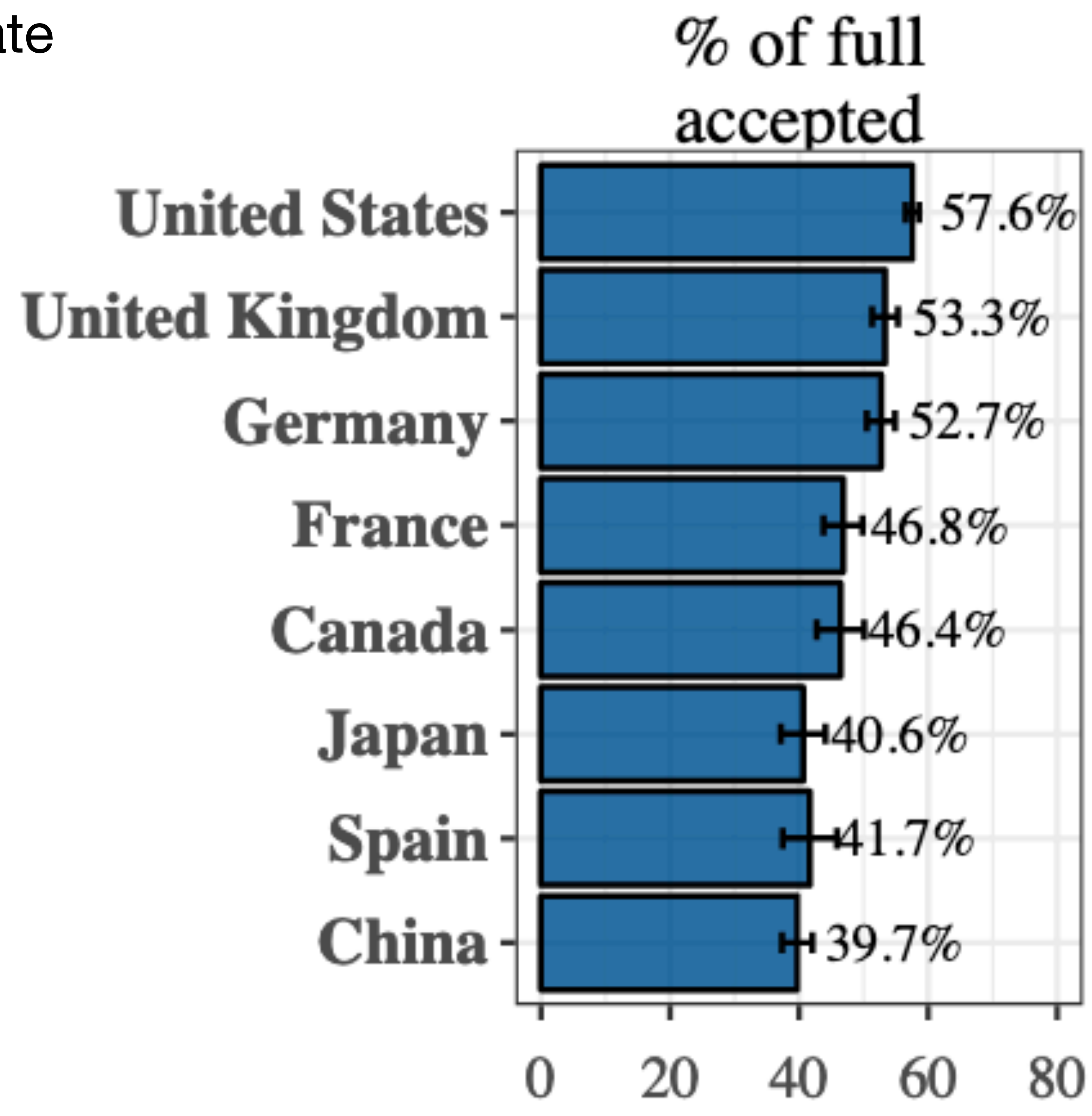


Evidence of *disparity*

Women authors and those from peripheral nations accepted less

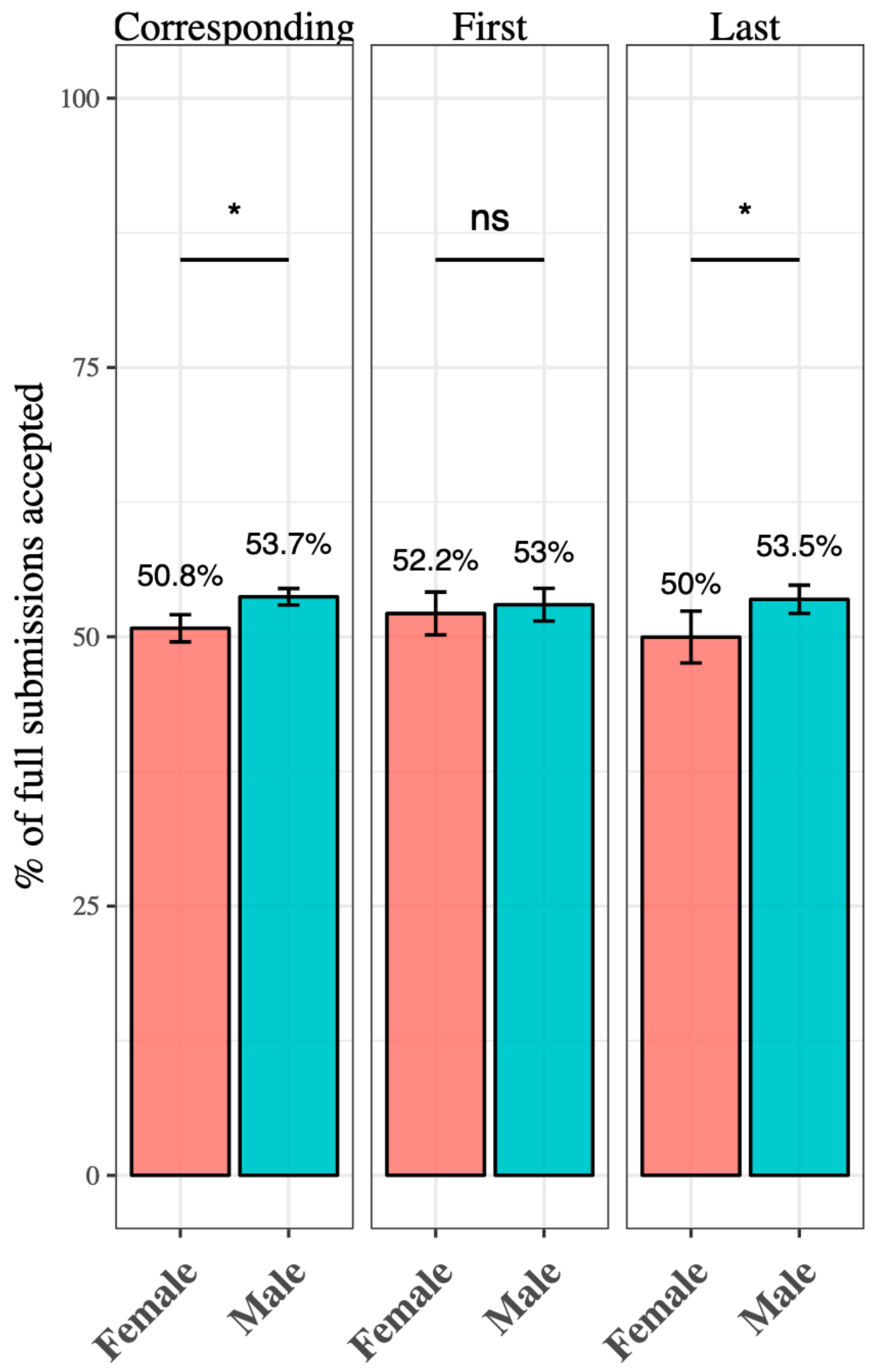


Women last authors are accepted at a 2.9 % point lower rate

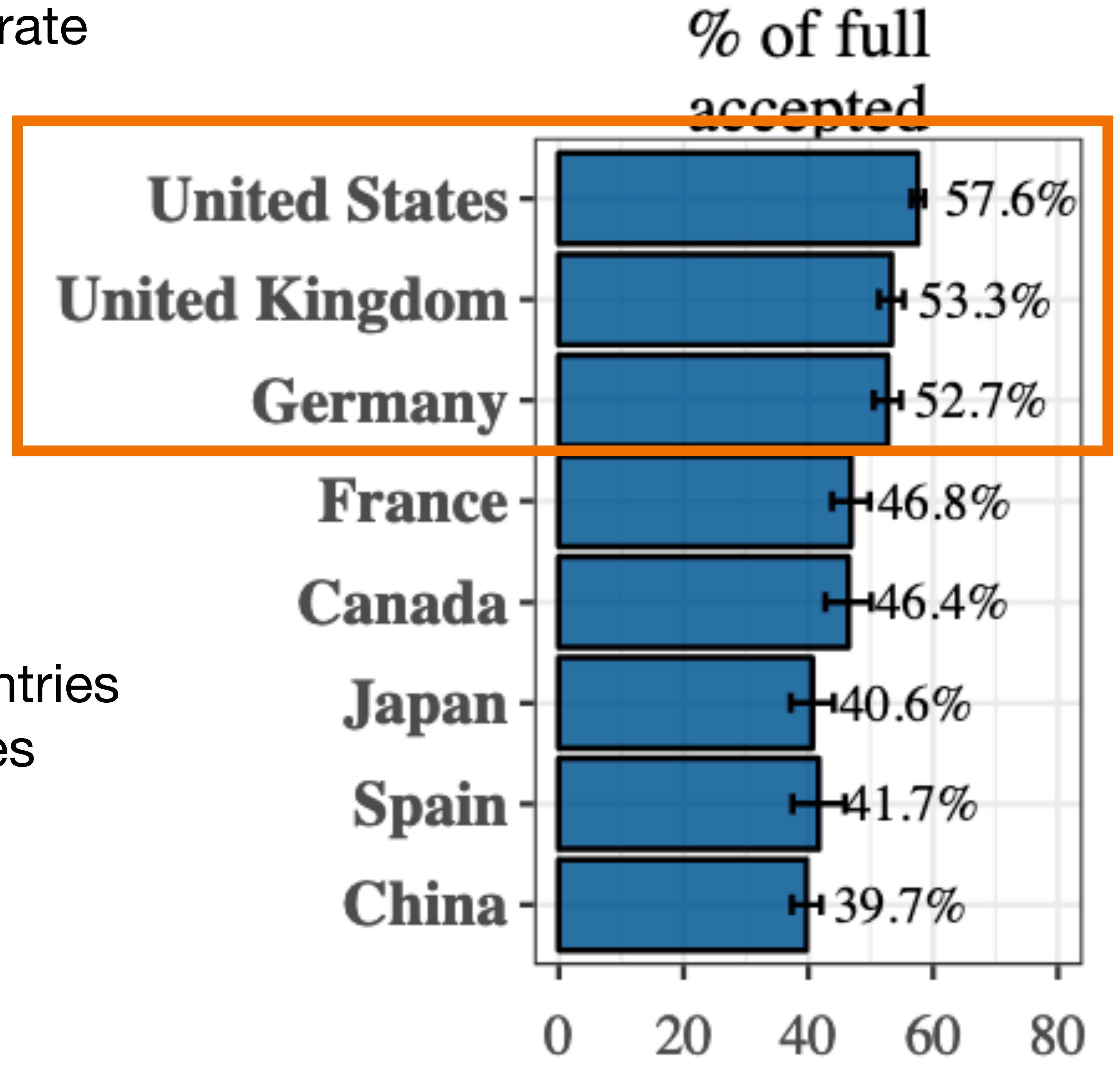


Evidence of *disparity*

Women authors and those from peripheral nations accepted less



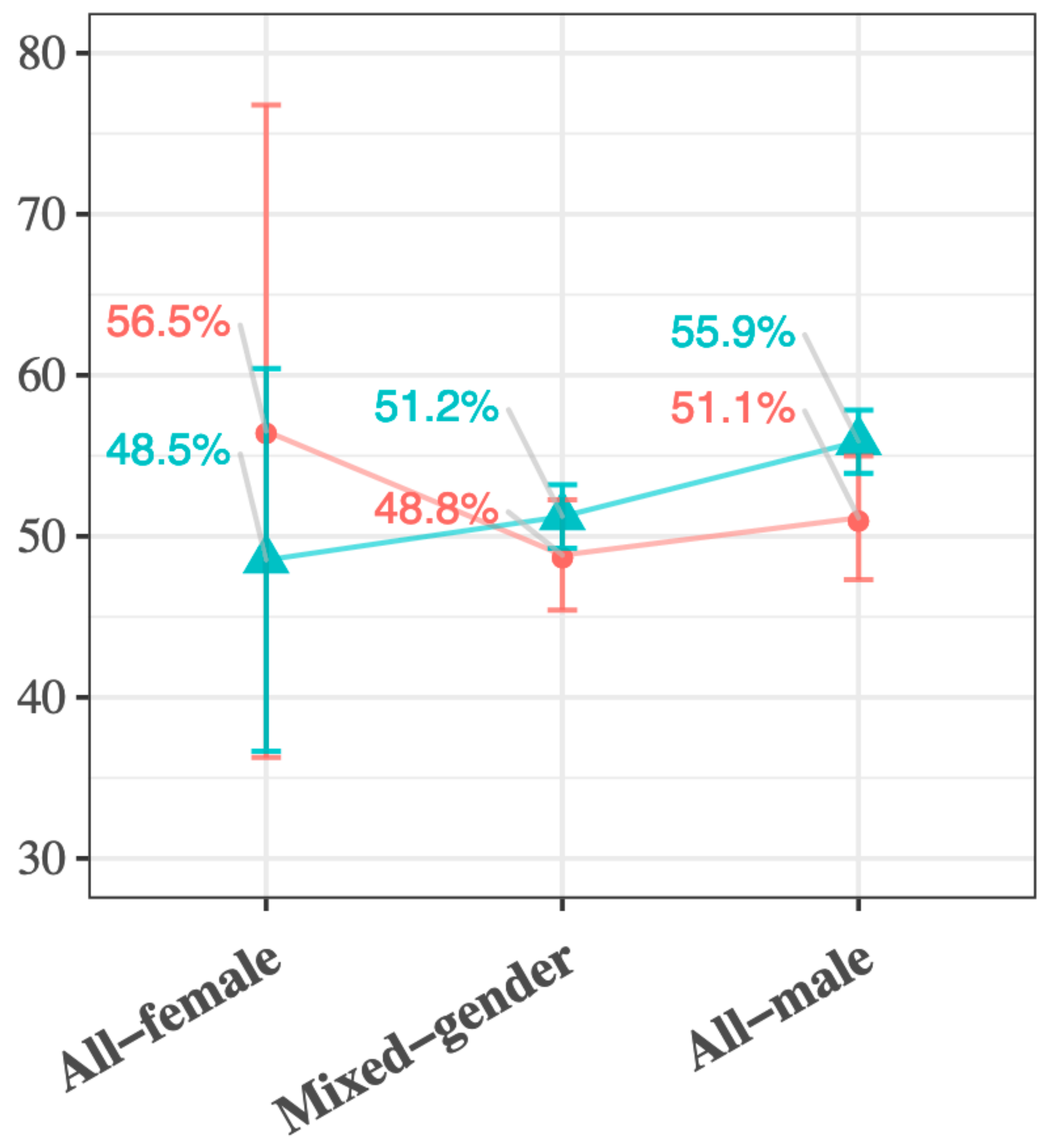
Women last authors are accepted at a 2.9 % point lower rate

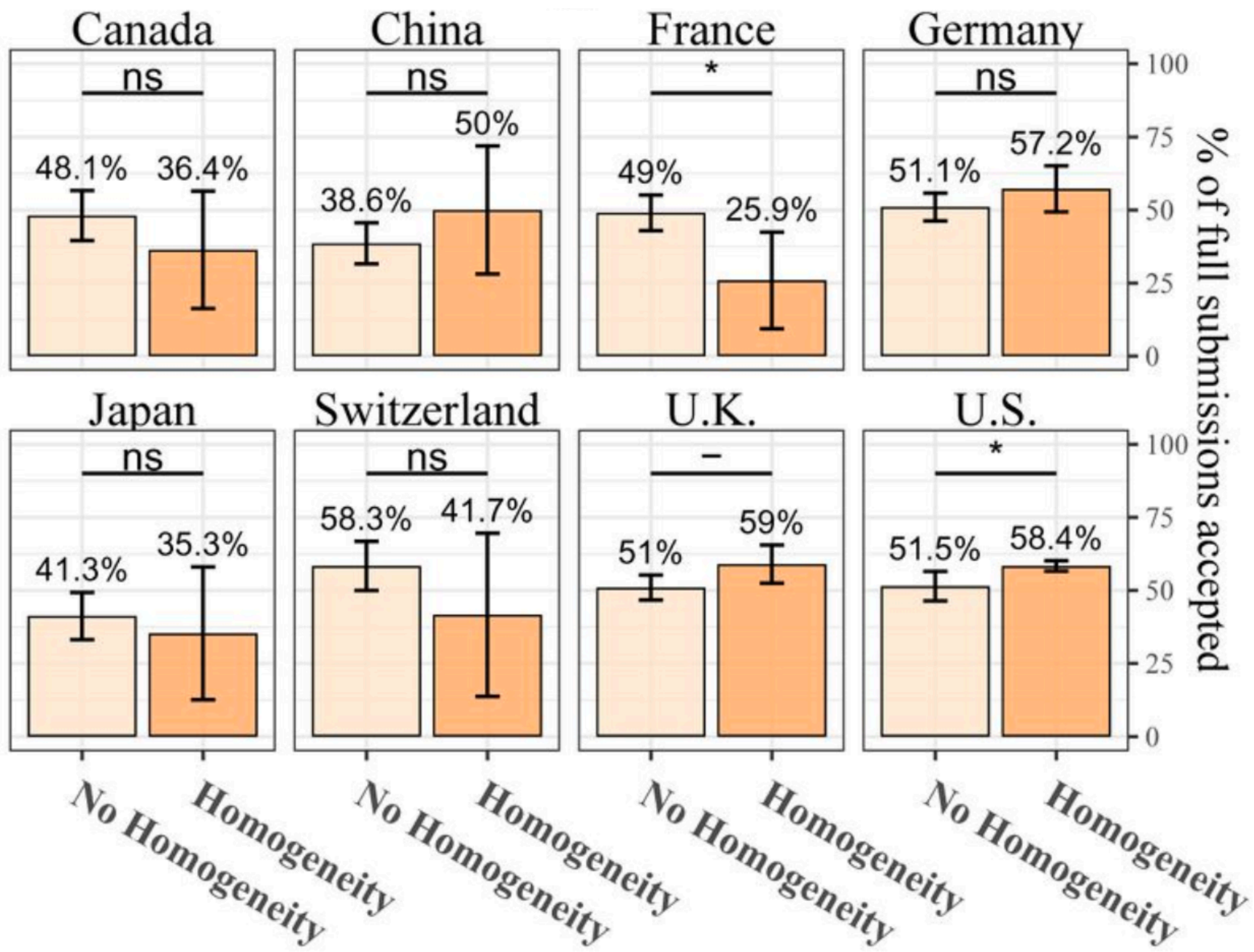


Authors from core countries accepted at higher rates

Evidence of *bias*

The composition of the peer reviewer team matters!

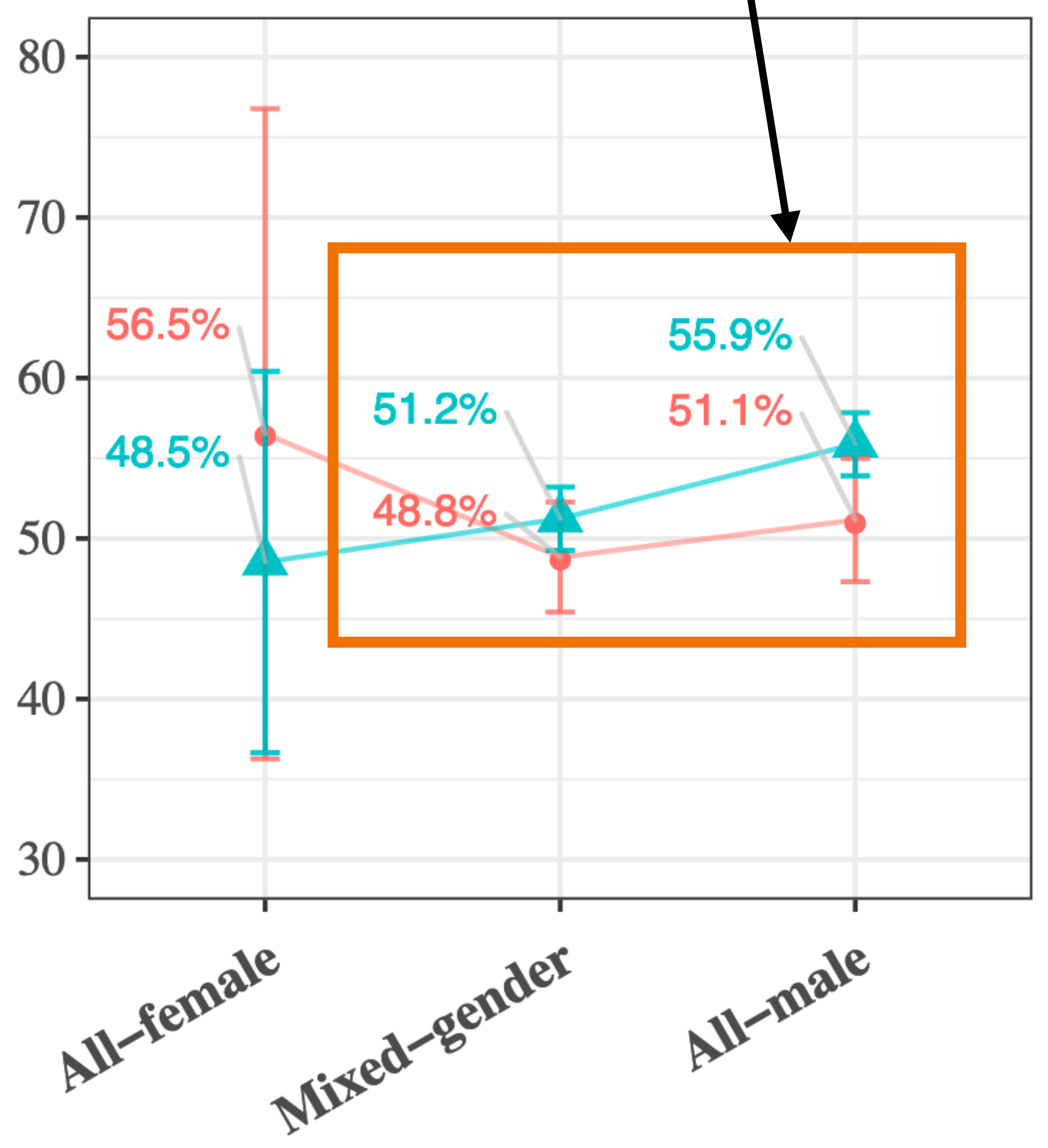


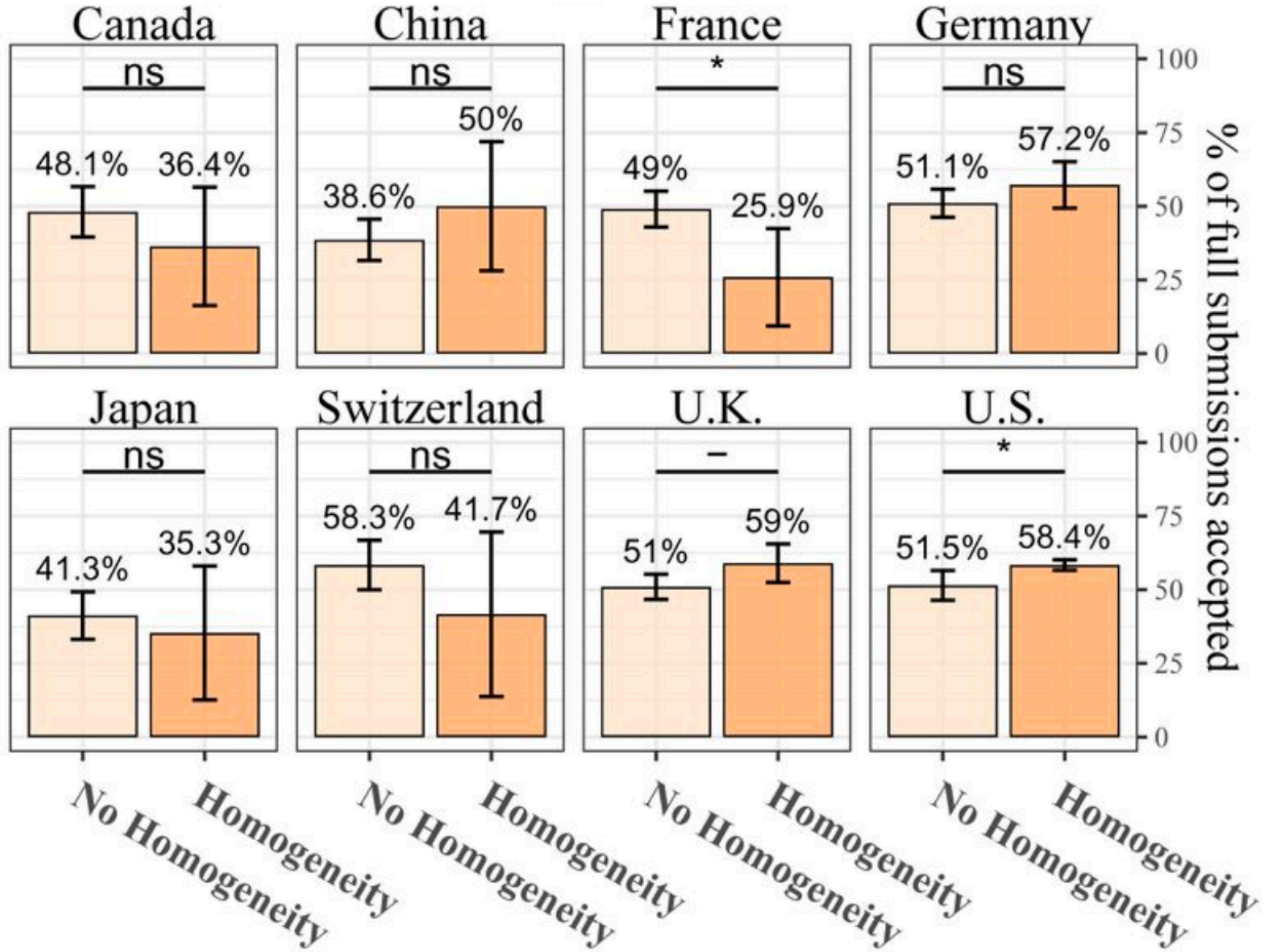


Evidence of *bias*

The composition of the peer reviewer team matters!

4.8% disparity for all-male teams, which shrinks in the case of mixed-gender reviewers

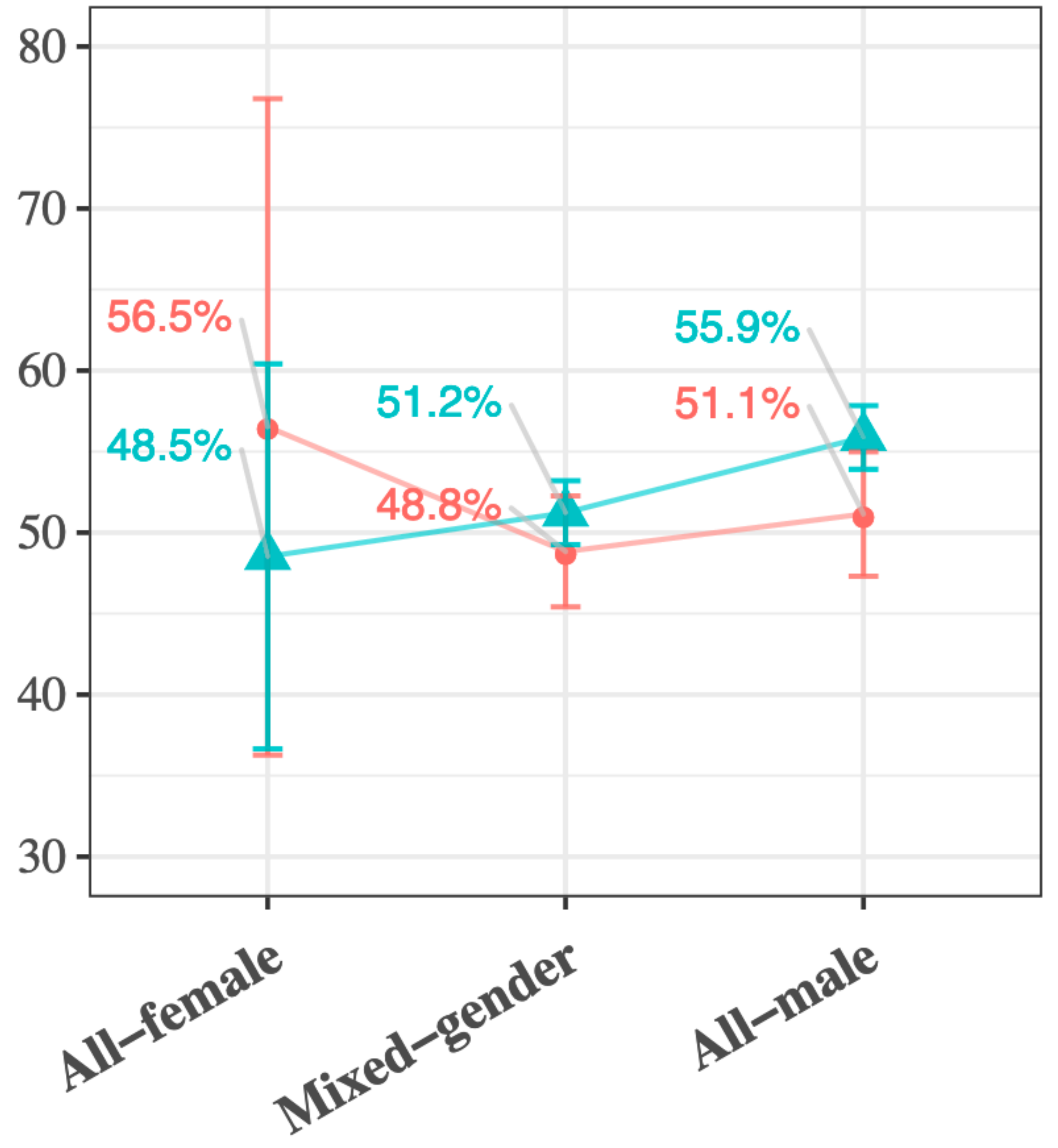




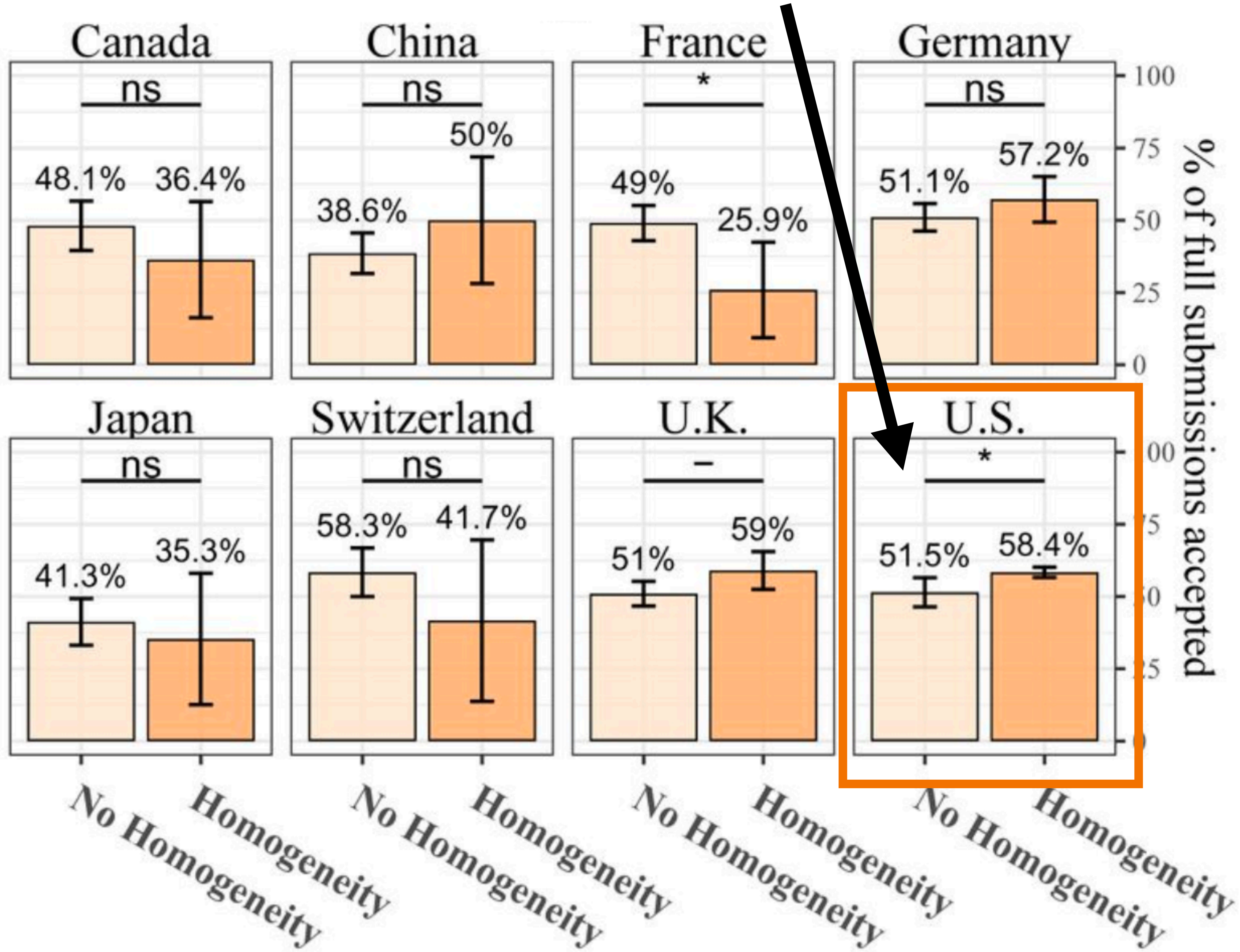
Evidence of *bias*

The composition of the peer reviewer team matters!

4.8% disparity for all-male teams, which shrinks in the case of mixed-gender reviewers

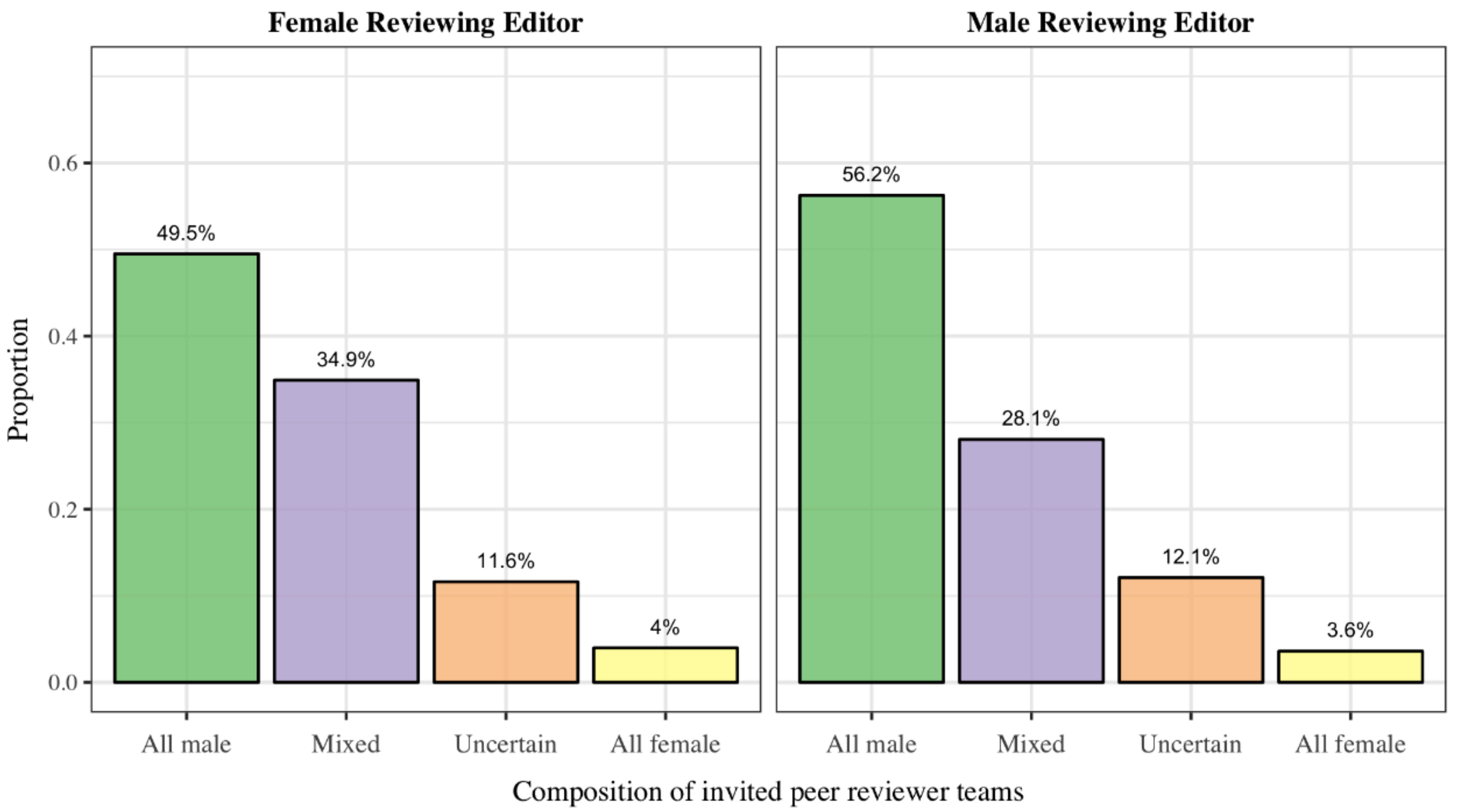


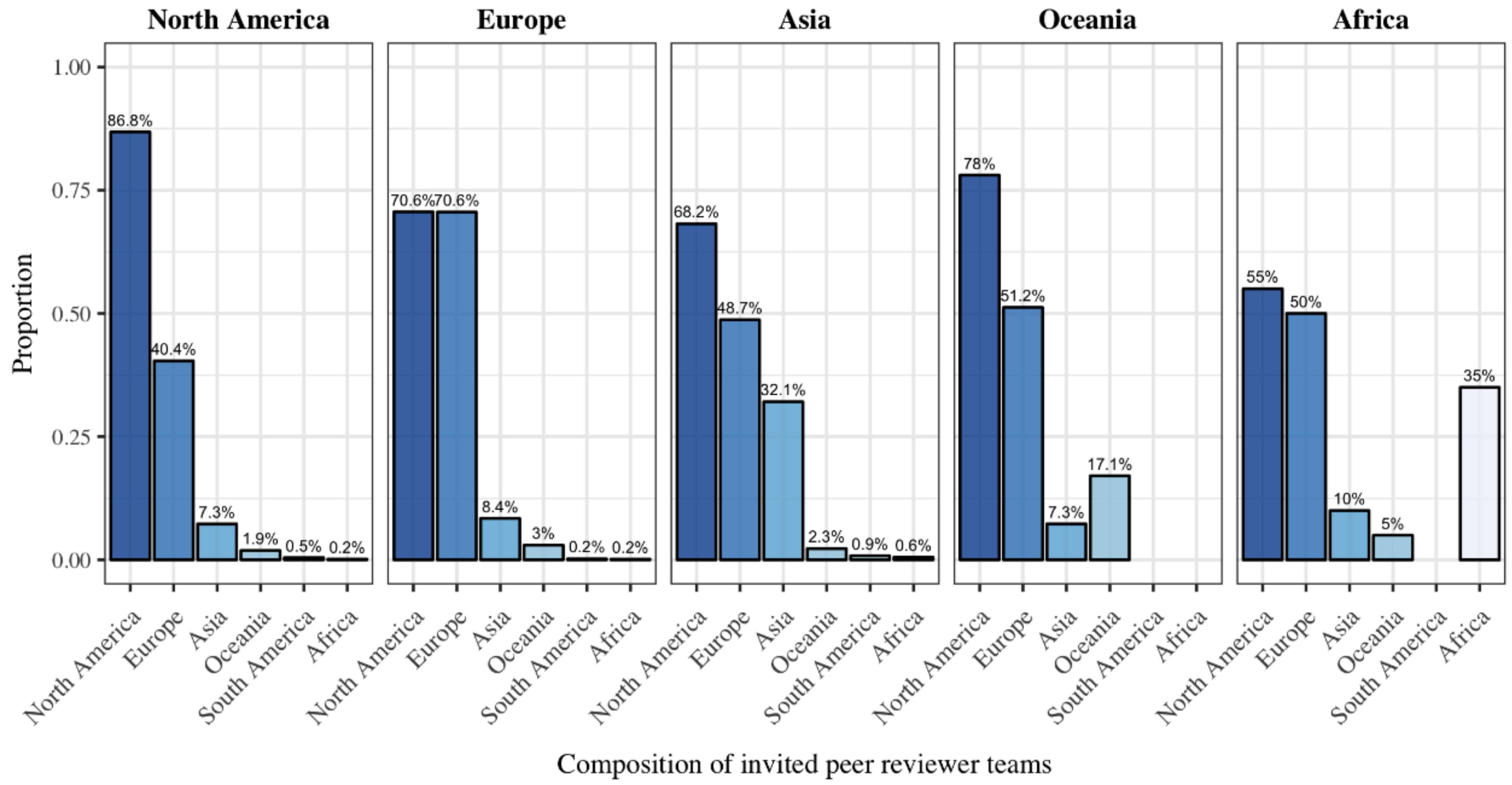
U.S. papers with U.S. reviewers more likely to be accepted, Not so with all countries



Homophily in team formation

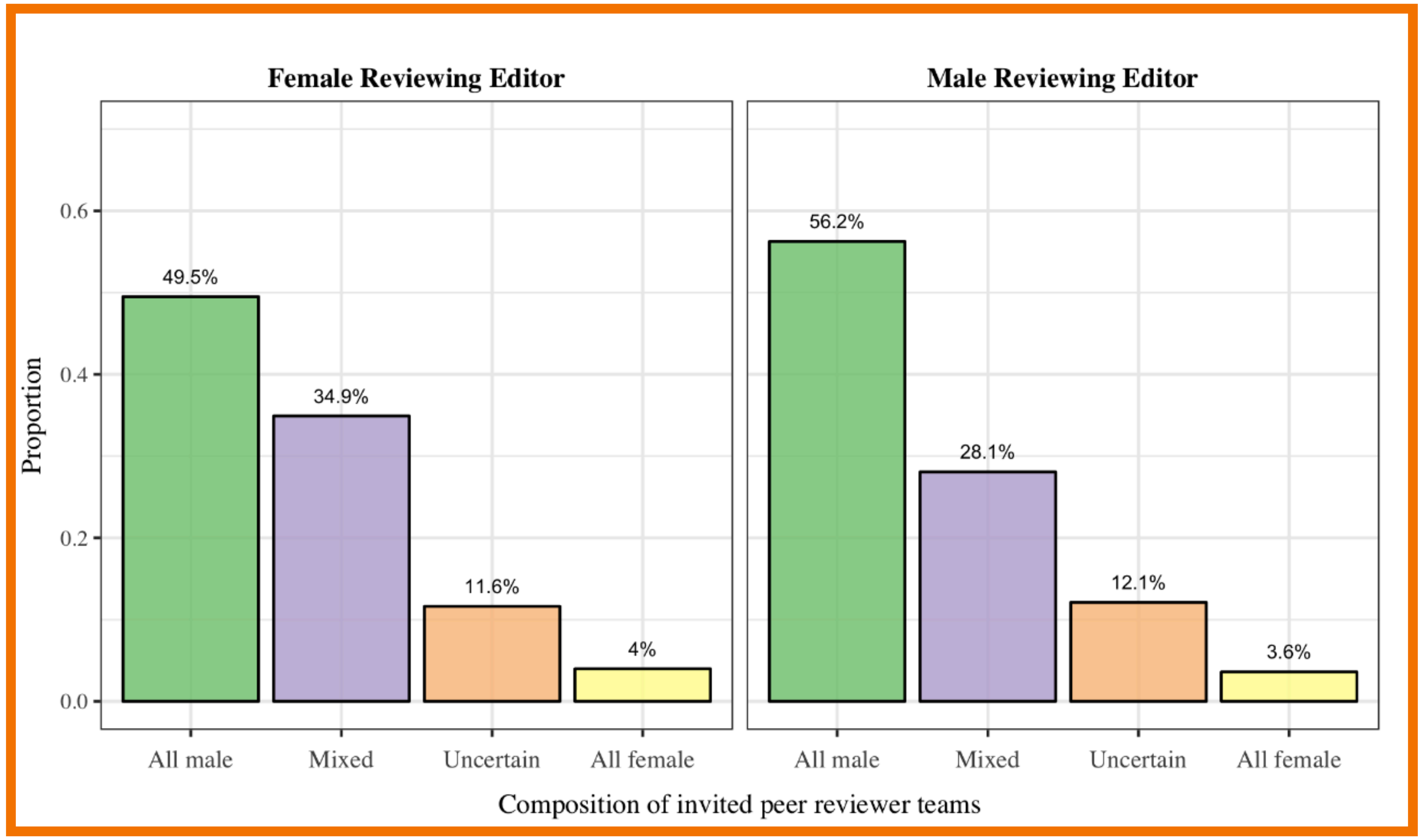
Fewer mixed-gender teams than we would expect

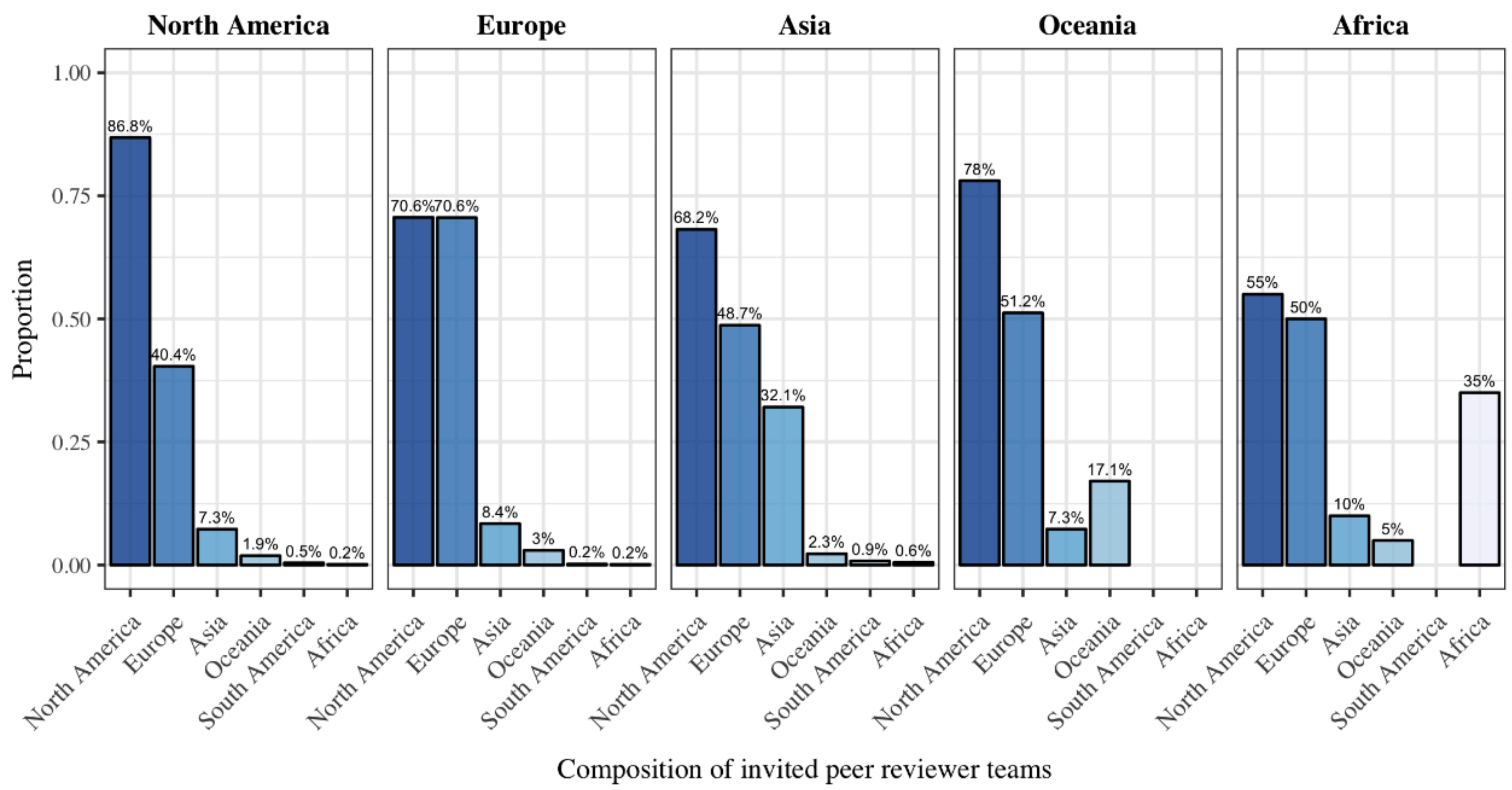




Homophily in team formation

Fewer mixed-gender teams than we would expect

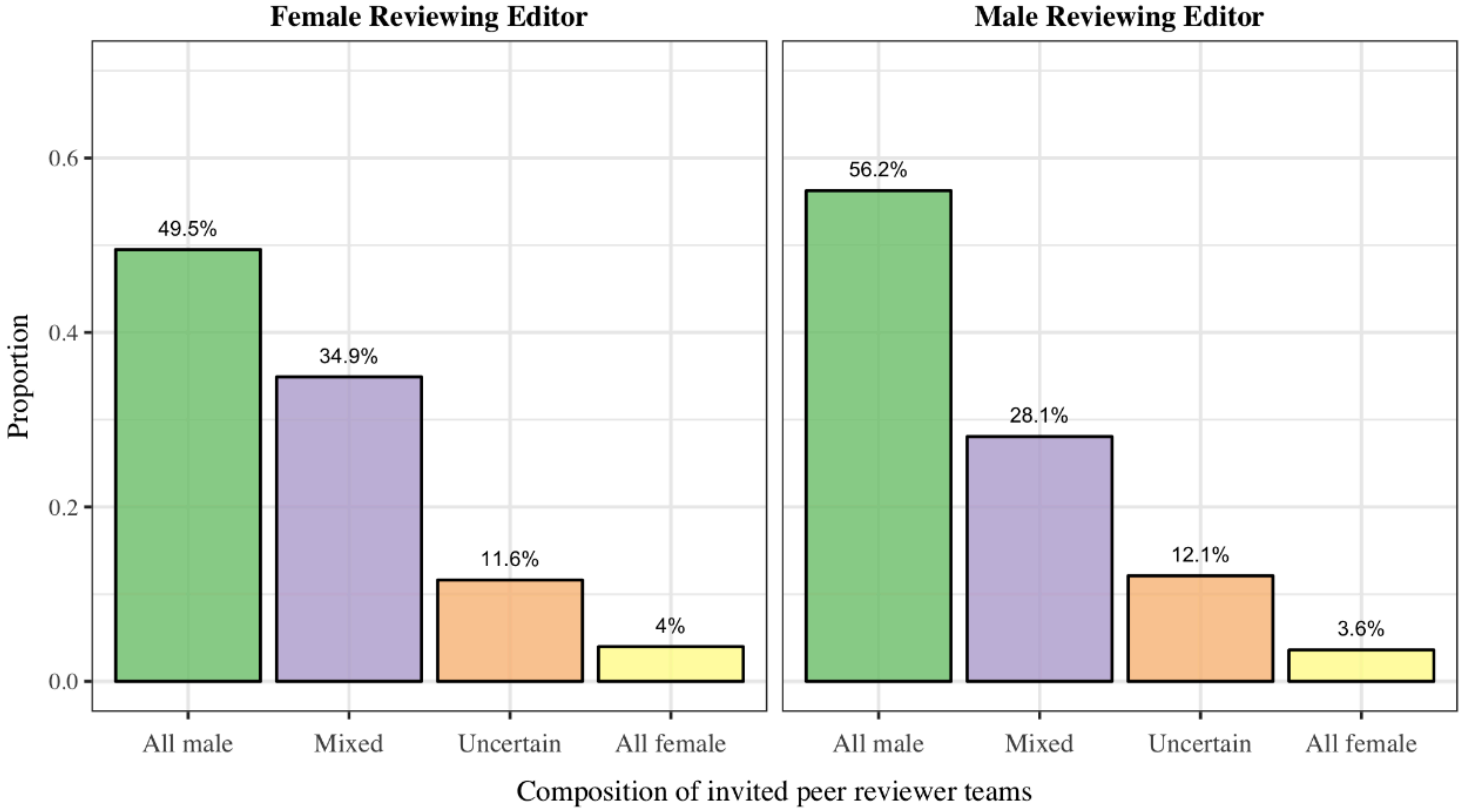




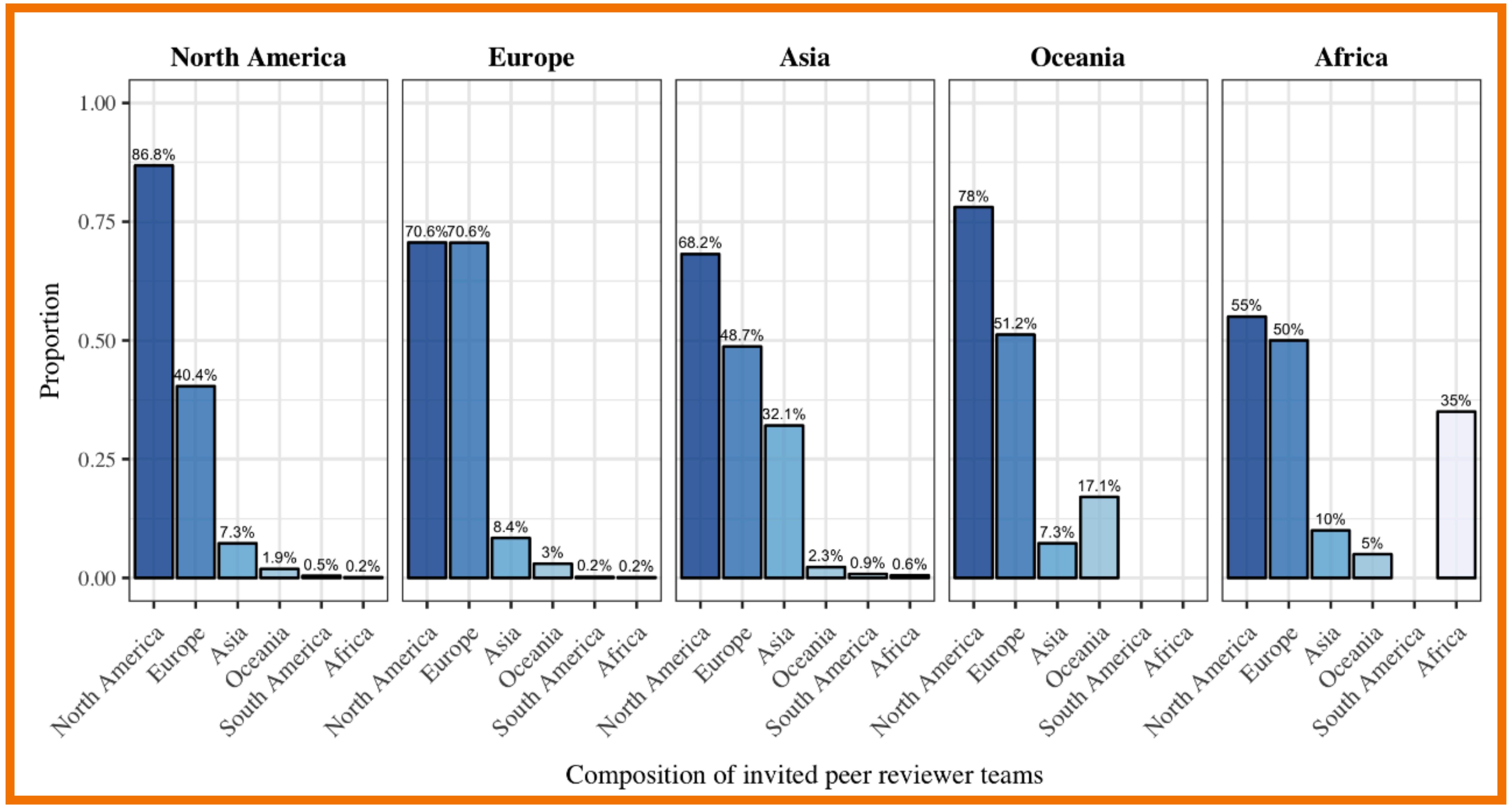
Male editors recruit more all-male teams, whereas women editors are more likely to recruit a mixed-gender team

Homophily in team formation

Fewer mixed-gender teams than we would expect



Male editors recruit more all-male teams, whereas women editors are more likely to recruit a mixed-gender team



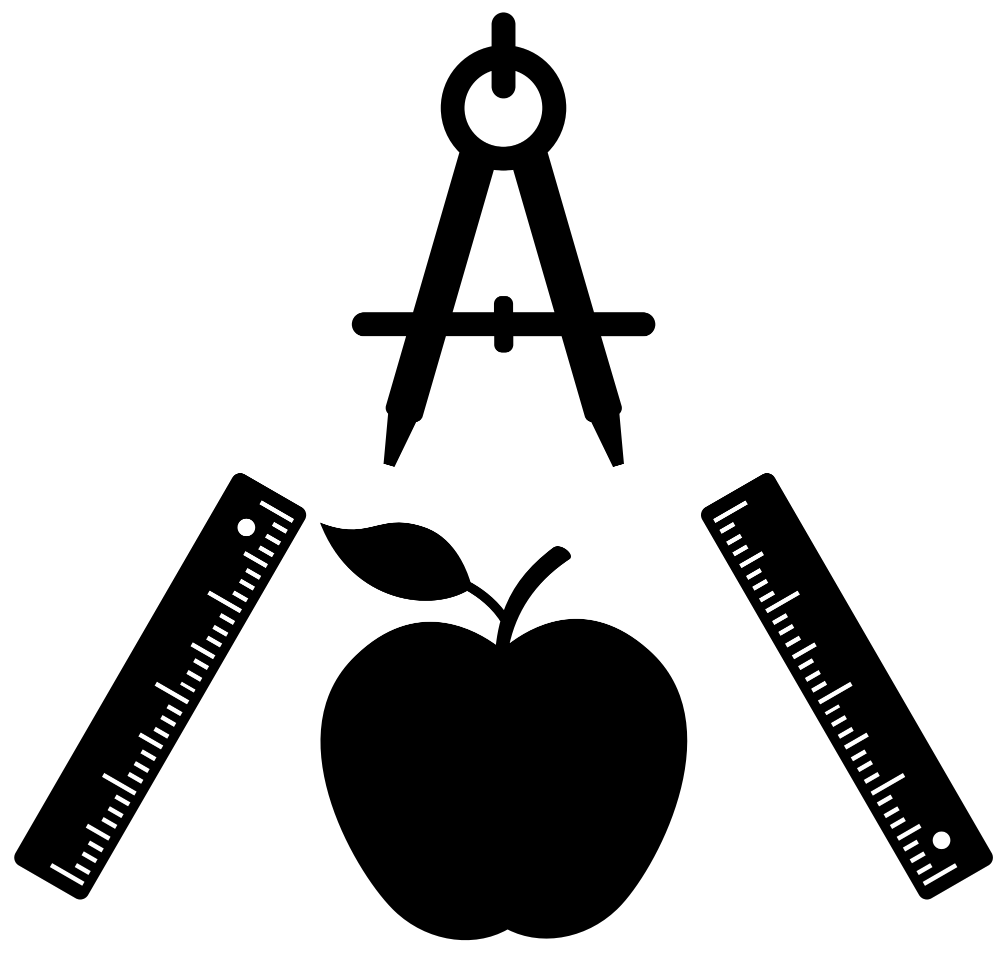
Similarity, editors recruit more reviewers from their own geographic area

Evidence of disparity and bias at *eLife*

Homophily drives both peer review outcomes, and the formation of teams

Study 2:

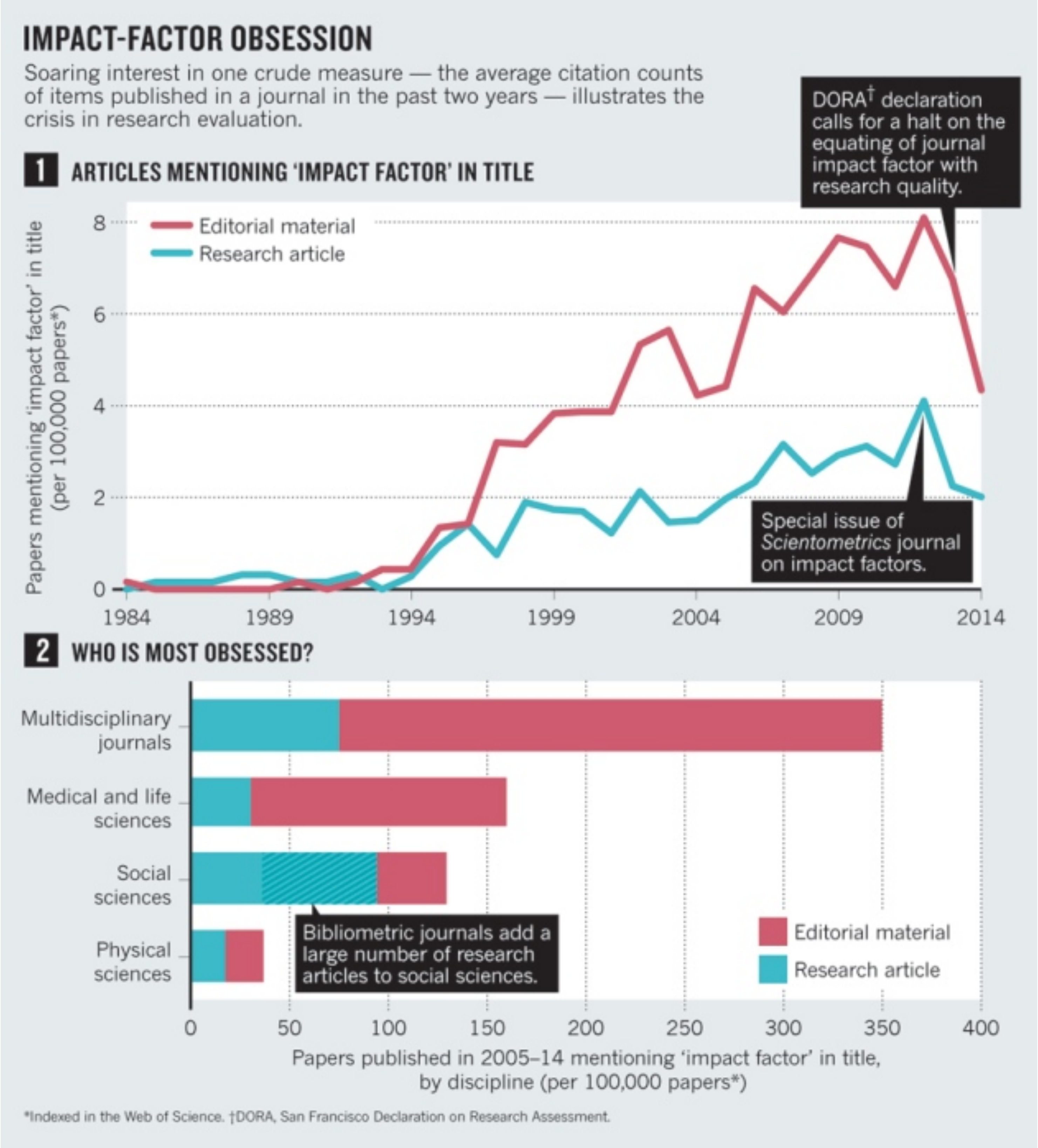
Student-teacher evaluations at U.S. Universities



Murray, D., Boothby, C., Zhao, H., Minik, V., Bérubé, N., Larivière, V., & Sugimoto, C. R. (2020). Exploring the personal and professional factors associated with student evaluations of tenure-track faculty. *PLOS ONE*, *15*(6), e0233515.

Metrics are ubiquitous in science evaluation

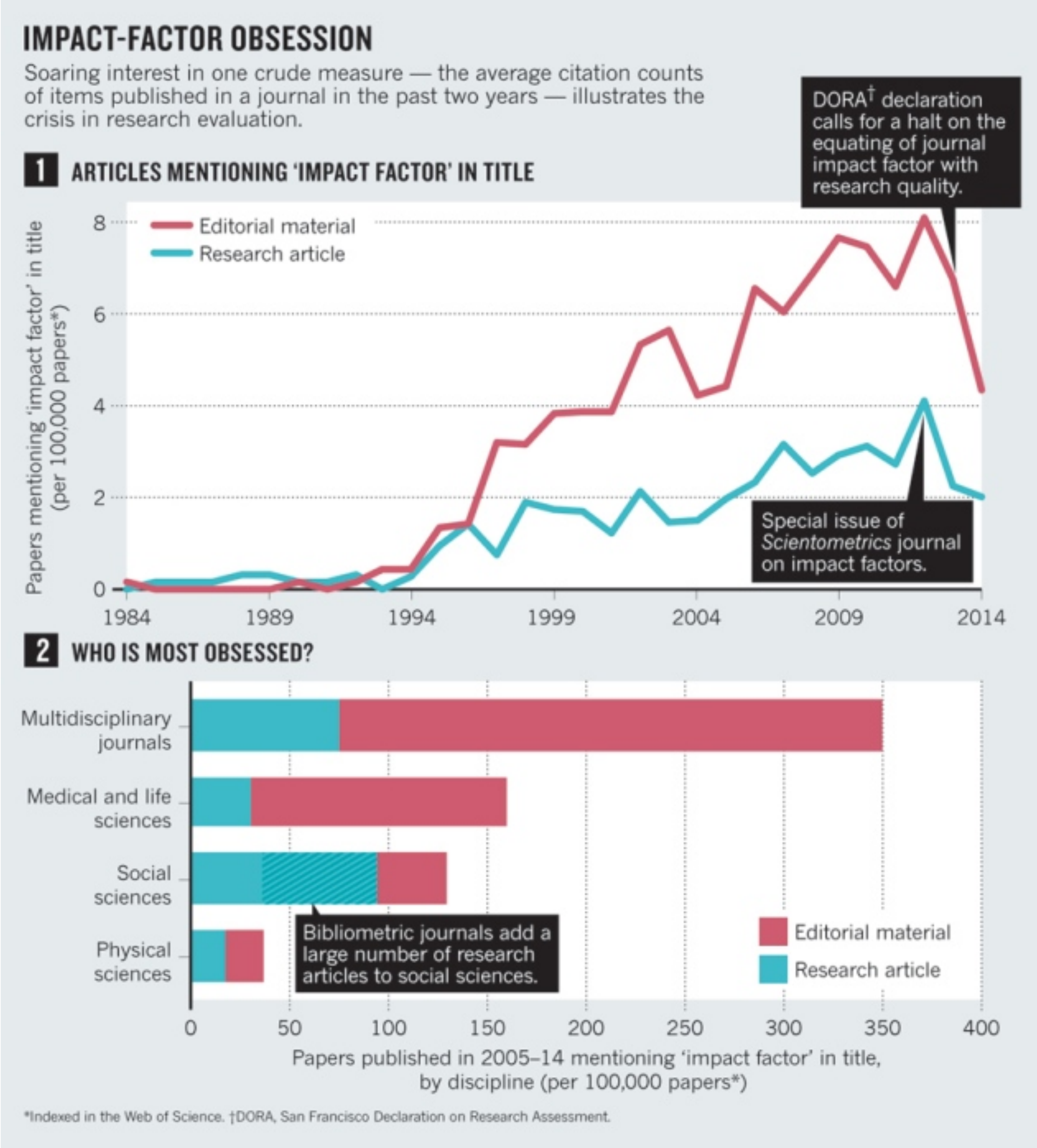
Citation-based metrics most discussed



Hicks, D., Wouters, P., Waltman, L., de Rijcke, S., & Rafols, I. (2015). Bibliometrics: The Leiden Manifesto for research metrics. *Nature News*, *520*(7548), 429.

Metrics are ubiquitous in science evaluation

Citation-based metrics most discussed, but teaching metrics can also shape a career



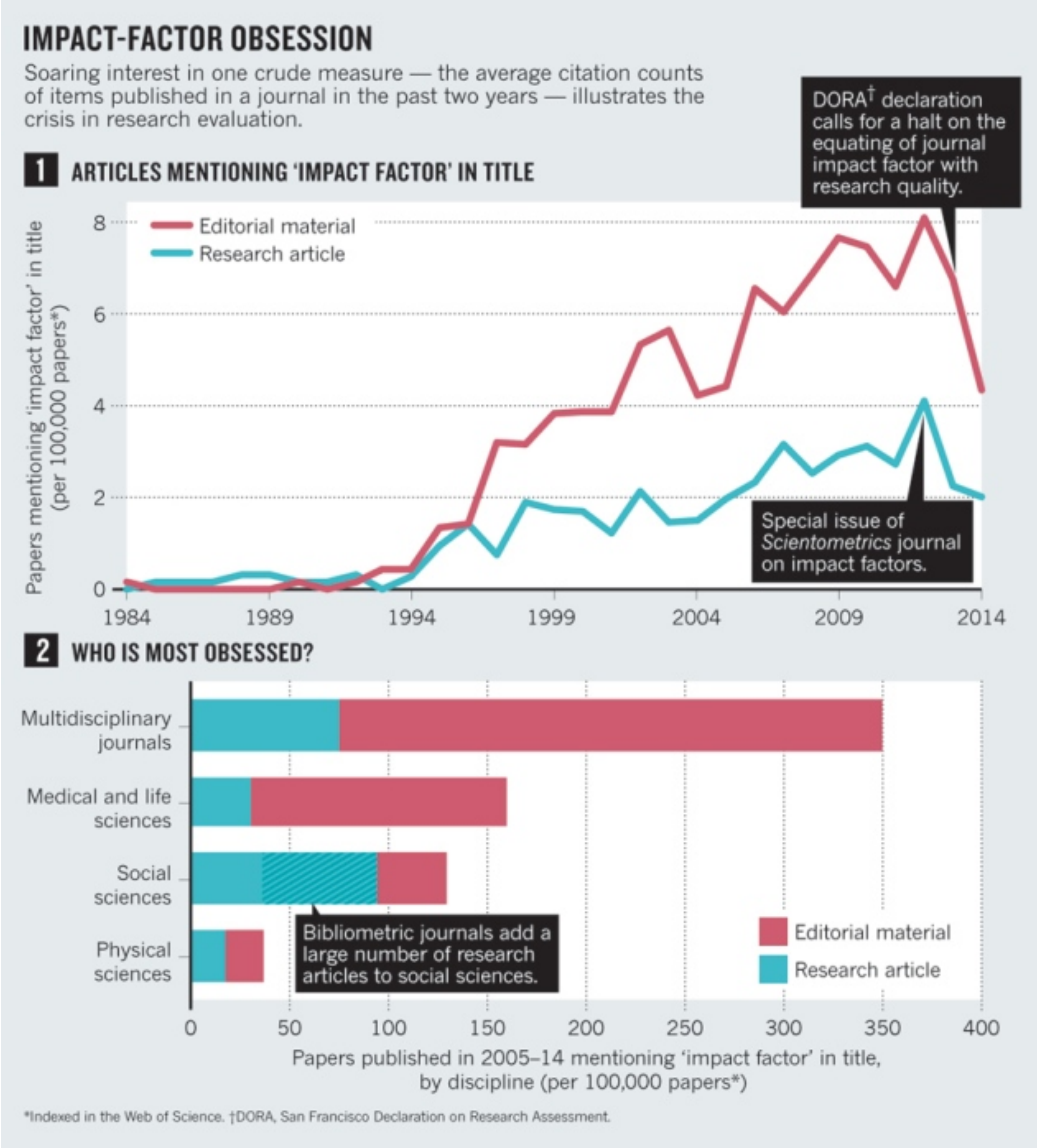
	Strongly agree	Agree	Disagree	Strongly disagree
1. My teacher in this class makes me feel that s/he really cares about me.	☐	☐	☐	☐
2. My teacher really tries to understand how students feel about things.	☐	☐	☐	☐
3. Students in this class treat the teacher with respect.	☐	☐	☐	☐
4. Our class stays busy and doesn't waste time.	☐	☐	☐	☐
5. My teacher has several good ways to explain each topic that we cover in this class.	☐	☐	☐	☐
6. My teacher explains difficult things clearly.	☐	☐	☐	☐
7. In this class, we learn a lot almost every day.	☐	☐	☐	☐
8. In this class, we learn to correct our mistakes.	☐	☐	☐	☐
9. My teacher makes lessons interesting.	☐	☐	☐	☐
10. I like the ways we learn in this class.	☐	☐	☐	☐
11. Students speak up and share their ideas about class work.	☐	☐	☐	☐
12. My teacher respects my ideas and suggestions.	☐	☐	☐	☐
13. My teacher checks to make sure we understand what s/he is teaching us.	☐	☐	☐	☐
14. The comments that I get on my work in this class help me understand how to improve.	☐	☐	☐	☐

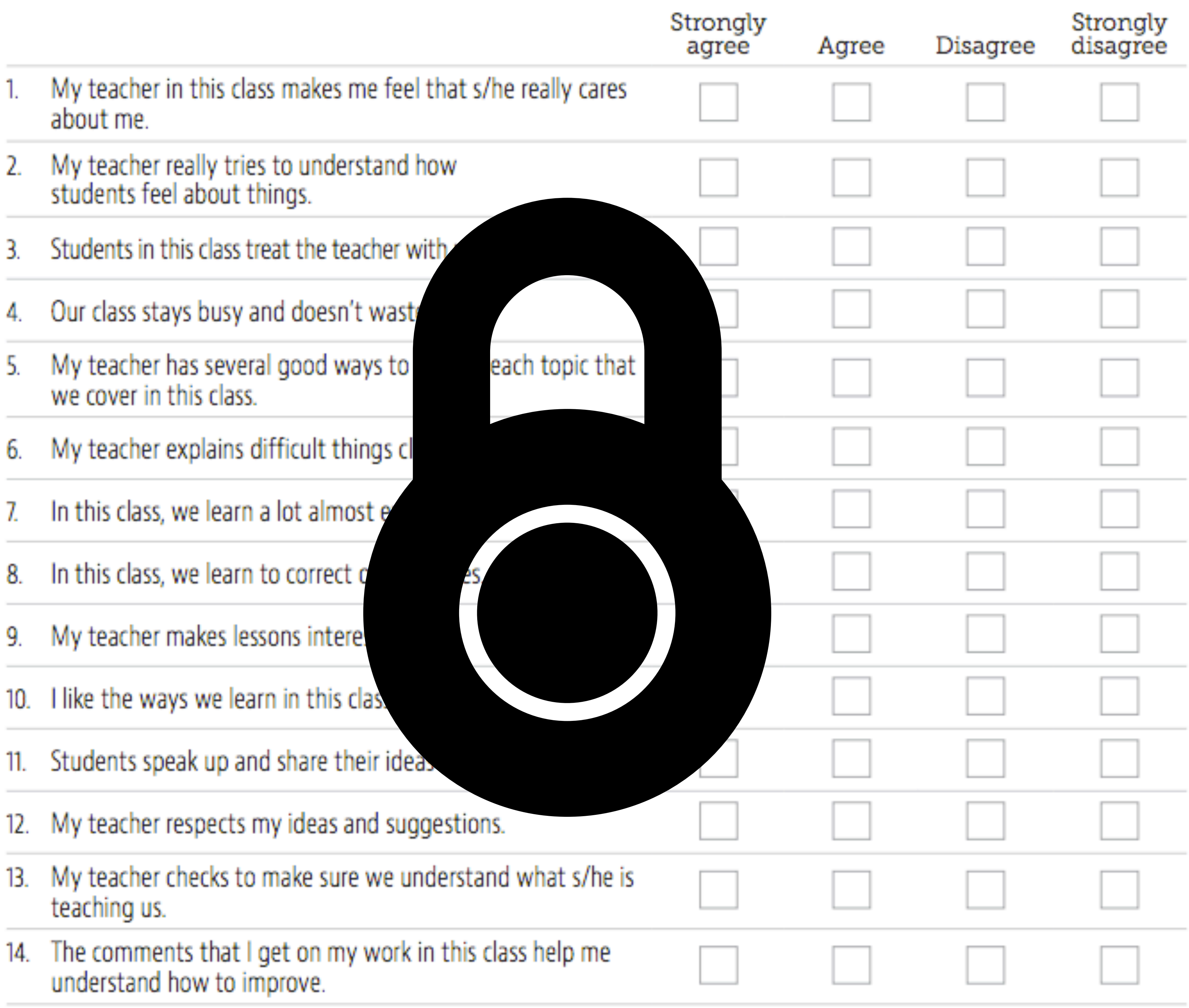
Hicks, D., Wouters, P., Waltman, L., de Rijcke, S., & Rafols, I. (2015). Bibliometrics: The Leiden Manifesto for research metrics. *Nature News*, *520*(7548), 429.

Metrics are ubiquitous in science evaluation

Citation-based metrics most discussed, but teaching metrics can also shape a career

But their sensitivity makes them difficult to study, at scale



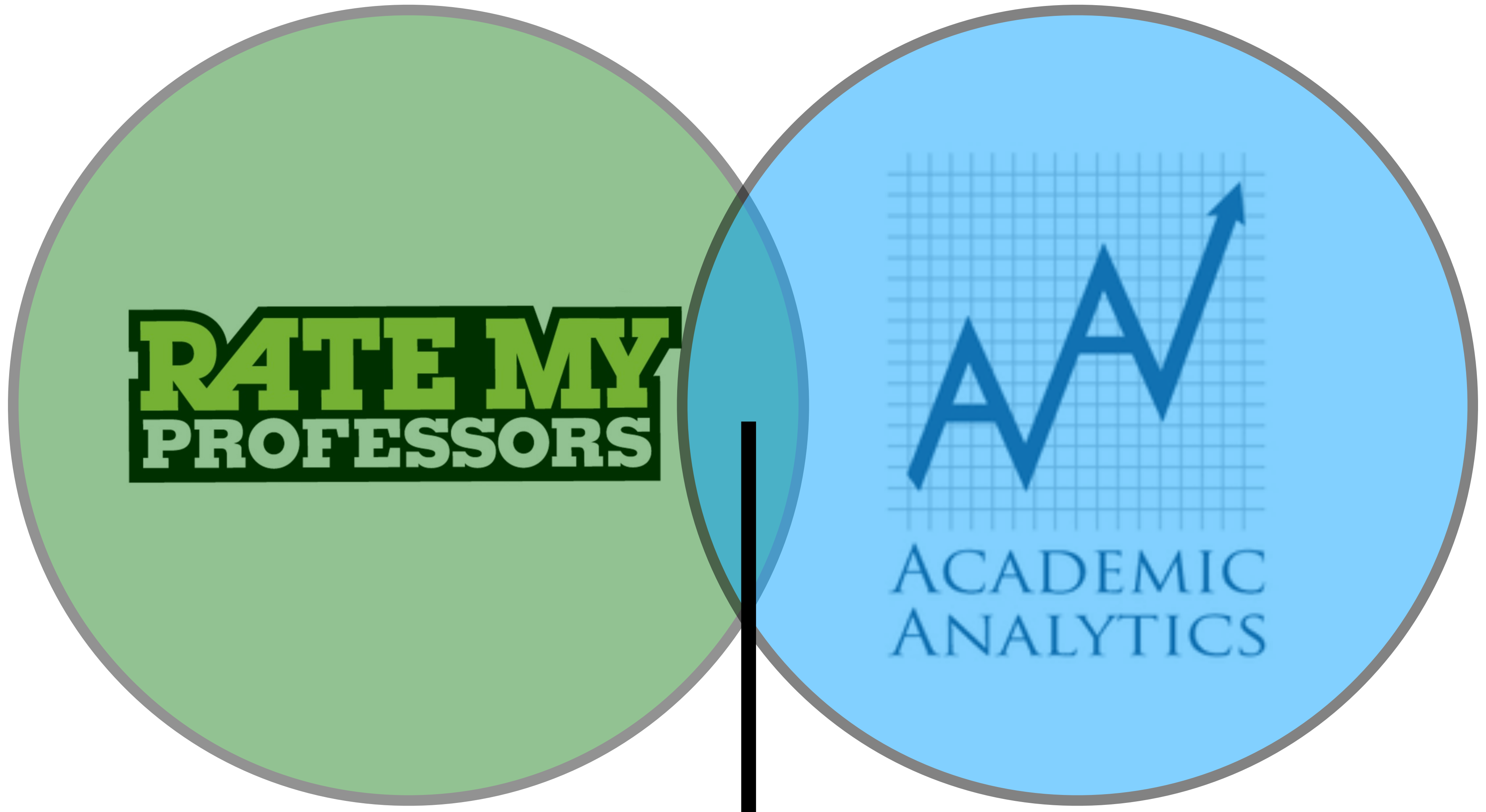


Hicks, D., Wouters, P., Waltman, L., de Rijcke, S., & Rafols, I. (2015). Bibliometrics: The Leiden Manifesto for research metrics. *Nature News*, *520*(7548), 429.

Create a massive and diverse dataset on teaching evaluations

Merging public teaching evaluations with private records of TT faculty, resulting in almost 19,000 records

RateMyProfessors.com has been shown to be an effective proxy for institutional evaluations



Academic Analytics curates data on tenure and tenure-track faculty performance

Merge to build a rich dataset of teaching ratings and demographics for U.S. faculty

ID	Rating	Gender	Race	...
...	...	...	...	...
...	...	...	...	...
...	...	...	...	...
...	...	...	...	...

Student ratings of teachers

Factors relating to ratings on 19,000 TT faculty on RateMyProfessors.com

Student ratings of teachers

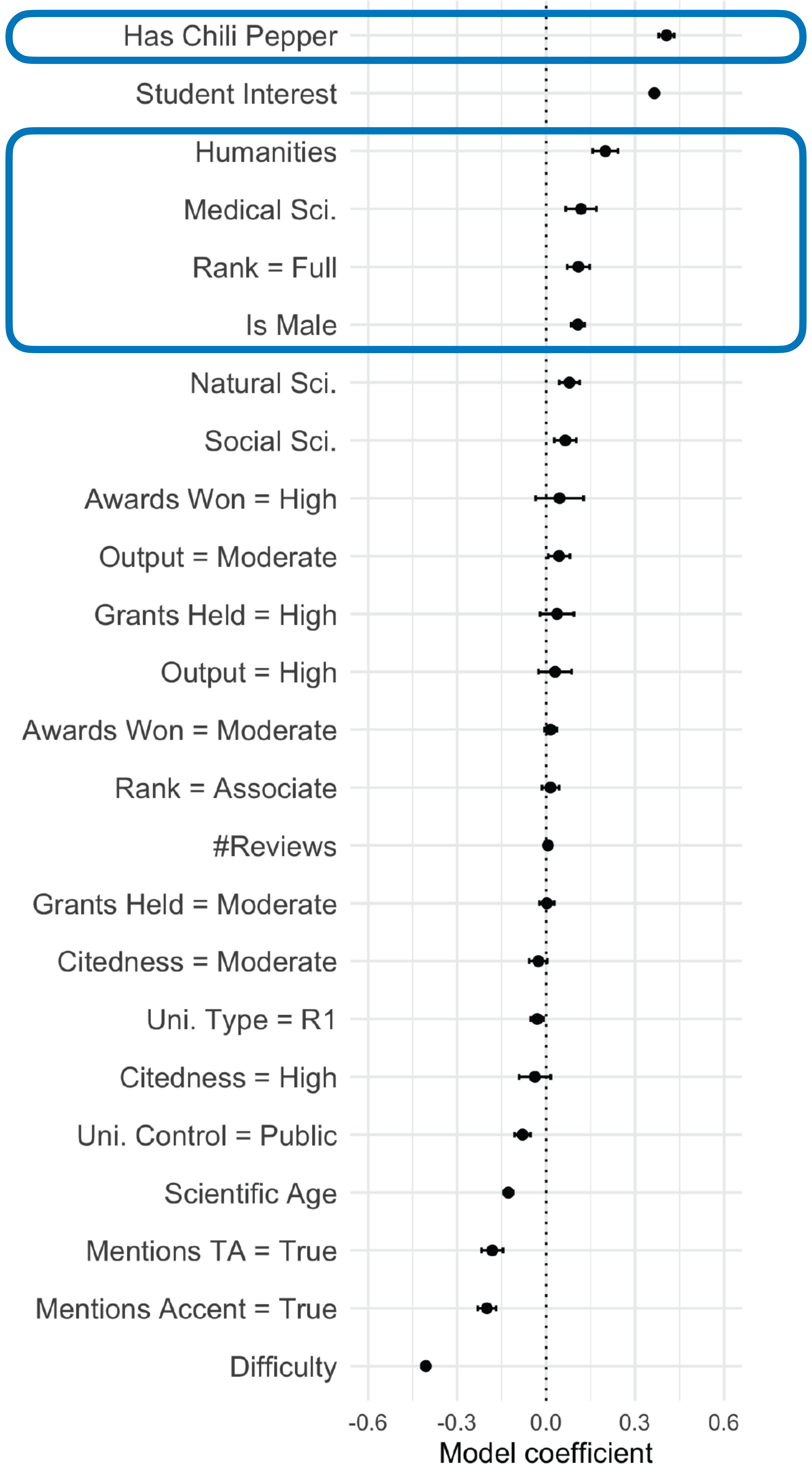
Factors relating to ratings on 19,000 TT faculty on RateMyProfessors.com

- Attractive profs rated higher

Student ratings of teachers

Factors relating to ratings on 19,000 TT faculty on RateMyProfessors.com

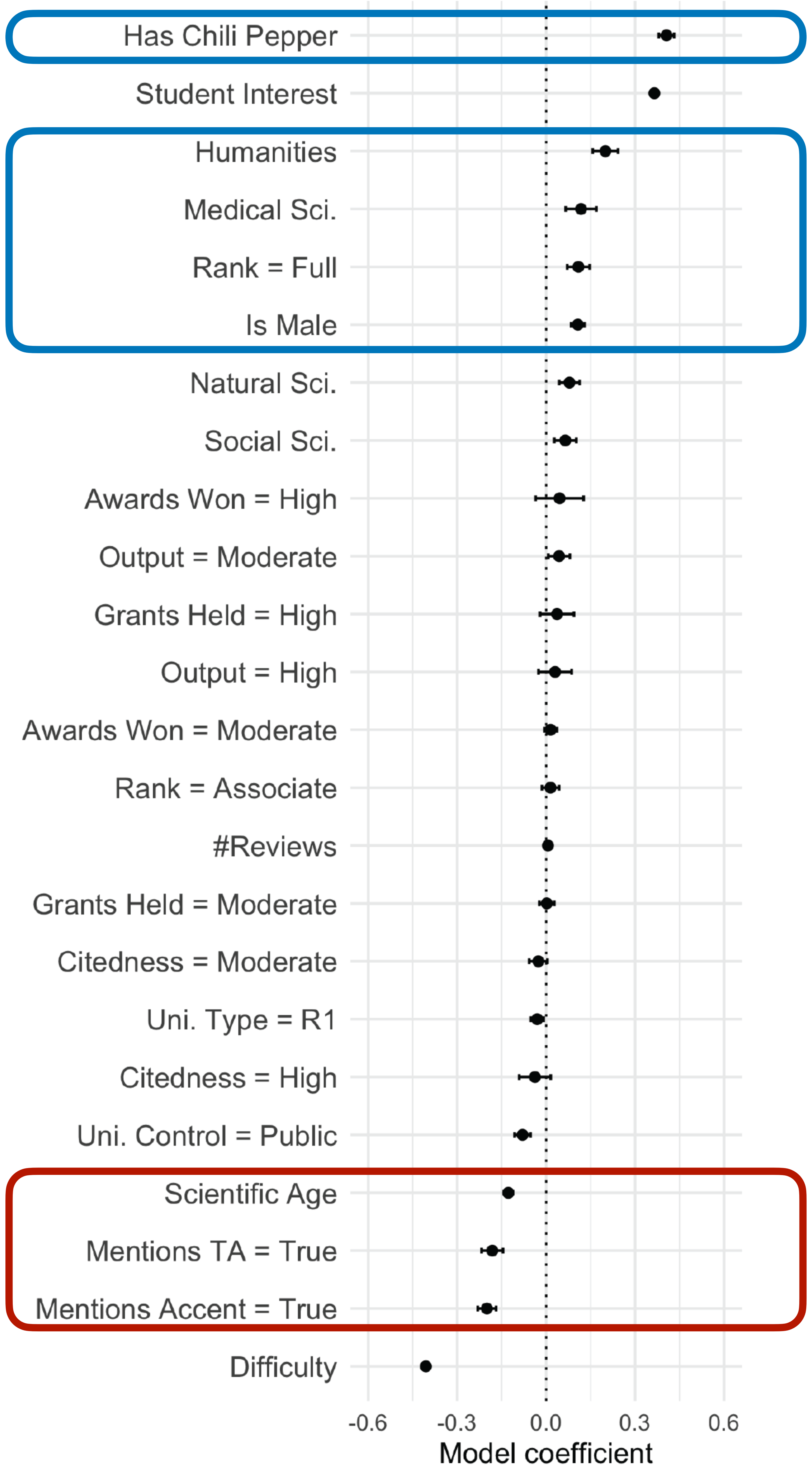
- Attractive profs rated higher
- Profs in humanities, men, and full professors have more positive reviews



Student ratings of teachers

Factors relating to ratings on 19,000 TT faculty on **[RateMyProfessors.com]**

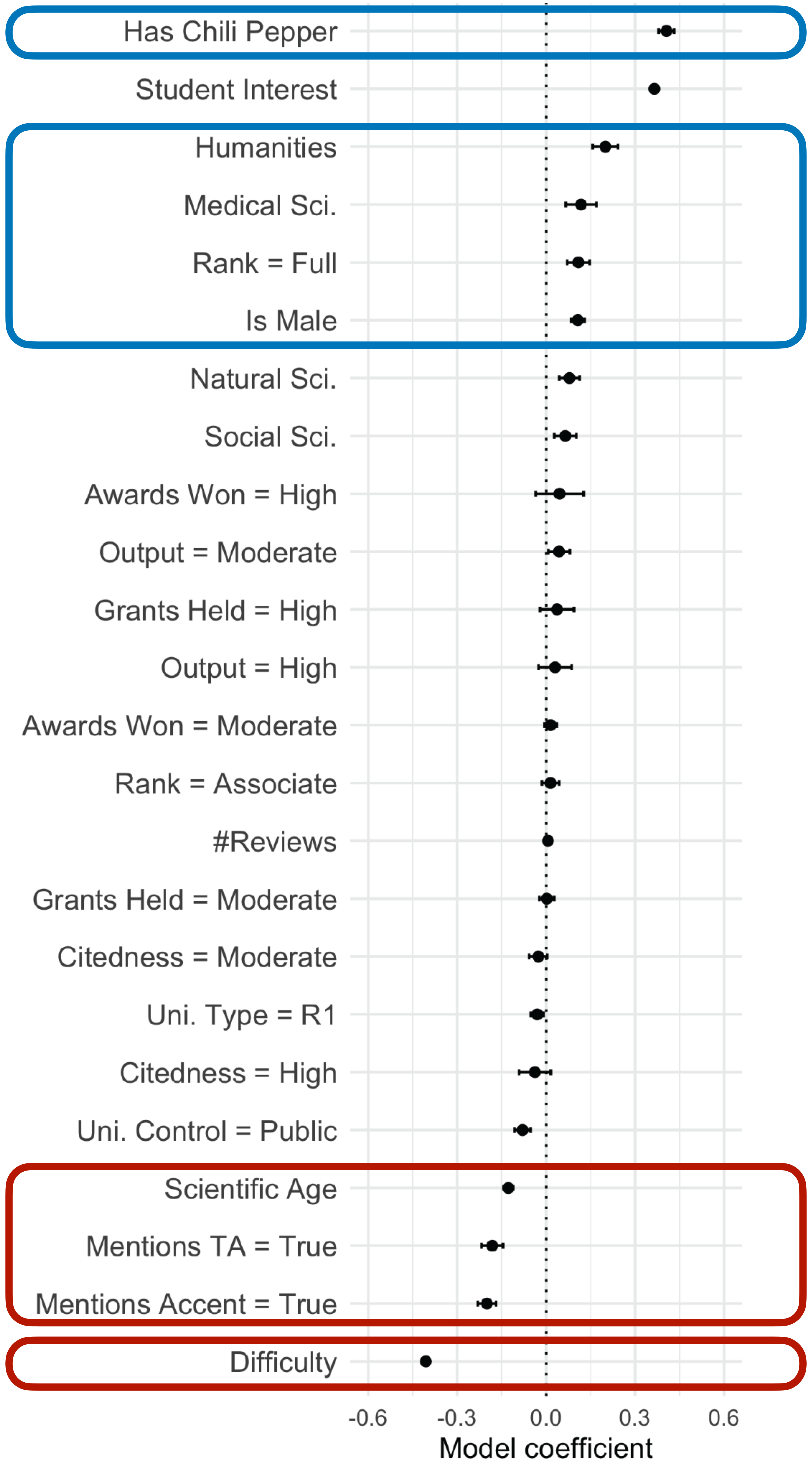
- Attractive profs rated higher
- Profs in humanities, men, and full professors have more positive reviews
- Older Profs, and those for whom an accent or TA was mentioned were rated lower



Student ratings of teachers

Factors relating to ratings on 19,000 TT faculty on RateMyProfessors.com

- Attractive profs rated higher
- Profs in humanities, men, and full professors have more positive reviews
- Older Profs, and those for whom an accent or TA was mentioned were rated lower
- The worst offense is teaching a difficult class



Student-teacher evaluations have deep demographic biases

Women and non-white faculty at a disadvantage

Do they even measure teaching ability?

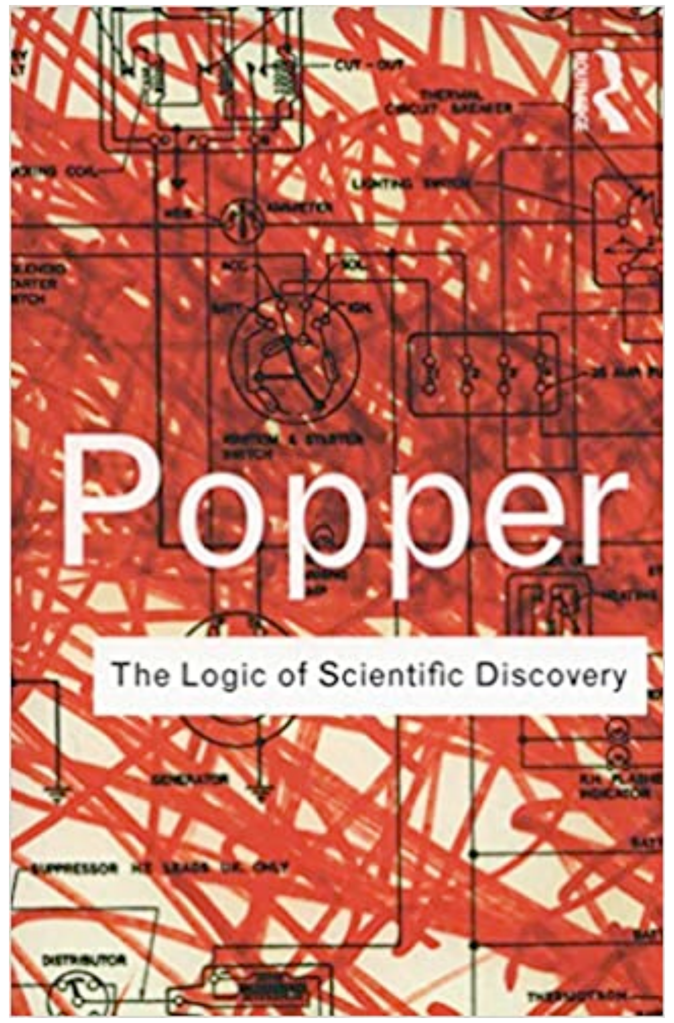
Study 3:

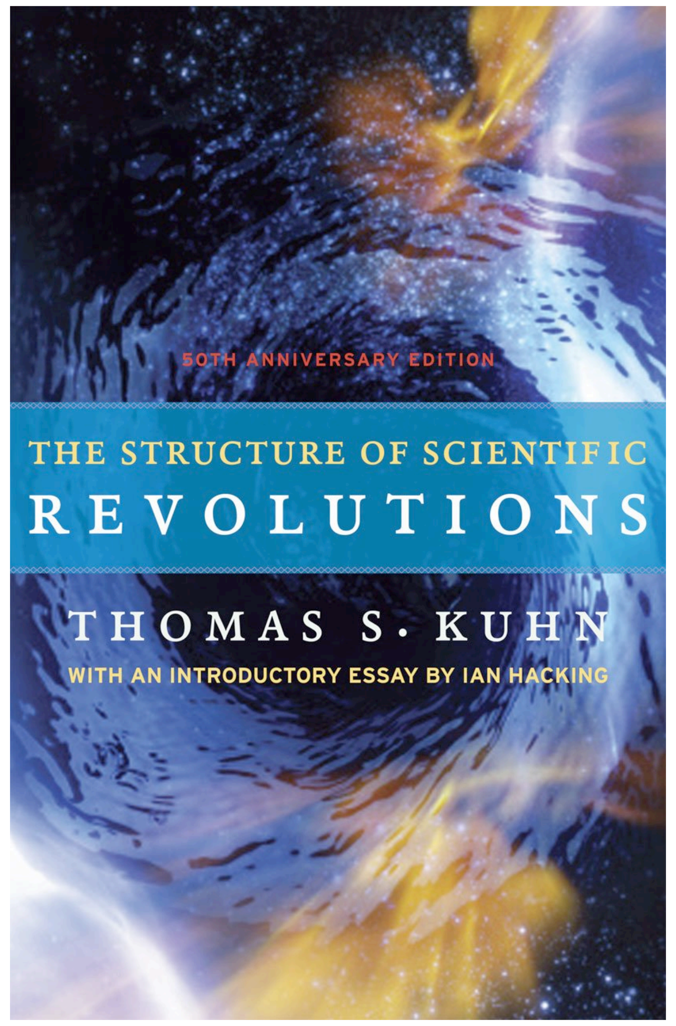
Measuring disagreement across science

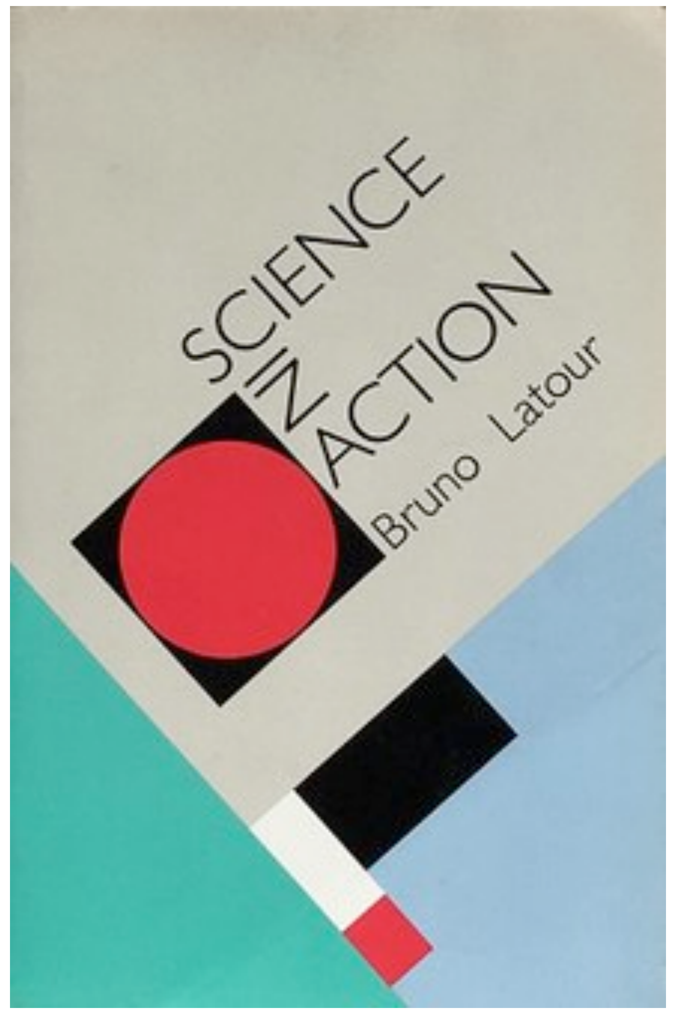
Lamers, W. S., Boyack, K., Larivière, V., Sugimoto, C. R., van Eck, N. J., Waltman, L., & **Murray, D**. (2021). *Measuring Disagreement in Science*.

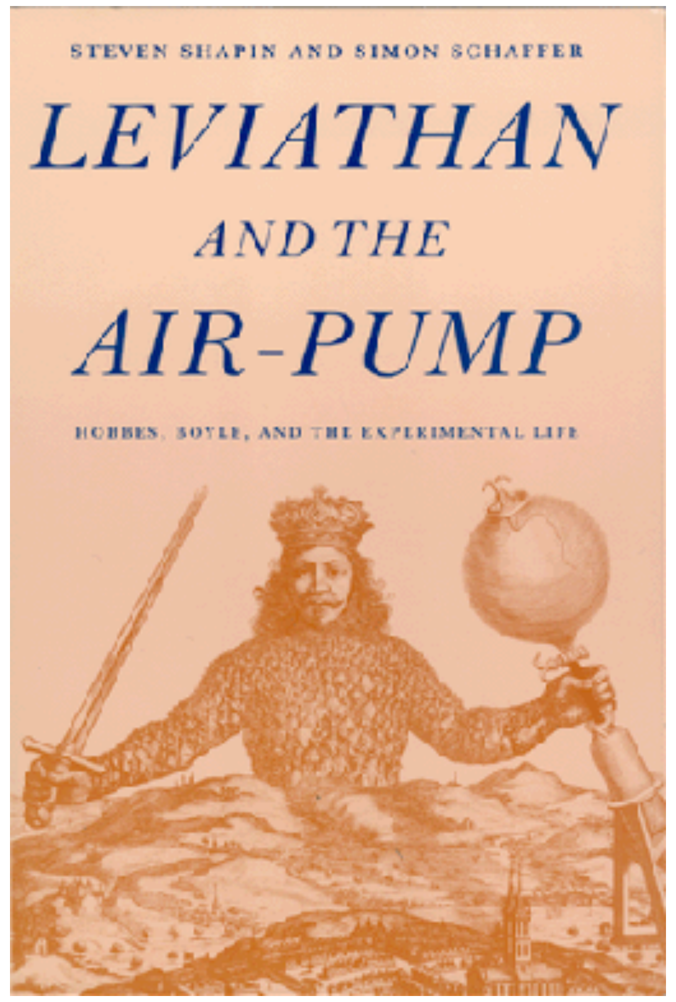
Disagreement is pervasive, and maybe even essential to science

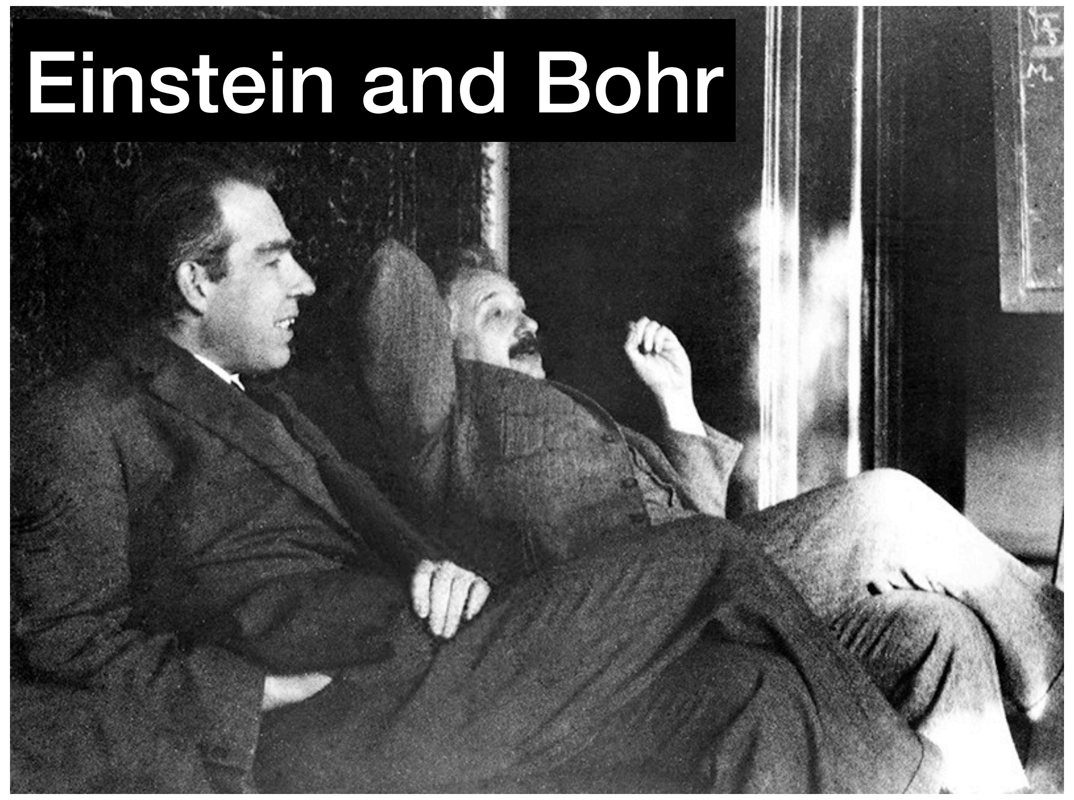
Featuring prominently in theories and histories of science







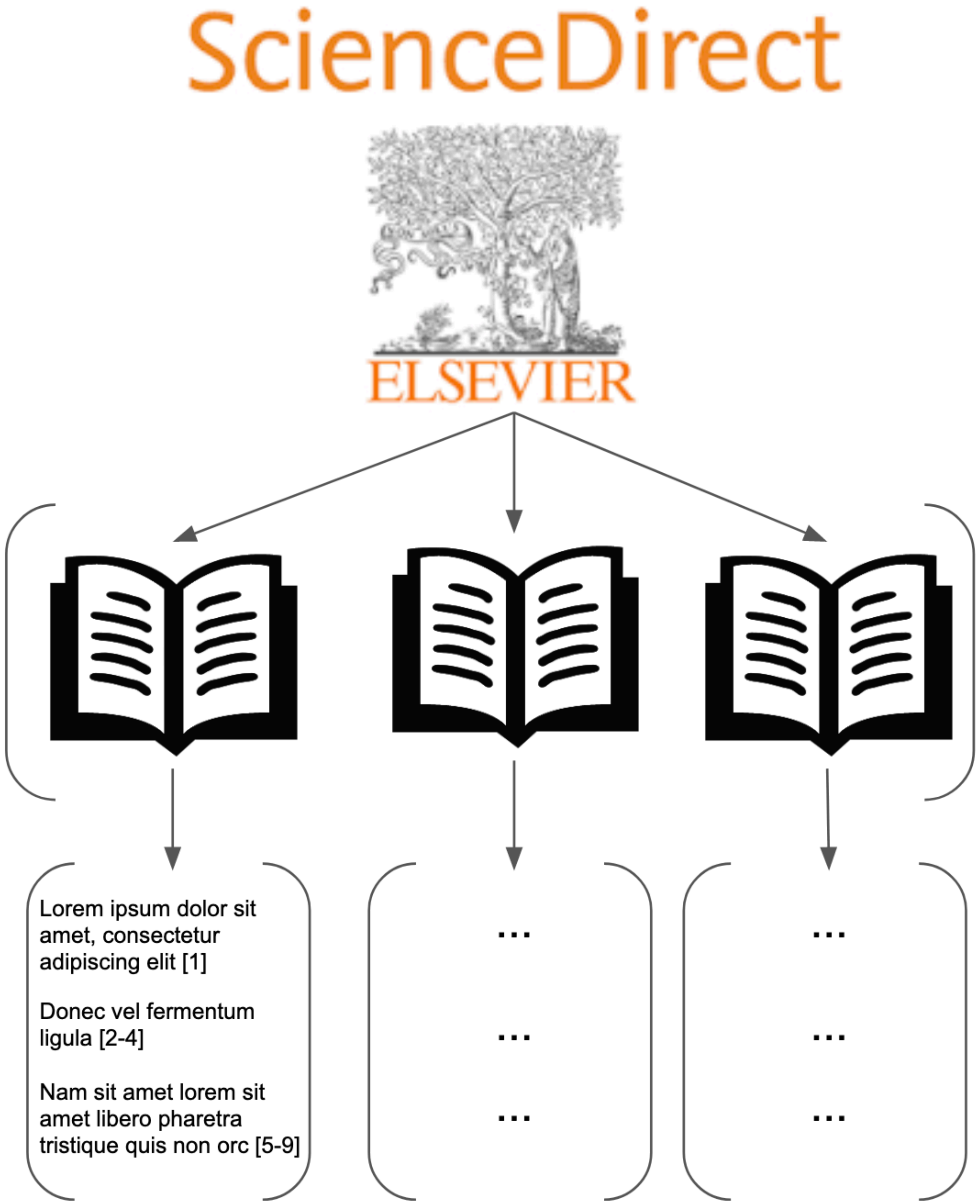




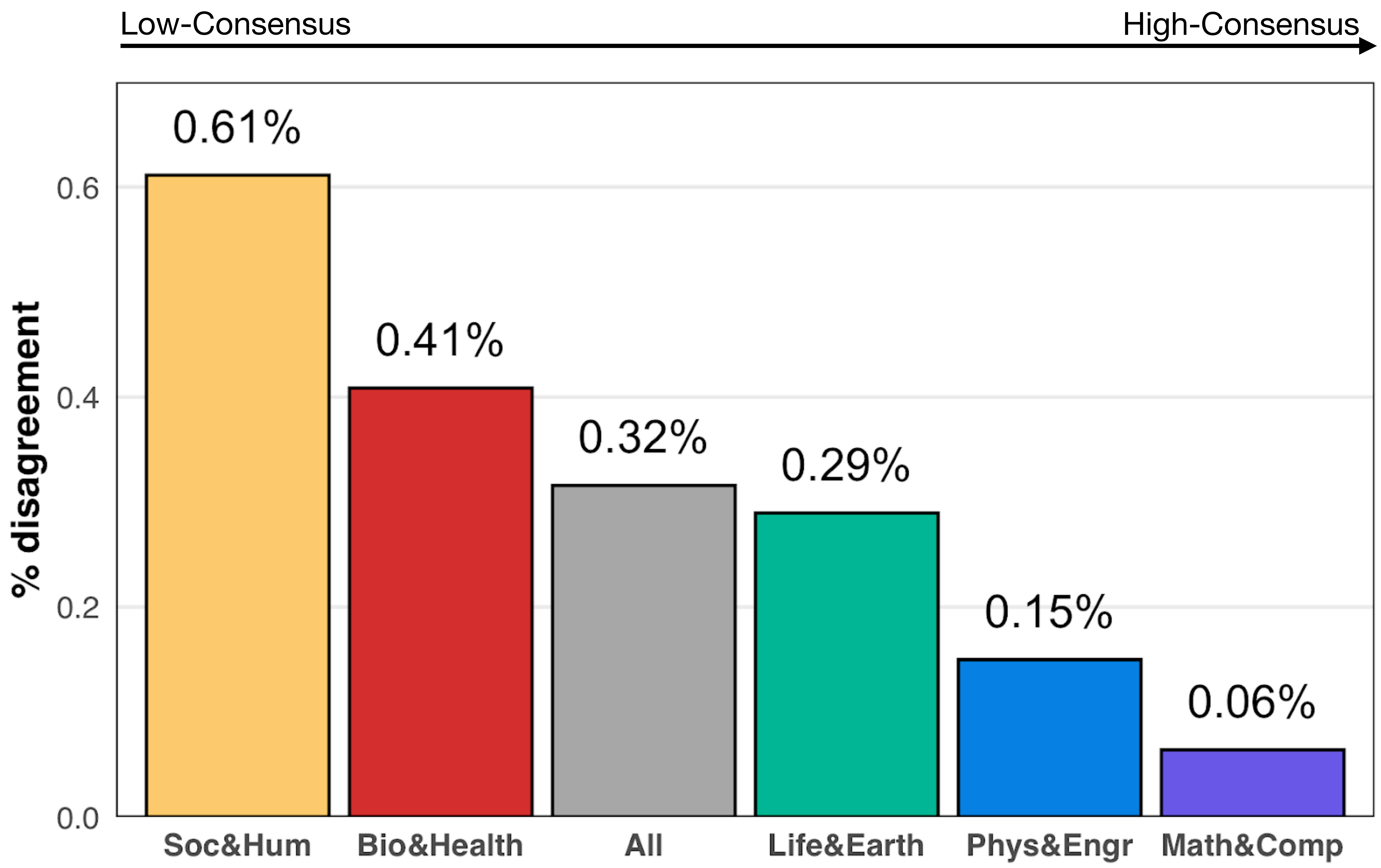
Data

Text-mining disagreement

- Disagreement between texts
- Extract citation sentences
- Over 3 million full-text English-language article
- Develop a cue-word based approach for identifying instances of disagreement
- Do fields disagree…differently?

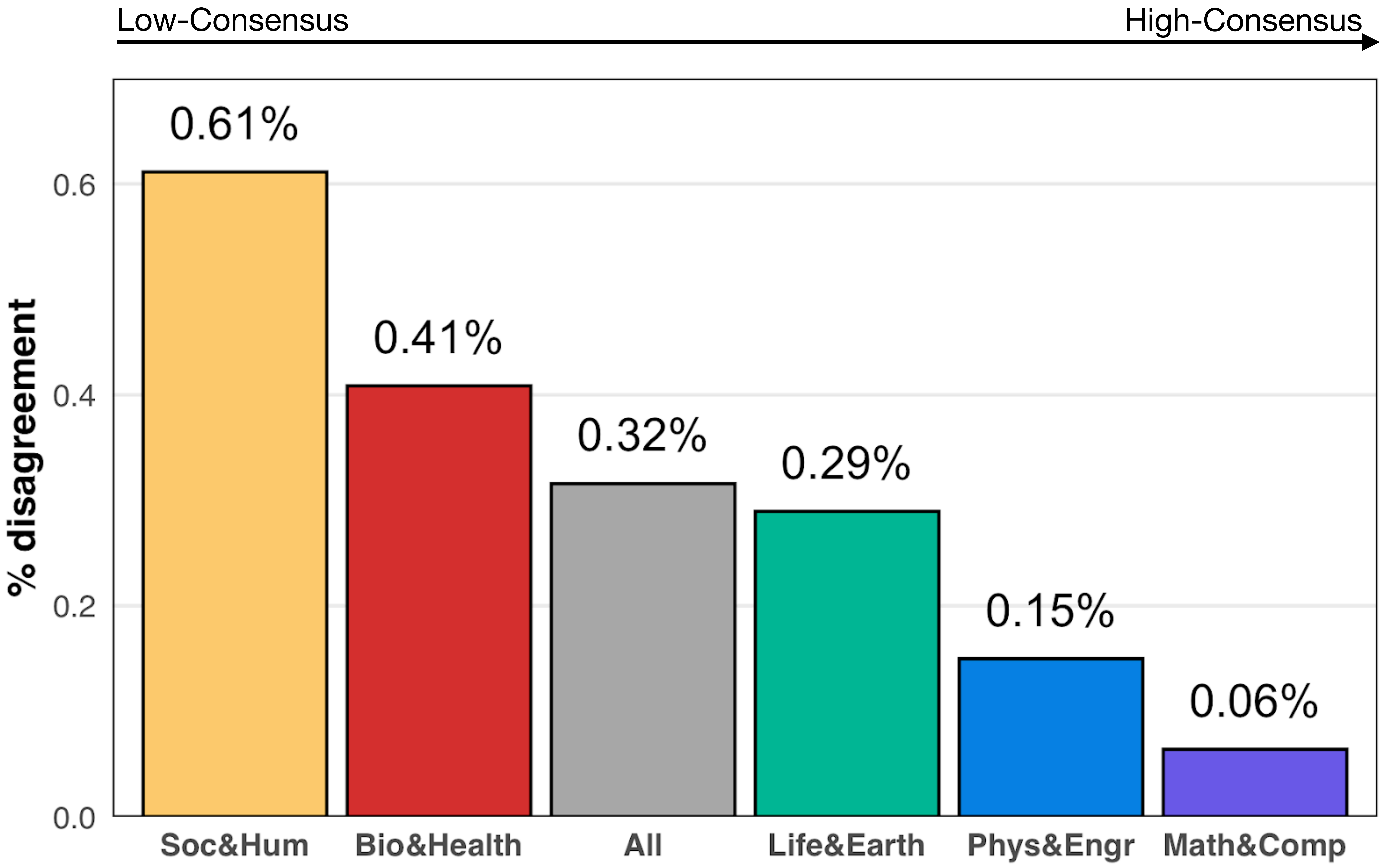


Disagreement by discipline



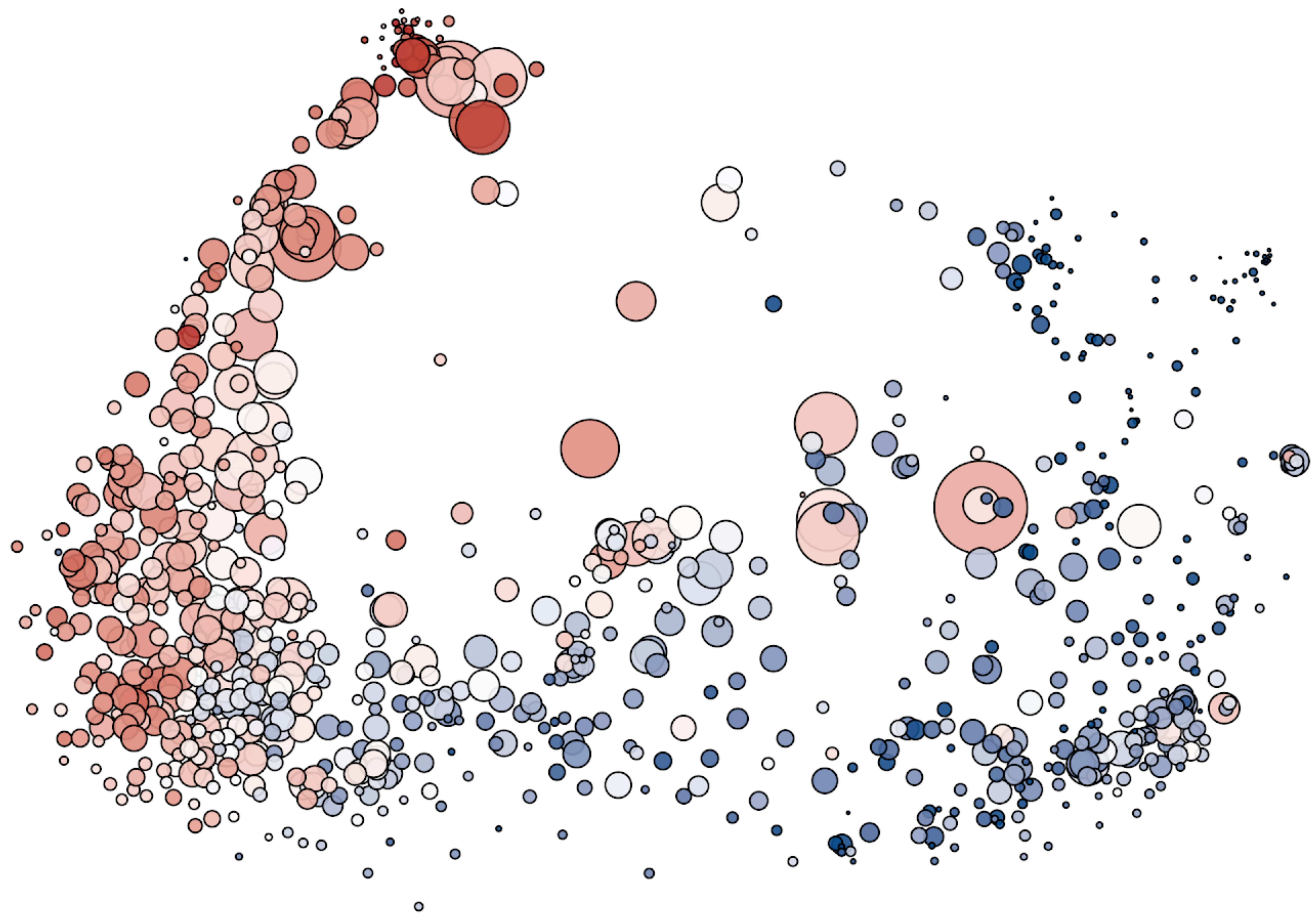
Disagreement by discipline

Too coarse?



More detail

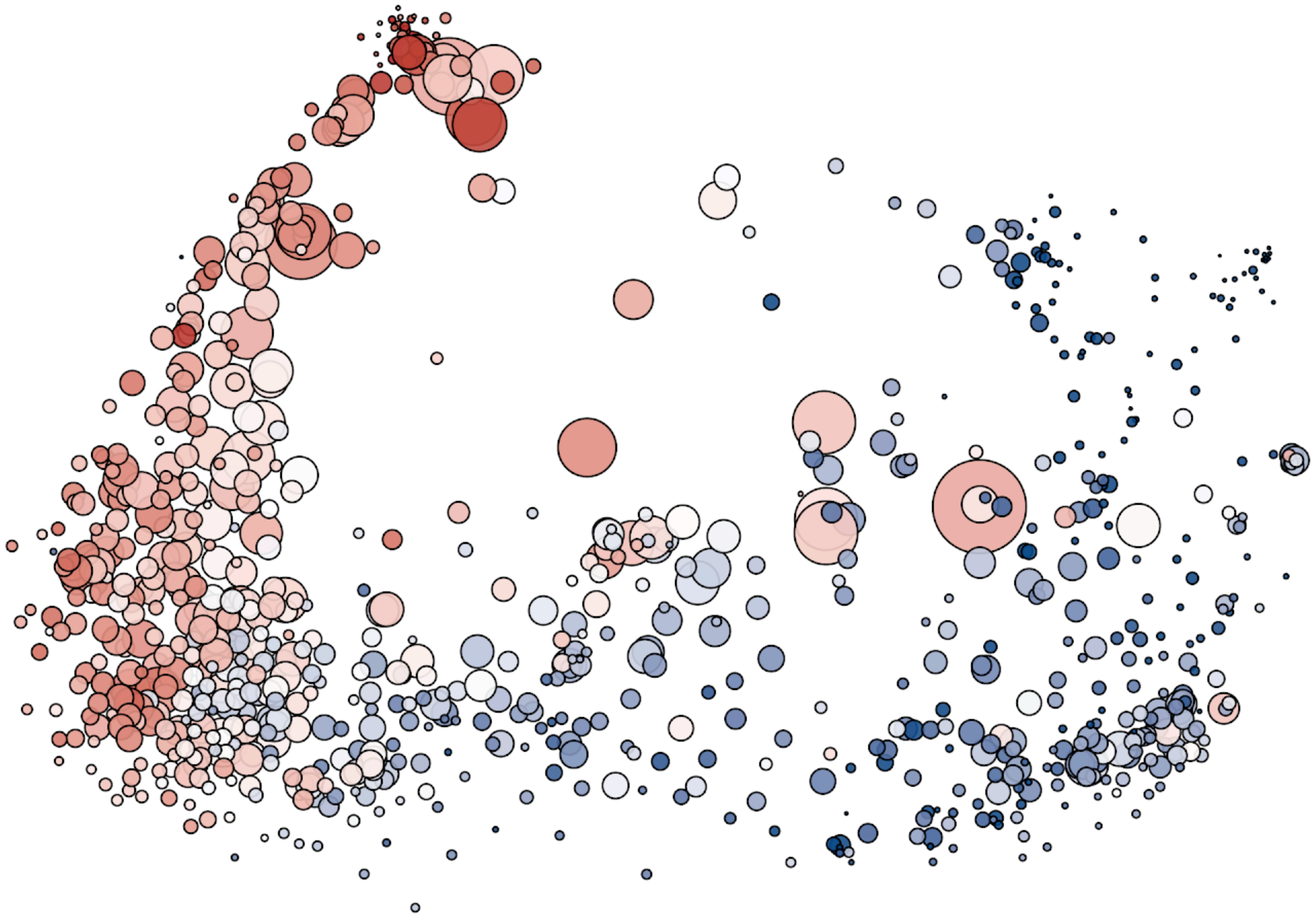
886 meso-level fields



More detail

886 meso-level fields

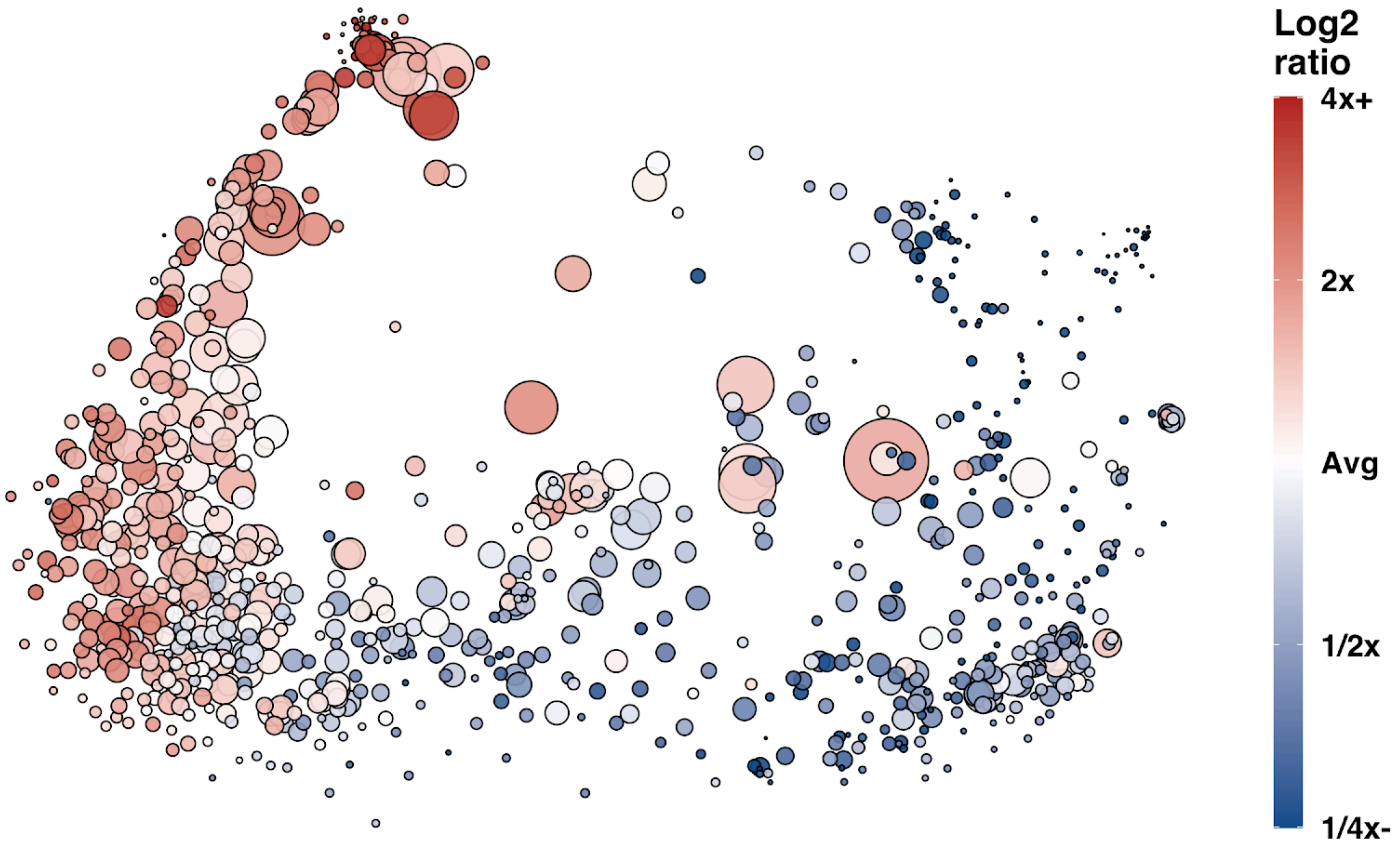
- Each dot is a cluster of papers
- Area maps to size of field
- Distance reflects relatedness



More detail

886 meso-level fields

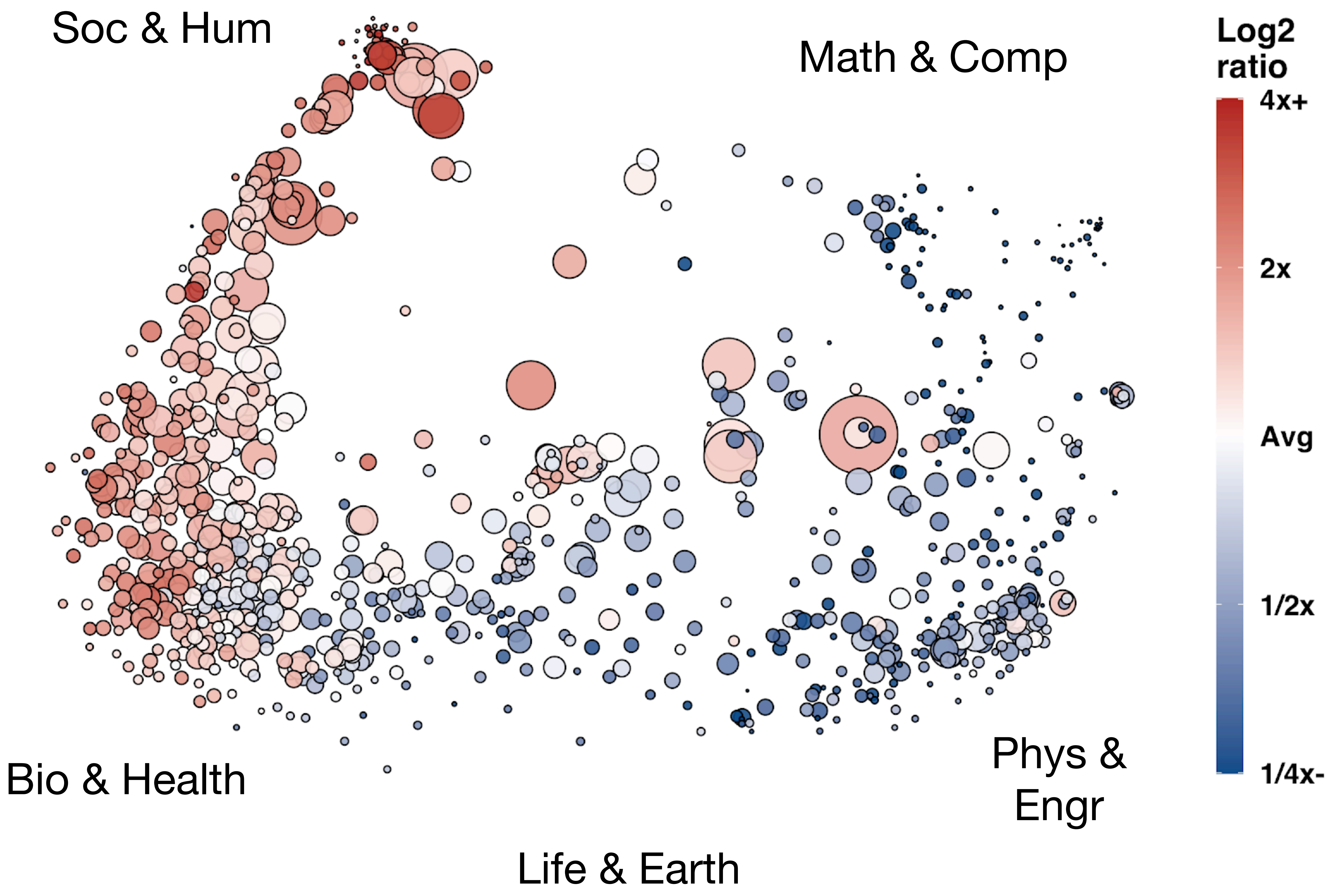
- Each dot is a cluster of papers
- Area maps to size of field
- Distance reflects relatedness
- Color reflects ratio of disagreement to overall average



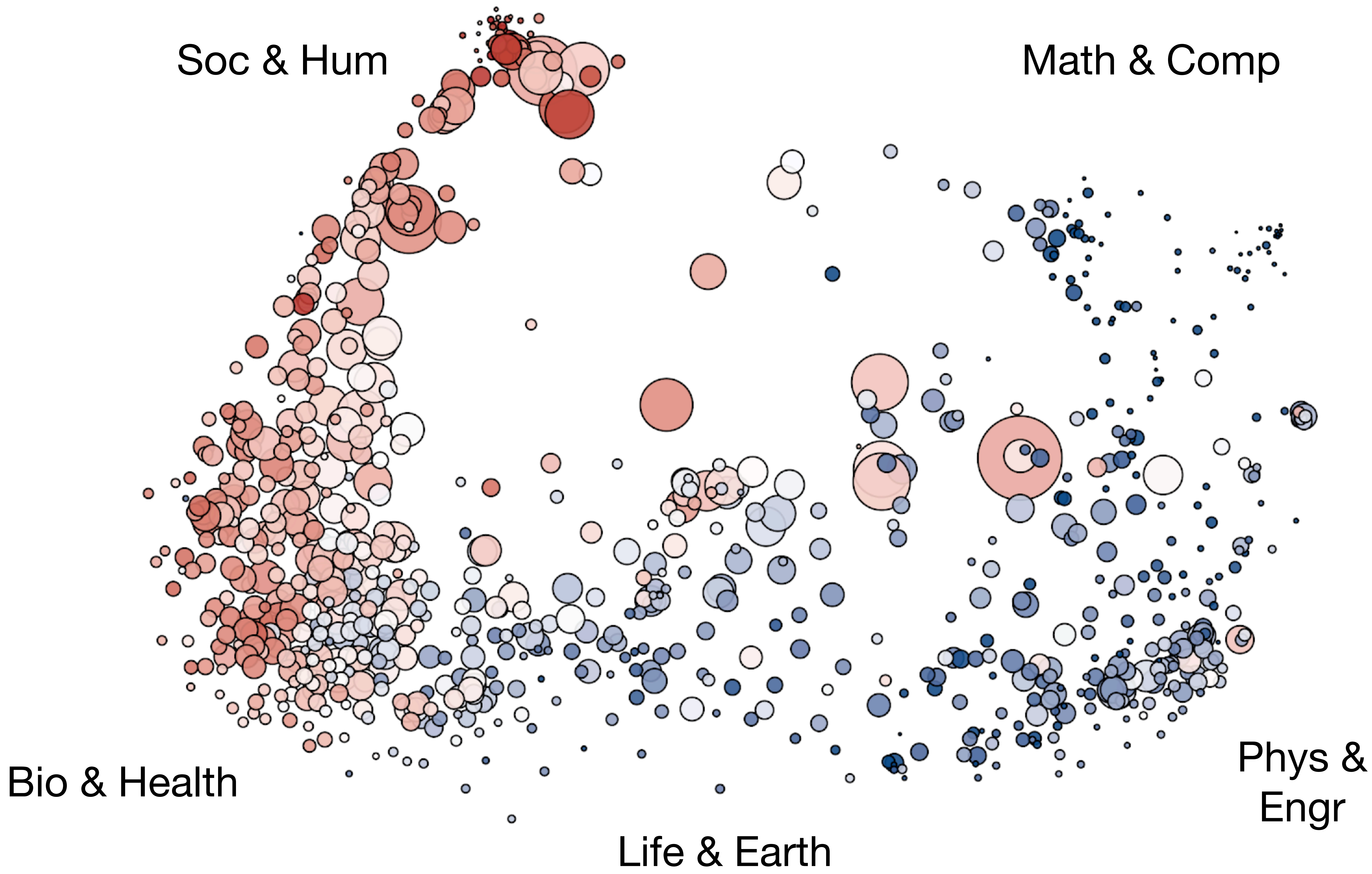
More detail

886 meso-level fields

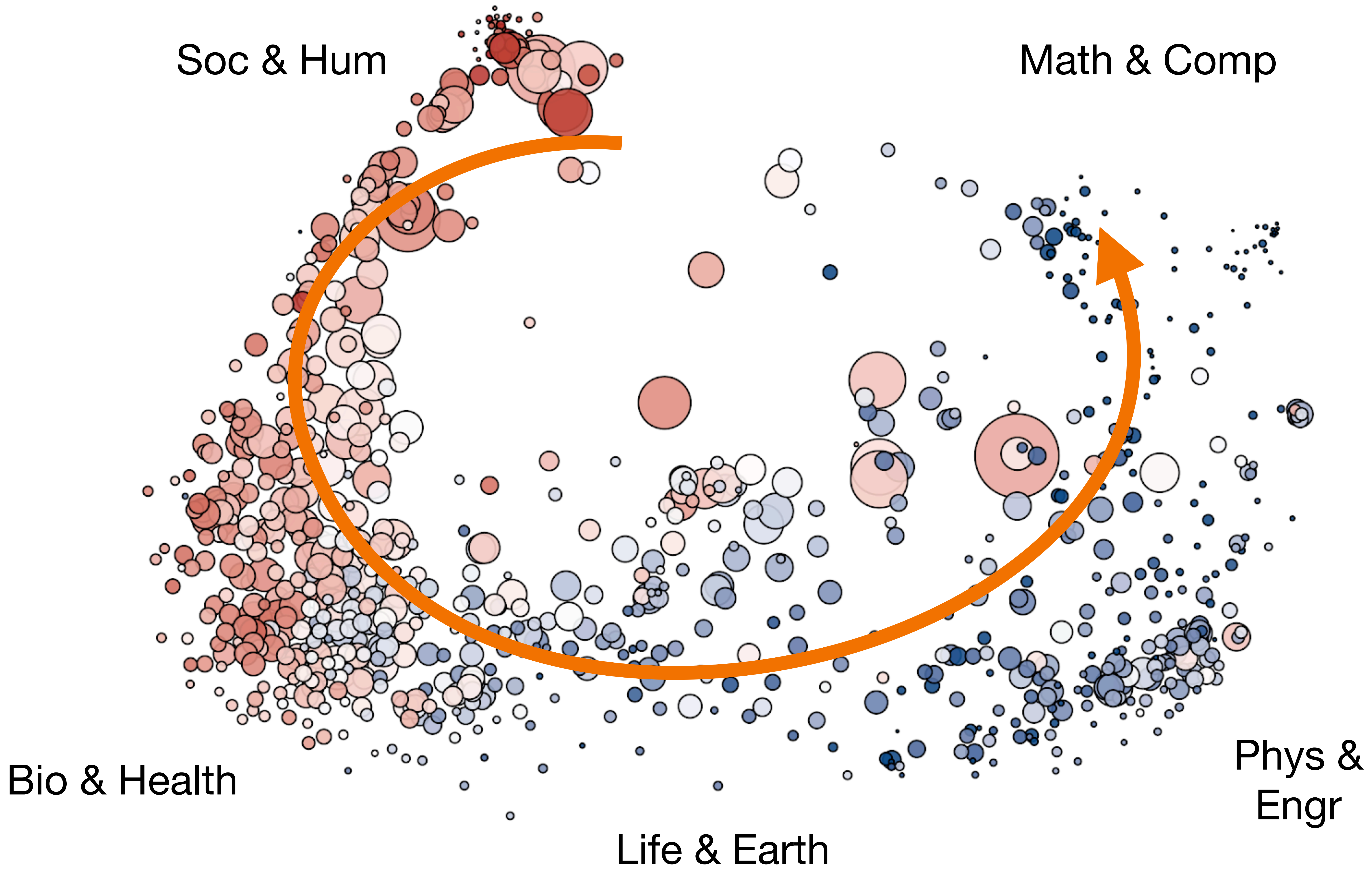
- Each dot is a cluster of papers
- Area maps to size of field
- Distance reflects relatedness
- Color reflects ratio of disagreement to overall average



The pattern repeats

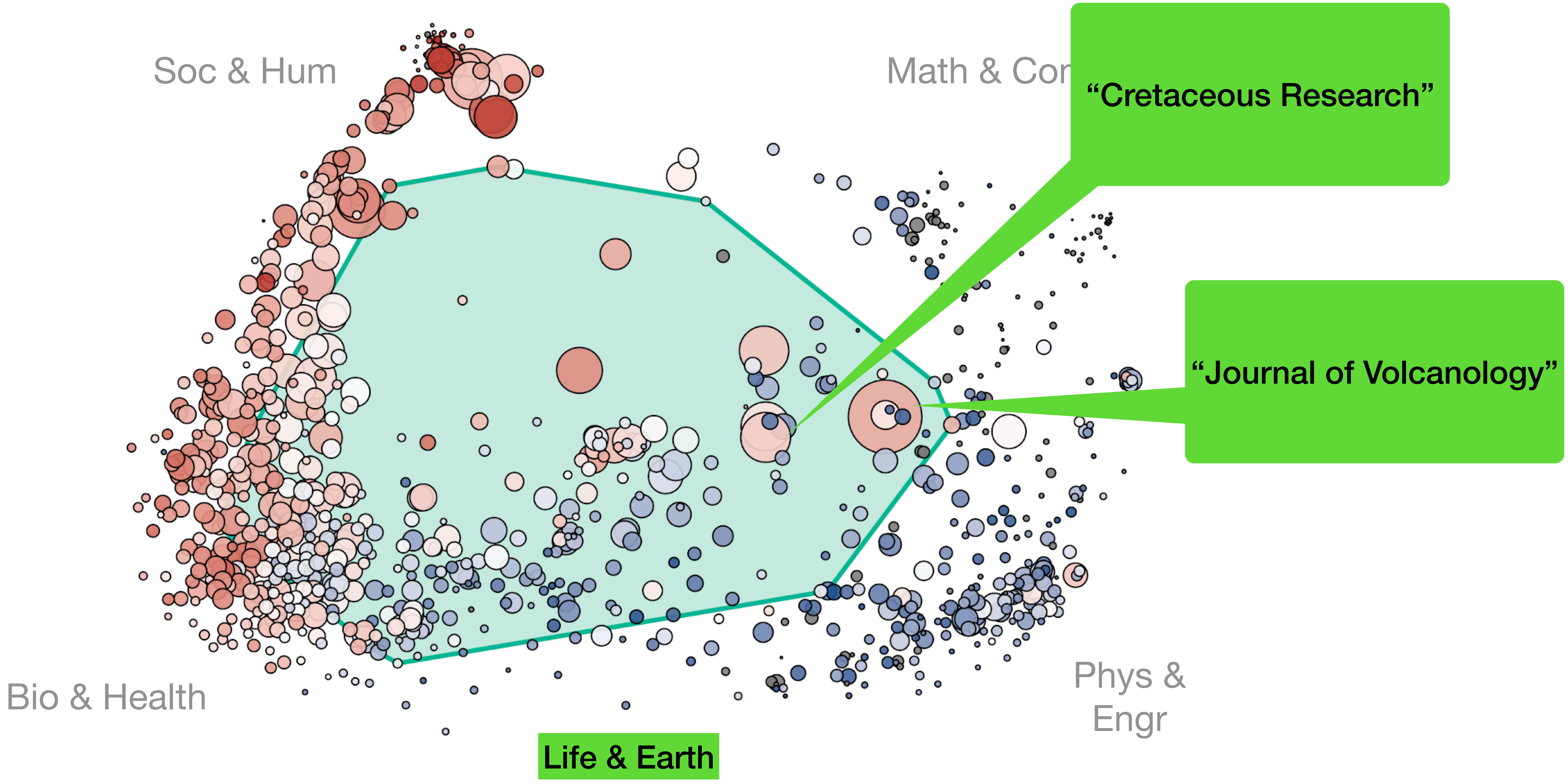


The pattern repeats



Heterogeneity

Fields depending on historical records



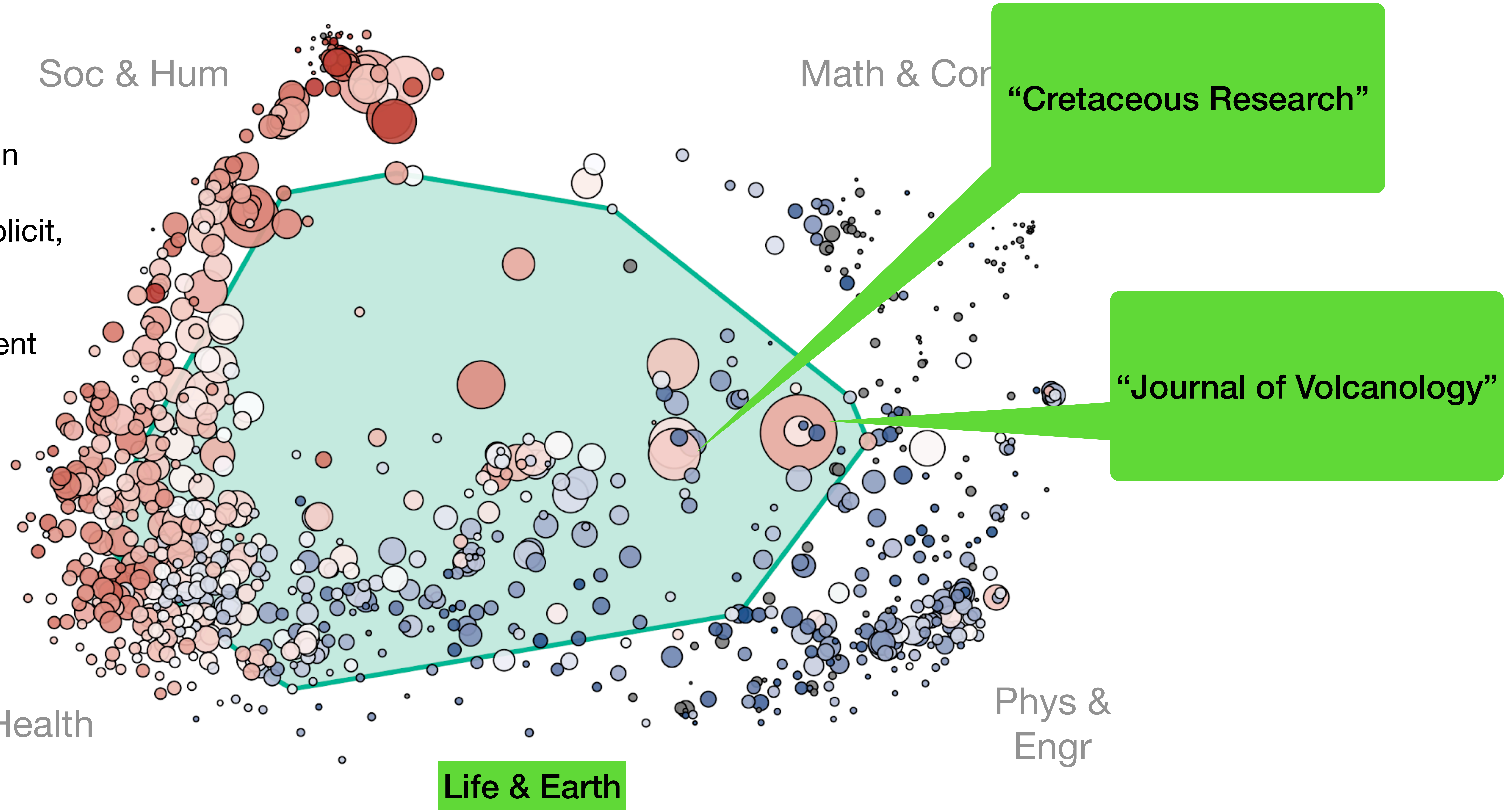
Heterogeneity

Fields depending on historical records

Historical records: Incomplete, subjective, open to interpretation

Experiments: reproducible, explicit, more *“objective”*

May govern rates of disagreement



Disagreement in science follows a spectrum

Yet there is great heterogeneity depending on the characteristics of specific fields

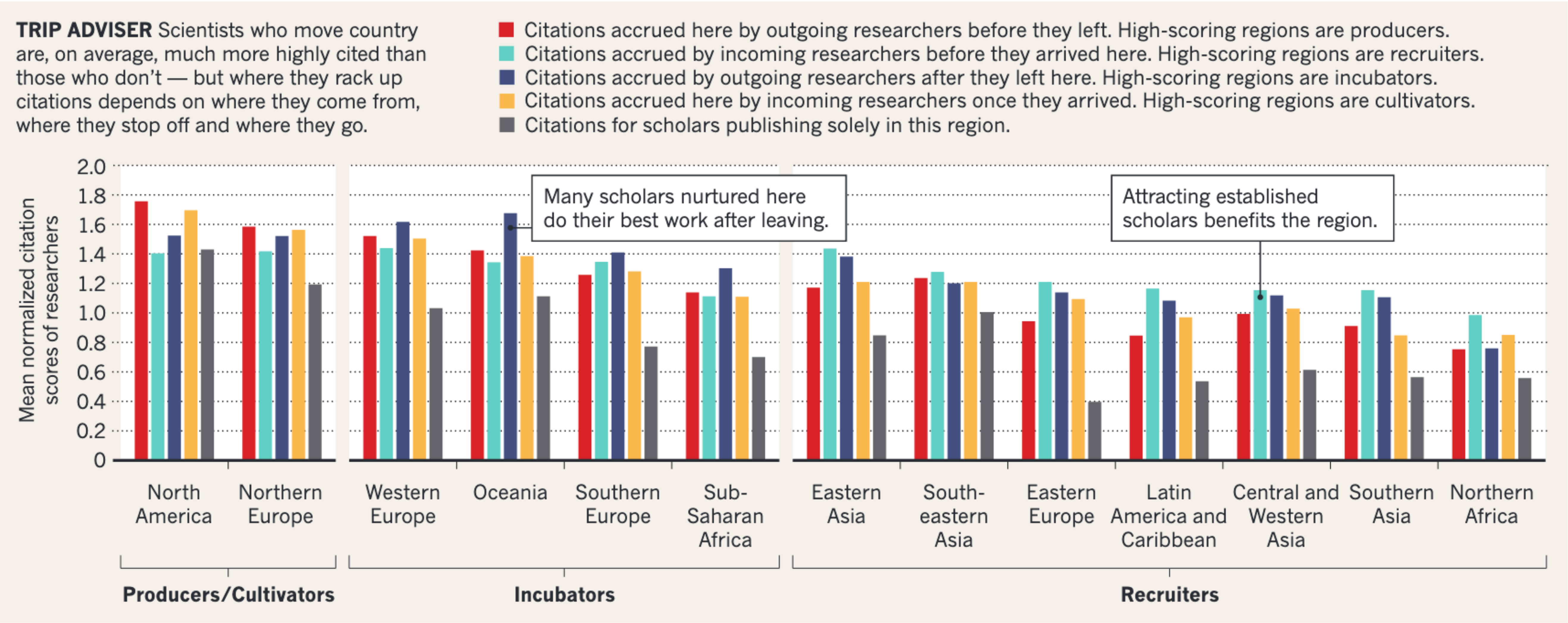
Study 4:

The landscape of global scientific mobility

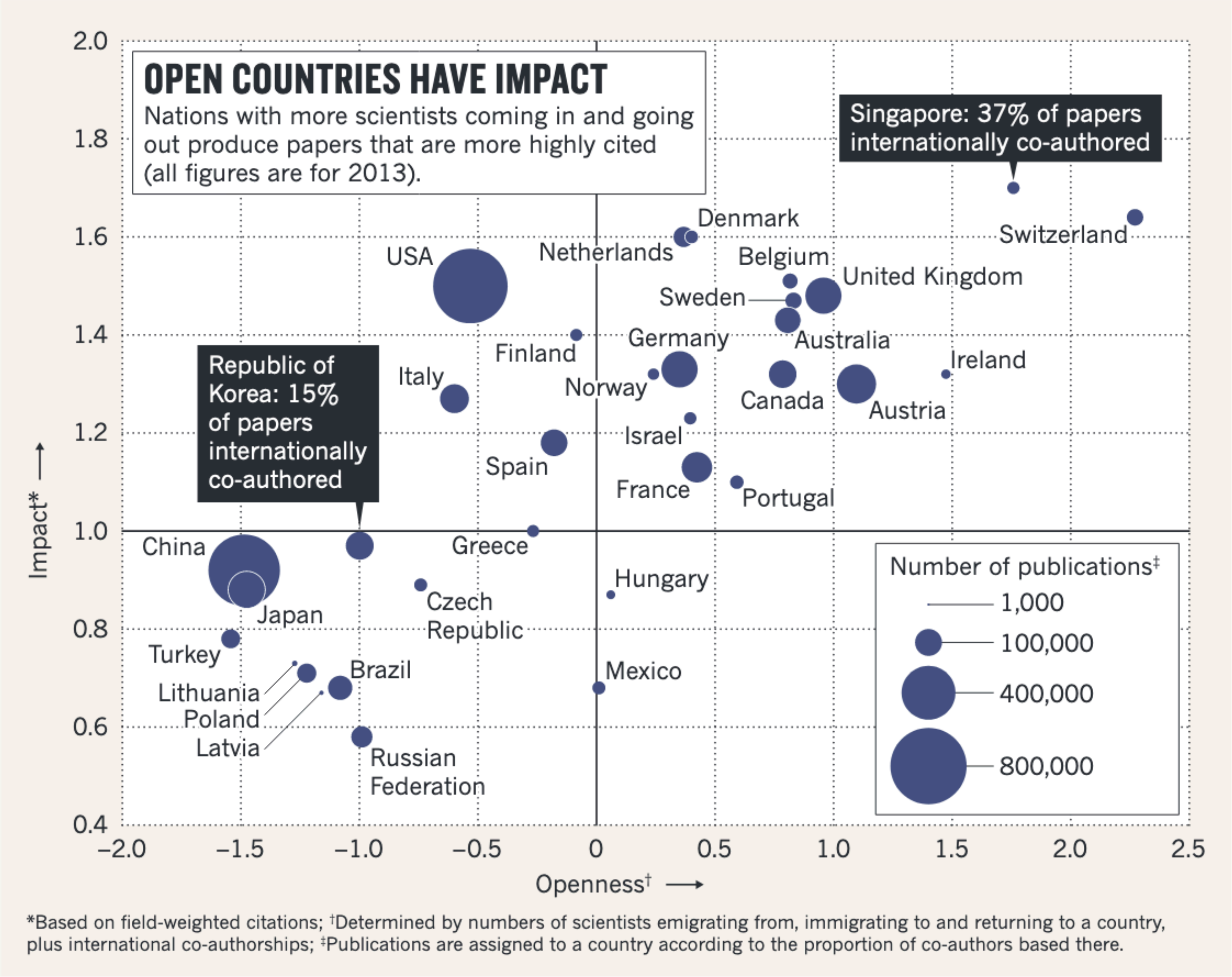
Murray, D., Yoon, J., Kojaku, S., Costas, R., Jung, W.-S., Milojević, S., & Ahn, Y.-Y. (2020). Unsupervised embedding of trajectories captures the latent structure of mobility. *ArXiv:2012.02785 [Physics]*.

Scientific mobility drives impact, collaboration, innovation, and the diffusion of ideas

Mobile scientists have more impact than their non-mobile counterparts



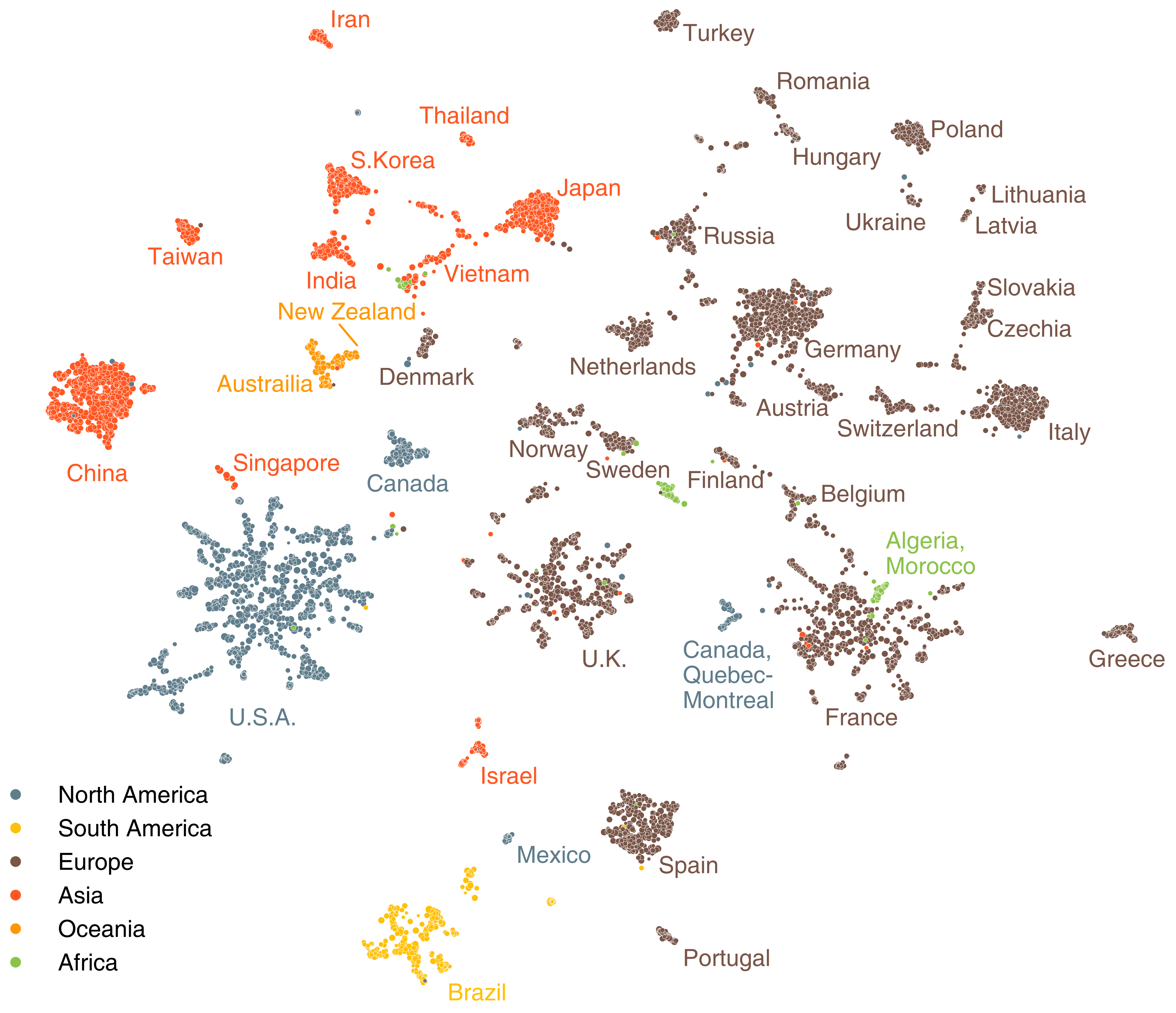
"Open" countries have greater scientific impact



Sugimoto, C. R., Robinson-Garcia, N., Murray, D. S., Yegros-Yegros, A., Costas, R., & Larivière, V. (2017). Scientists have most impact when they're free to move. *Nature News*, *550*(7674), 29.

Wagner, C. S., & Jonkers, K. (2017). Open countries have strong science. *Nature News*, *550*(7674), 32.

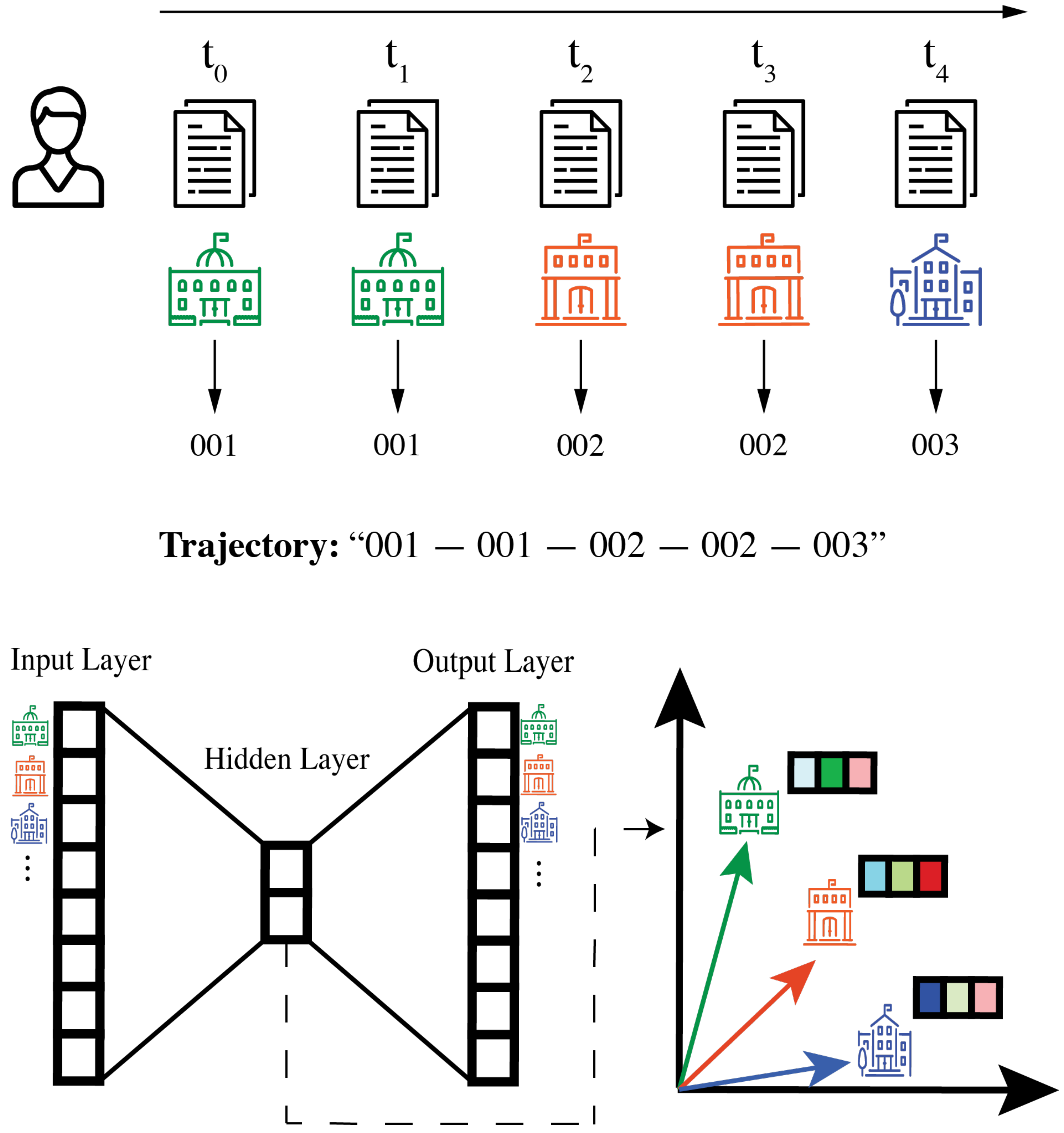
Can we leverage new computational techniques to learn an “embedding” of mobility?



Embedding mobility

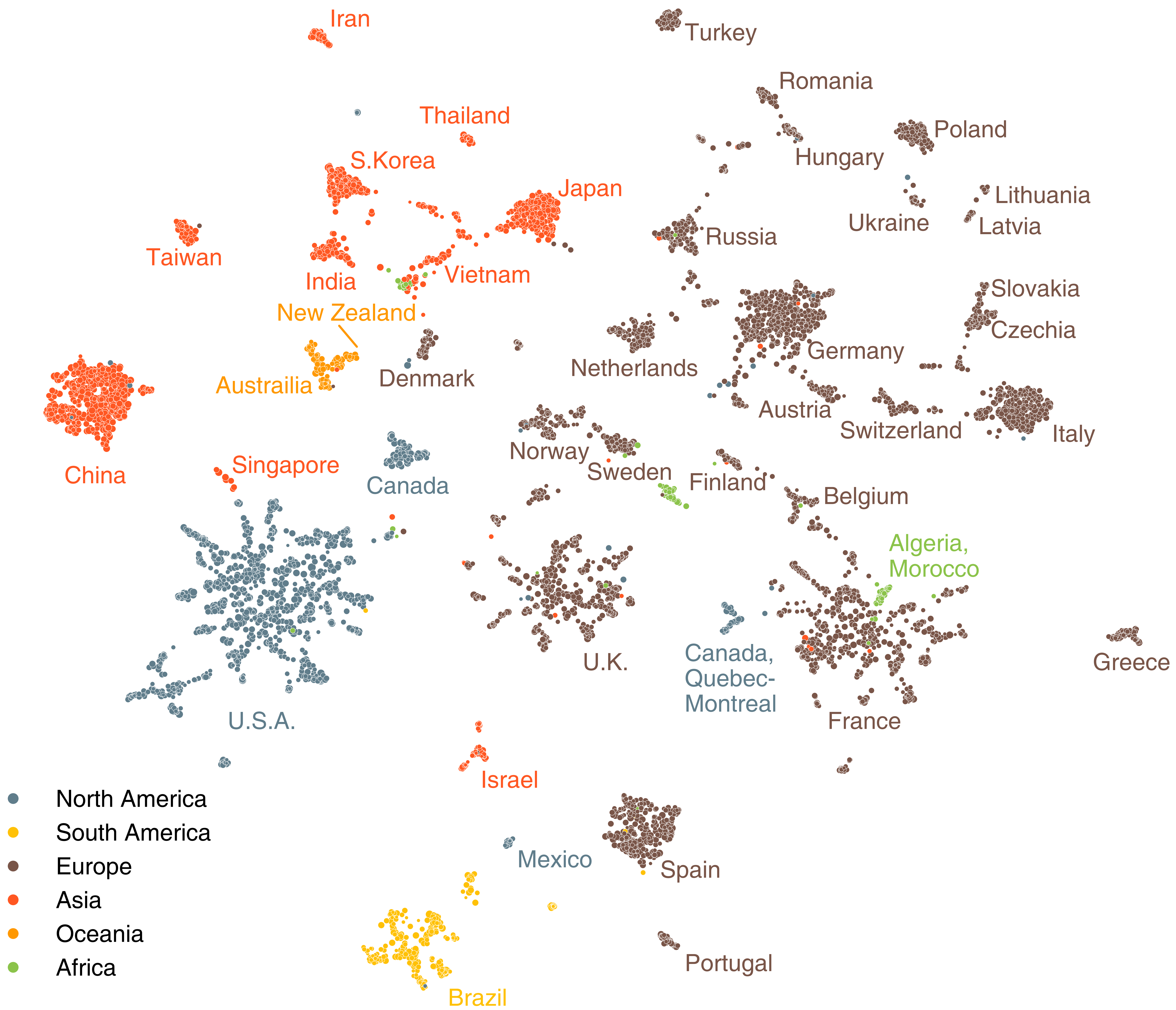
Using NLP tools

- Career trajectories of 3 million scientists derived from publications
- Give as input to *word2vec*
- Produce a dense, vector-space representation of organizations
- Distance reflects affinity, or co-appearence in career trajectories
- Valid representation of mobility



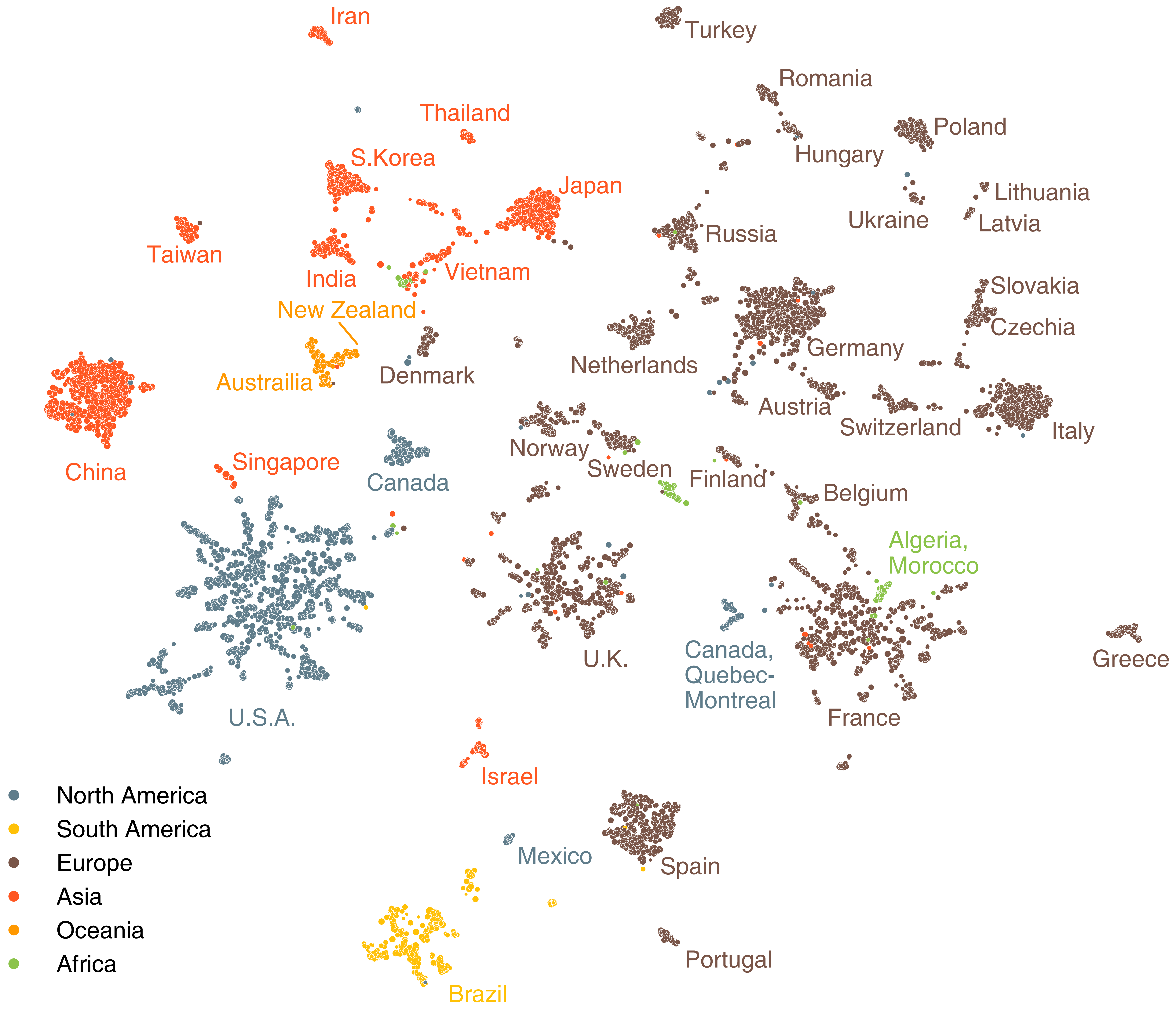
Visualizing the embedding space

UMAP projection of organizations



Visualizing the embedding space

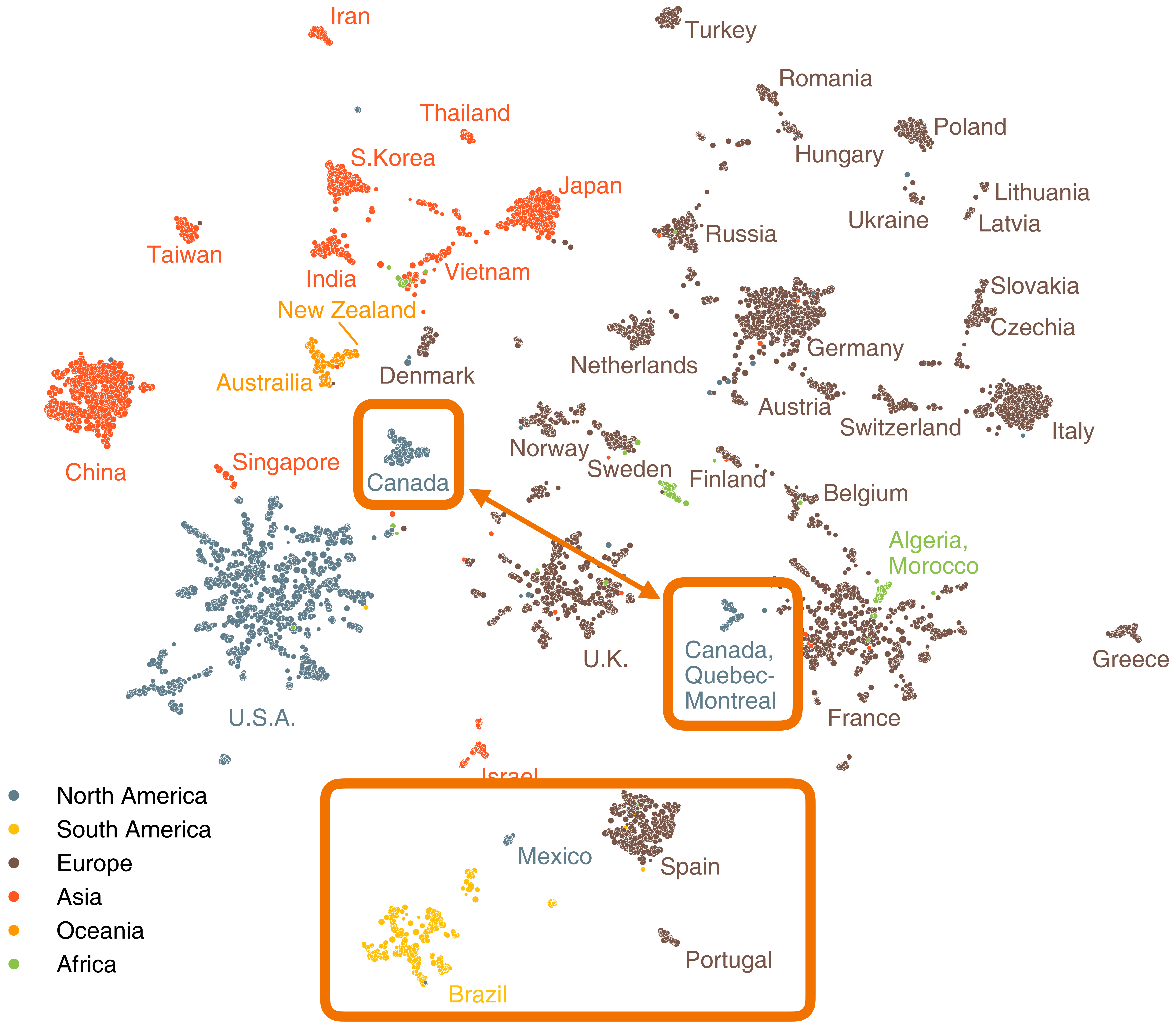
Geography matters



Visualizing the embedding space

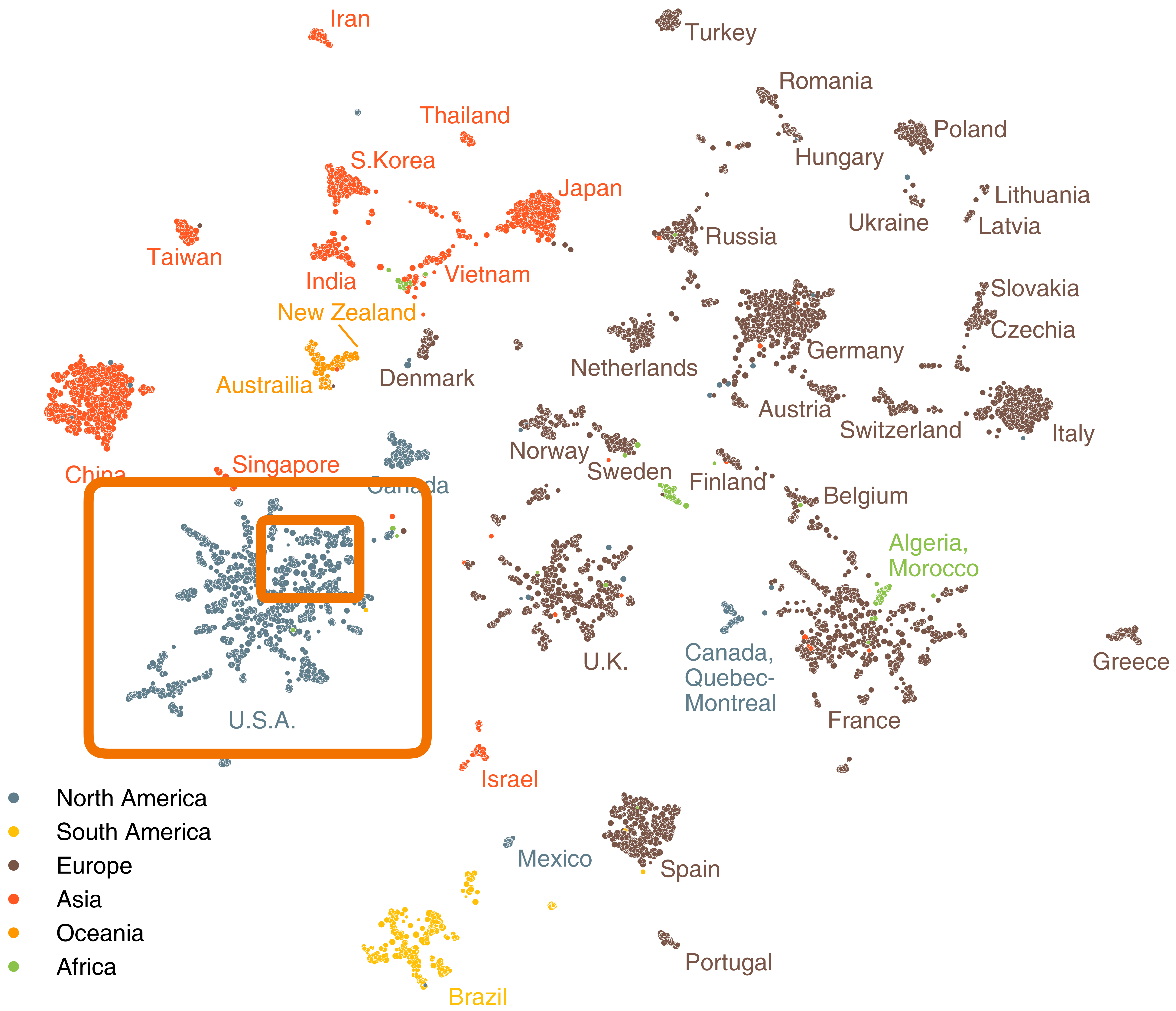
Geography matters

So does language, culture, and shared history



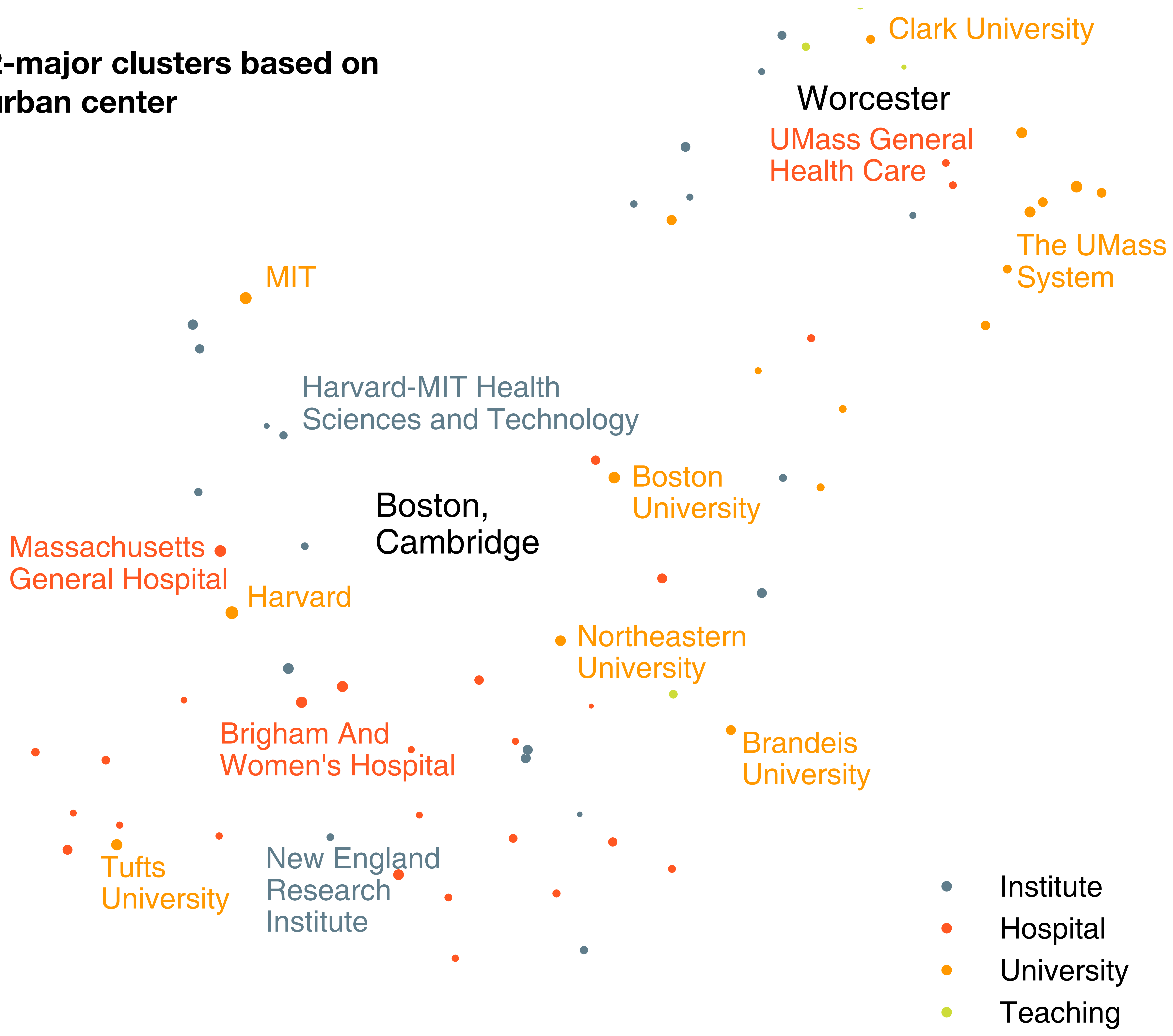
Visualizing the embedding space

We can “zoom in”



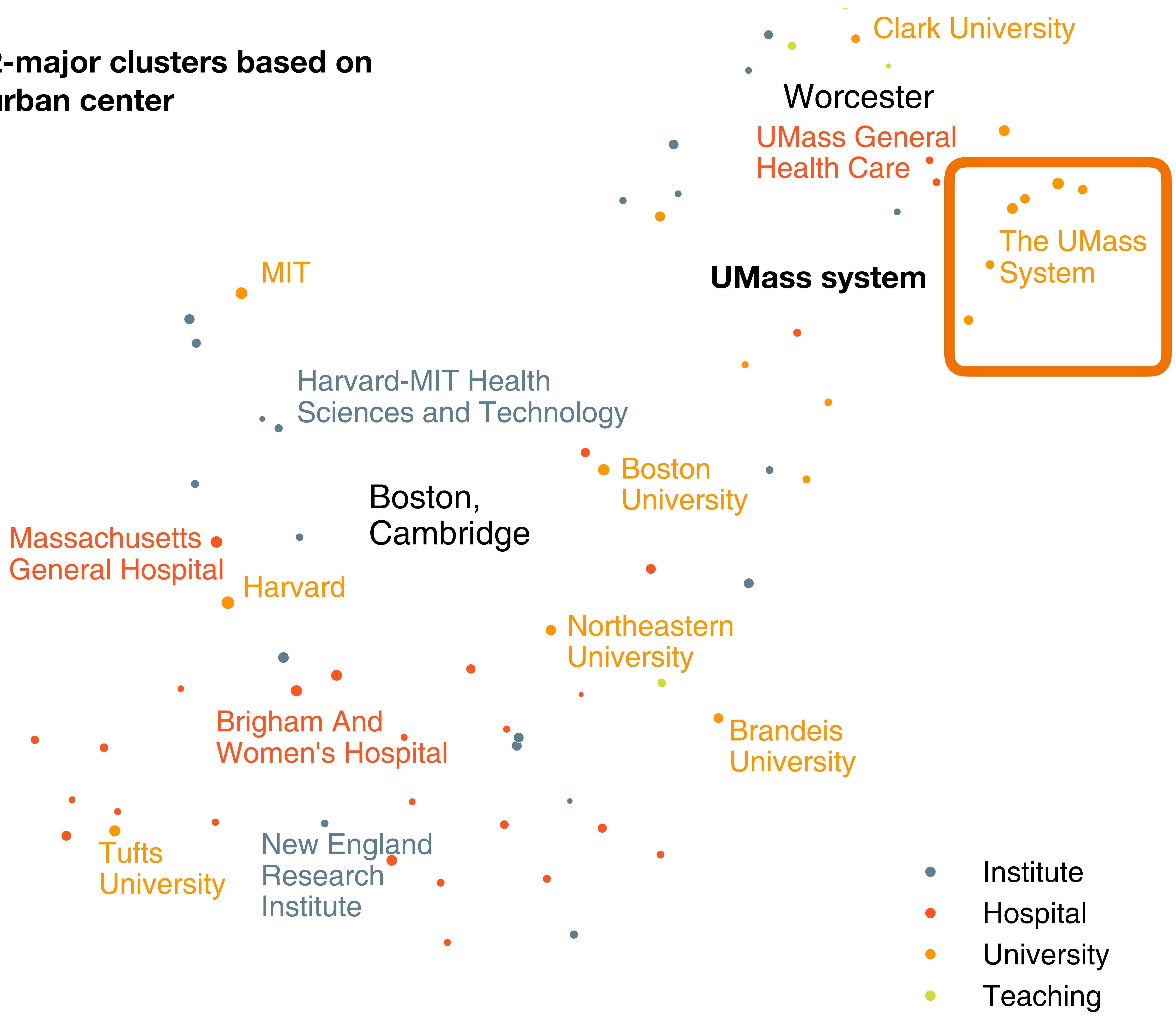
Massachusetts

2-major clusters based on urban center

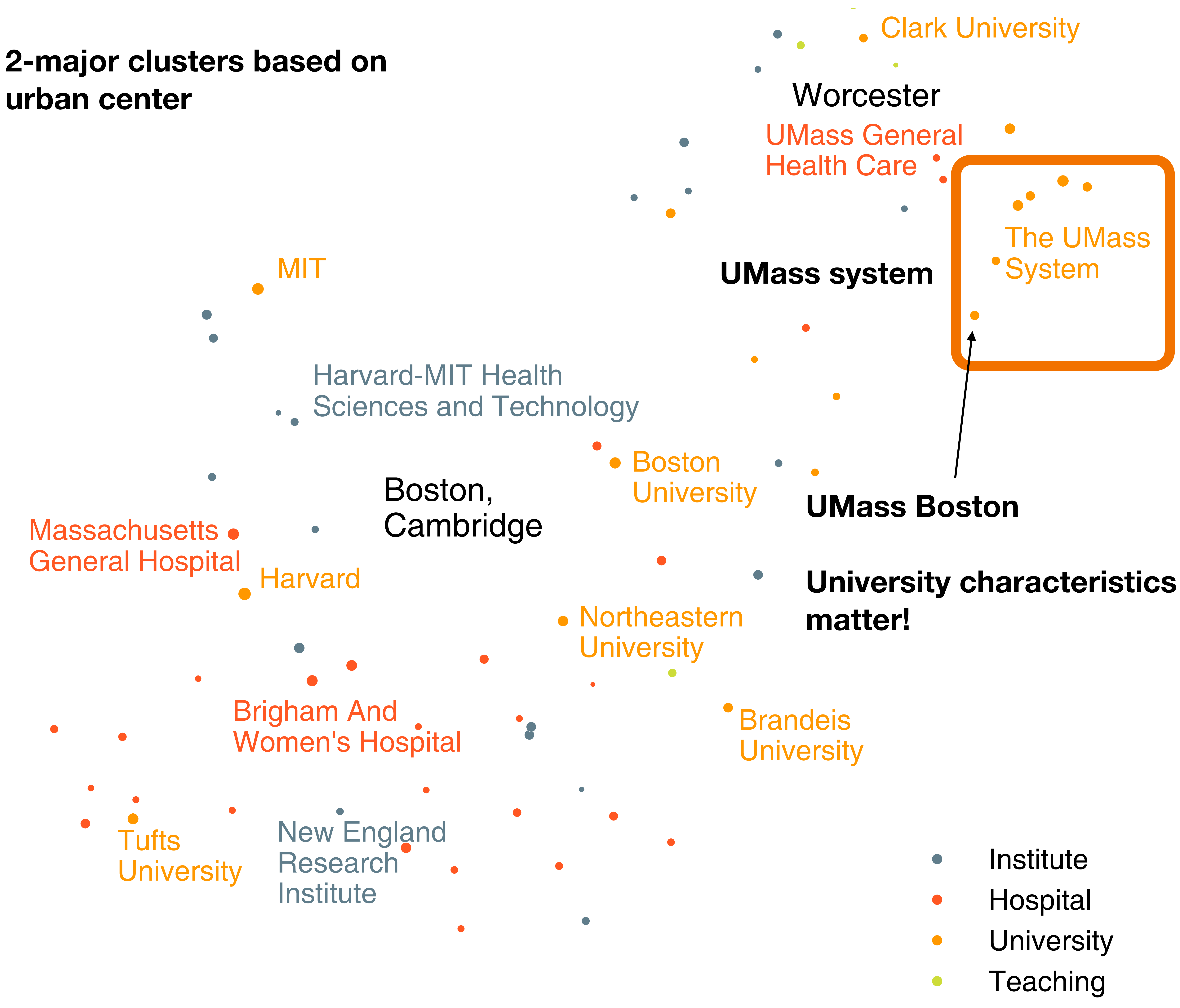


Massachusetts

2-major clusters based on urban center

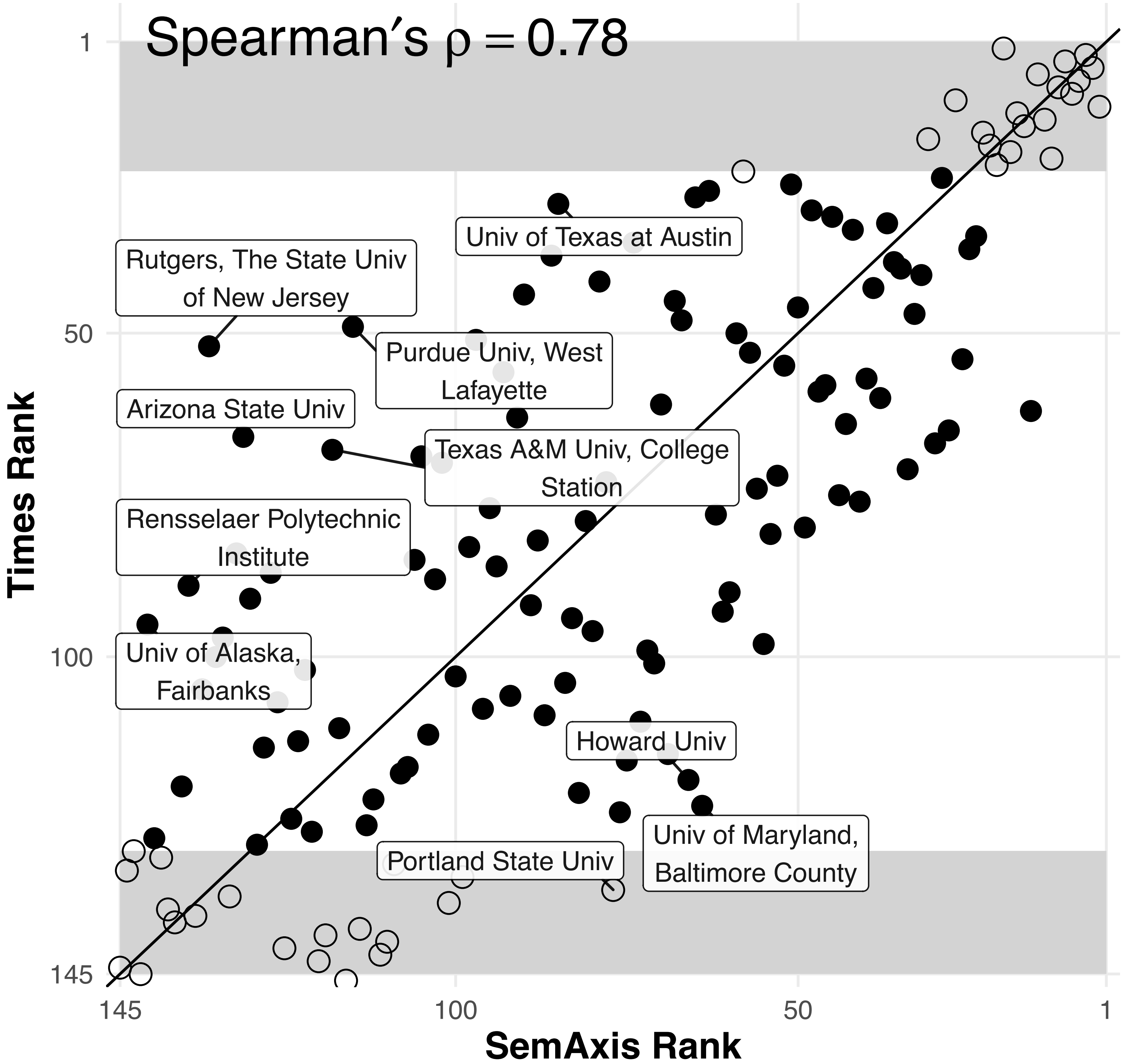


Massachusetts

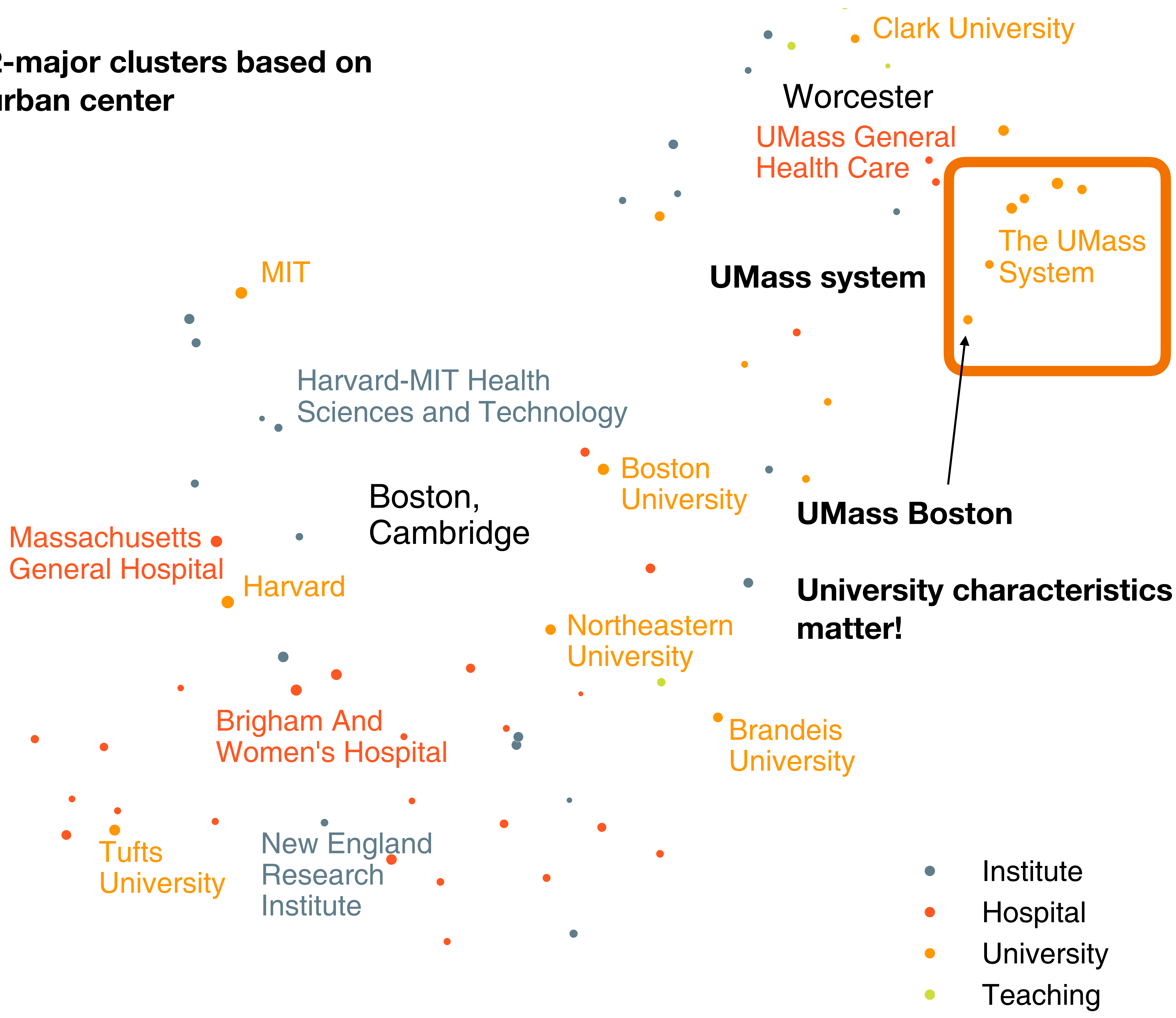


Massachusetts

Rankings can be recovered from the embedding space



2-major clusters based on urban center

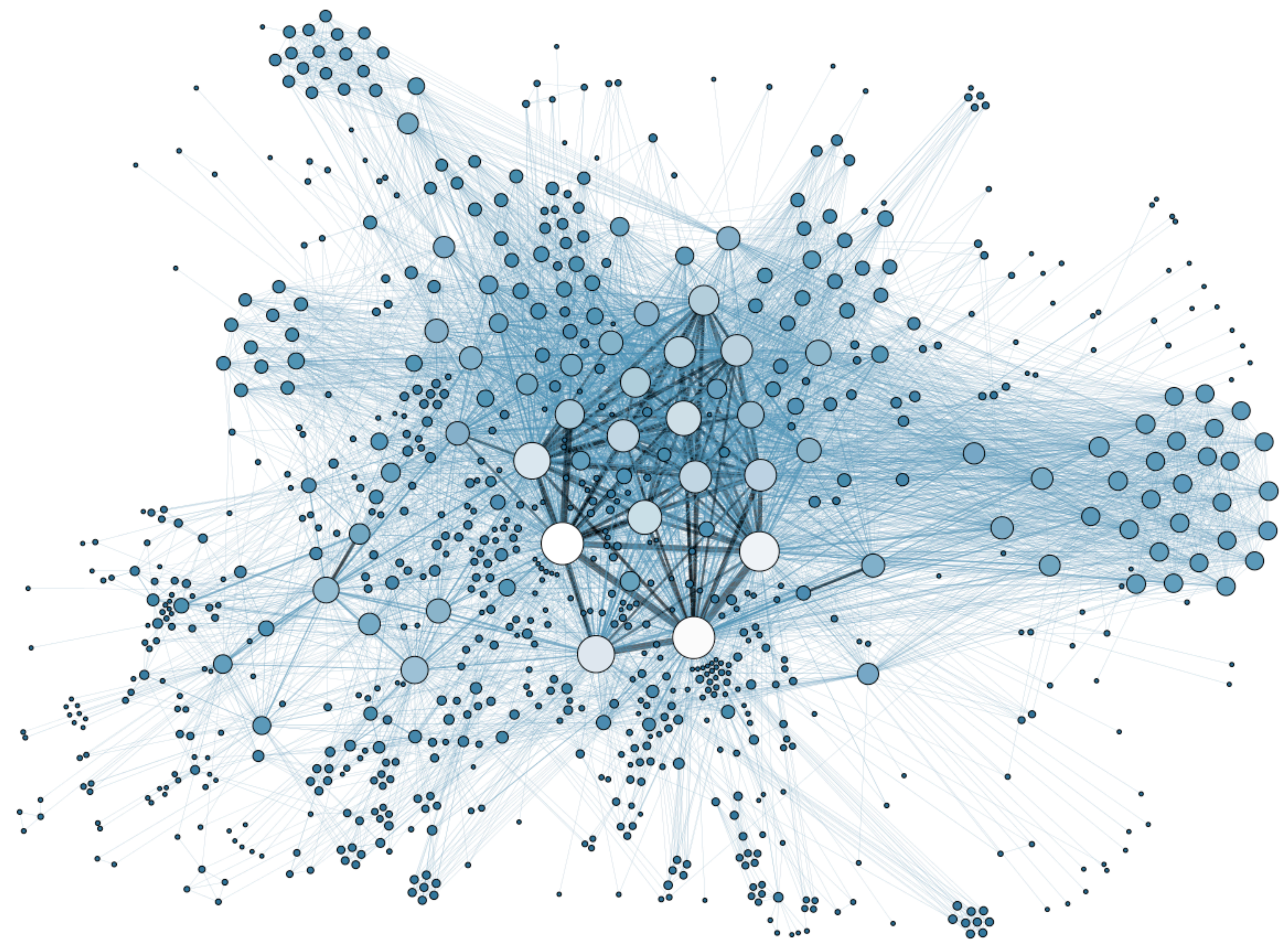


Scientific mobility is driven by a host of factors

Geography, language, culture, politics, prestige, and more all contribute to the mobility decisions of individual scientists

Discussion

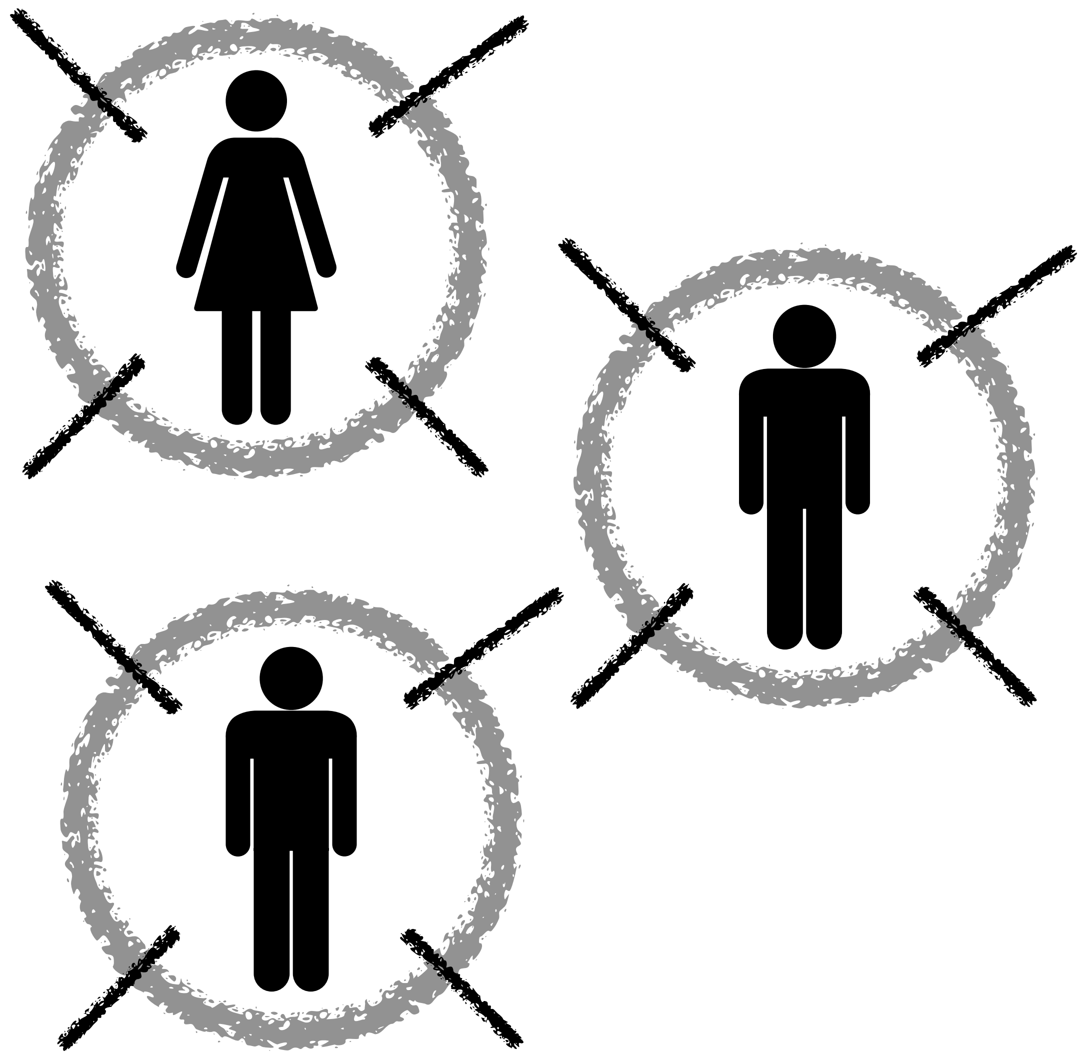
What can we learn about the organization and behavior of science given the complexity perspective?



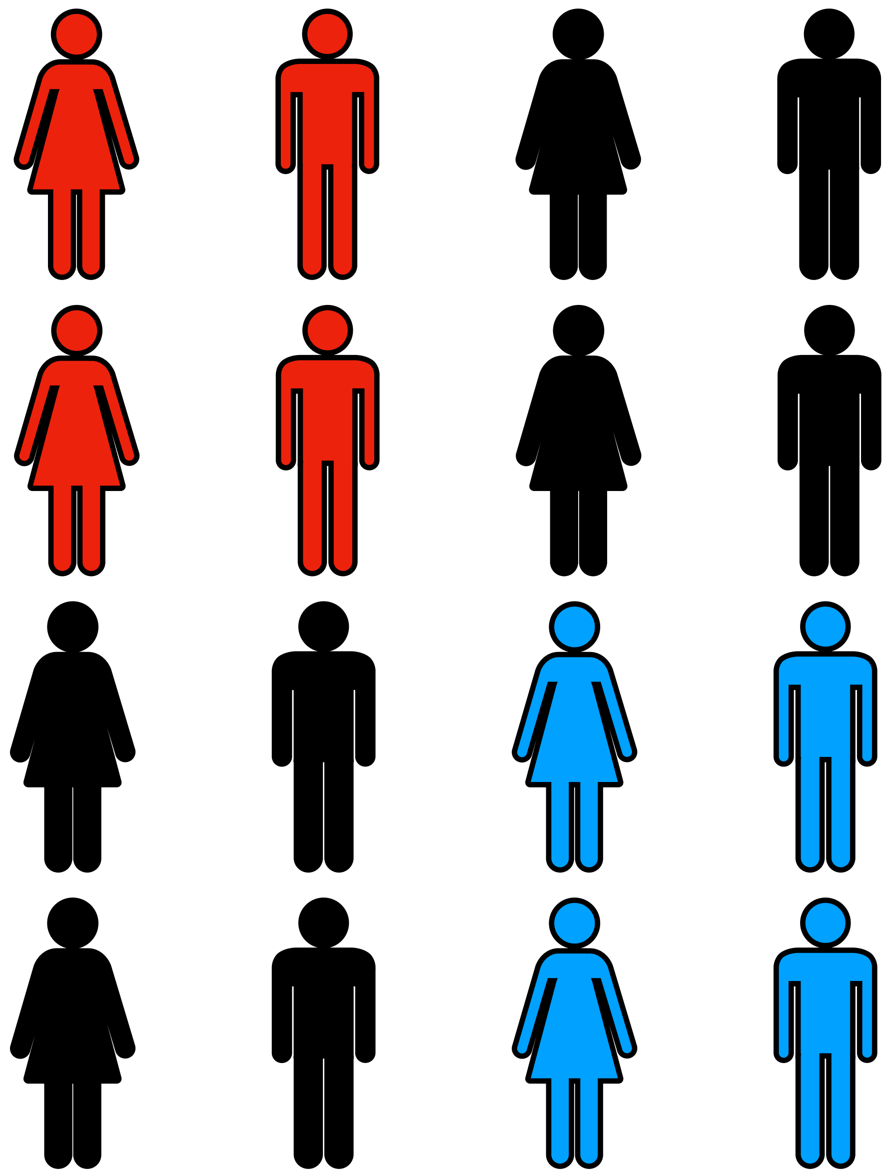
First: Bottom-up forces

Forces that govern the interactions and self-organization of mostly autonomous individuals

Autonomy



Self-organization

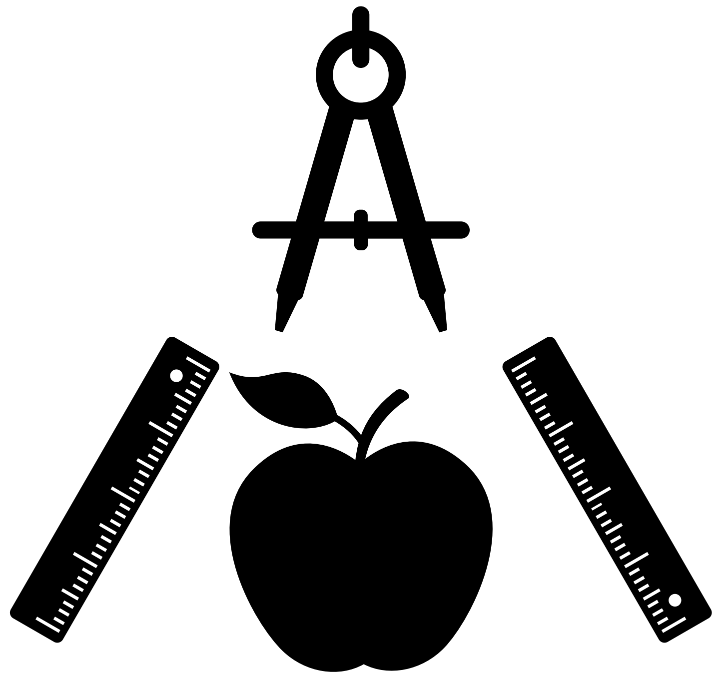


Bottom-up forces

- **Demographic bias**

Peer review at *eLife*

Student-teacher evaluations



Some are **advantaged** and others **disadvantaged** in important interactions such as peer review

Dictates who gets **capital**

Contributes to **success**, or career **exit**

Bottom-up forces

- Demographic bias
- **Homophily**

Peer review at *eLife*

Social: More interactions with similar people

Preferences: more favorable interactions with similar people

Determines both **success** and **social networks**

Bottom-up forces

- Demographic bias
- Homophily
- **Proximity**

Interactions more common between **nearby** scientists

Not just geography!

Also **linguistic**, **cultural**, **disciplinary**, and **social** proximity

Determines who interacts with who, social networks

Bottom-up forces

- Demographic bias
- Homophily
- Proximity
- **Physical realities**

Disagreement in science

Research topics demand certain behaviors and interactions

Determines the nature and quantity of interactions

Experimental fields allow more consensus, less disagreement, than non-experimental fields

Bottom-up forces

Global scientific mobility

- Demographic bias
- Homophily
- Proximity
- Physical realities
- **Reputation**

The ***currency*** of science

Institutional prestige shapes mobility decisions

Bottom-up forces

- Demographic bias
- Homophily
- Proximity
- Physical realities
- **Reputation**

Global scientific mobility

Peer review at *eLife*

Disagreement in science

The ***currency*** of science

Institutional prestige shapes mobility decisions

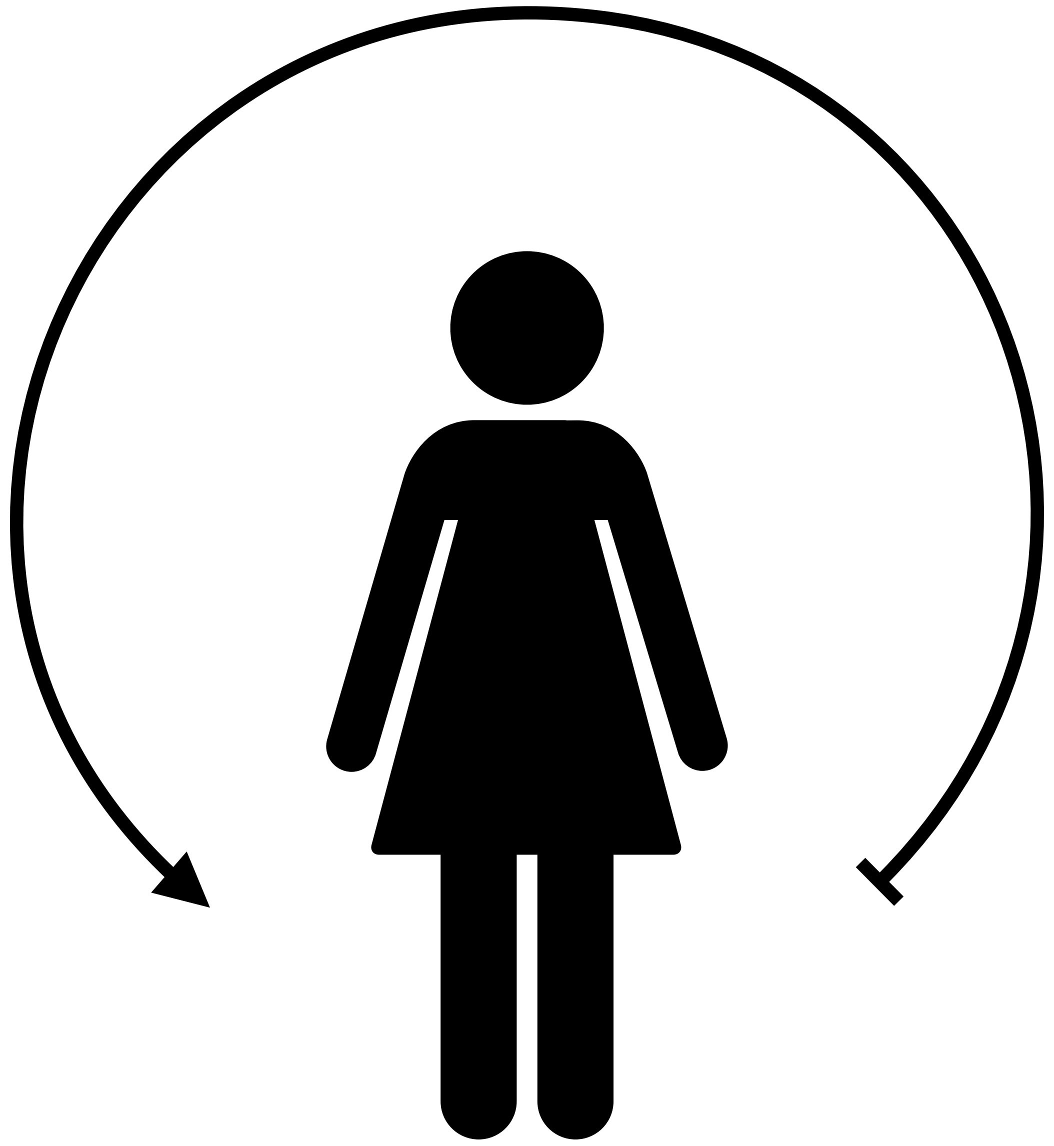
Might affect peer review decisions, disagreement

How does structure result from these bottom-up forces?

Feedback

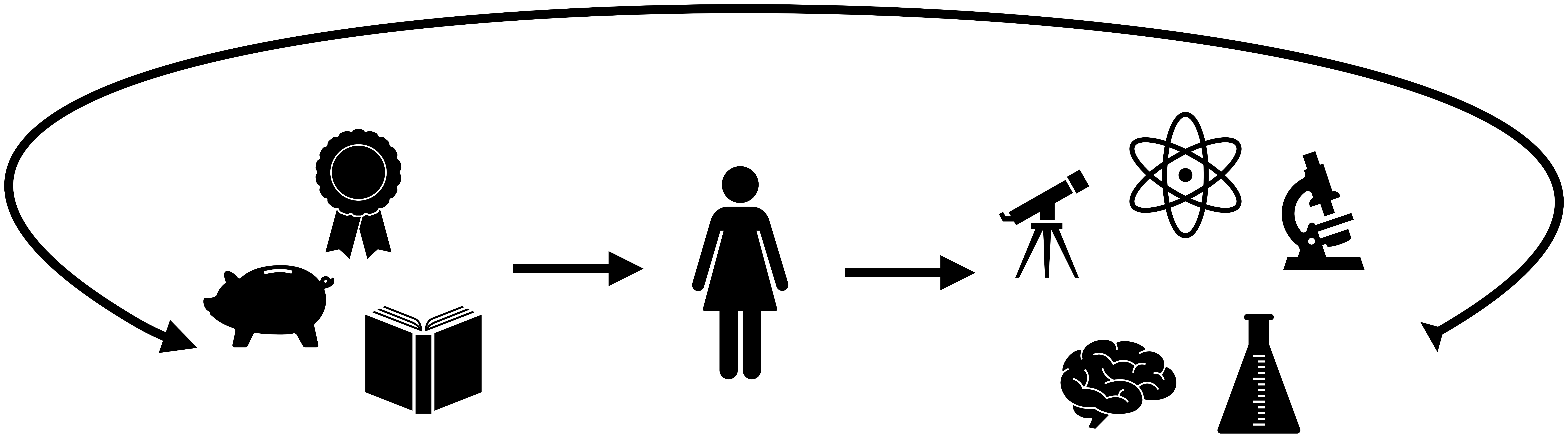
A central idea about complex systems

Key to the structure of science



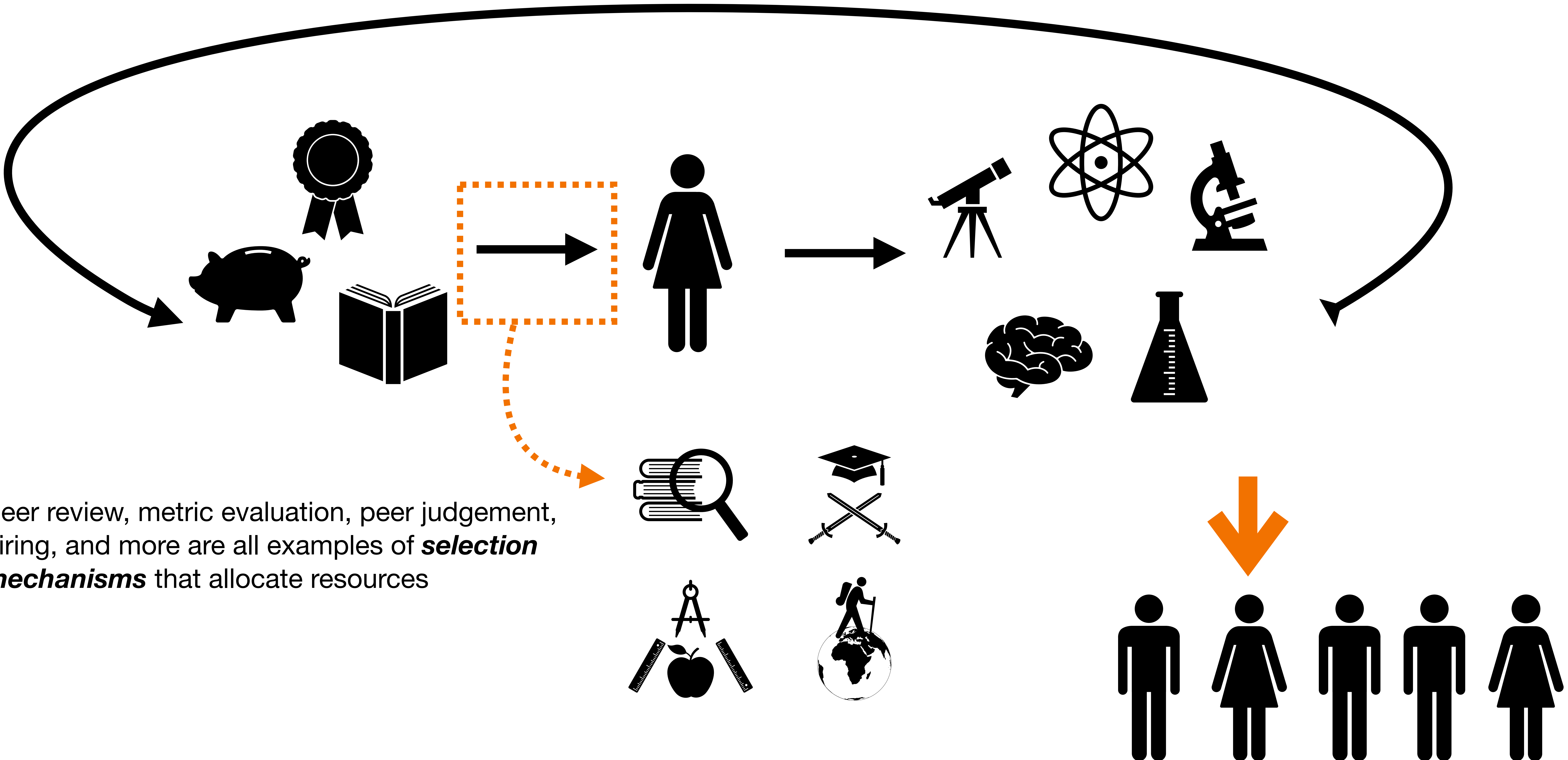
Feedback and success

Resources allow success, which allows accumulation of more resources



Feedback and success

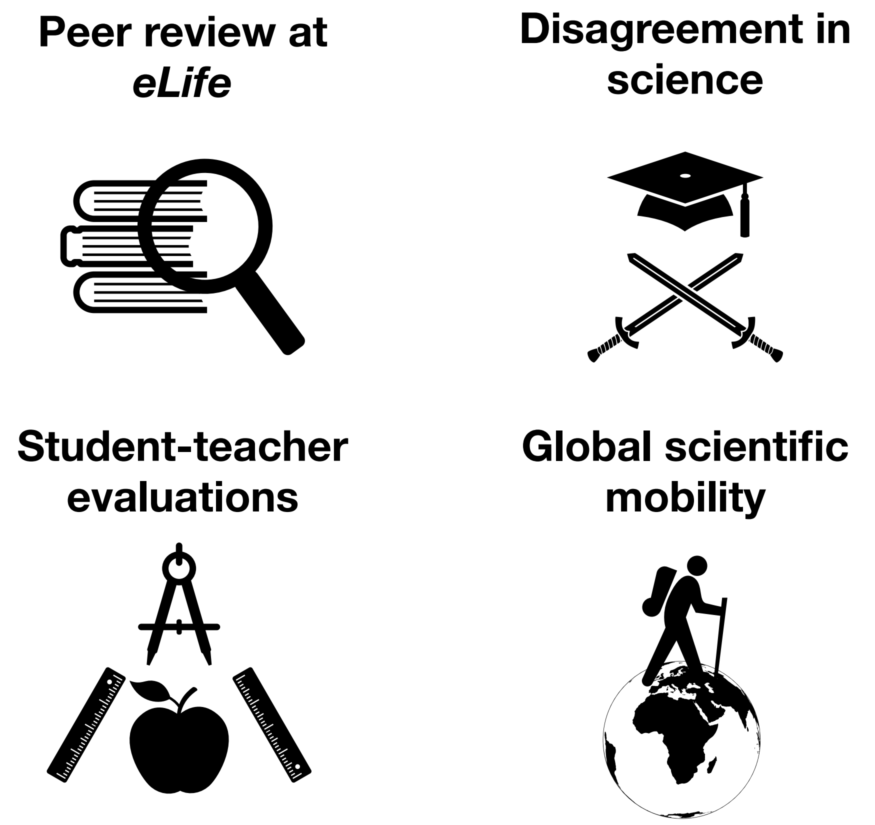
Resources allow success, which allows accumulation of more resources



Peer review, metric evaluation, peer judgement, hiring, and more are all examples of ***selection mechanisms*** that allocate resources

Feedback and success

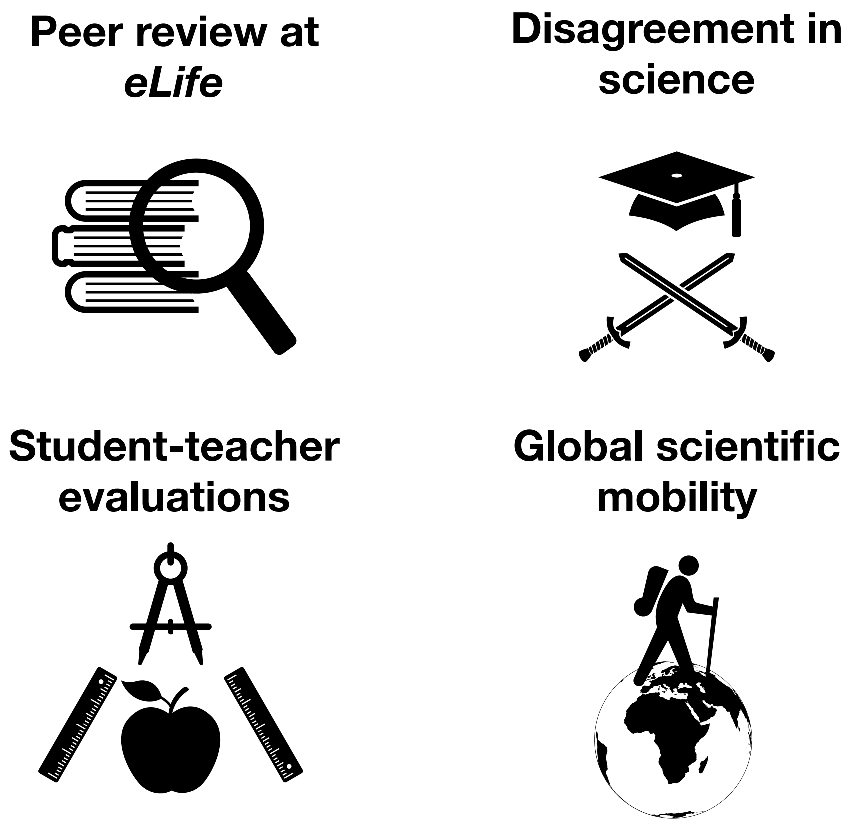
Ideally, selections are made according to *merit*



Feedback and success

Ideally, selections are made according to ***merit***

But bias, homophily, proximity, reputation, and more, affect selections



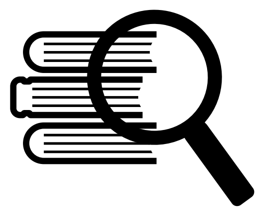
Feedback and success

Ideally, selections are made according to *merit*

But bias, homophily, proximity, reputation, and more, affect selections

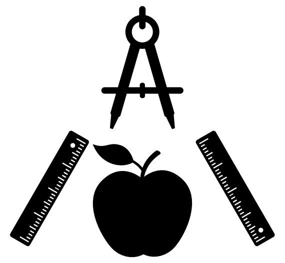
The advantaged or already-successful will find more success, those disadvantaged will not

Peer review at *eLife*



Disagreement in science

Student-teacher evaluations



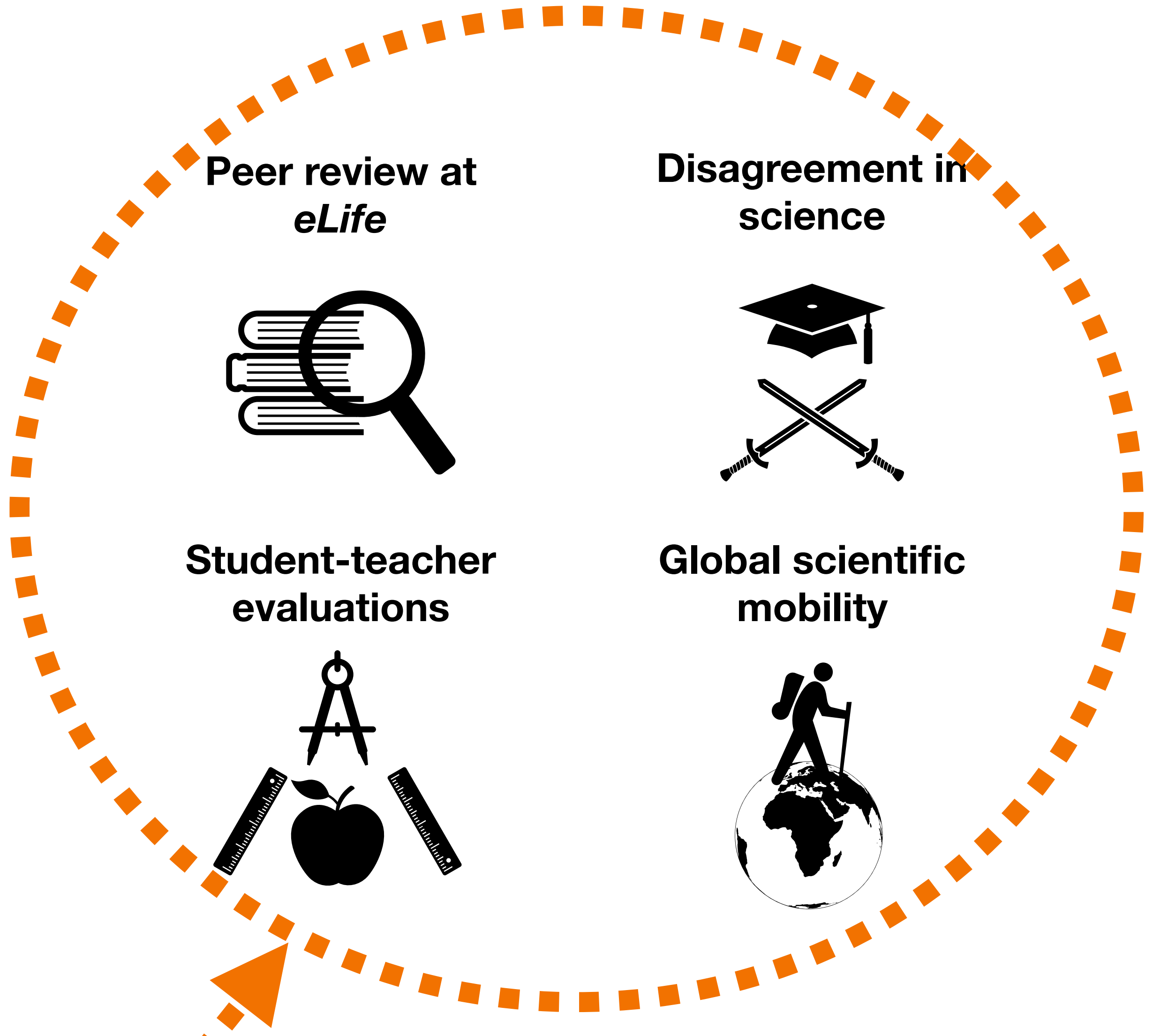
Global scientific mobility

Feedback and success

Ideally, selections are made according to *merit*

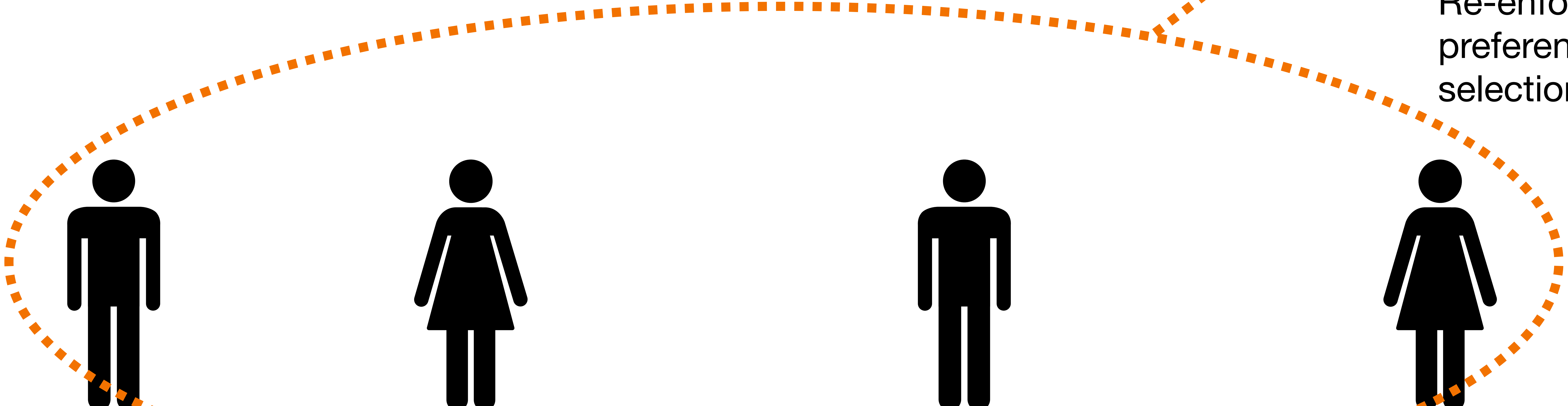
But bias, homophily, proximity, reputation, and more, affect selections

The advantaged or already-successful will find more success, those disadvantaged will not



Successful individuals become the new reviewers, committee members, and evaluators

Re-enforce their biases and preferences back into future selections



Any mechanism that allocates resources that benefit one's career can create a feedback loop

Judgements based on social criteria can disadvantage marginalized groups, resulting in their underrepresentation and perpetuating biases and inequalities into the future

But if the rich only ever get richer, then how does anything change?

But if the rich only ever get richer, then how does anything change?

Feedback also offers a solution

Feedback as amplification

The effects of some events can rapidly compound, growing stronger and more pervasive

Early advantage is likely amplified…but not always

Chaos + amplification

The triggers of amplification can seem **random** and **unexplainable**, and in **unexpected places**, with deep and far-reaching consequences

Chaos + amplification

The triggers of amplification can seem **random** and **unexplainable**, and in **unexpected places**, with deep and far-reaching consequences

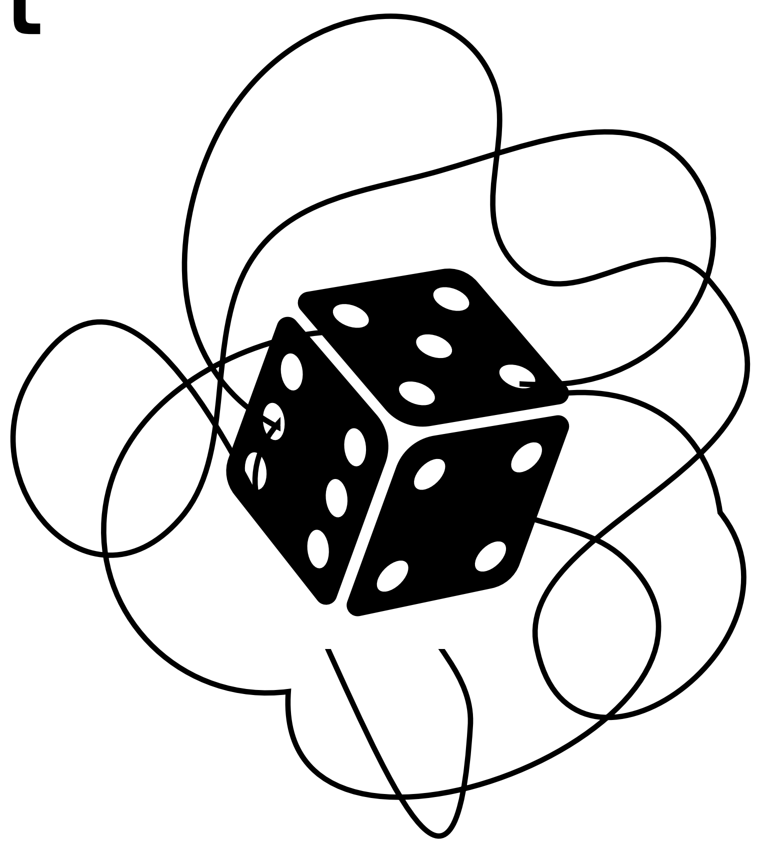
Black swan-like events, such as major breakthroughs and paradigm shifts, that form crucial moments in the history of science

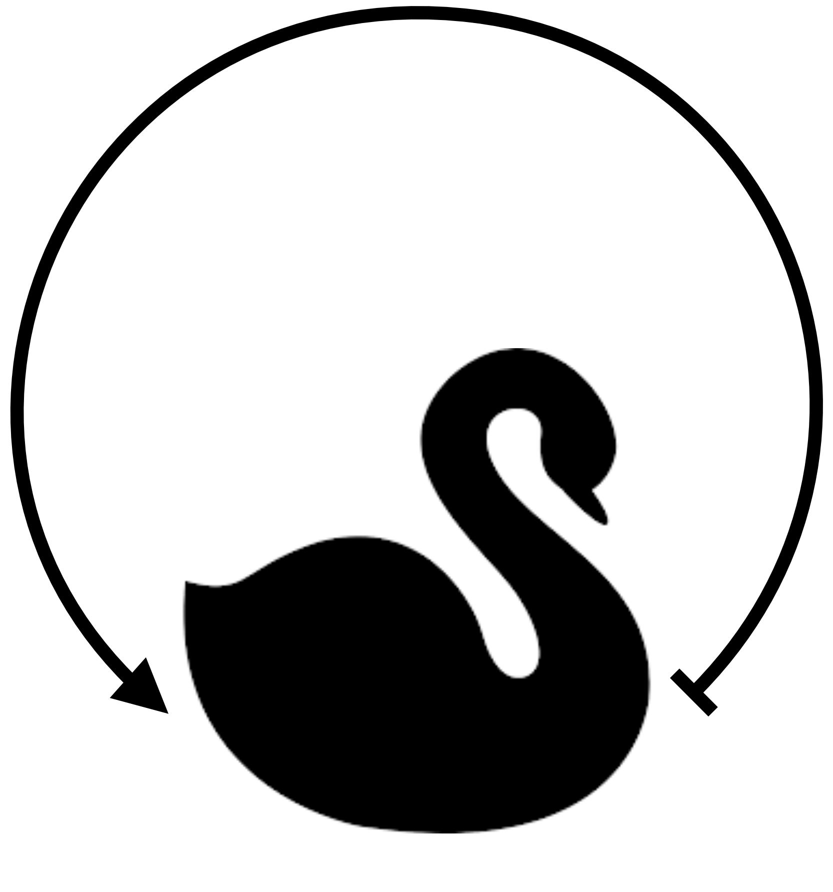
Chaos + amplification

The triggers of amplification can seem **random** and **unexplainable**, and in **unexpected places**, with deep and far-reaching consequences

Black swan-like events, such as major breakthroughs and paradigm shifts, that form crucial moments in the history of science

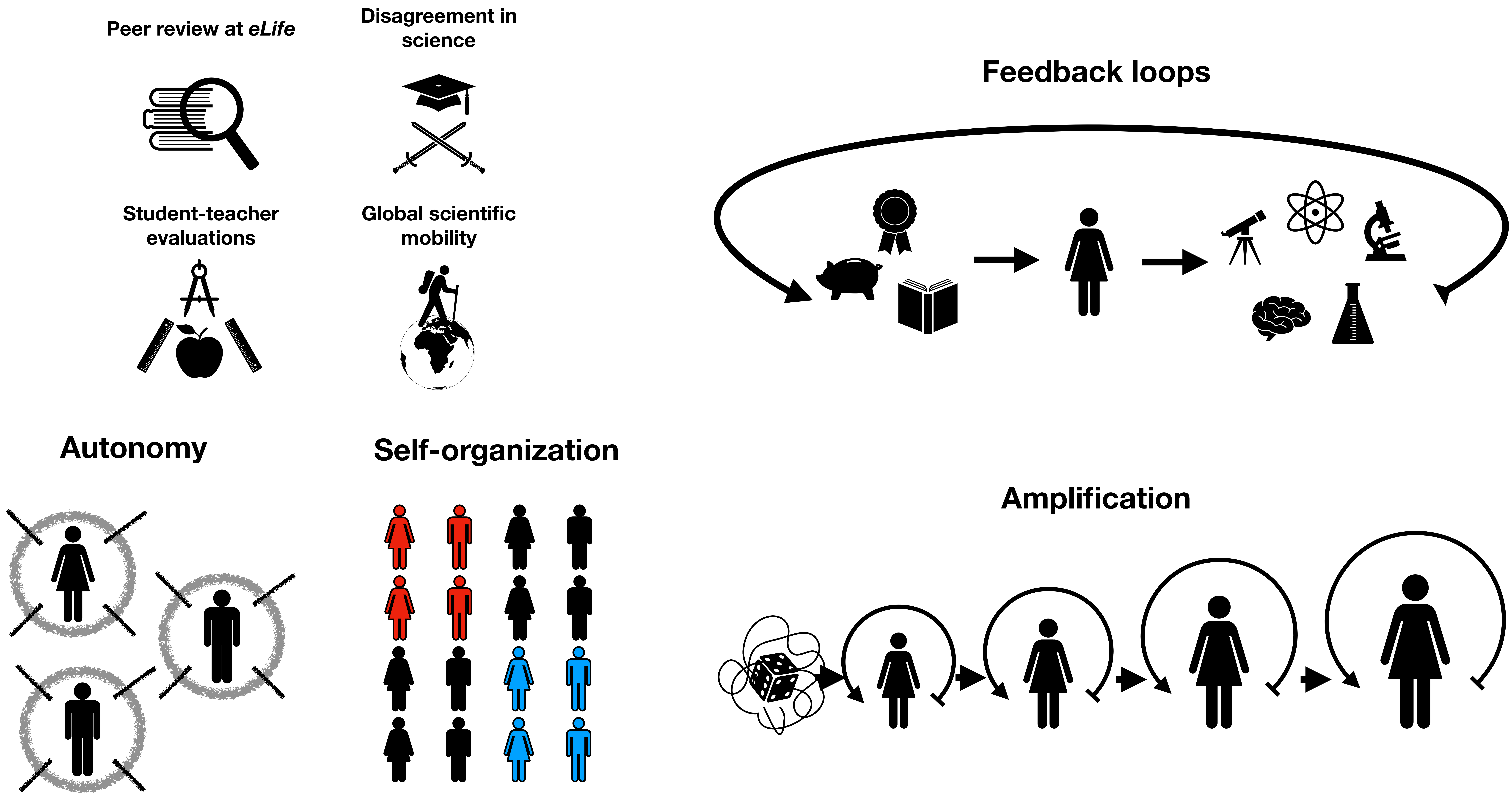
Predicting and studying these events is difficult, but they are often the most important





Conclusion

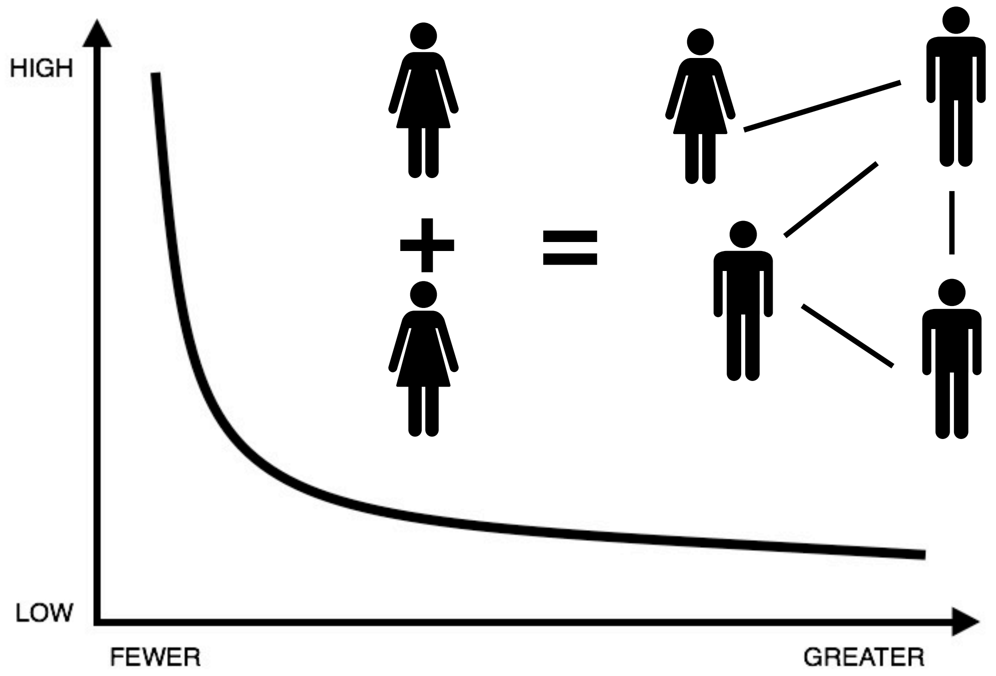
The complexity perspective



Complexity concepts for future research

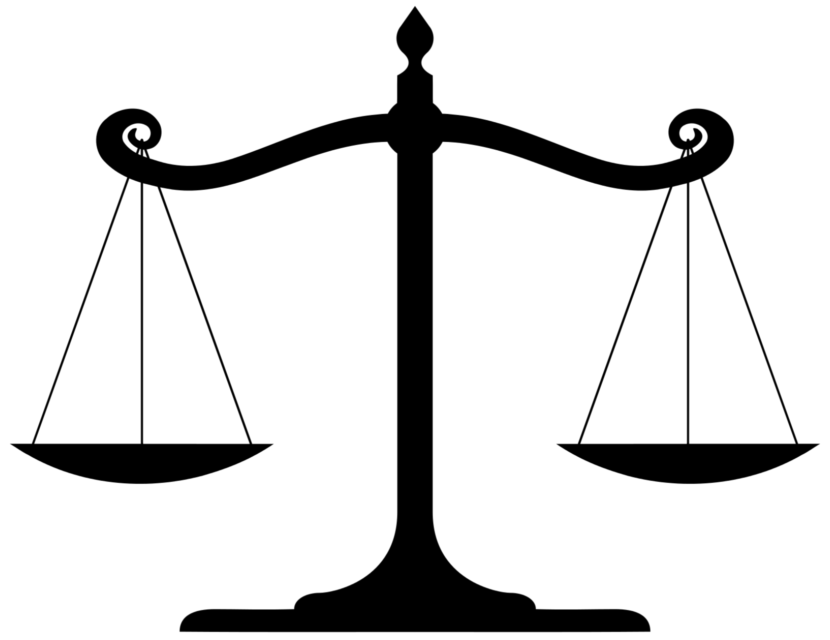
Emergence

Scaling effects

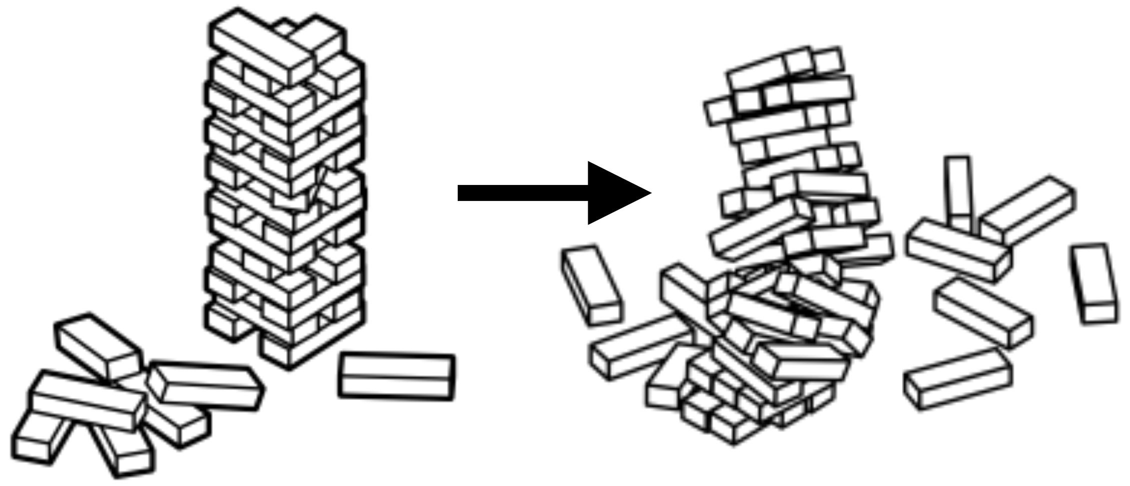


Resilience

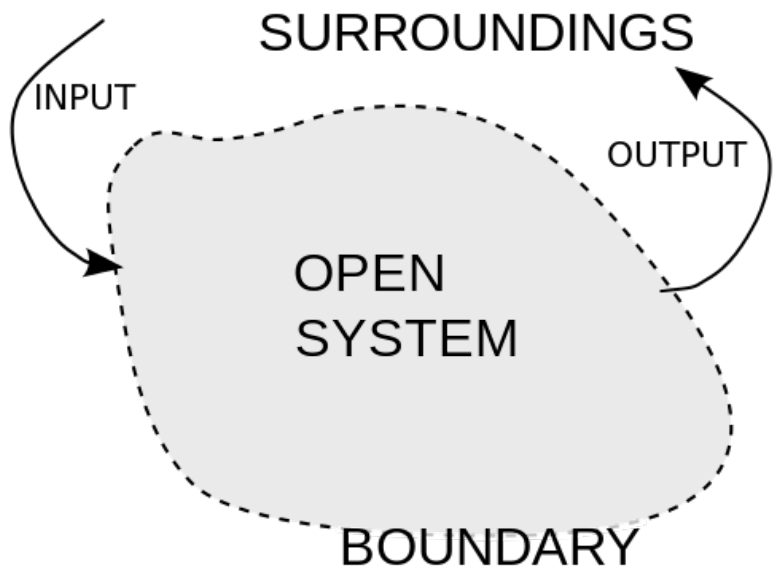
Stable equilibriums



Tipping Points



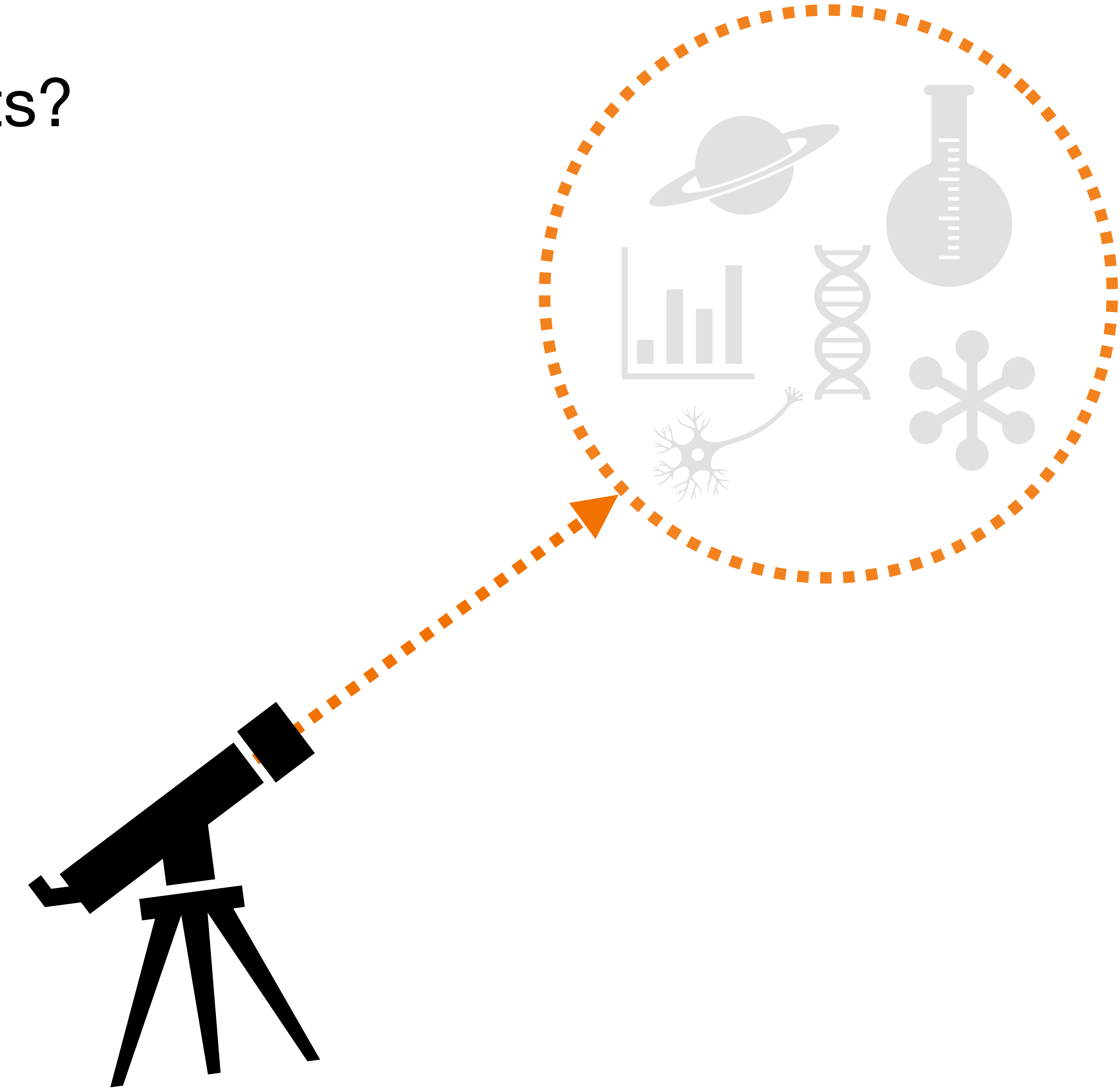
Environmental adaptation



Implications to the Science of Science

Limits posed by the size, heterogeneity, interconnectivity, and chaos of science

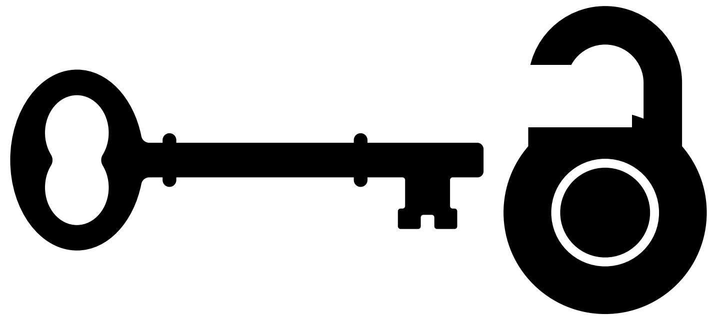
- How to craft valid and meaningful measurements?
- Are there even universal laws of science?
- Can causal mechanisms be identified?
- Is prediction even possible?



Implications to Science Policy

Crafting policy for a complex world

- Are we measuring what's really important?
- Can policy outcomes be anticipated?
- Will successful policies in one context apply to others?
- How to tell if policies do more harm than good?

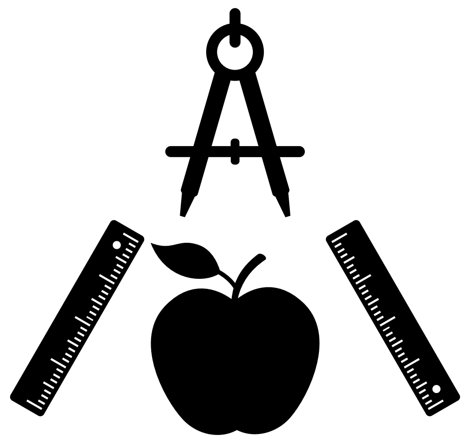


The need for more, and better data

Open science practices can fuel development of the complexity perspective

Peer review at *eLife*

Data provided by *eLife*



Student-teacher evaluations

Public & private data on U.S. faculty

Disagreement in science

Increasing availability of full-text data

Global scientific mobility

Name-disambiguated bibliographic records

In a complex world, new perspectives are needed

Complexity science has the potential to offer new insights into the inner workings of science

Thanks!

Peer review at *eLife*

Kyle Siler, Vincent Larivière, Wei Mun Chan, Andrew M. Collings, Jennifer Raymond, and Cassidy R. Sugimoto

Student-teacher evaluations

Clara Boothby, Huimeng Zhao, Vanessa Minik, Nicolas Bérubé, Dr. Vincent Larivière, and Dr. Cassidy R. Sugimoto

Disagreement in science

Wout Lamers, Dr. Kevin Boyack, Dr. Vincent Larivière, Dr. Cassidy R. Sugimoto, Dr. Nees Jan van Eck, and Dr. Ludo Waltman

Global scientific mobility

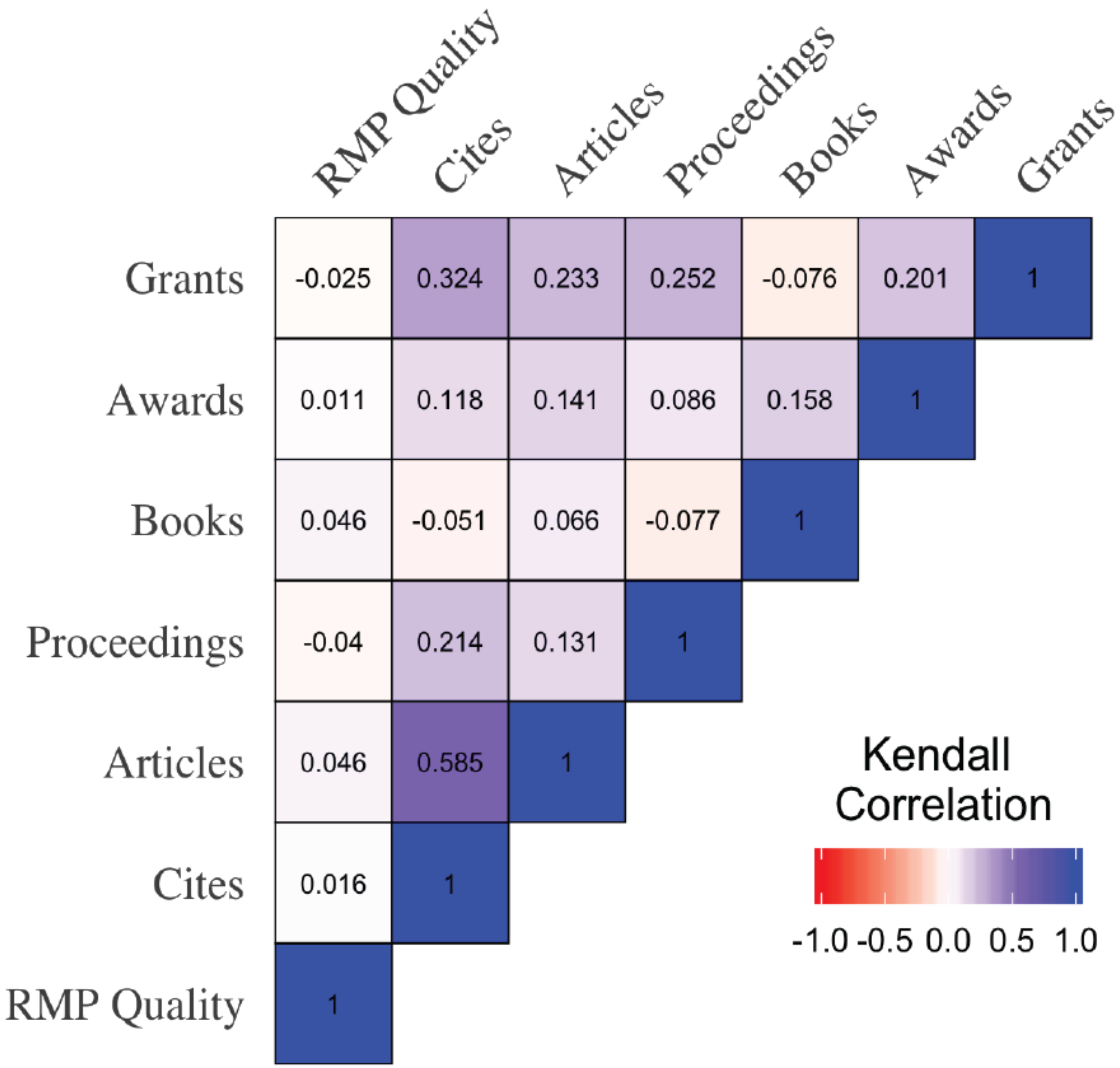
Jisung Yoon, Sadamori Kojaku, Rodrigo Costas, Woo-Sung Jung, Staša Milojević, and Yong-Yeol Ahn

And my wonderful committee, including Cassidy R. Sugimoto, Yong-Yeol Ahn, Staša Milojević, Santo Fortunato, and Guillaume Cabanac

Appendix

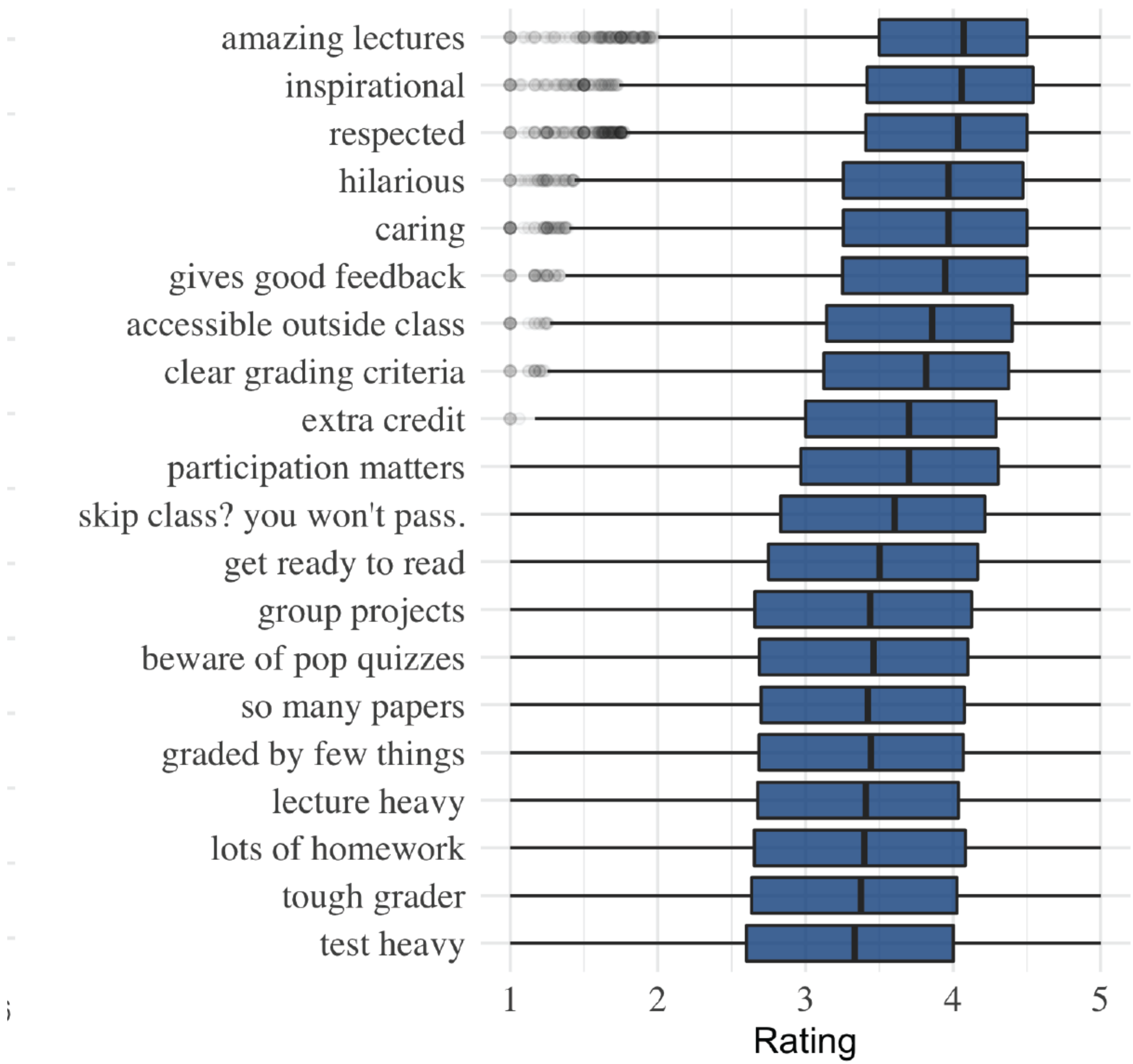
Digging deeper—teaching and research

NO strong correlation with non-parametric analysis



So what are students rating?

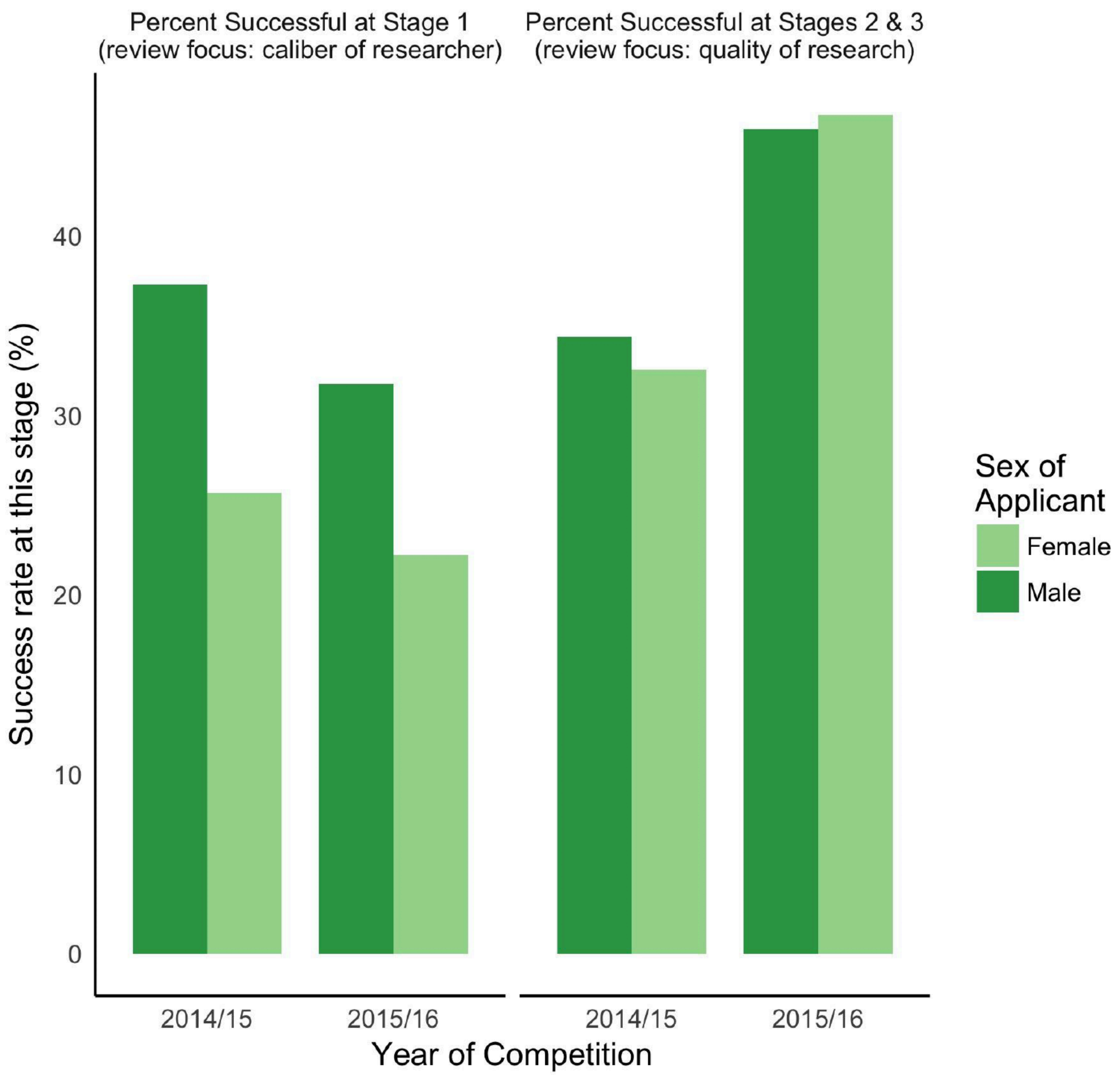
Emotive tags (positive review) and workload related tags (low reviews)



Gender bias

Grant peer review at CIHR

- 2-stage review process

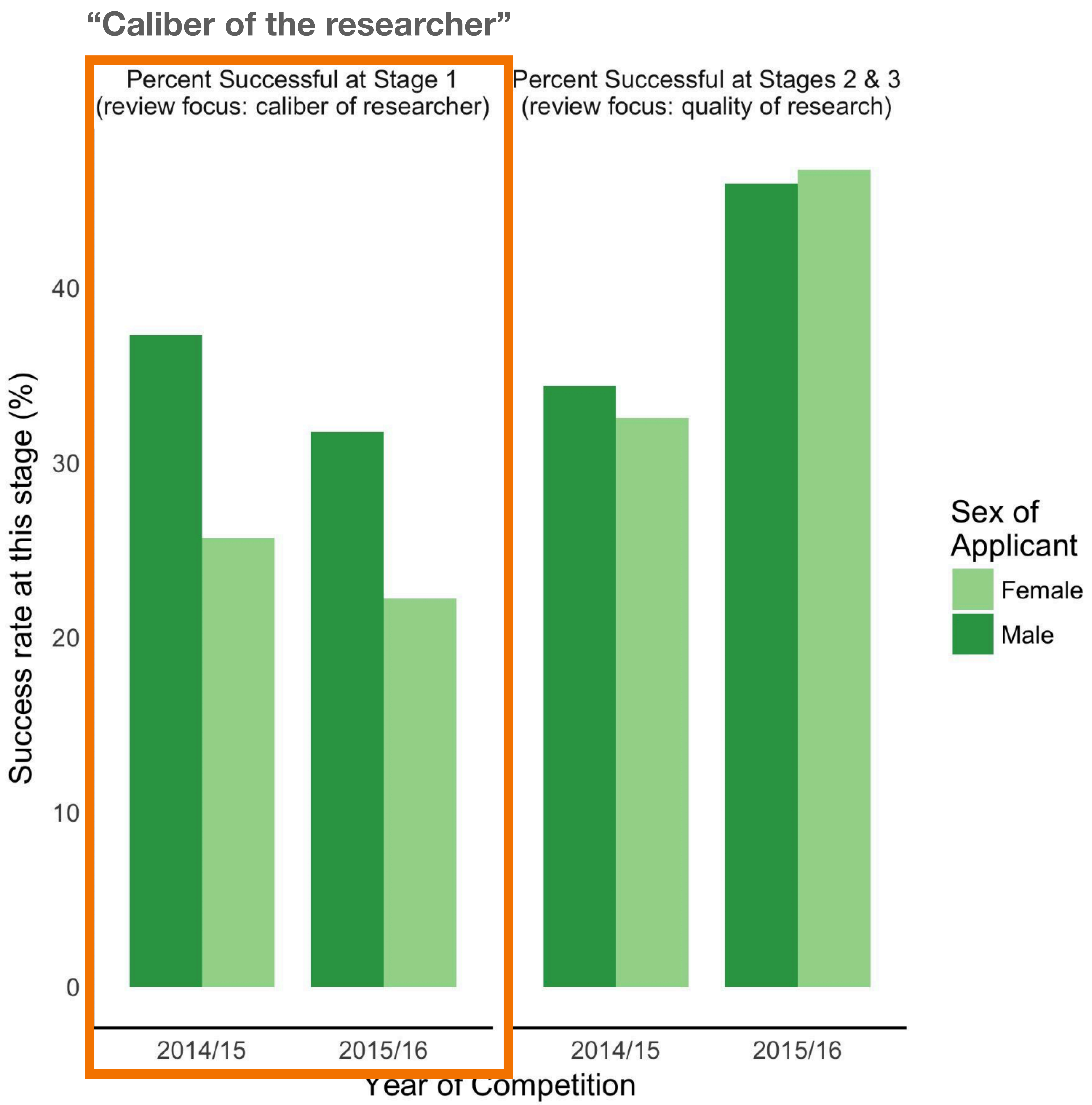


Witteman, H. O., Hendricks, M., Straus, S., & Tannenbaum, C. (2017). Female grant applicants are equally successful when peer reviewers assess the science, but not when they assess the scientist. *BioRxiv*, 232868.

Gender bias

Grant peer review at CIHR

- 2-stage review process
- Women rated lower in “Caliber of researcher”

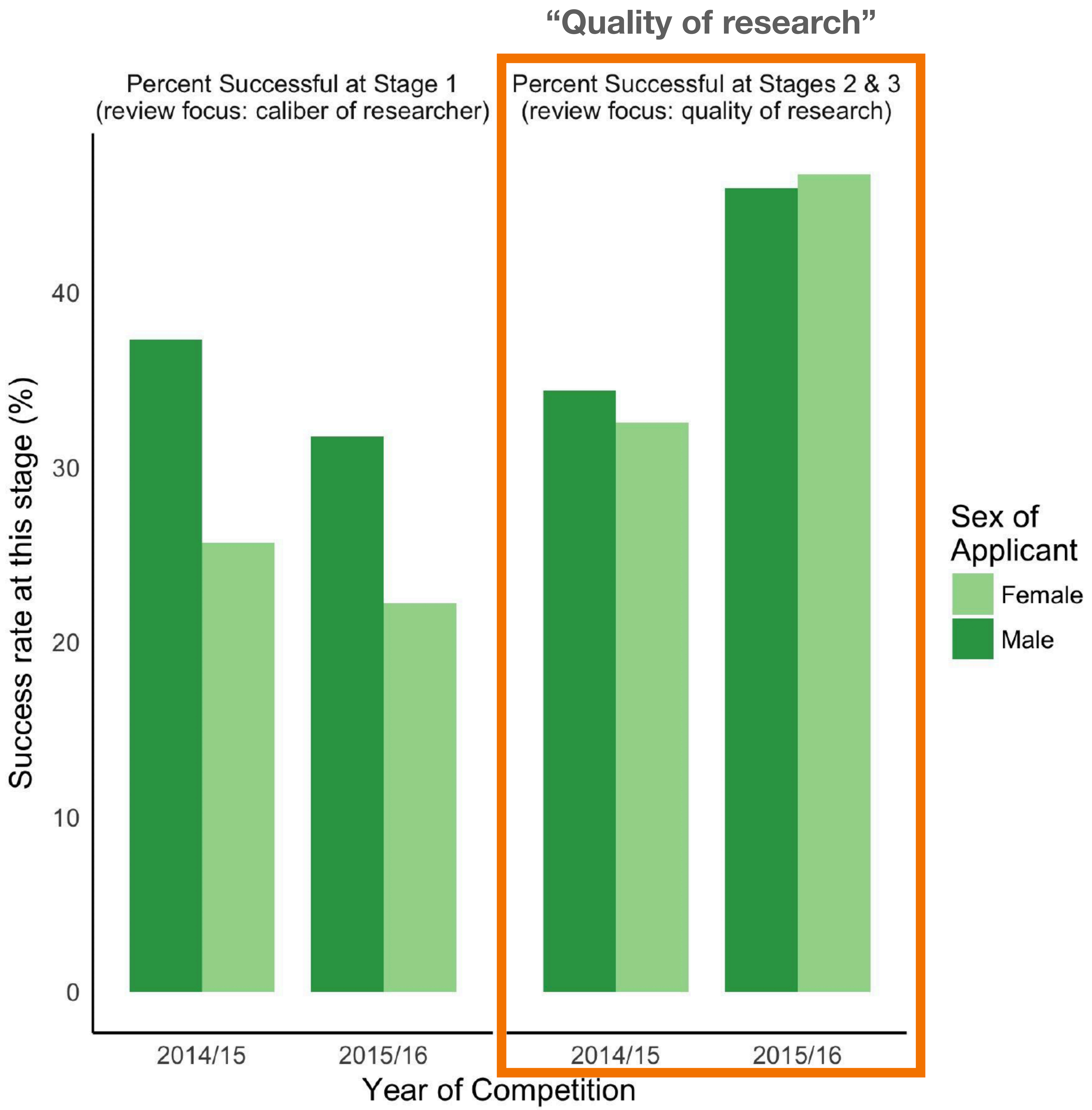


Witteman, H. O., Hendricks, M., Straus, S., & Tannenbaum, C. (2017). Female grant applicants are equally successful when peer reviewers assess the science, but not when they assess the scientist. *BioRxiv*, 232868.

Gender bias

Grant peer review at CIHR

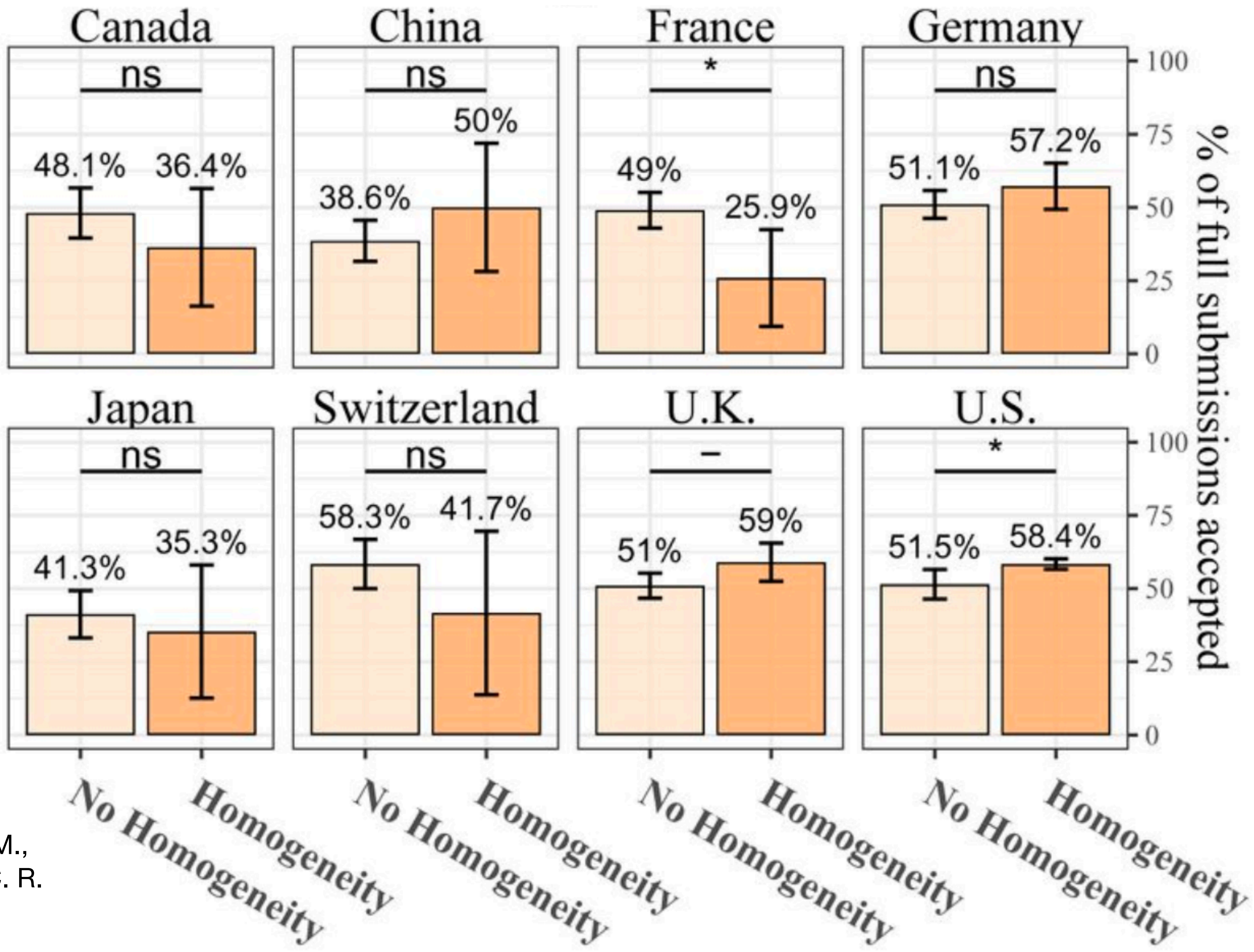
- 2-stage review process
- Women rated lower in “Caliber of researcher”
- Gender equity when rating on “quality of research”



Witteman, H. O., Hendricks, M., Straus, S., & Tannenbaum, C. (2017). Female grant applicants are equally successful when peer reviewers assess the science, but not when they assess the scientist. *BioRxiv*, 232868.

Reviewer nationality also matters

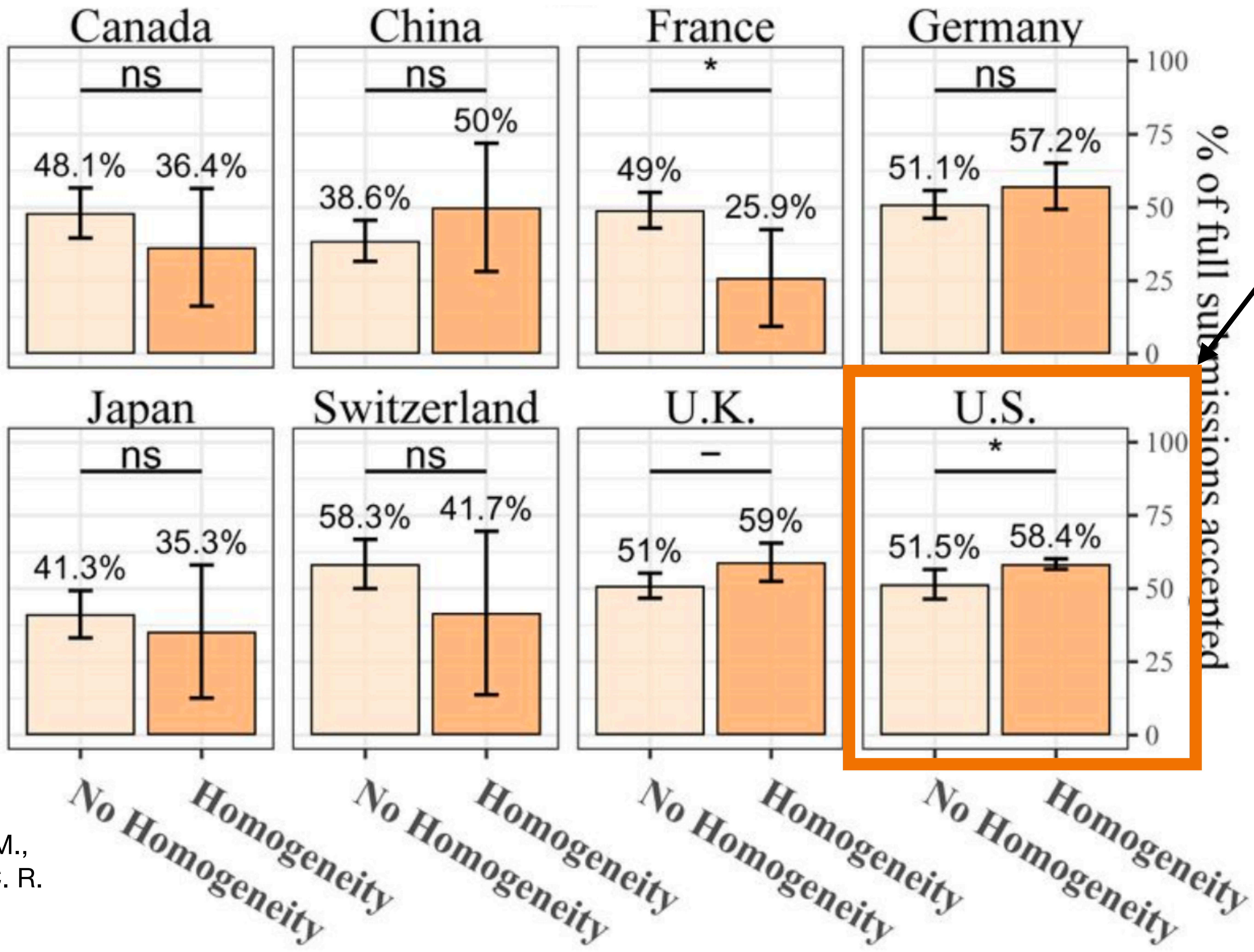
Review outcomes based on homogeny between author and reviewers



Murray, D., Siler, K., Larivière, V., Chan, W. M., Collings, A. M., Raymond, J., & Sugimoto, C. R. (2019). Author-Reviewer Homophily in Peer Review. *BioRxiv*, 400515.

Reviewer nationality also matters

Review outcomes based on homogeny between author and reviewers

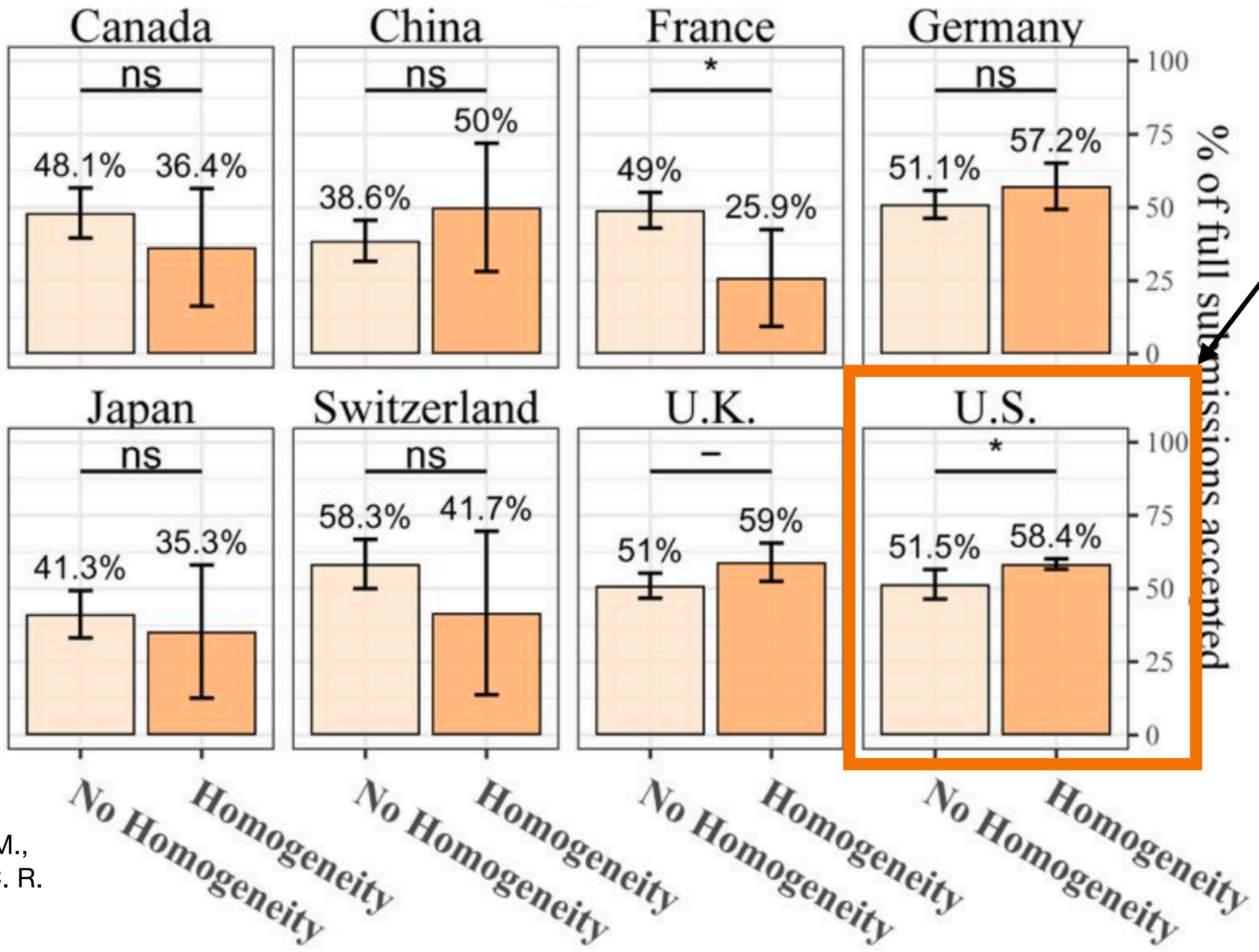


U.S. authors reviewed by U.S. reviewer accepted more often

Murray, D., Siler, K., Larivière, V., Chan, W. M., Collings, A. M., Raymond, J., & Sugimoto, C. R. (2019). Author-Reviewer Homophily in Peer Review. *BioRxiv*, 400515.

Reviewer nationality also matters

Review outcomes based on homogeny between author and reviewers

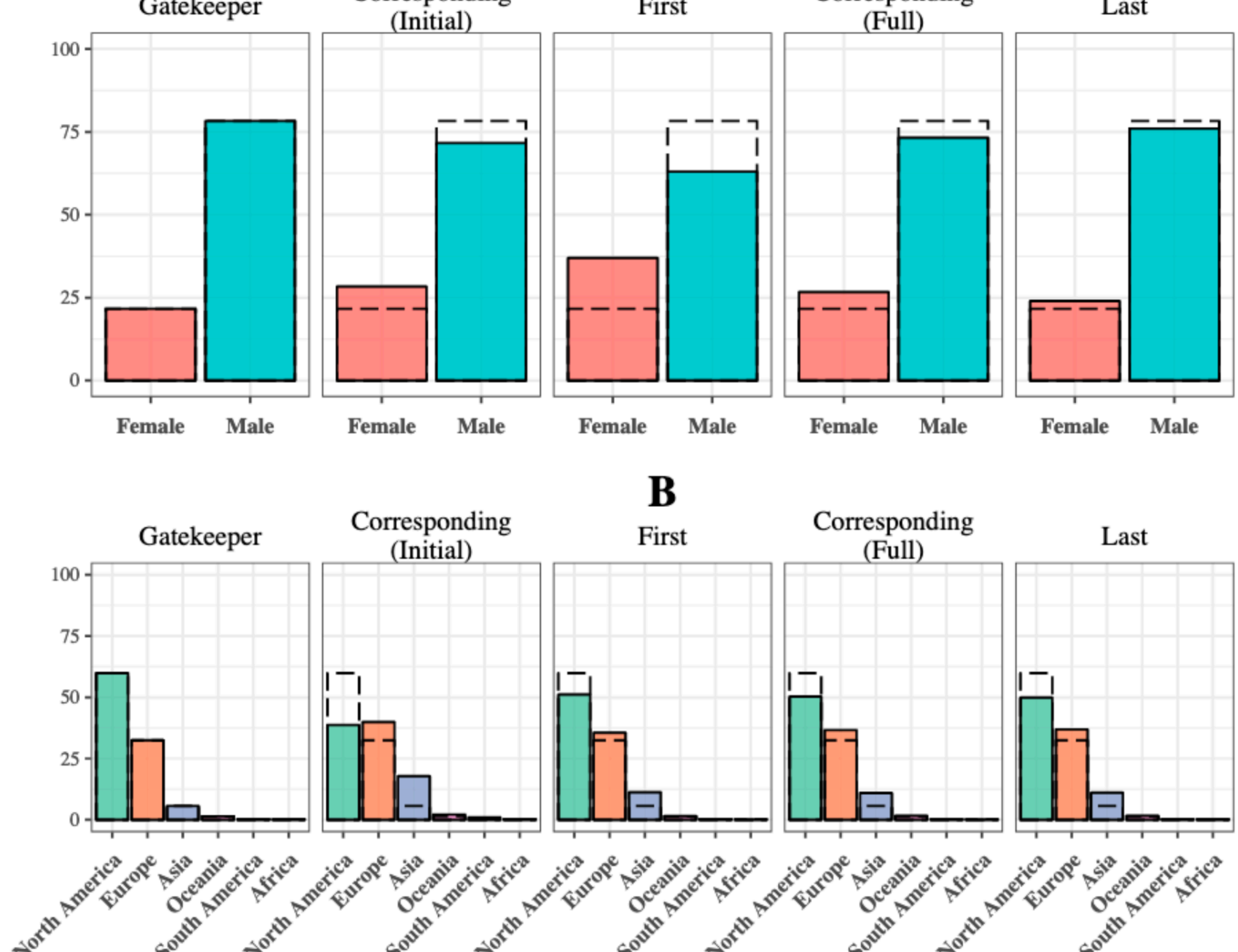


U.S. authors reviewed by U.S. reviewer accepted more often

A pattern that is not as clear across other countries

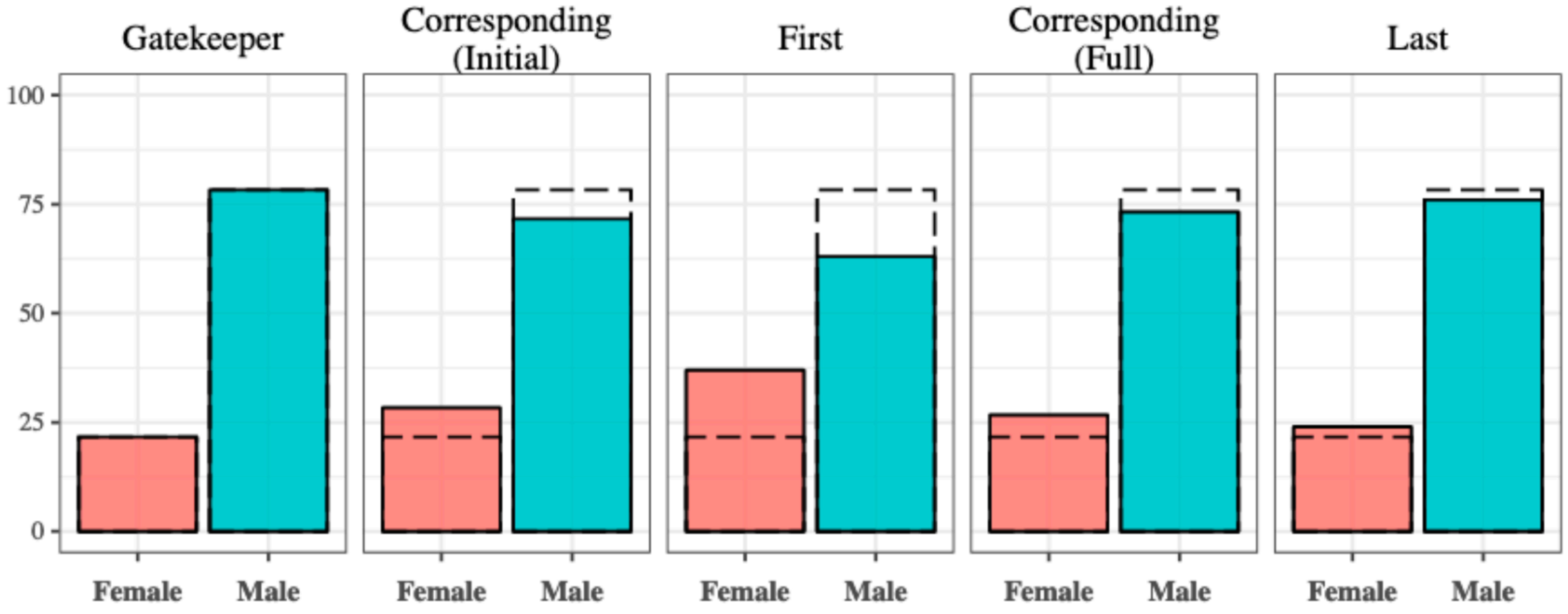
Murray, D., Siler, K., Larivière, V., Chan, W. M., Collings, A. M., Raymond, J., & Sugimoto, C. R. (2019). Author-Reviewer Homophily in Peer Review. *BioRxiv*, 400515.

Ma



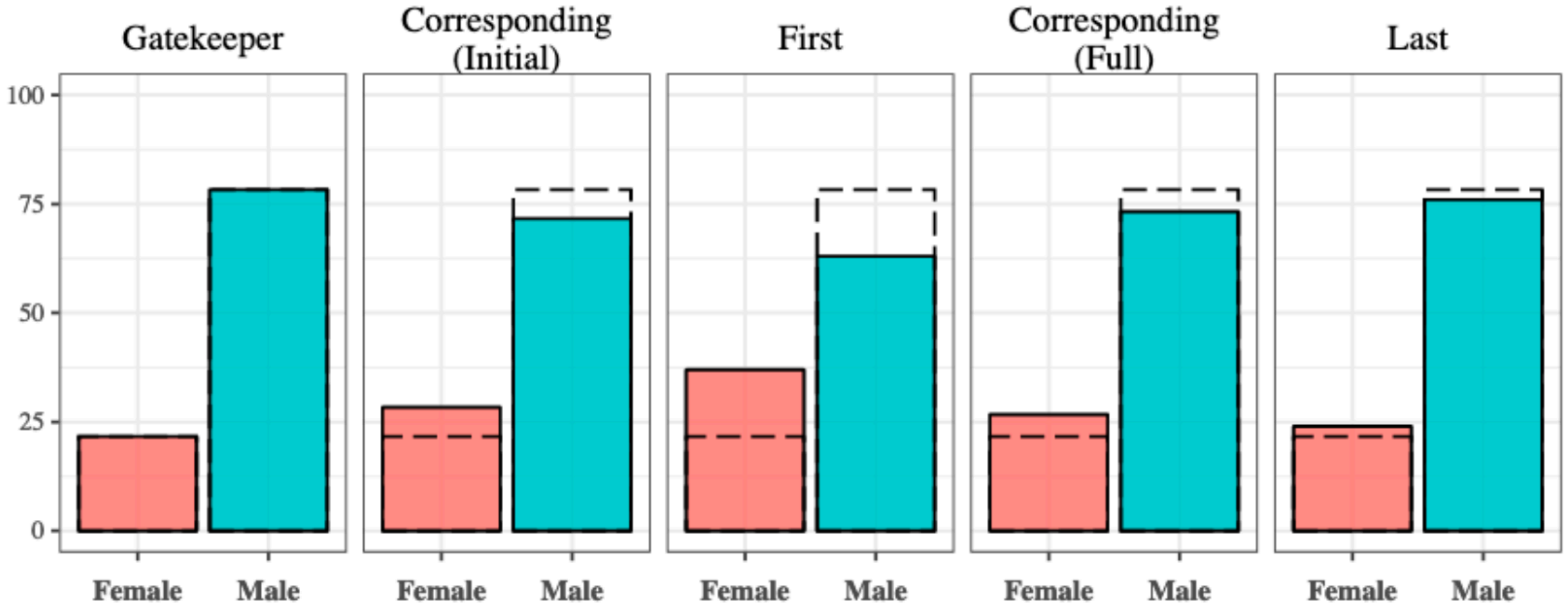
Gatekeeper
Corresponding (Initial)
First
Corresponding (Full)
Last
100
75
50
25
0
Female
Male
B
North America
Europe
Asia
Oceania
South America
Africa

Make reviewers represent the authorship



This is something that eLife does well!

Make reviewers represent the authorship

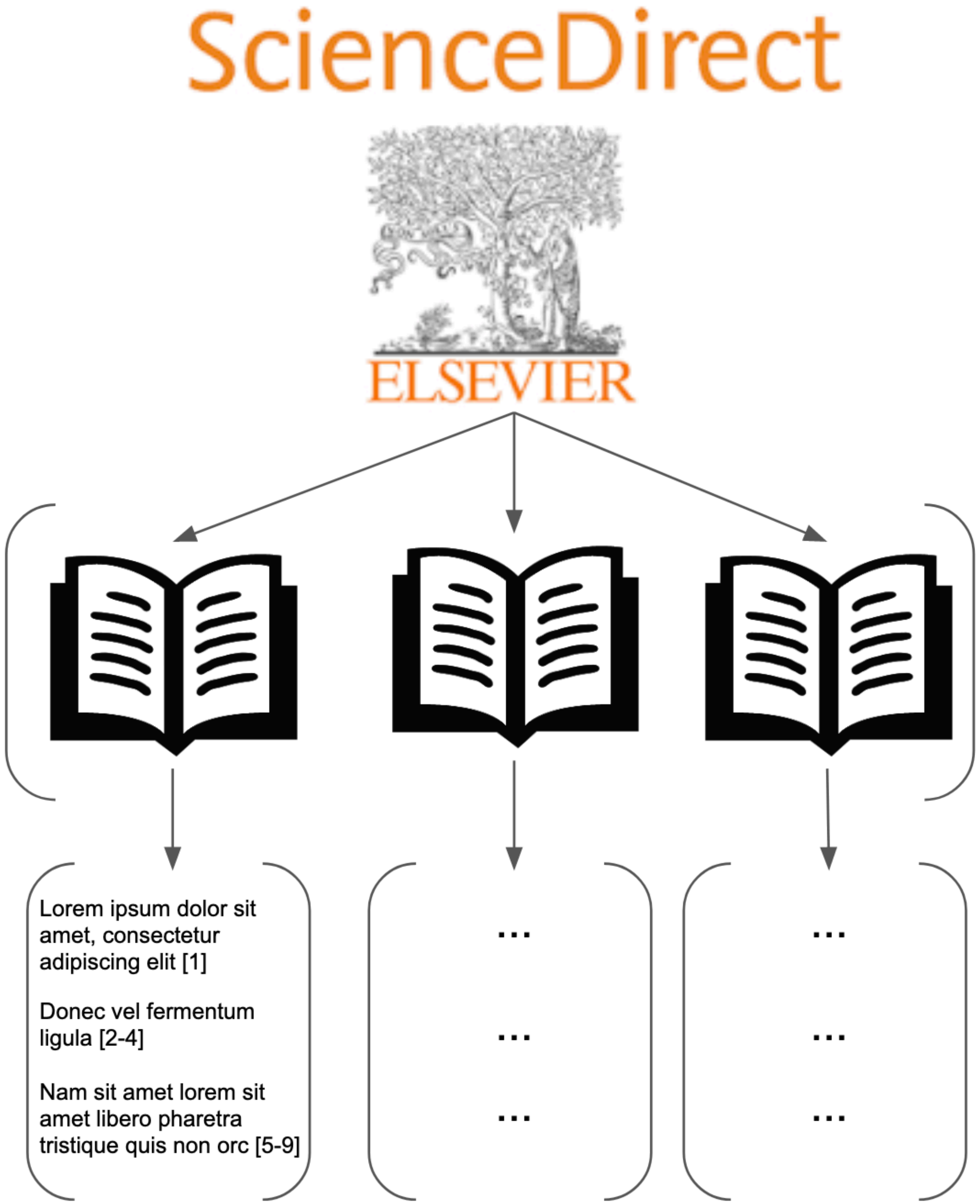


This is something that eLife does well!

Only about half of papers were reviewed by mixed-gender teams

Data

- Disagreement between texts
- Extract citation sentences
- Over 3 million full-text English-language article
- Identify disagreement citations



Signal & filter terms

	*_standalone_*	+studies	+ideas	+methods	+results
Challenge*					
Conflict*					
Contradict*					
Contrary					
Contrast*					
Contravers*					
Debat*					
Differ*					
Disagree*					
Disprov*					
No consensus					
Questionable*					
Refut*					

Signal & filter terms

	standalone	+studies	+ideas	+methods	+results
Challenge*					
Conflict*					
Contradict*					
Contrary					
Contrast*					
Contravers*					
Debat*					
Differ*					
Disagree*					
Disprov*					
No consensus					
Questionable*					
Refut*					

"...recruiting participants was **challenging**..."

"However, recent studies have **disagreed** with this approach"

Signal & filter terms

	*_standalone_*	+studies	+ideas	+methods	+results
Challenge*	*50 citation sentences*	*50 citation sentences*	*50 citation sentences*	*50 citation sentences*	*50 citation sentences*
Conflict*	*50 citation sentences*	*50 citation sentences*	*50 citation sentences*	*50 citation sentences*	*50 citation sentences*
Contradict*	*50 citation sentences*	*50 citation sentences*	*50 citation sentences*	*50 citation sentences*	*50 citation sentences*
Contrary	*50 citation sentences*	*50 citation sentences*	*50 citation sentences*	*50 citation sentences*	*50 citation sentences*
Contrast*	*50 citation sentences*	*50 citation sentences*	*50 citation sentences*	*50 citation sentences*	*50 citation sentences*
Contravers*	*50 citation sentences*	*50 citation sentences*	*50 citation sentences*	*50 citation sentences*	*50 citation sentences*
Debat*	*50 citation sentences*	*50 citation sentences*	*50 citation sentences*	*50 citation sentences*	*50 citation sentences*
Differ*	*50 citation sentences*	*50 citation sentences*	*50 citation sentences*	*50 citation sentences*	*50 citation sentences*
Disagree*	*50 citation sentences*	*50 citation sentences*	*50 citation sentences*	*50 citation sentences*	*50 citation sentences*
Disprov*	*50 citation sentences*	*50 citation sentences*	*50 citation sentences*	*50 citation sentences*	*50 citation sentences*
No consensus	*50 citation sentences*	*50 citation sentences*	*50 citation sentences*	*50 citation sentences*	*50 citation sentences*
Questionable*	*50 citation sentences*	*50 citation sentences*	*50 citation sentences*	*50 citation sentences*	*50 citation sentences*
Refut*	*50 citation sentences*	*50 citation sentences*	*50 citation sentences*	*50 citation sentences*	*50 citation sentences*

Which combinations are most valid?

Sampled 50 citation sentences for every combination

Two coders independently labeled them as Valid disagreement or Invalid

Take the most valid as our indicator of disagreement

Signal & filter terms

	standalone	+studies	+ideas	+methods	+results
Challenge*	50 citation sentences	50 citation sentences	50 citation sentences	50 citation sentences	50 citation sentences
Conflict*	50 citation sentences	50 citation sentences	50 citation sentences	50 citation sentences	50 citation sentences
Contradict*	50 citation sentences	50 citation sentences	50 citation sentences	50 citation sentences	50 citation sentences
Contrary	50 citation sentences	50 citation sentences	50 citation sentences	50 citation sentences	50 citation sentences
Contrast*	50 citation sentences	50 citation sentences	50 citation sentences	50 citation sentences	50 citation sentences
Contravers*	50 citation sentences	50 citation sentences	50 citation sentences	50 citation sentences	50 citation sentences
Debat*	50 citation sentences	50 citation sentences	50 citation sentences	50 citation sentences	50 citation sentences
Differ*	50 citation sentences	50 citation sentences	50 citation sentences	50 citation sentences	50 citation sentences
Disagree*	50 citation sentences	50 citation sentences	50 citation sentences	50 citation sentences	50 citation sentences
Disprov*	50 citation sentences	50 citation sentences	50 citation sentences	50 citation sentences	50 citation sentences
No consensus	50 citation sentences	50 citation sentences	50 citation sentences	50 citation sentences	50 citation sentences
Questionable*	50 citation sentences	50 citation sentences	50 citation sentences	50 citation sentences	50 citation sentences
Refut*	50 citation sentences	50 citation sentences	50 citation sentences	50 citation sentences	50 citation sentences

Which combinations are most valid?

Sampled 50 citation sentences for every combination

Two coders independently labeled them as Valid disagreement or Invalid

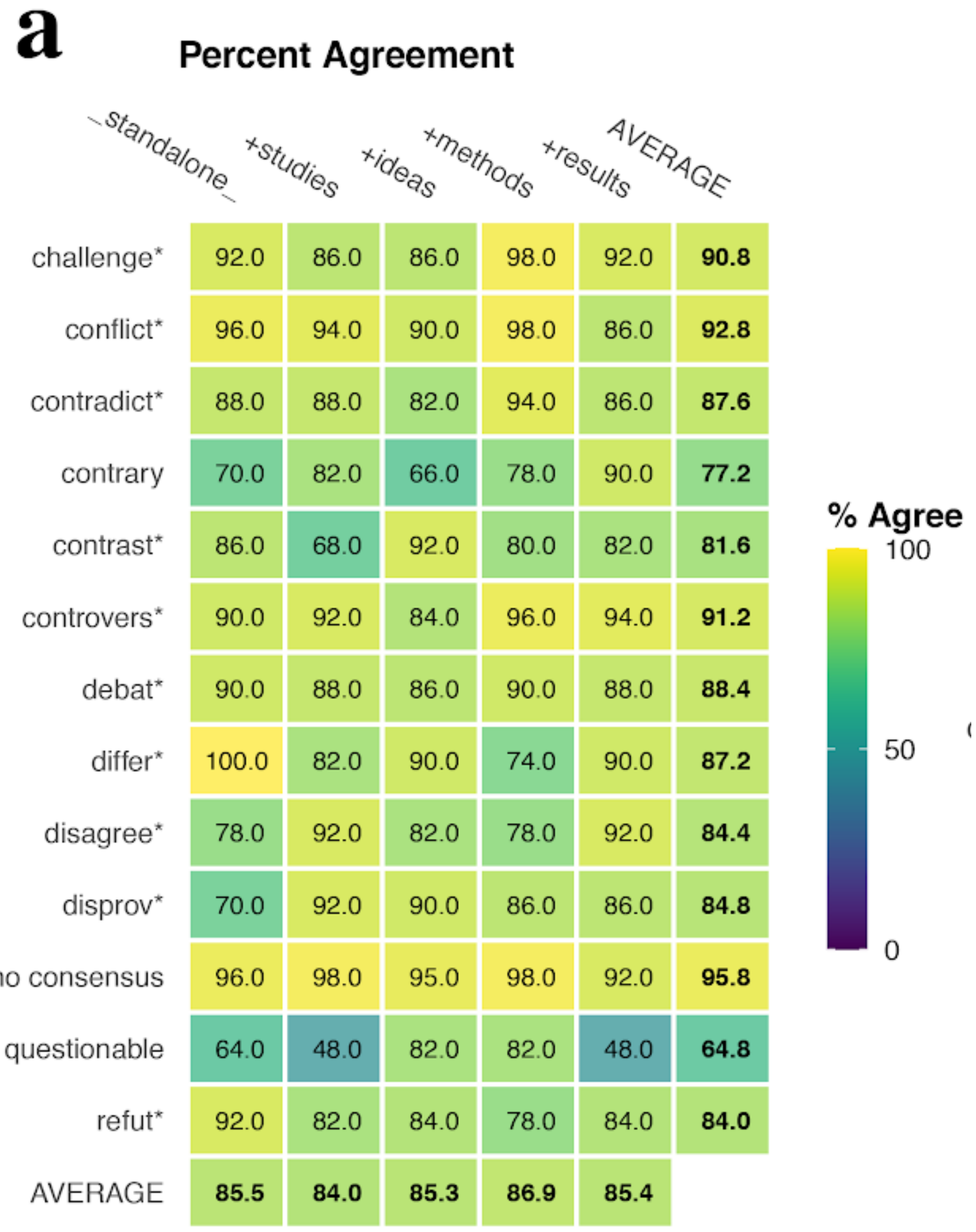
Take the most valid as our indicator of disagreement

23 queries representing ~450,000 citation sentences

Non-exhaustive, but precise!

Validity

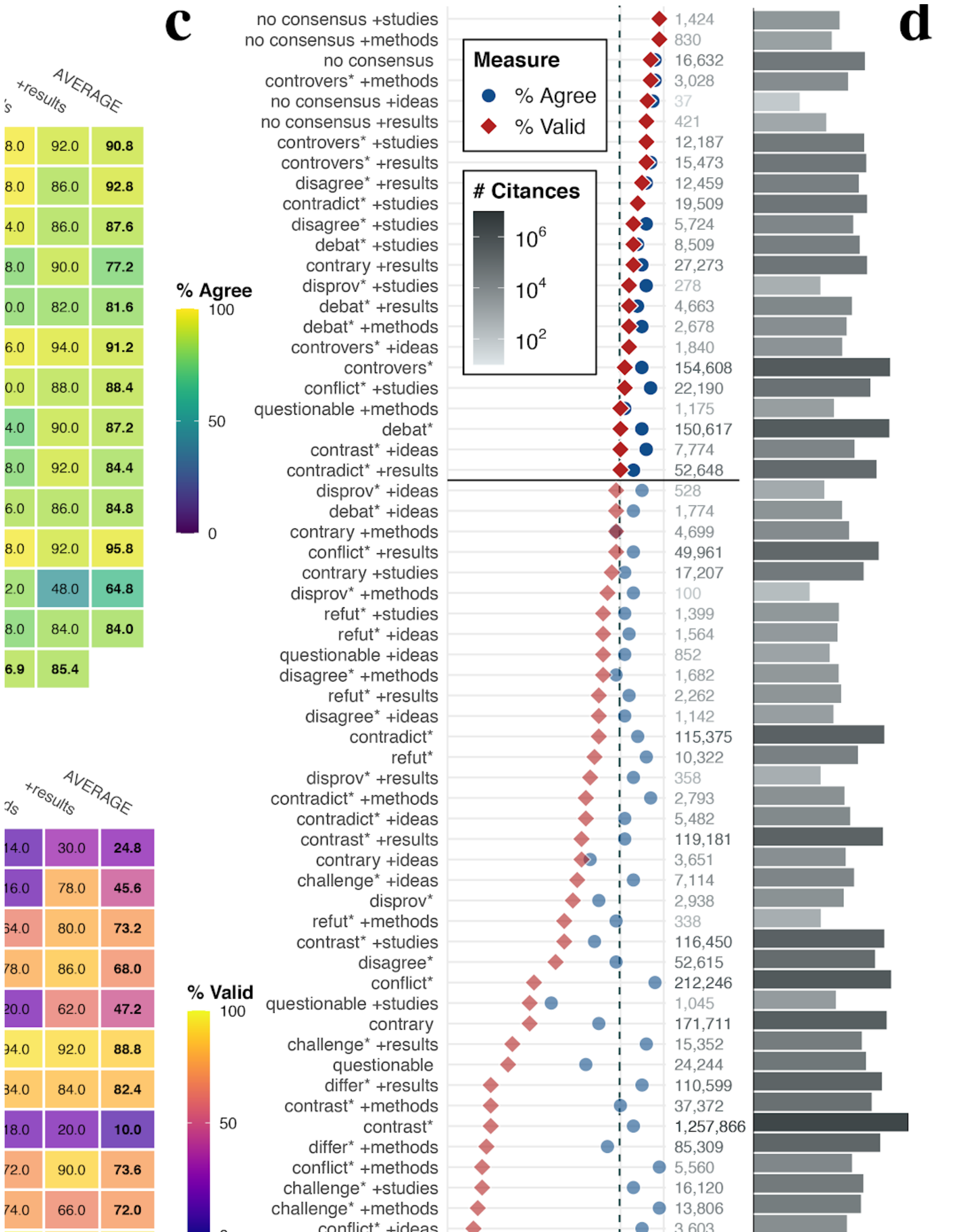
Coder 1: ✓ Valid **Coder 2: ✓ Valid**	Coder 1: ✓ Valid Coder 2: X Invalid
Coder 1: X Invalid Coder 2: ✓ Valid	**Coder 1: X Invalid** **Coder 2: X Invalid**



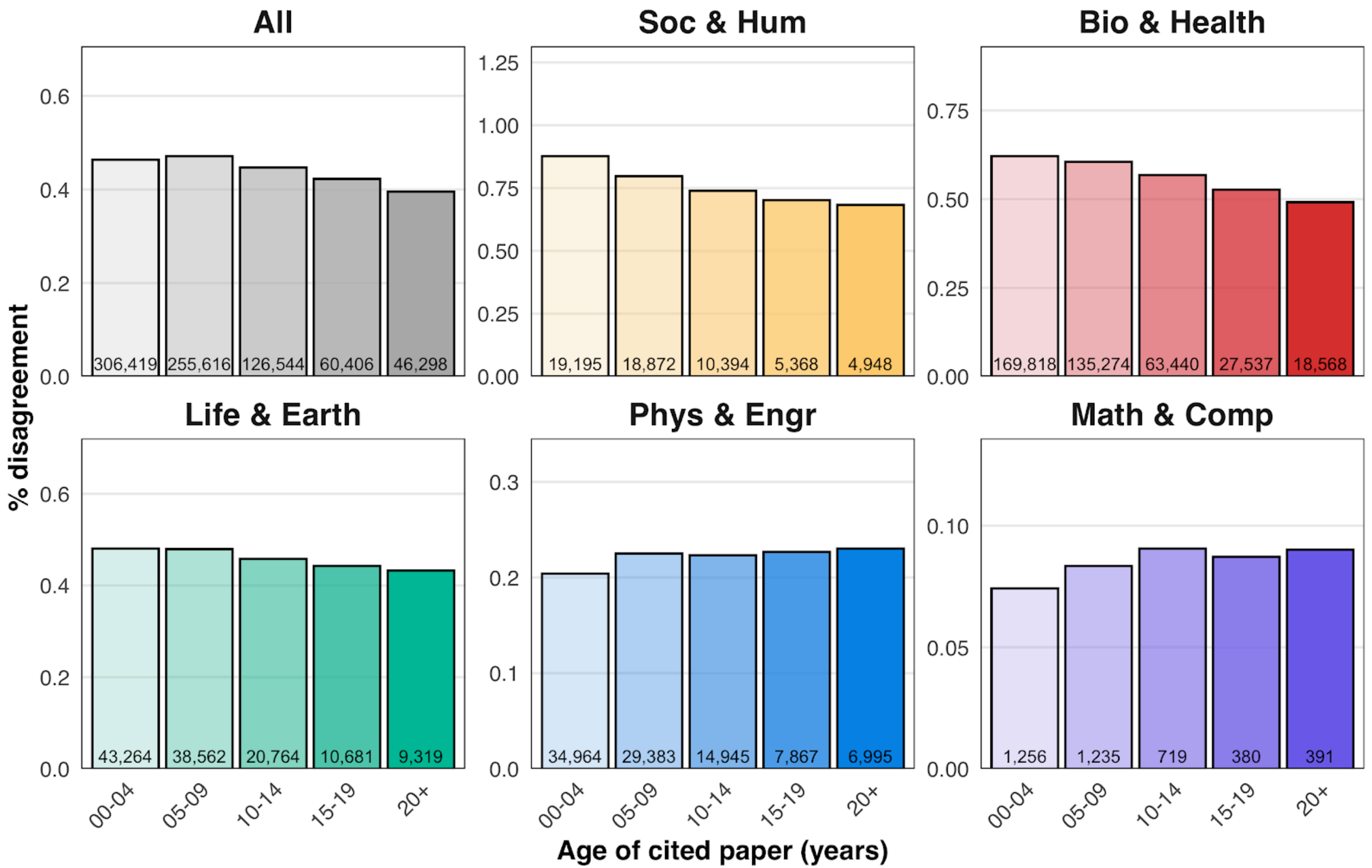
	standalone	+studies	+ideas	+methods	+results	AVERAGE
challenge*	92.0	86.0	86.0	98.0	92.0	**90.8**
conflict*	96.0	94.0	90.0	98.0	86.0	**92.8**
contradict*	88.0	88.0	82.0	94.0	86.0	**87.6**
contrary	70.0	82.0	66.0	78.0	90.0	**77.2**
contrast*	86.0	68.0	92.0	80.0	82.0	**81.6**
controvers*	90.0	92.0	84.0	96.0	94.0	**91.2**
debat*	90.0	88.0	86.0	90.0	88.0	**88.4**
differ*	100.0	82.0	90.0	74.0	90.0	**87.2**
disagree*	78.0	92.0	82.0	78.0	92.0	**84.4**
disprov*	70.0	92.0	90.0	86.0	86.0	**84.8**
no consensus	96.0	98.0	95.0	98.0	92.0	**95.8**
questionable	64.0	48.0	82.0	82.0	48.0	**64.8**
refut*	92.0	82.0	84.0	78.0	84.0	**84.0**
AVERAGE	**85.5**	**84.0**	**85.3**	**86.9**	**85.4**	

Validity

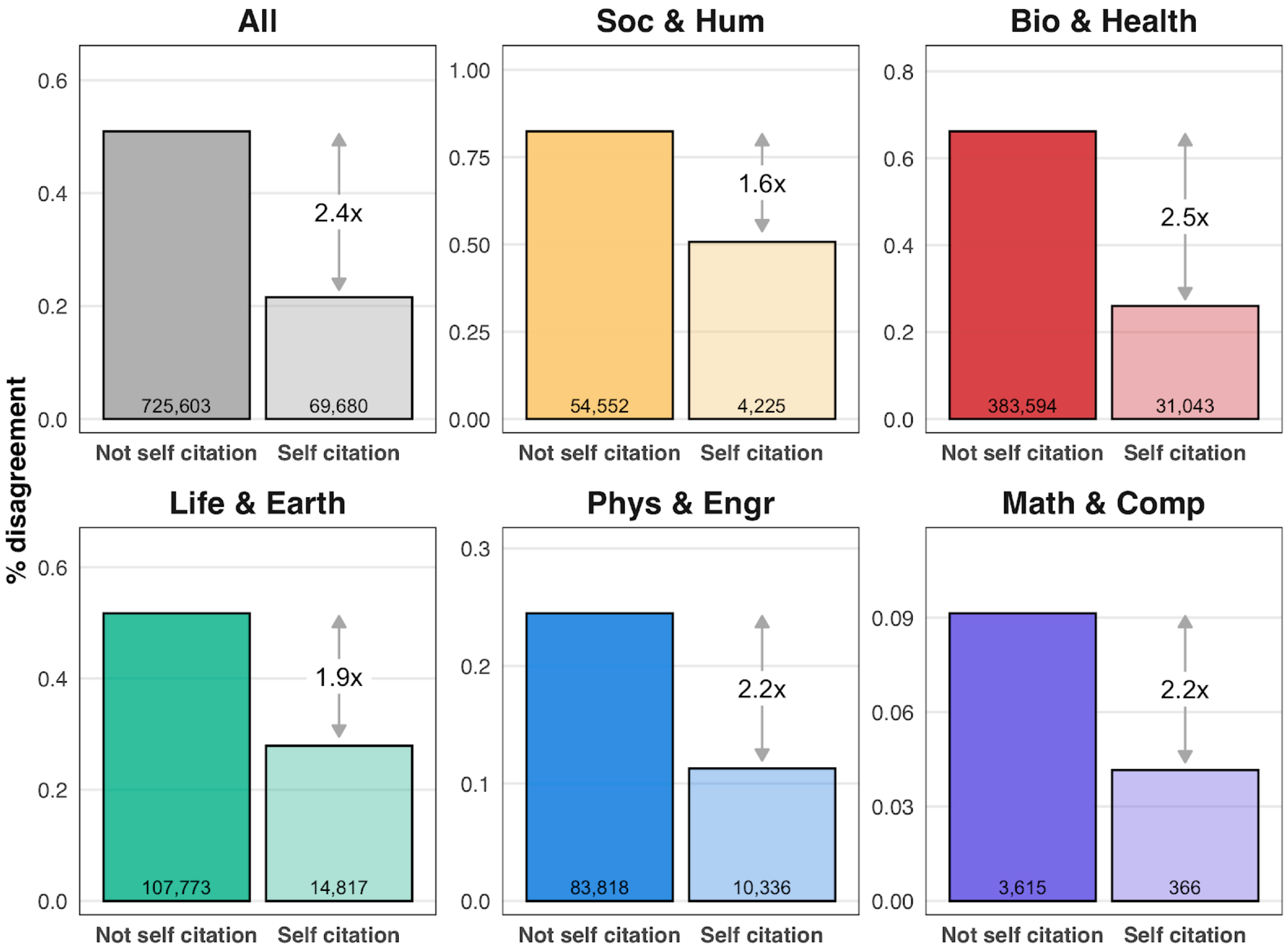
Coder 1: ✓ Valid **Coder 2: ✓ Valid**	Coder 1: ✓ Valid Coder 2: X Invalid
Coder 1: X Invalid Coder 2: ✓ Valid	Coder 1: X Invalid Coder 2: X Invalid



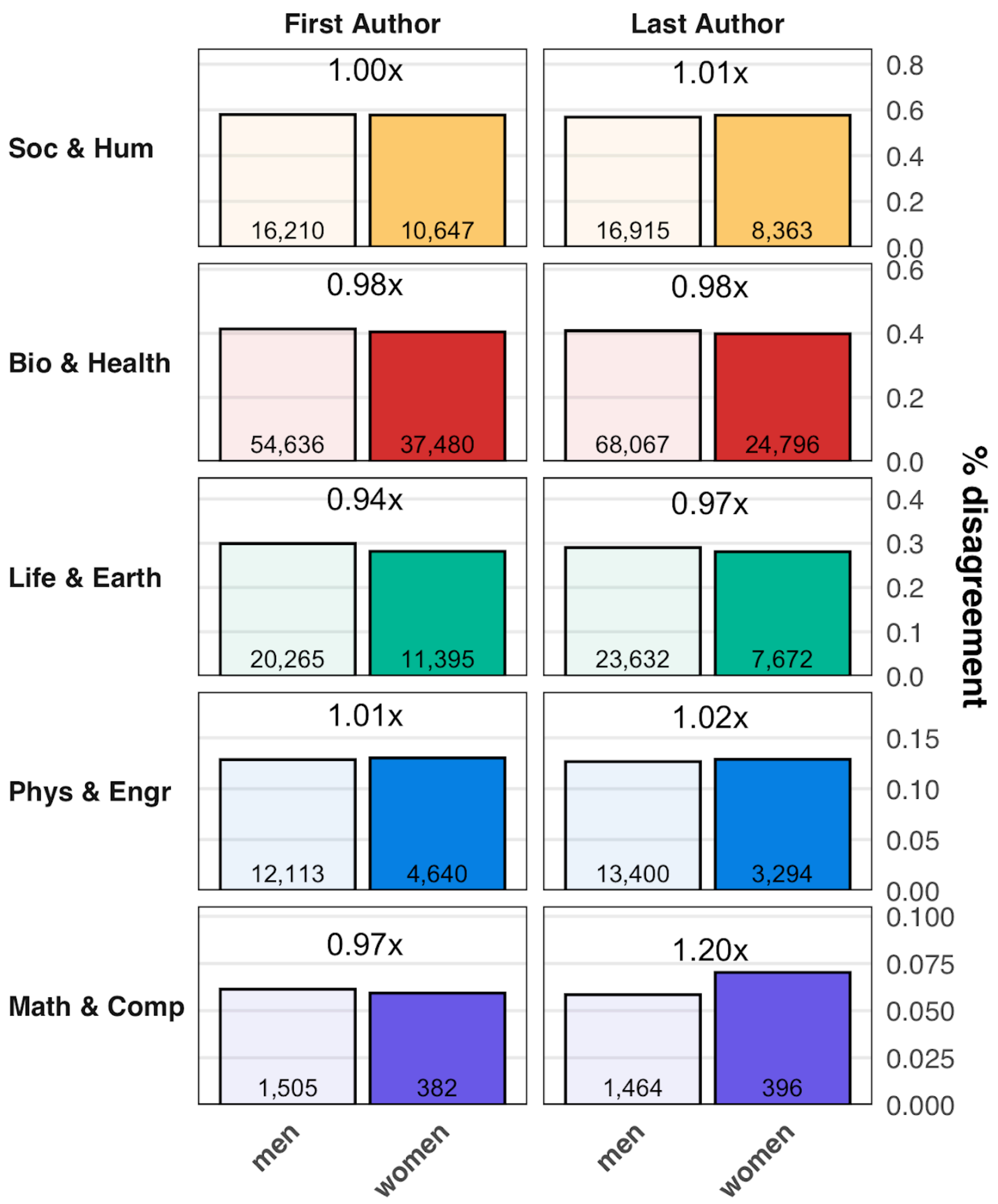
Less disagreement for older papers



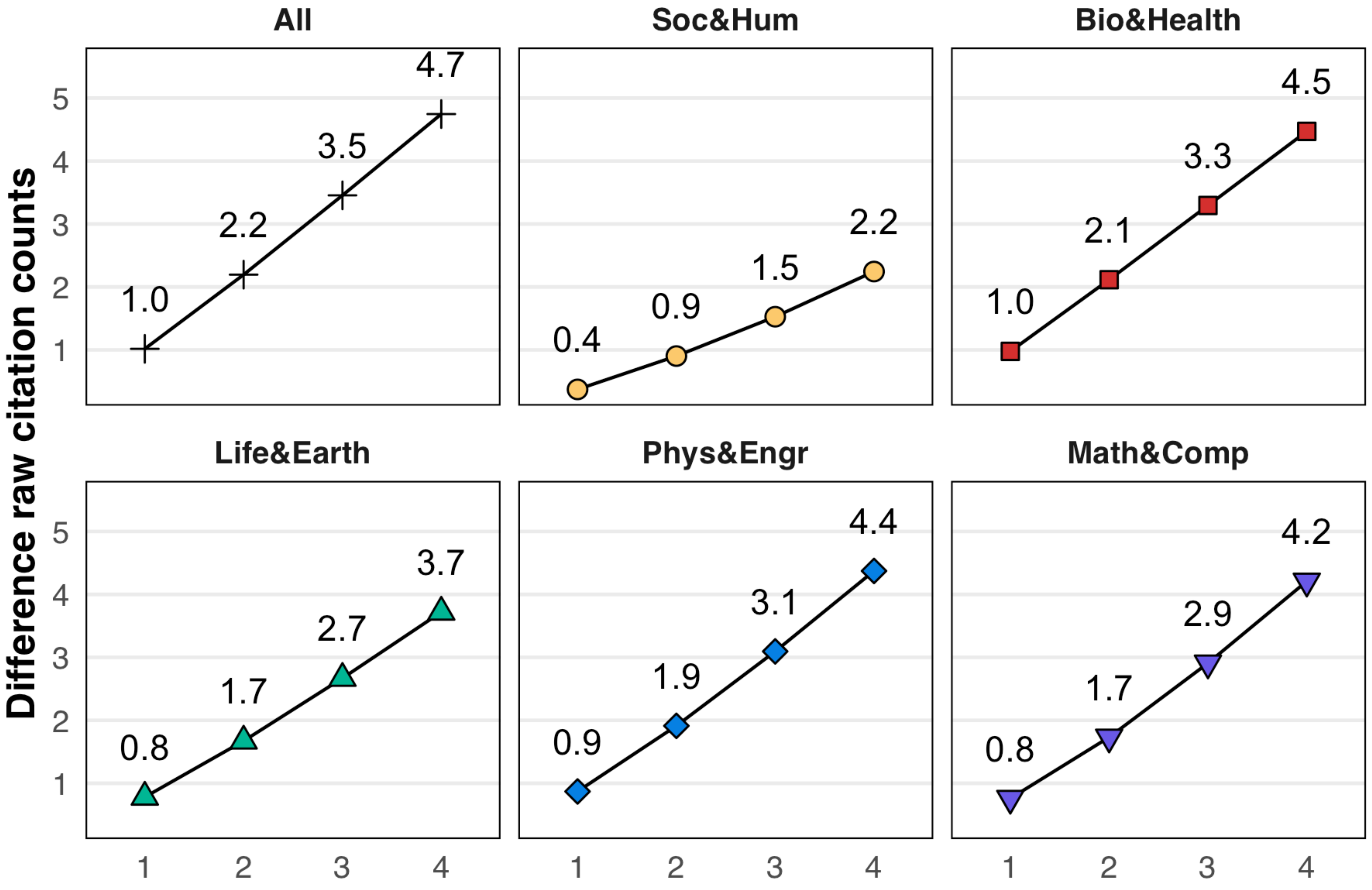
Disagreement and self-citation



Disagreement and gender

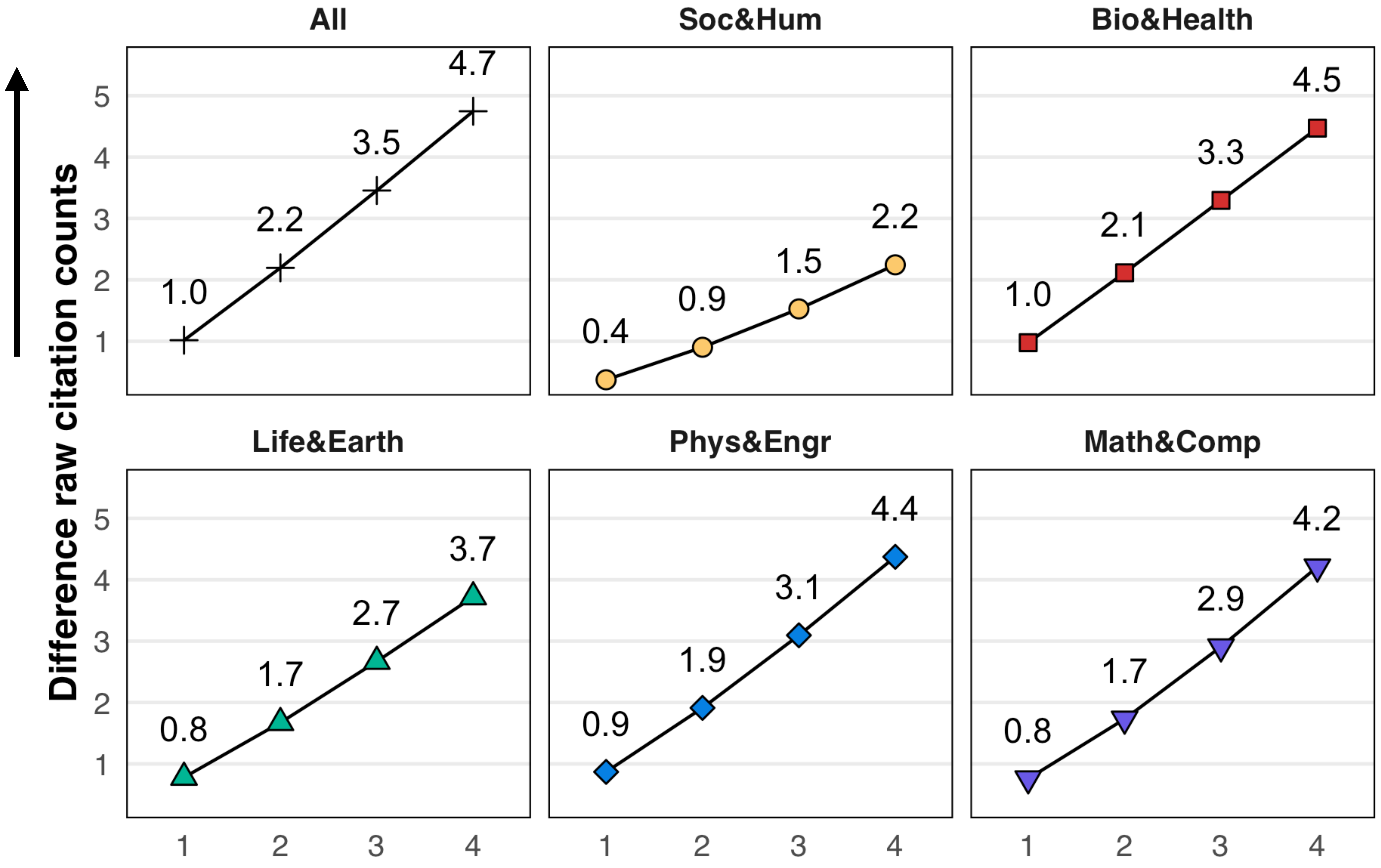


Compared papers with a disagreement citation to those without



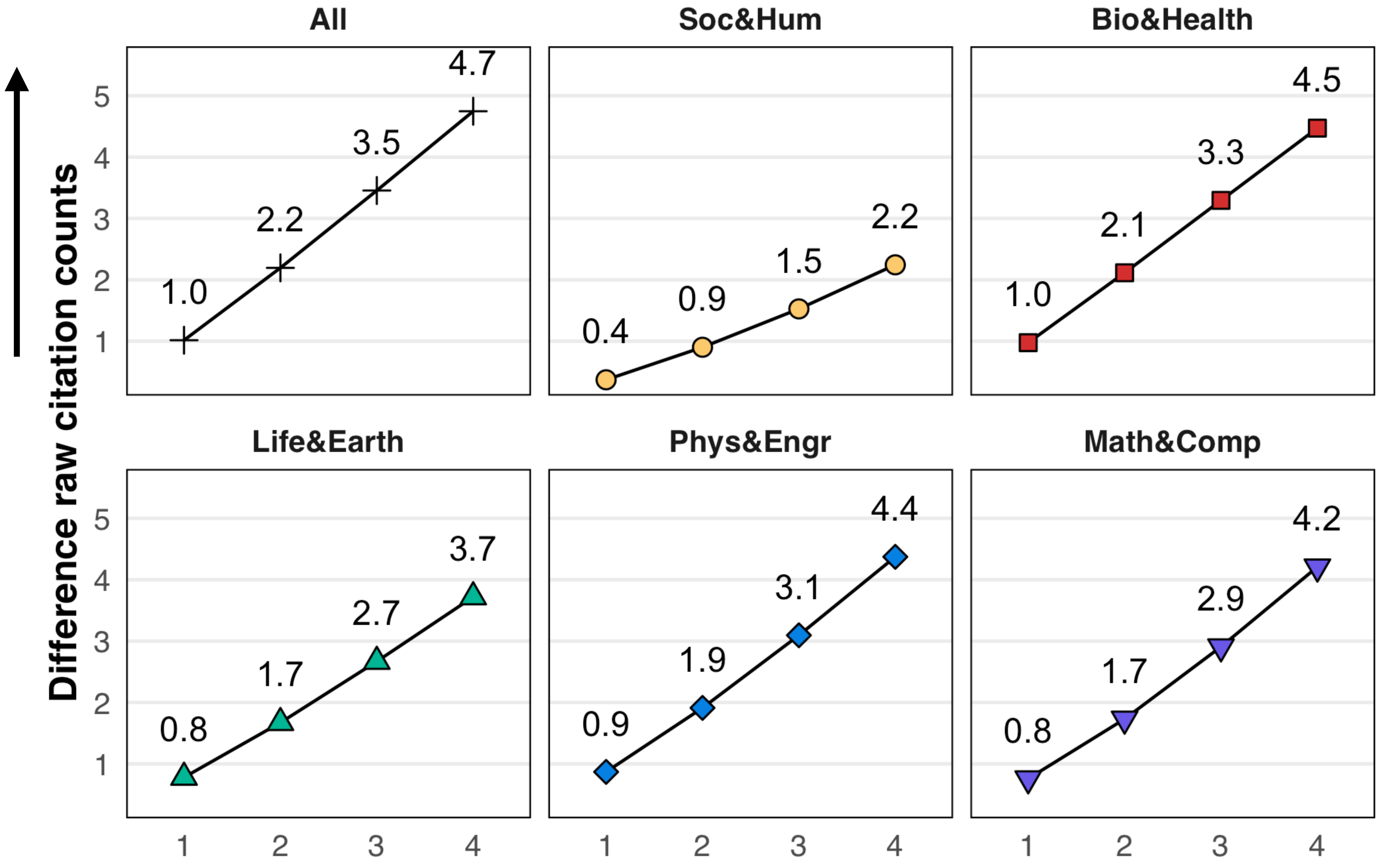
Compared papers with a disagreement citation to those without

Higher values indicate that papers with disagreement sentences received more citations



Compared papers with a disagreement citation to those without

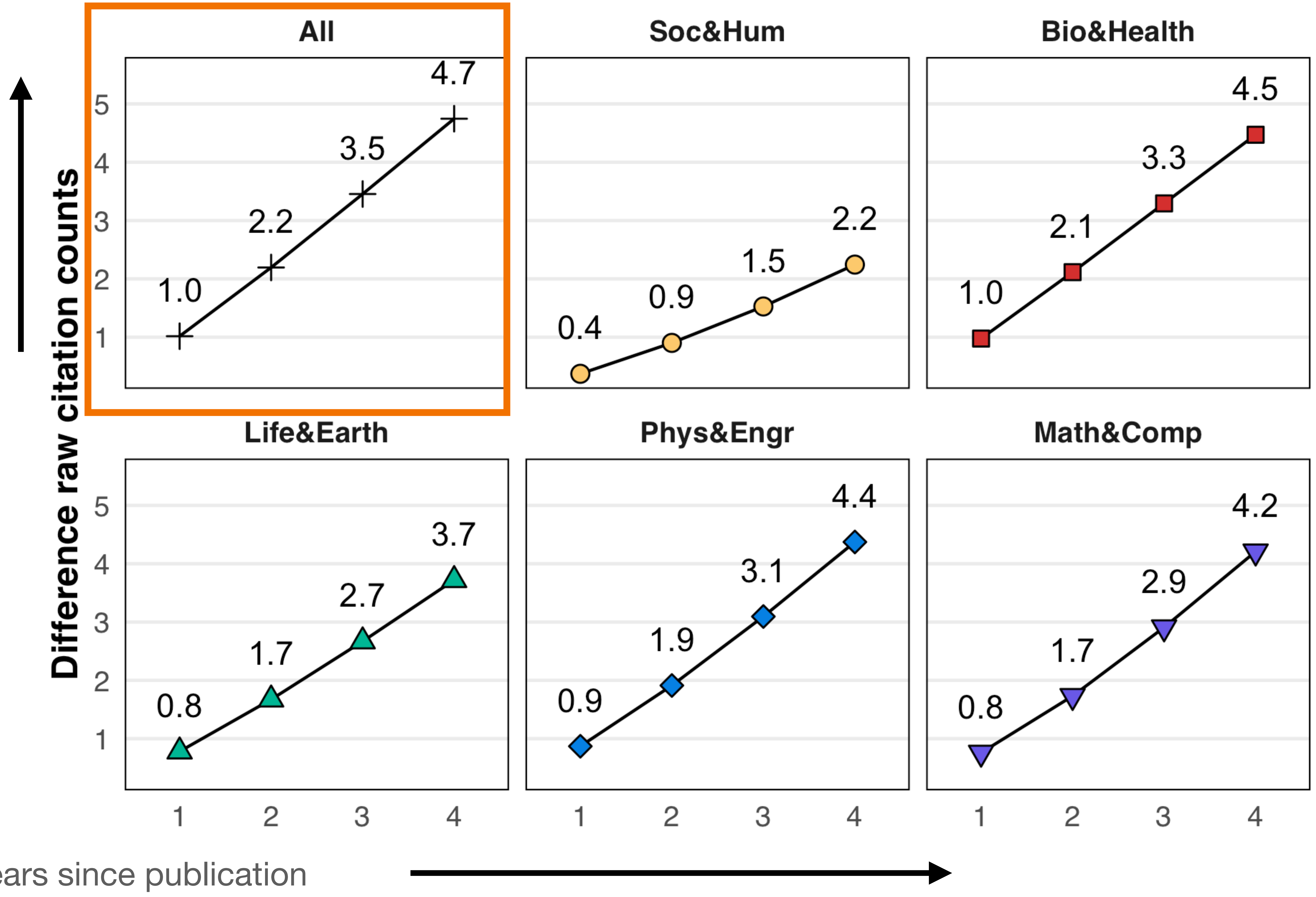
Higher values indicate that papers with disagreement sentences received more citations



Years since publication

Papers that disagree have more citations

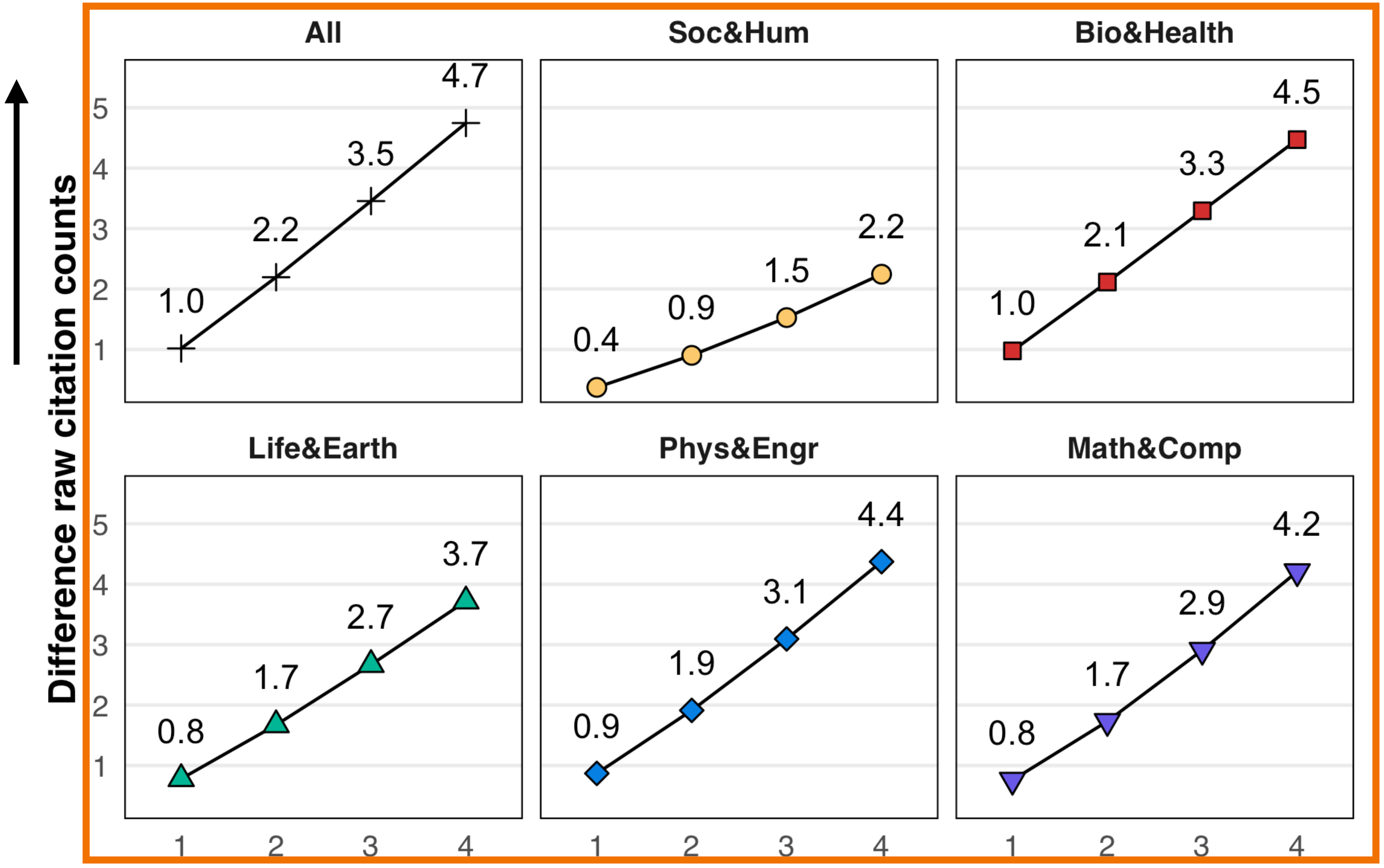
Higher values indicate that papers with disagreement sentences received more citations



Papers that disagree have more citations

True across every major field

Higher values indicate that papers with disagreement distances received more citations



Years since publication

What about being disagreed with?

Compare papers that were similar, when one received a disagreement citation

Disagreement

Received 10 citations in 5 years

Received ? Citations in following year

No Disagreement

Received 10 citations in 5 years

Received ? Citations in following year

What about being disagreed with?

Compare papers that were similar, when one received a disagreement citation

Disagreement

Received 10 citations in 5 years

Received _?_ Citations in following year

Compare citations received by the two populations

No Disagreement

Received 10 citations in 5 years

Received _?_ Citations in following year

Being disagreed with has little effect

Field	Avg. citations in year following disagreement	Expected citations in year following disagreement	Difference
All	3.03	3.08	-0.05
Bio & Health	2.73	2.81	-0.08
Life & Earth	3.43	3.35	+0.08
Math & Comp	3.36	3.34	+0.02
Phys & Engr	3.55	3.52	+0.03
Soc & Hum	3.04	3.11	-0.07

Being disagreed with has little effect

Less than one tenth of a citation difference, in all cases

Field	Avg. citations in year following disagreement	Expected citations in year following disagreement	Difference
All	3.03	3.08	-0.05
Bio & Health	2.73	2.81	-0.08
Life & Earth	3.43	3.35	+0.08
Math & Comp	3.36	3.34	+0.02
Phys & Engr	3.55	3.52	+0.03
Soc & Hum	3.04	3.11	-0.07

Difference in citations between those that received a disagreement citation, and those that didn't

Appendix — Mobility

Mobility is central to science

Institutionalized in evaluation

Article 19: Requirement for International Visits
When applying for promotion to full professor or equivalent rank, the applicants who were born after January 1, 1970, must complete at least a 6-month international visit.

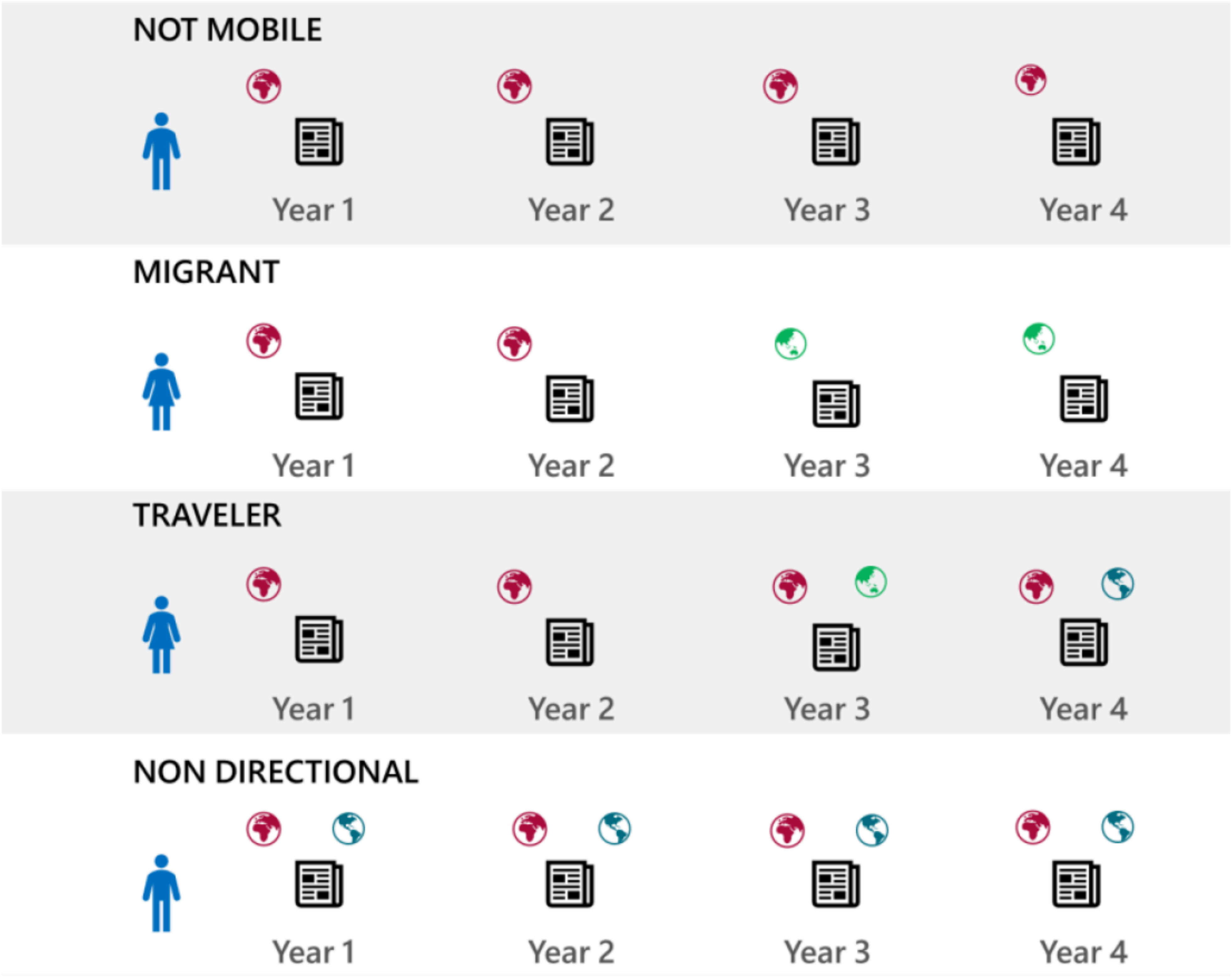
- Hangzhou Dianzi University

Thanks to Fei She

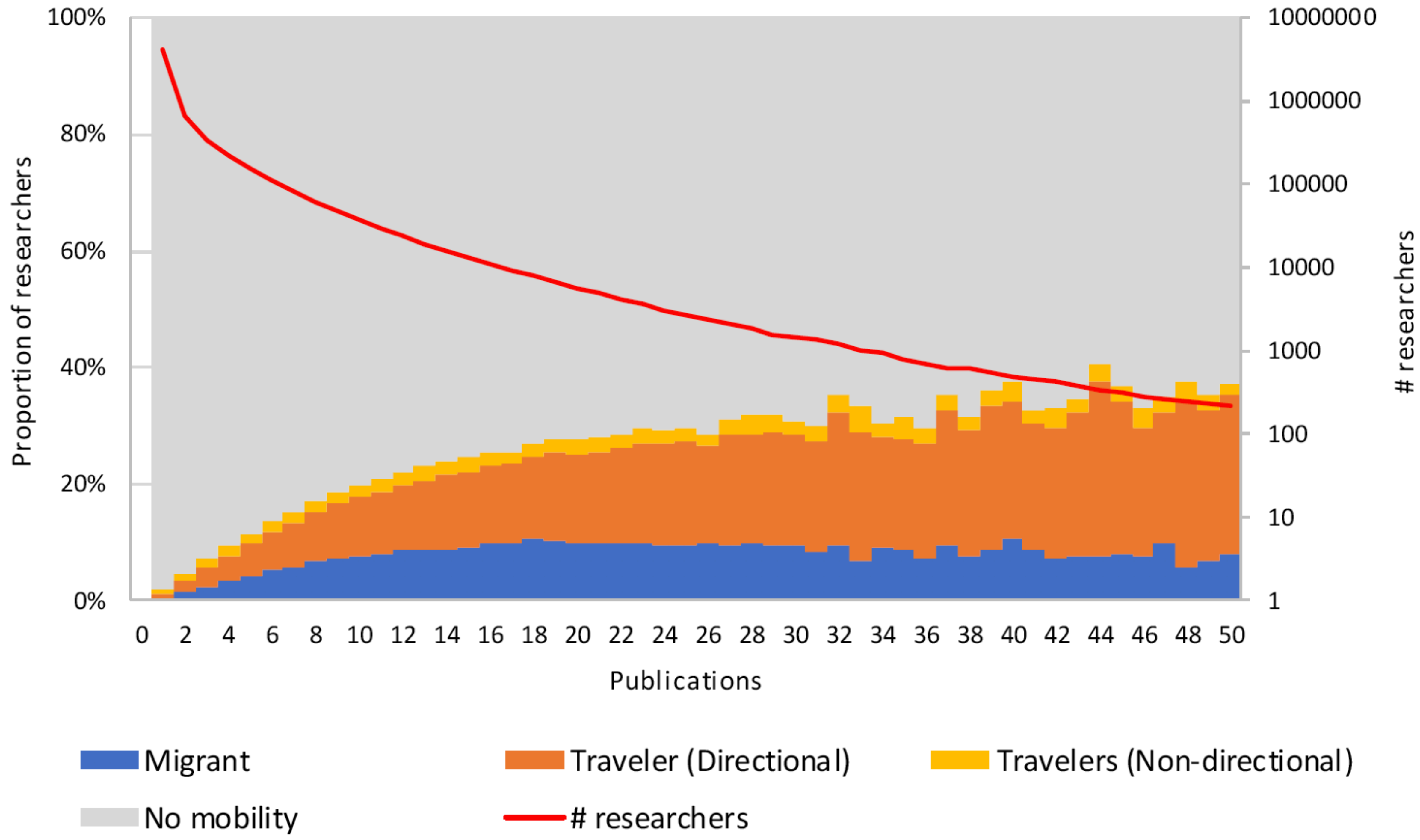
Mobility drives economies, cultural exchange, epidemics

Hanson, R., Mouton, C. A., Grissom, A. R., & Godges, J. P. (2020). *COVID-19 Air Traffic Visualization: Decisionmakers Should Base Travel Restrictions on Infection Rates Per Capita and Air Traffic Levels*.

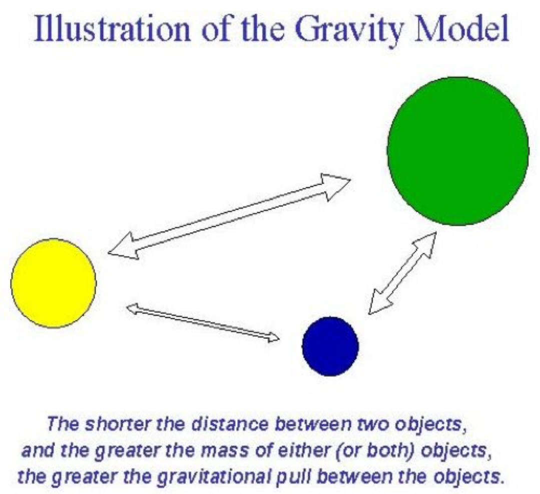
Mobility is complicated



Robinson-Garcia, N., Sugimoto, C. R., **Murray, D.**, Yegros-Yegros, A., Larivière, V., & Costas, R. (2018). Scientific mobility indicators in practice: International mobility profiles at the country level. *El Profesional de La Información*, *27*(3), 511.



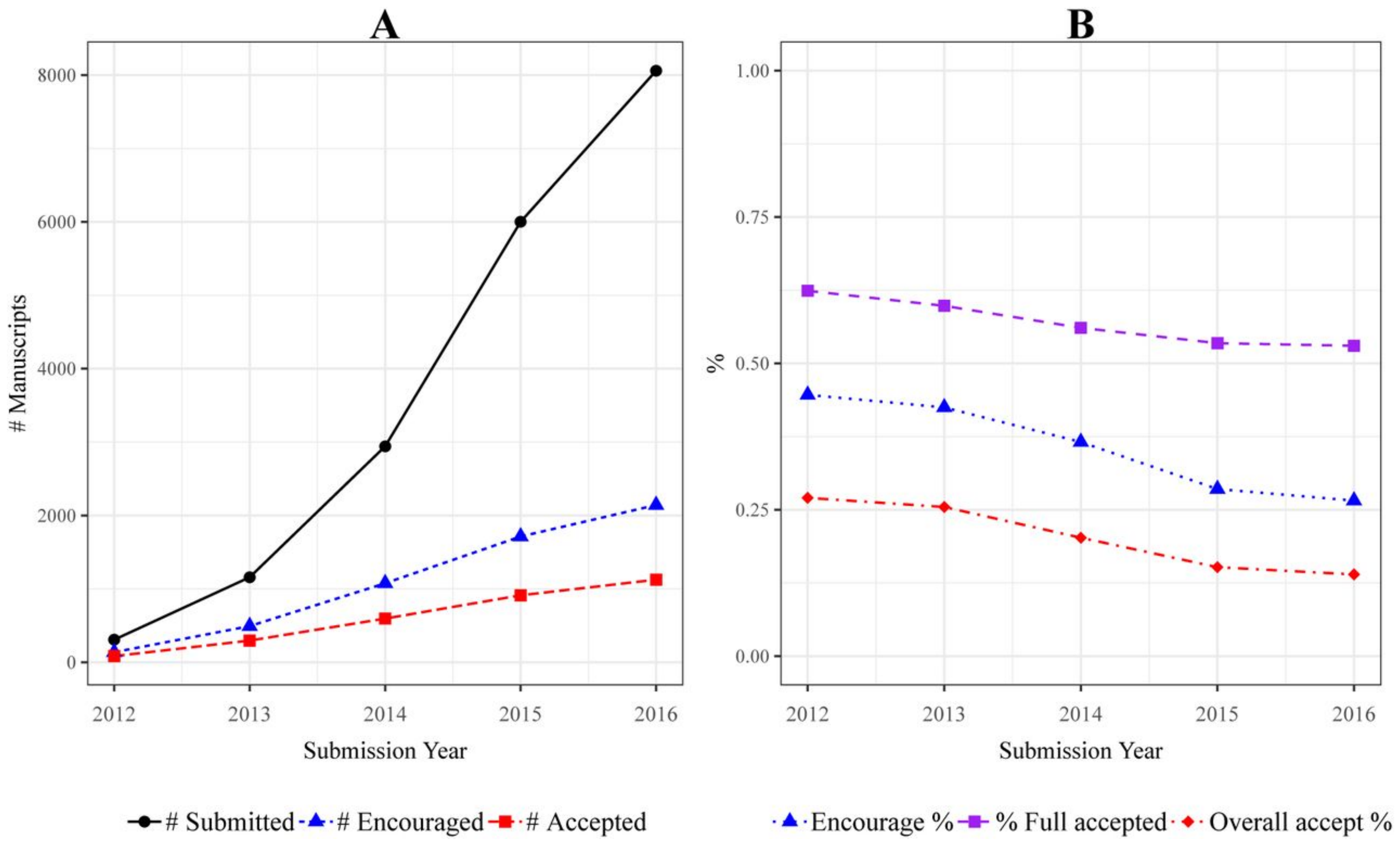
Robinson-Garcia, N., Sugimoto, C. R., **Murray, D.**, Yegros-Yegros, A., Larivière, V., & Costas, R. (2019). The many faces of mobility: Using bibliometric data to measure the movement of scientists. *Journal of Informetrics*, *13*(1), 50–63.



Gravity Model

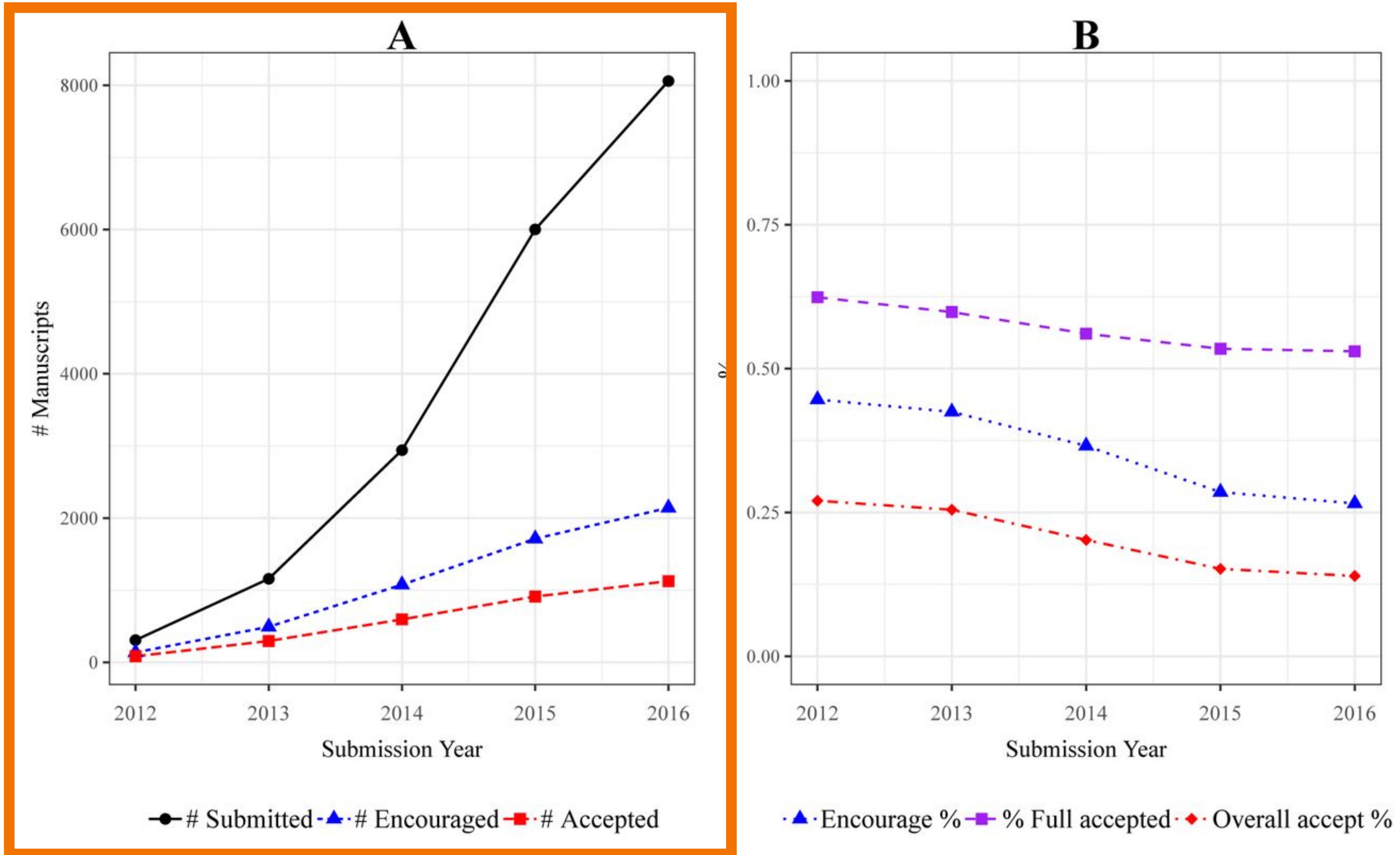
- Popular model of mobility
- ***Flows*** between places (co-affiliations) a function of their size and distance $T_{ij} = C\dfrac{m_i m_j}{f(r_{ij})}$
- We use two kinds of distance measure.
  1. Geographical distance
  2. Embedding distance $d_{ij} = 1 - \dfrac{v_i \cdot v_j}{|v_i||v_j|}$

eLife has evolved since its creation



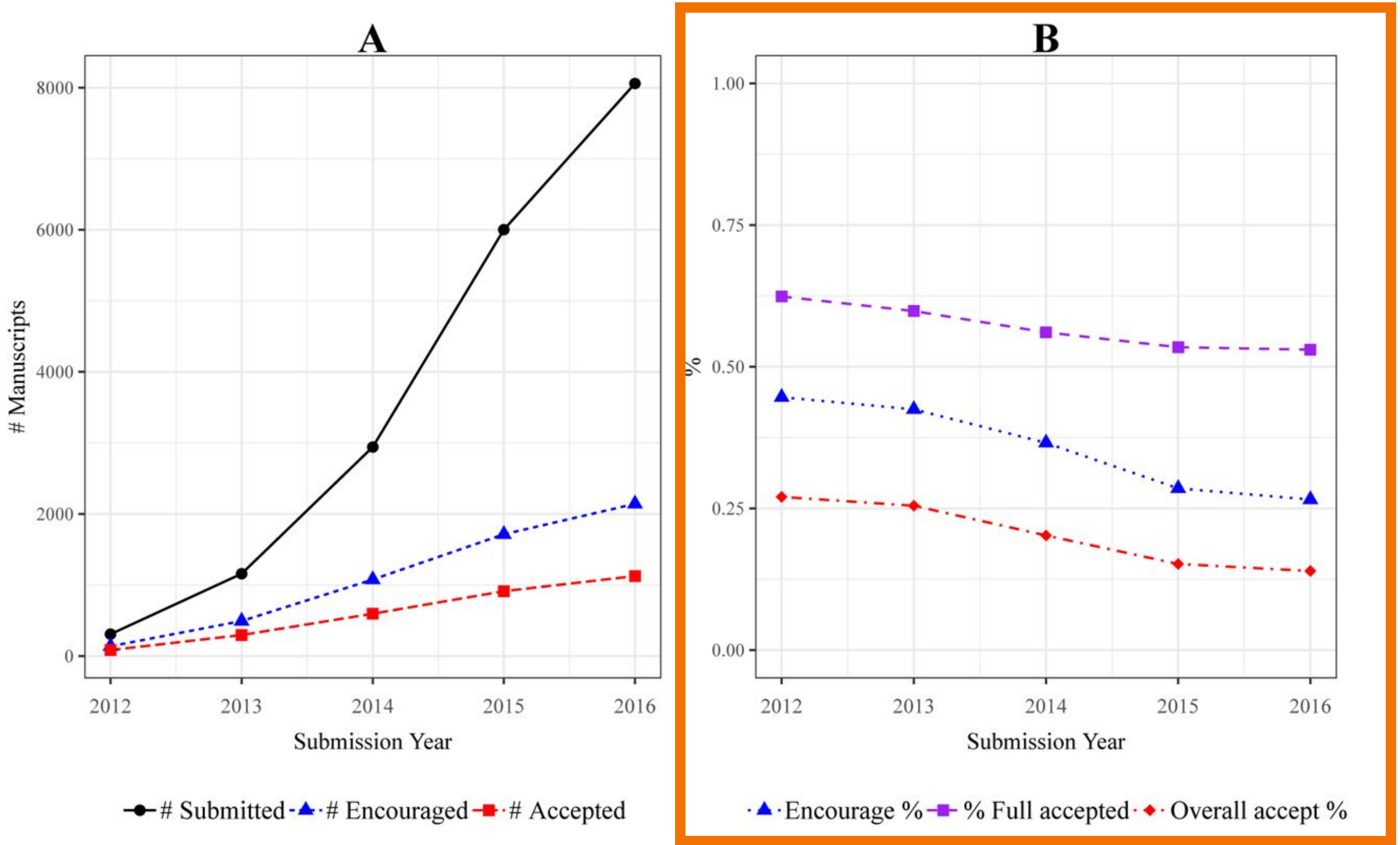
eLife has evolved since its creation

More submissions every year, outpacing published papers



eLife has evolved since its creation

Has become more selective over time

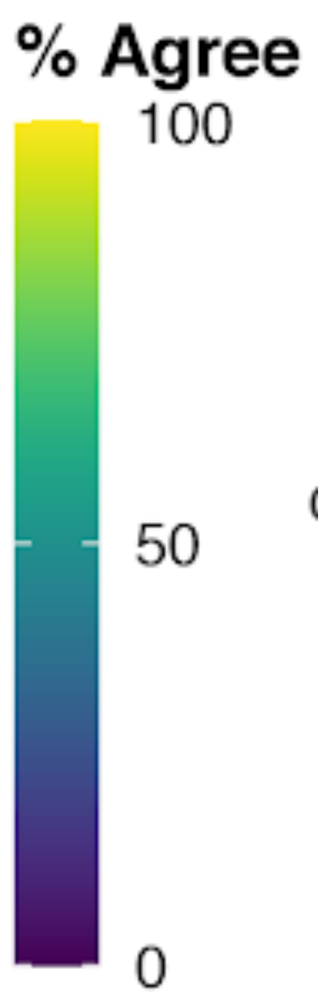


Validity

Coder 1: ✓ Valid **Coder 2: ✓ Valid**	Coder 1: ✓ Valid Coder 2: ✗ Invalid
Coder 1: ✗ Invalid Coder 2: ✓ Valid	**Coder 1: ✗ Invalid** **Coder 2: ✗ Invalid**

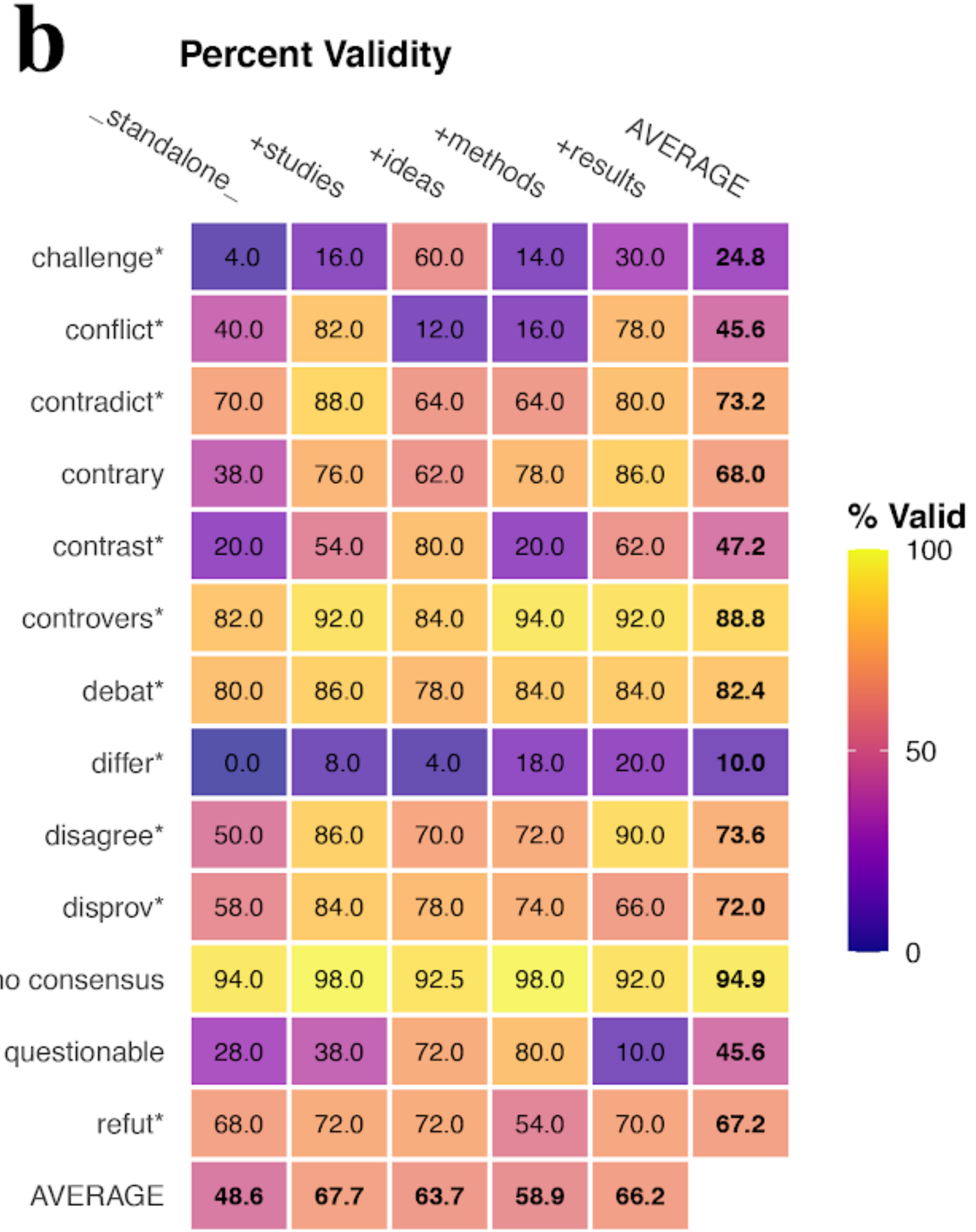
a **Percent Agreement**

	standalone	+studies	+ideas	+methods	+results	AVERAGE
challenge*	92.0	86.0	86.0	98.0	92.0	**90.8**
conflict*	96.0	94.0	90.0	98.0	86.0	**92.8**
contradict*	88.0	88.0	82.0	94.0	86.0	**87.6**
contrary	70.0	82.0	66.0	78.0	90.0	**77.2**
contrast*	86.0	68.0	92.0	80.0	82.0	**81.6**
controvers*	90.0	92.0	84.0	96.0	94.0	**91.2**
debat*	90.0	88.0	86.0	90.0	88.0	**88.4**
differ*	100.0	82.0	90.0	74.0	90.0	**87.2**
disagree*	78.0	92.0	82.0	78.0	92.0	**84.4**
disprov*	70.0	92.0	90.0	86.0	86.0	**84.8**
no consensus	96.0	98.0	95.0	98.0	92.0	**95.8**
questionable	64.0	48.0	82.0	82.0	48.0	**64.8**
refut*	92.0	82.0	84.0	78.0	84.0	**84.0**
AVERAGE	**85.5**	**84.0**	**85.3**	**86.9**	**85.4**	



Validity

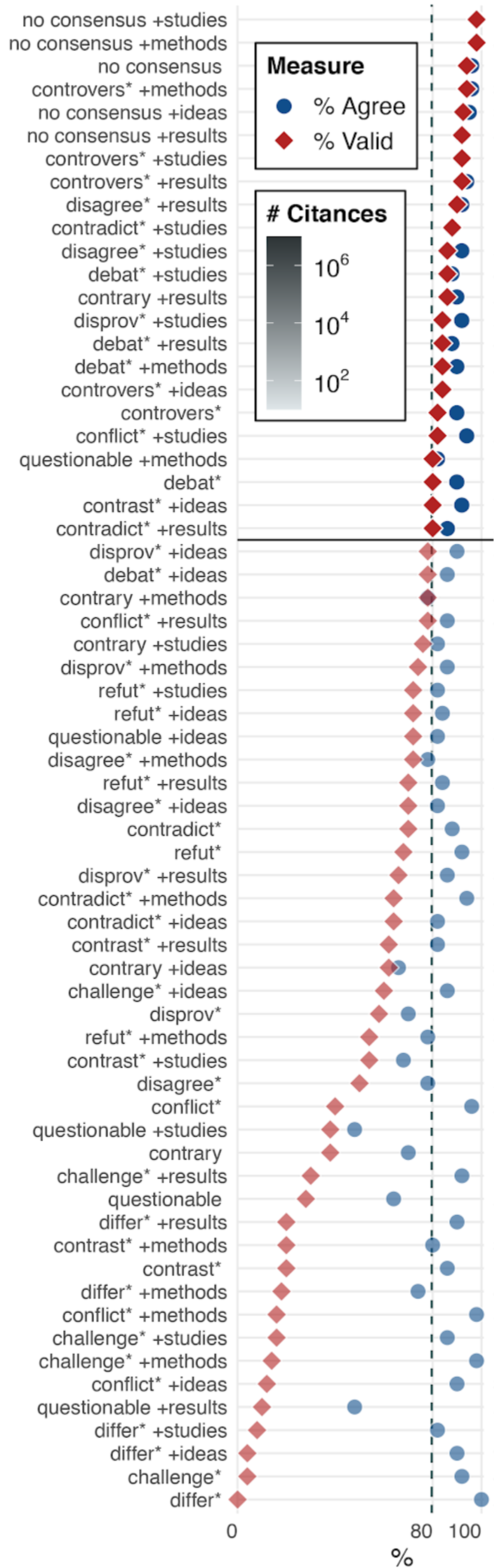
Coder 1: ✓ Valid **Coder 2: ✓ Valid**	Coder 1: ✓ Valid Coder 2: X Invalid
Coder 1: X Invalid Coder 2: ✓ Valid	Coder 1: X Invalid Coder 2: X Invalid



Both together

- % Agreement stable, %valid varied widely

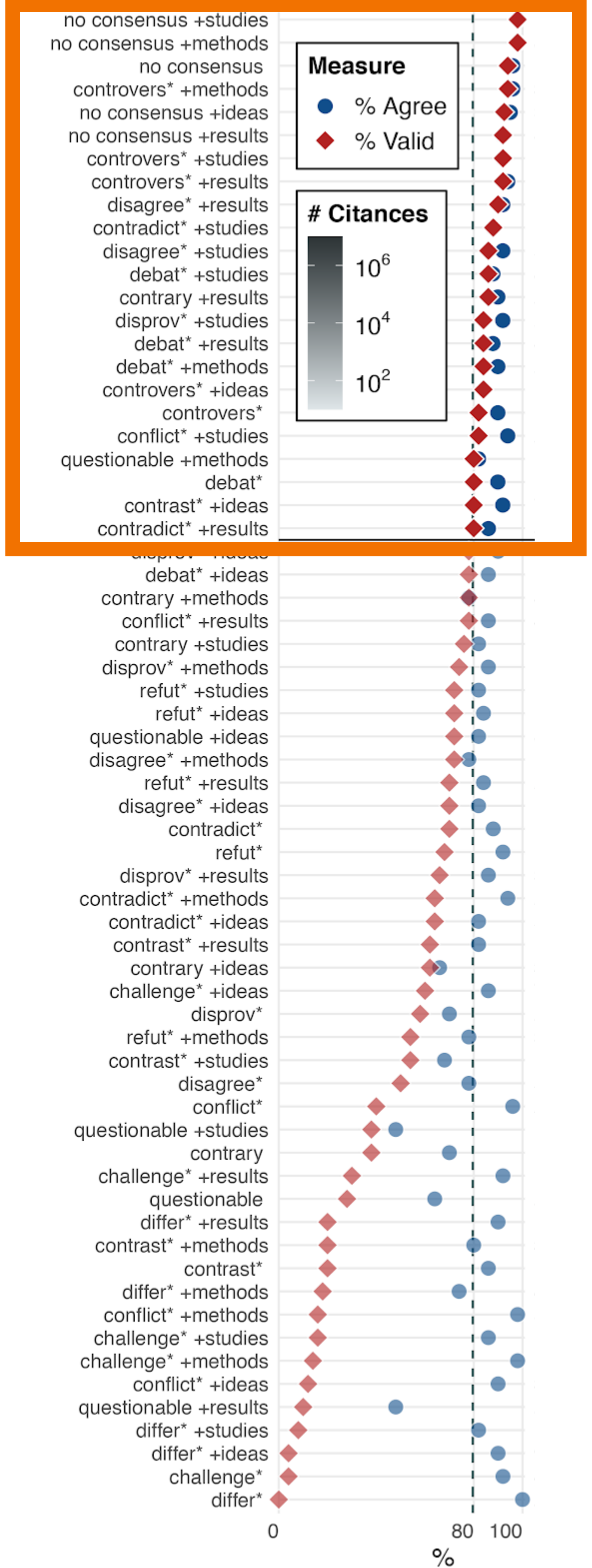
Ordered by validity



Both together

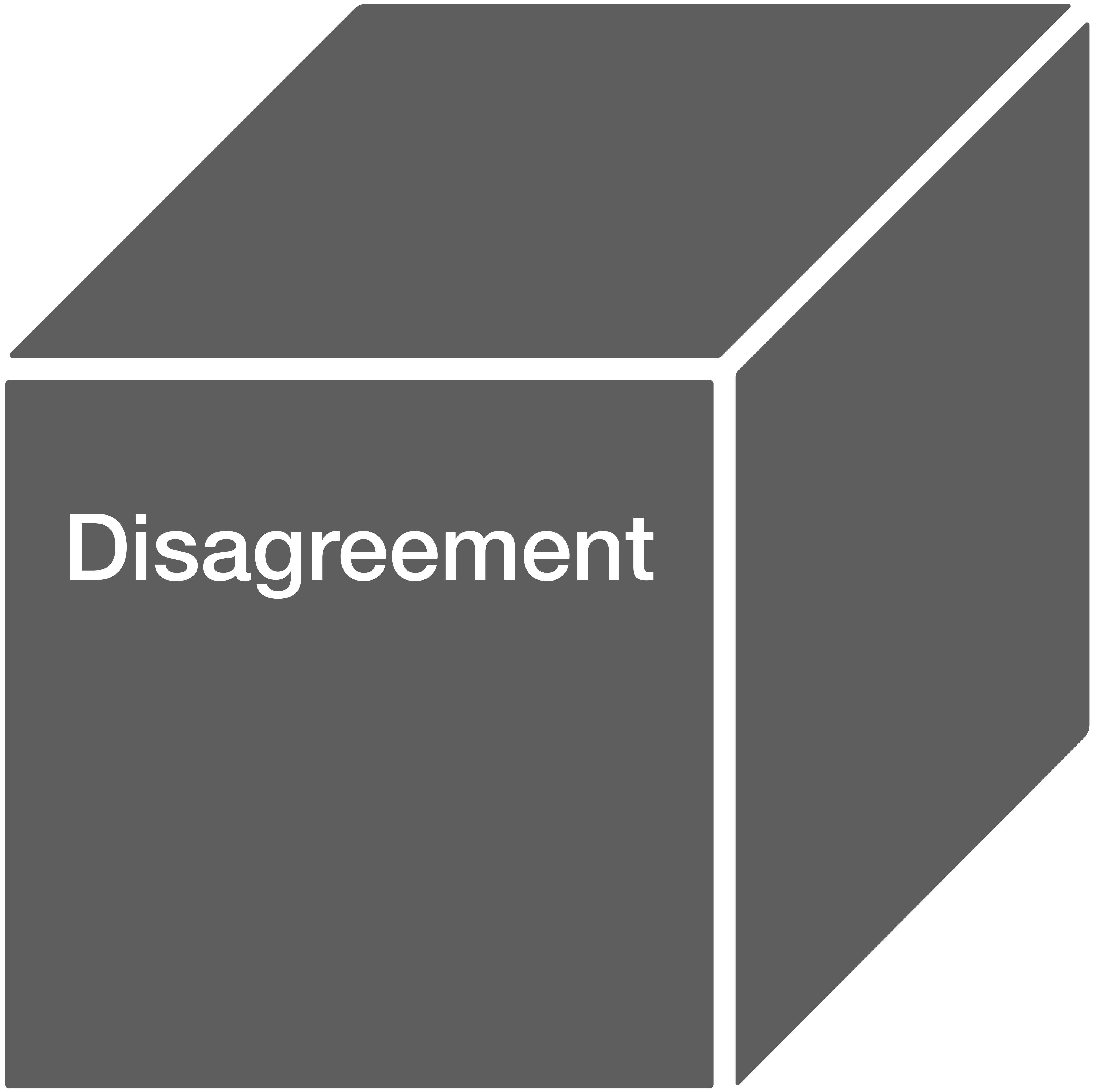
Ordered by validity

- % Agreement stable, %valid varied widely
- We don't want obviously bad queries
- Set a validity threshold, 80%



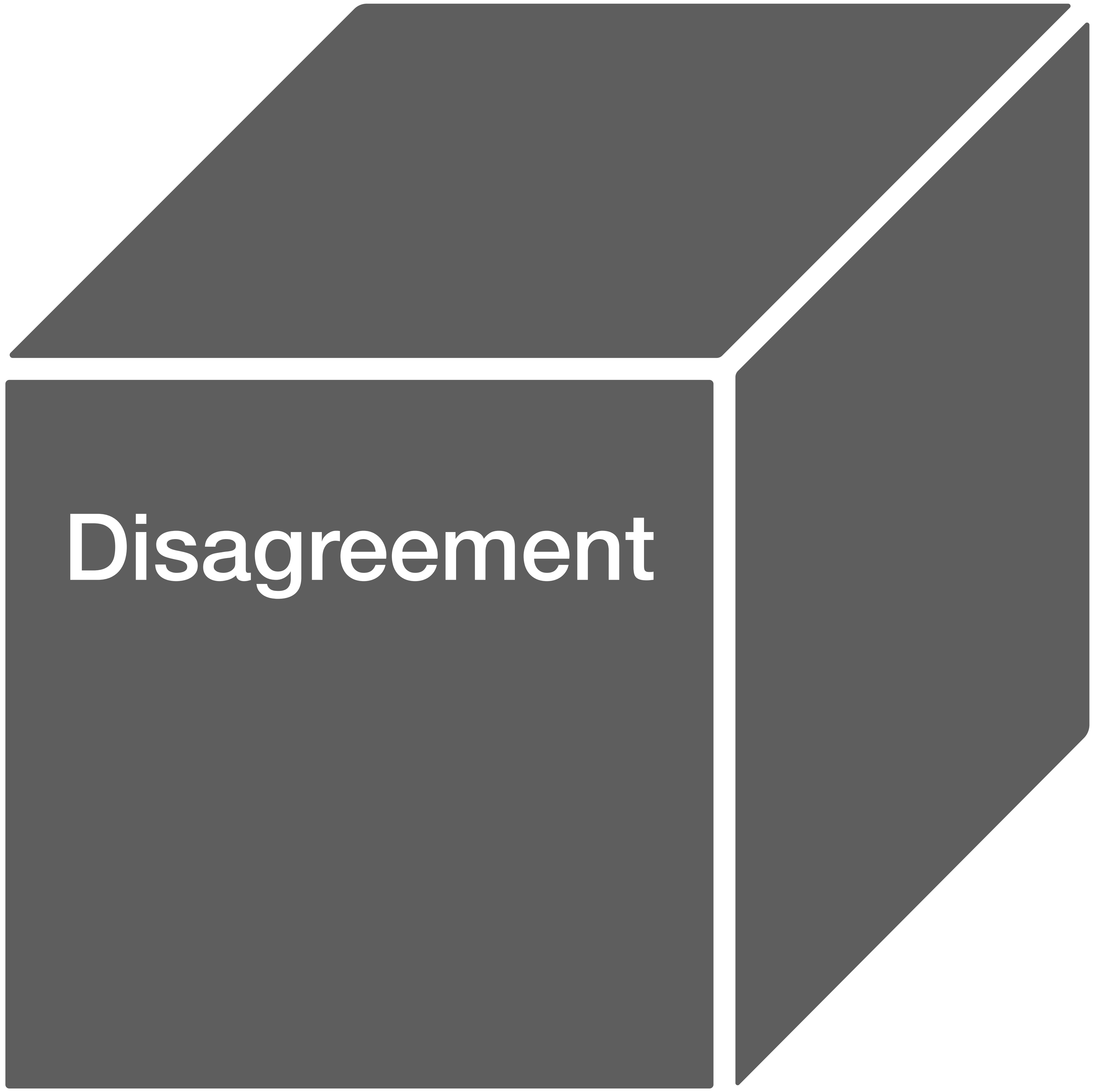
Both together

- % Agreement stable, %valid varied widely
- We don't want obviously bad queries
- Set a validity threshold, 80%
- 23 queries accounting for 450,000 citation sentences



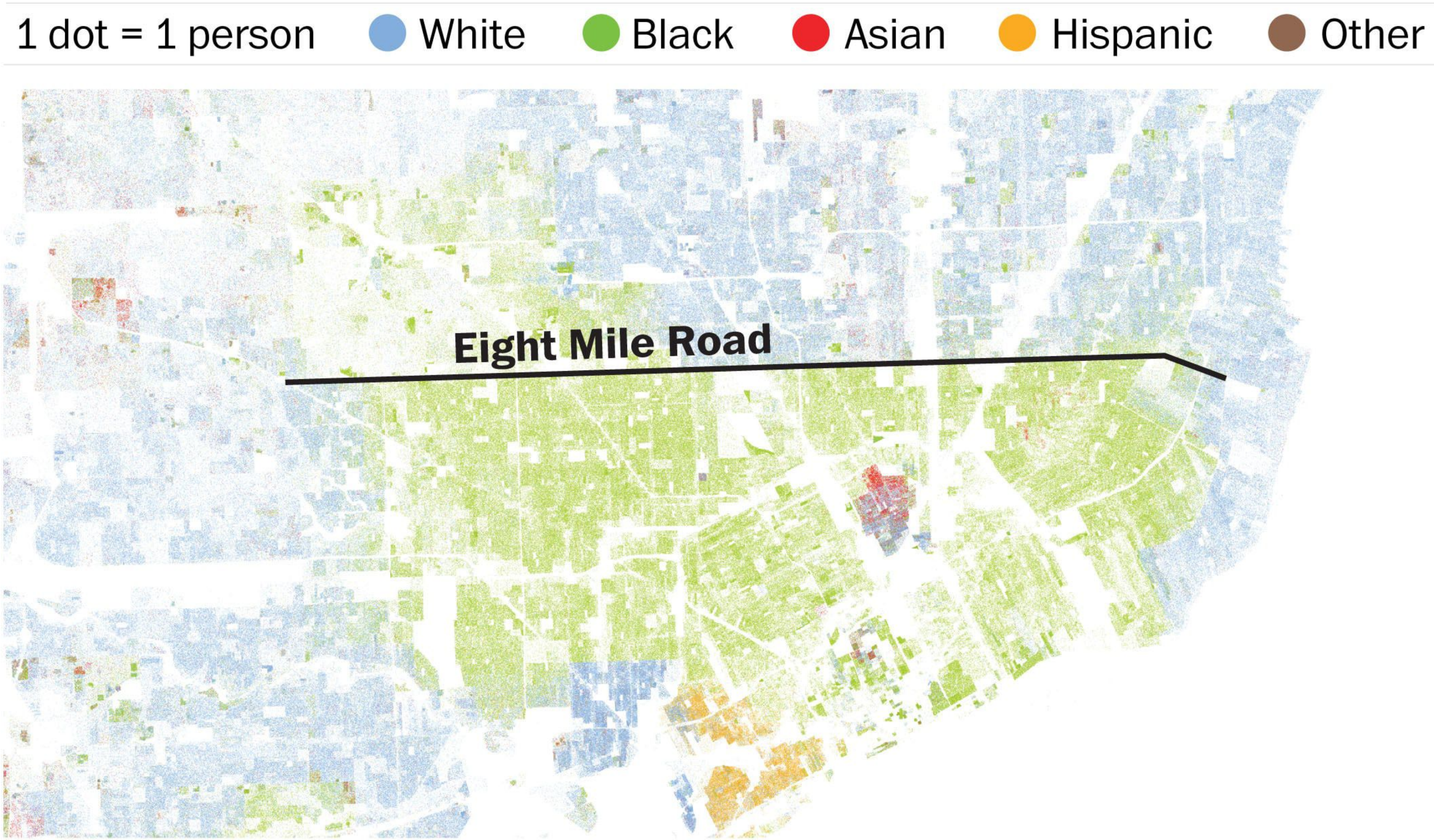
Both together

- % Agreement stable, %valid varied widely
- We don’t want obviously bad queries
- Set a validity threshold, 80%
- 23 queries accounting for 450,000 citation sentences. **Not exaustive, but preciese!**



Lines of segregation in Detroit

Economic opportunity and demographic makeup can wildly diverse in just a few city blocks



Source: U-Va. Cooper Center analysis of 2010 Census data

THE WASHINGTO

A basic principle: a good *representation* allows *prediction*.

$$P(w_t|w_c) \gg P(w_{\mathrm{random}}|w_c)$$

Target Context

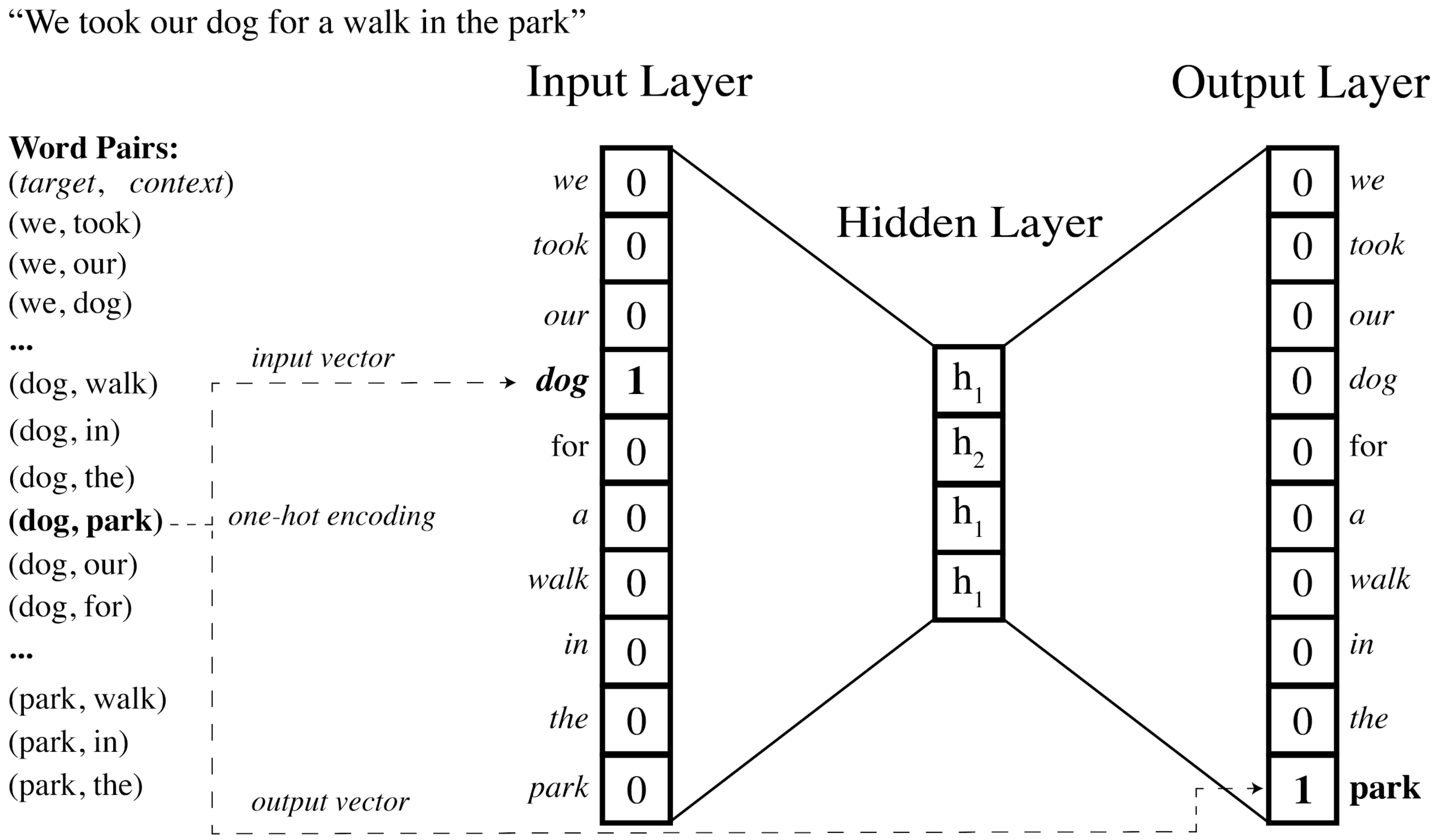
Let's assume that we can calculate the conditional probability with a function (e.g. dot product) of word **vectors** and **learn those vectors with a neural network**.

Learning word embeddings

Train a neural network to predict context words given a target

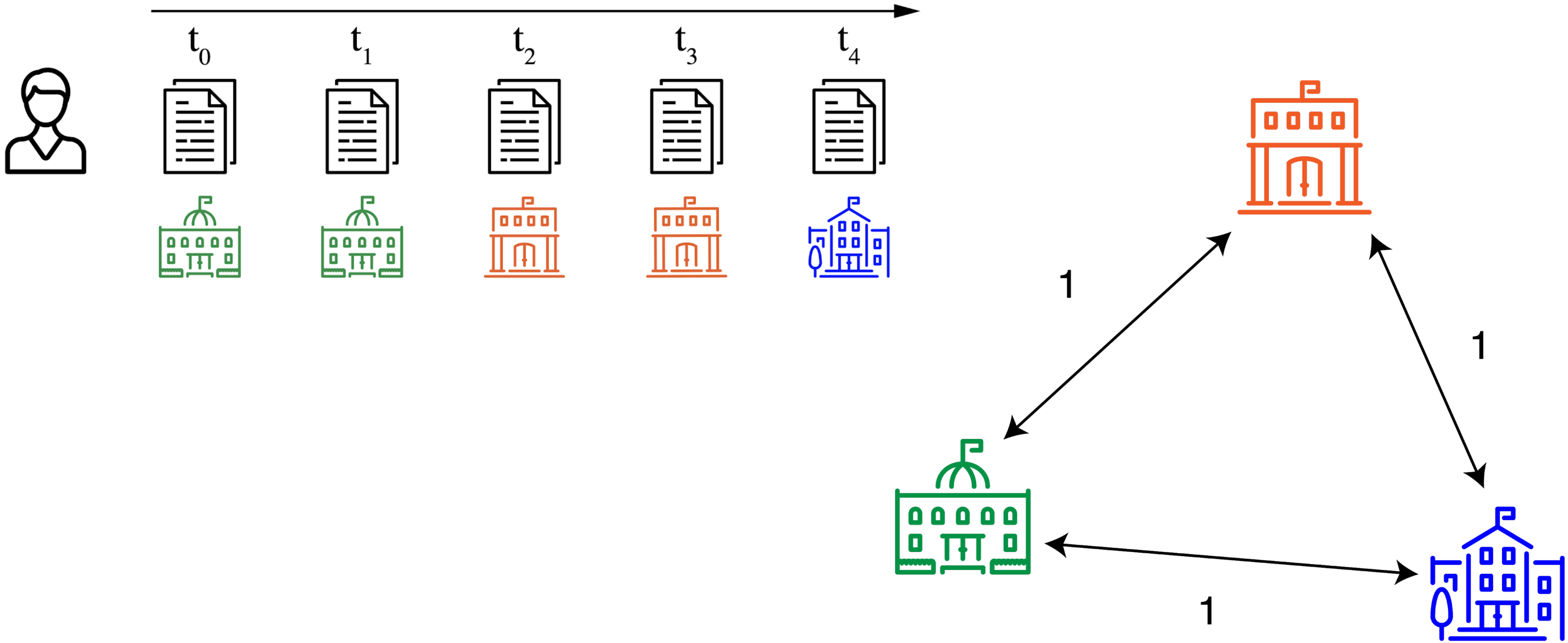
The hidden layer maps targets to contexts!

Words with similar contexts will have a similar "mapping" vector in the hidden layer



Each a trajectory of organizations

Derive “flux” from scientists career trajectories



Aside: why does it work?

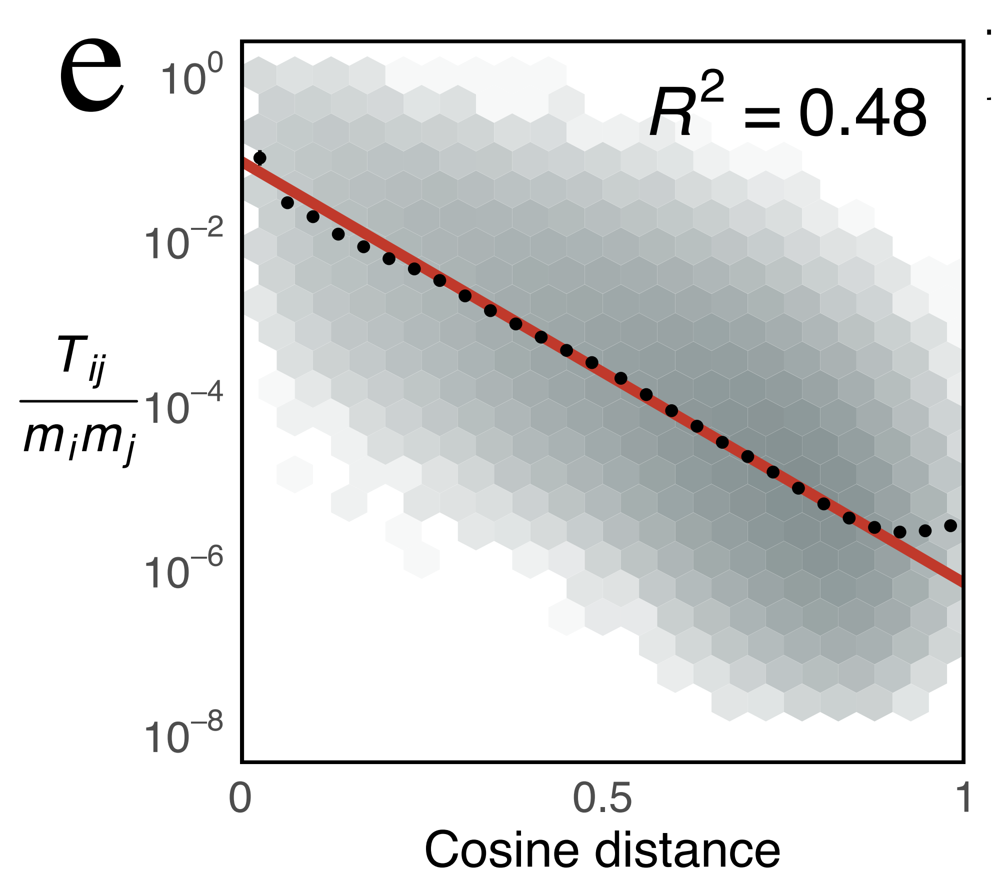
Relationship between word2vec and the Gravity Law

- Distance in the embedding space maps well to the gravity model
- Both correlation and prediction
- Word2vec finds a representation to predict adjacent words, but the gravity law emerges
- Deep connection between them?
- Preliminary work, but ideas welcome!

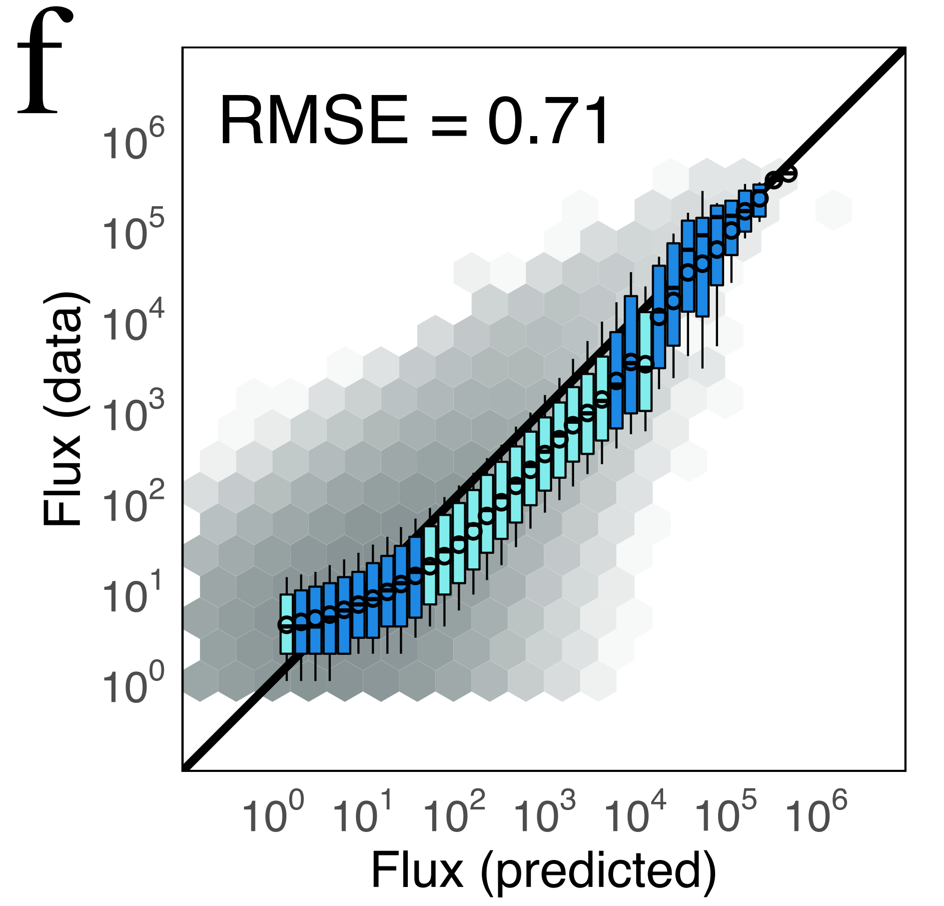
$$T_{ij} = C m_i m_j f(r_{ij})$$

$$\frac{T_{ij}}{m_i m_j} = f(r_{ij})$$

Correlation



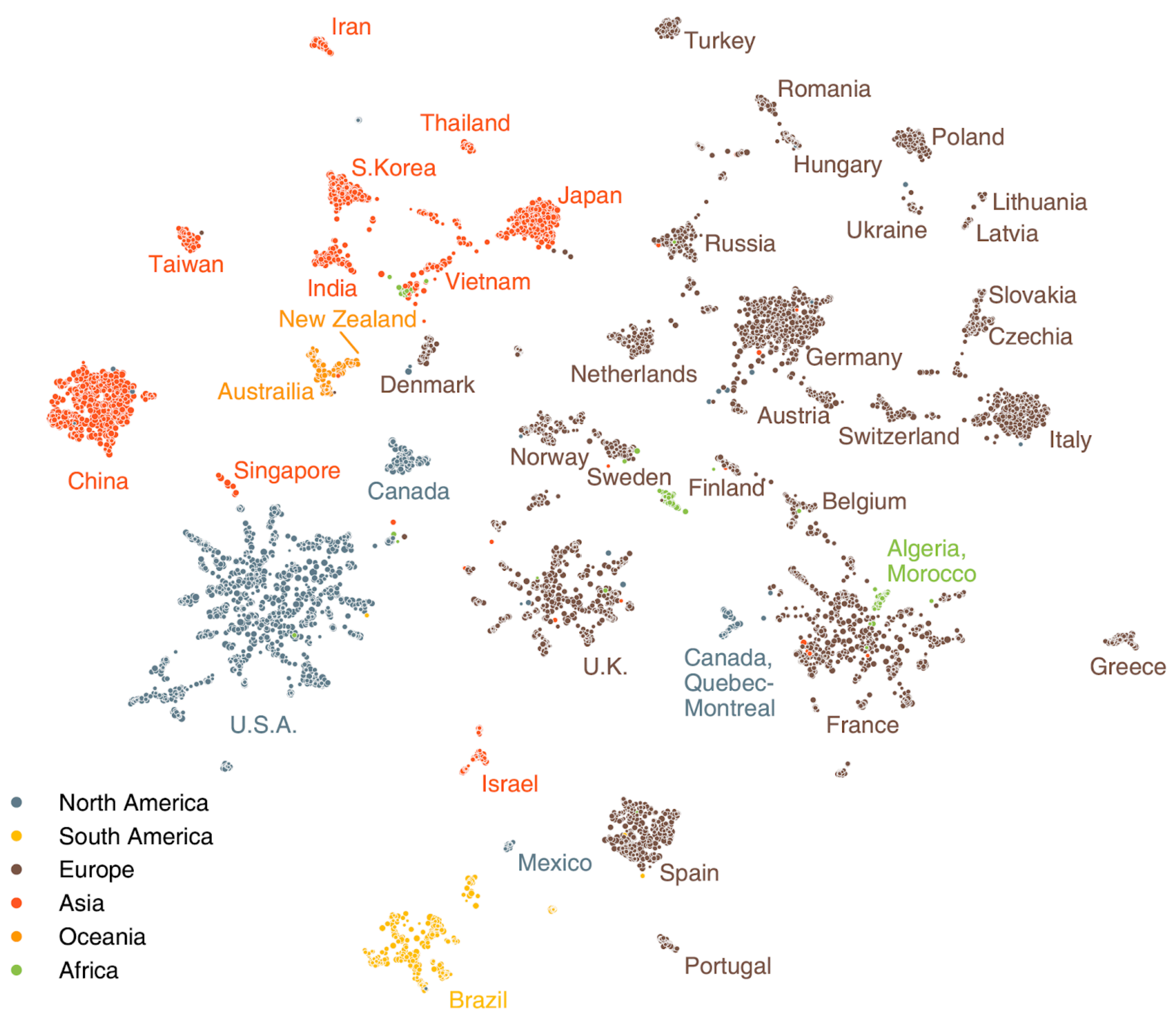
Predicted vs. Actual

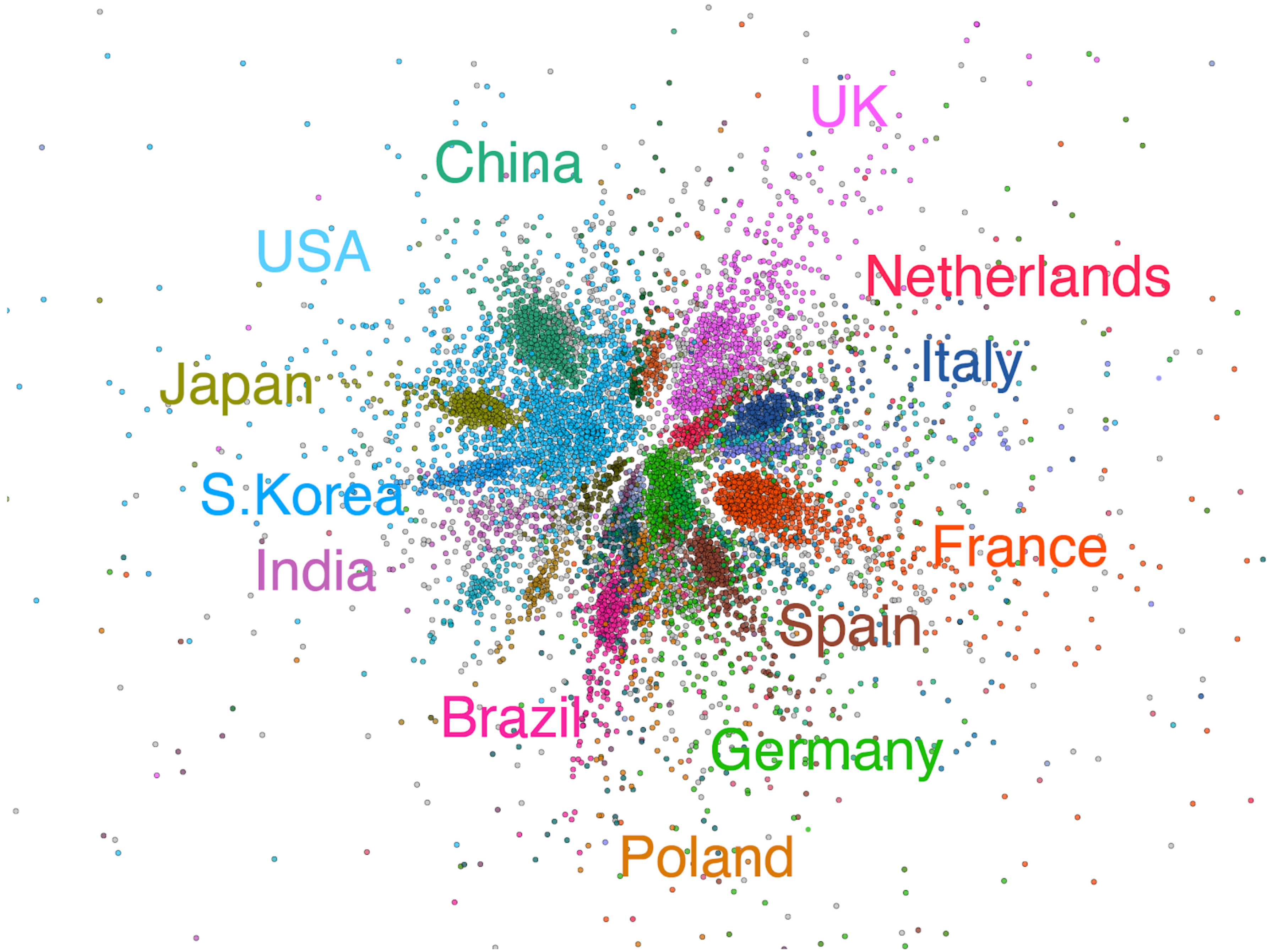


Murray, D., Yoon, J., Kojaku, S., Costas, R., Jung, W.-S., Milojević, S., & Ahn, Y.-Y. (2020). Unsupervised embedding of trajectories captures the latent structure of mobility. *(Under review, Nature)*

Why not networks?

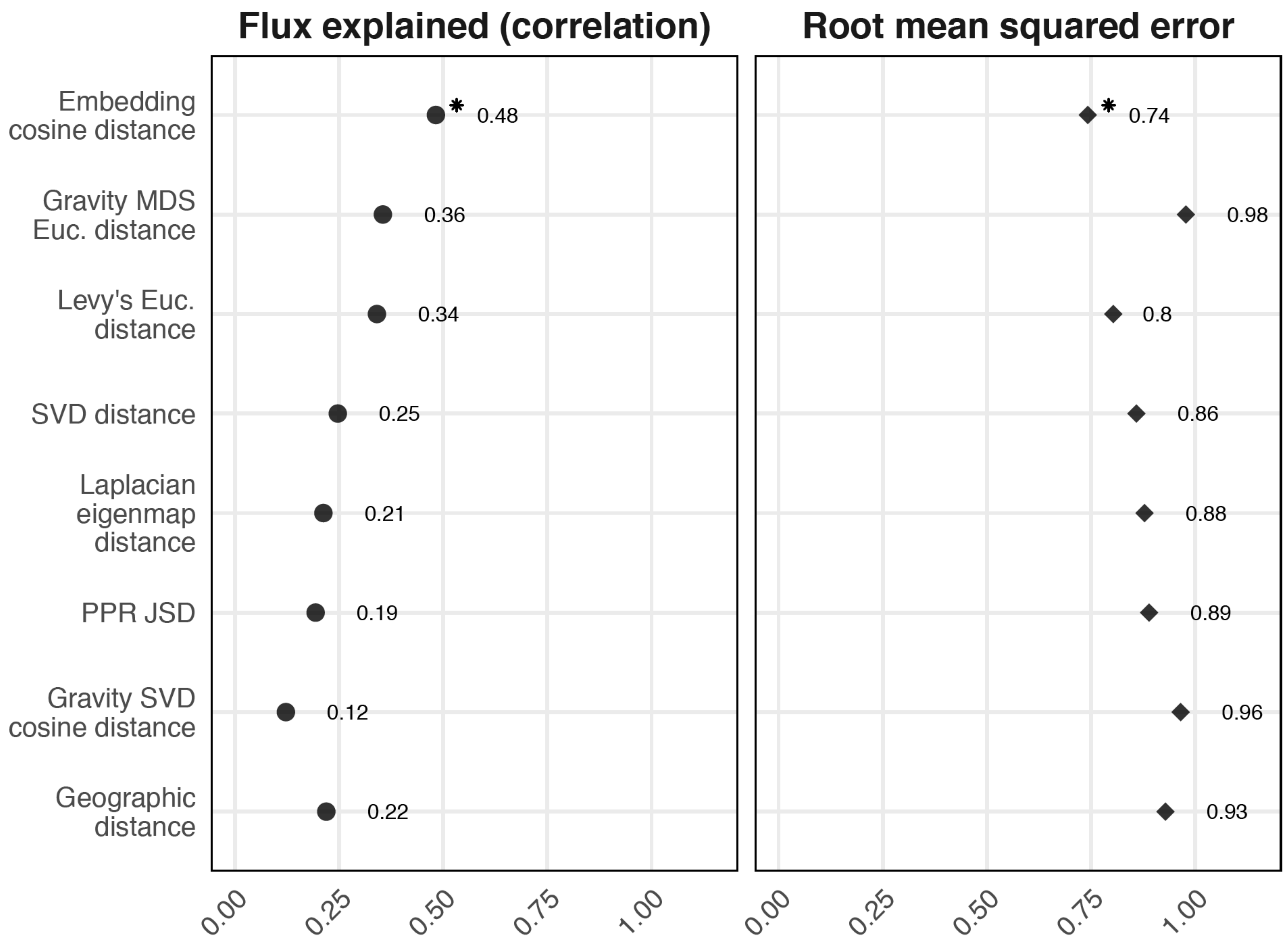
- Structure obfuscated in visualizations
- Poor performance (reported in the paper)
- Many techniques, tuning required
- Missing edges
- Embeddings provide access to many interesting techniques





These embeddings better capture mobility than alternatives

The representation can be used to better understand the landscape of global mobility

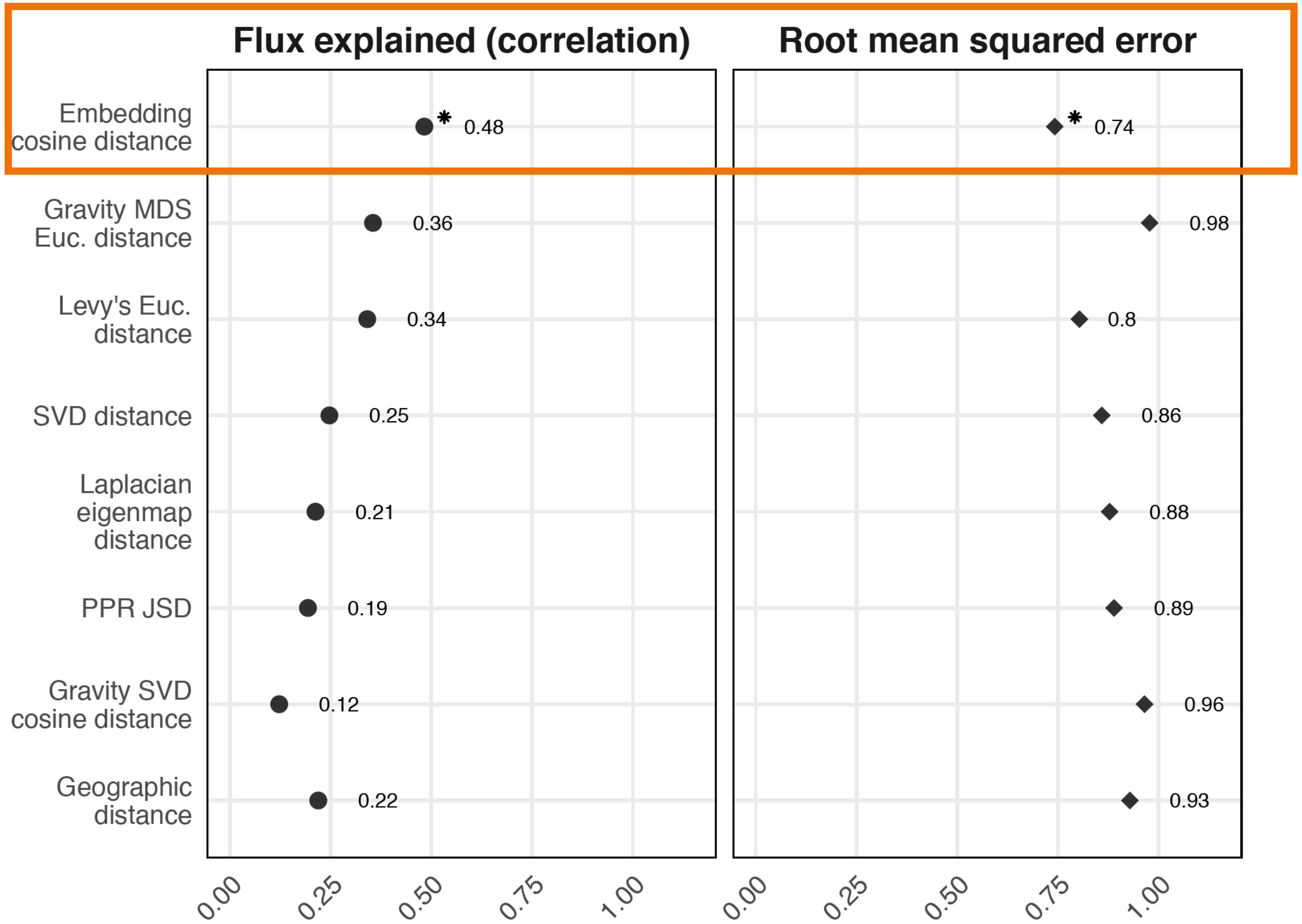


These embeddings better capture mobility than alternatives

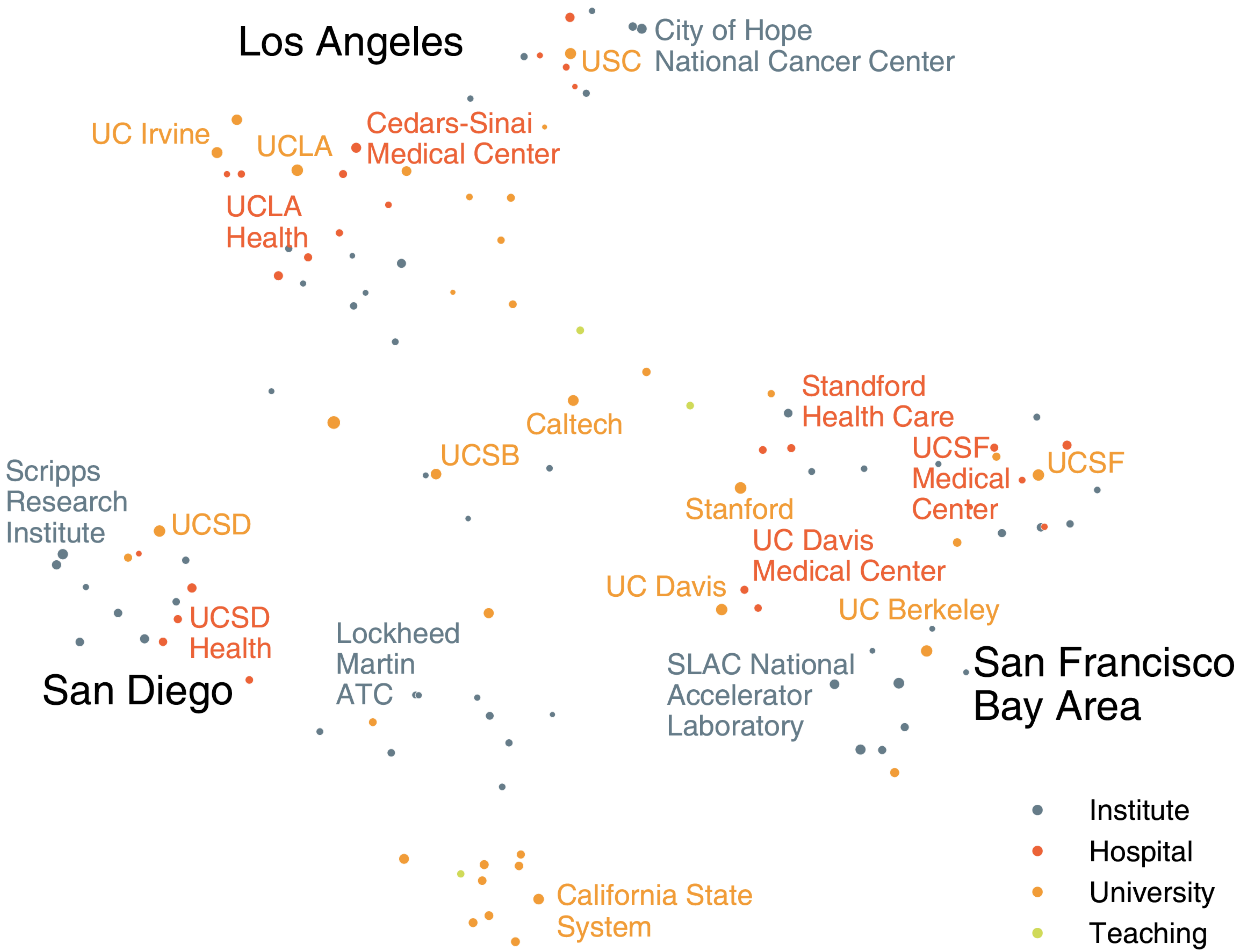
The representation can be used to better understand the landscape of global mobility

Better explains and better predicts mobility than baselines

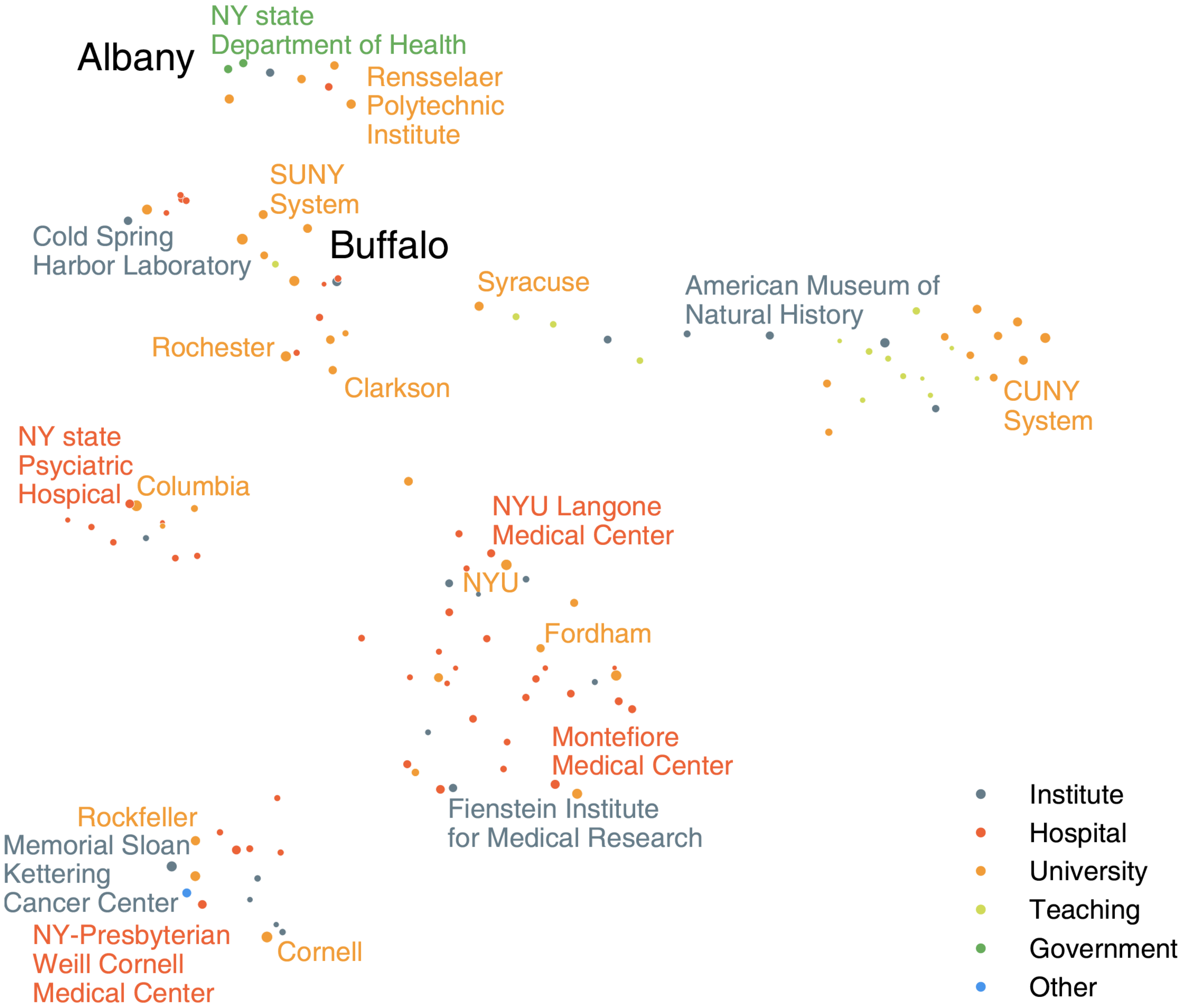
High validity!



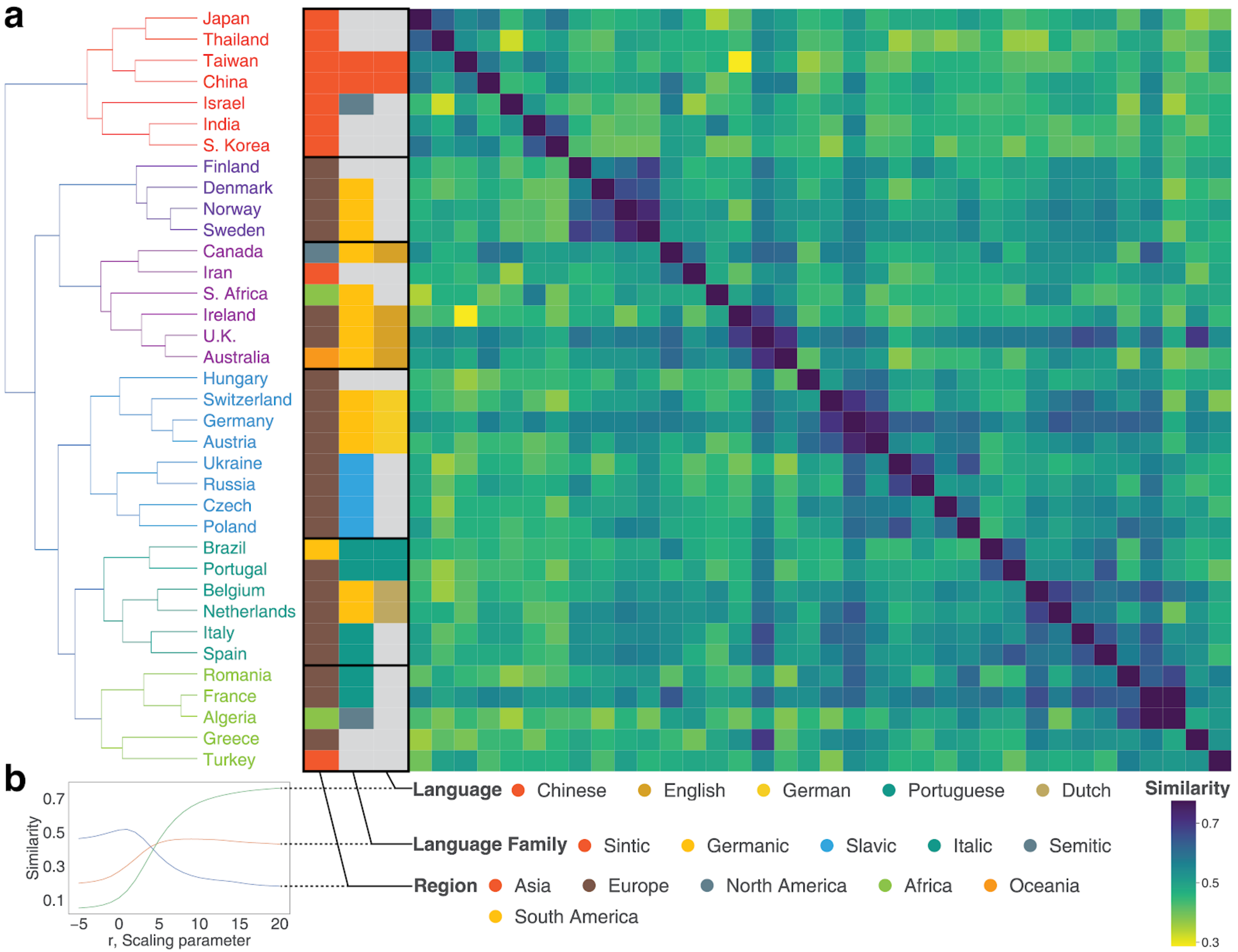
California Structure



New York Structure

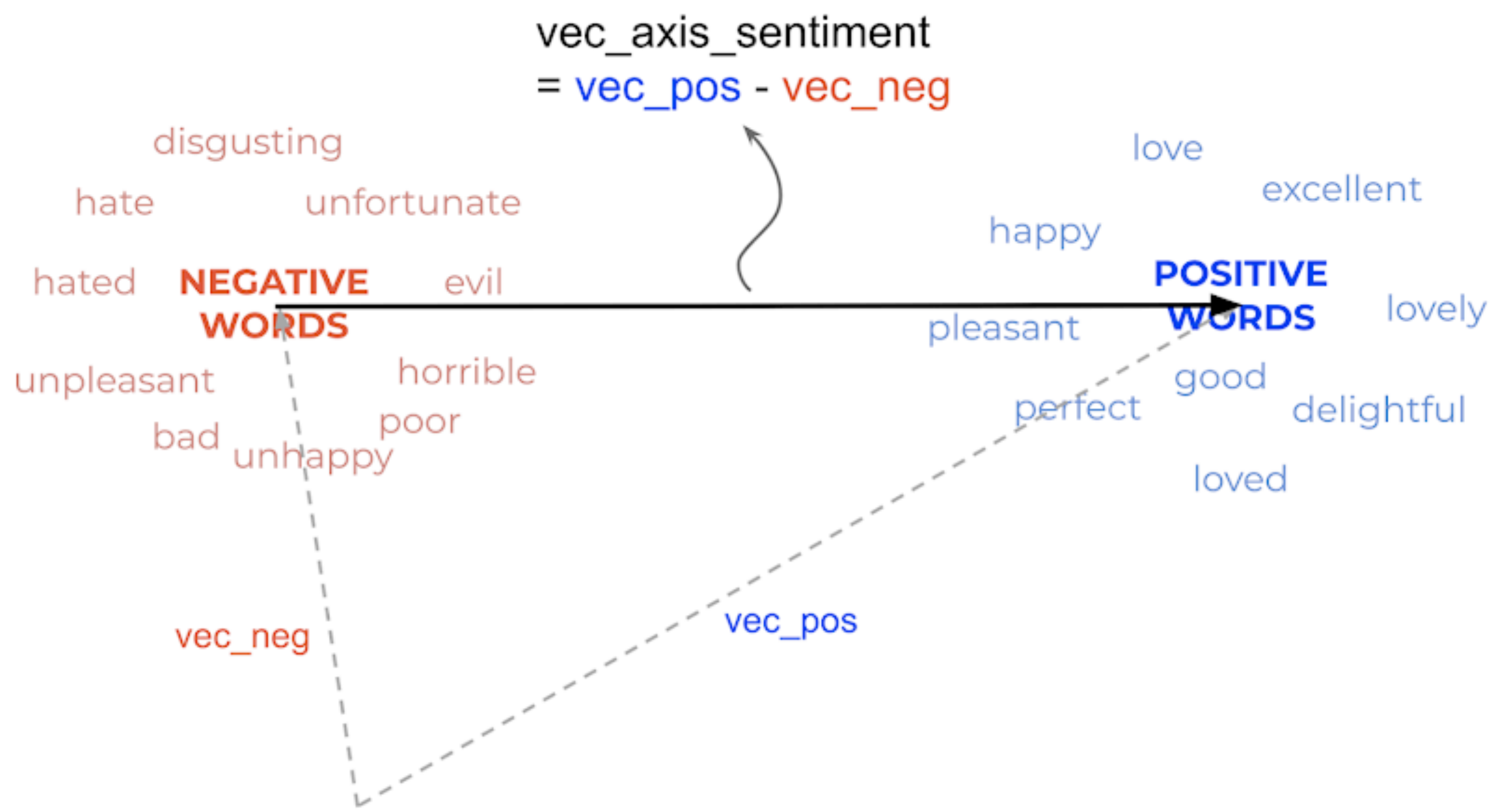


Geography, then language, structure the vector space

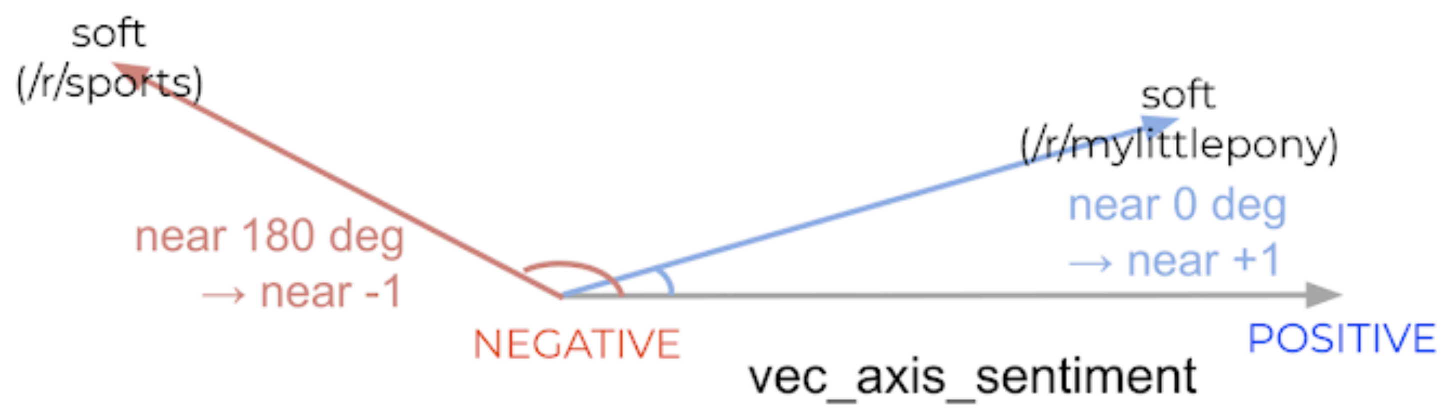


Leverage the semantic properties of the embedding

Step 1: Define a semantic axis

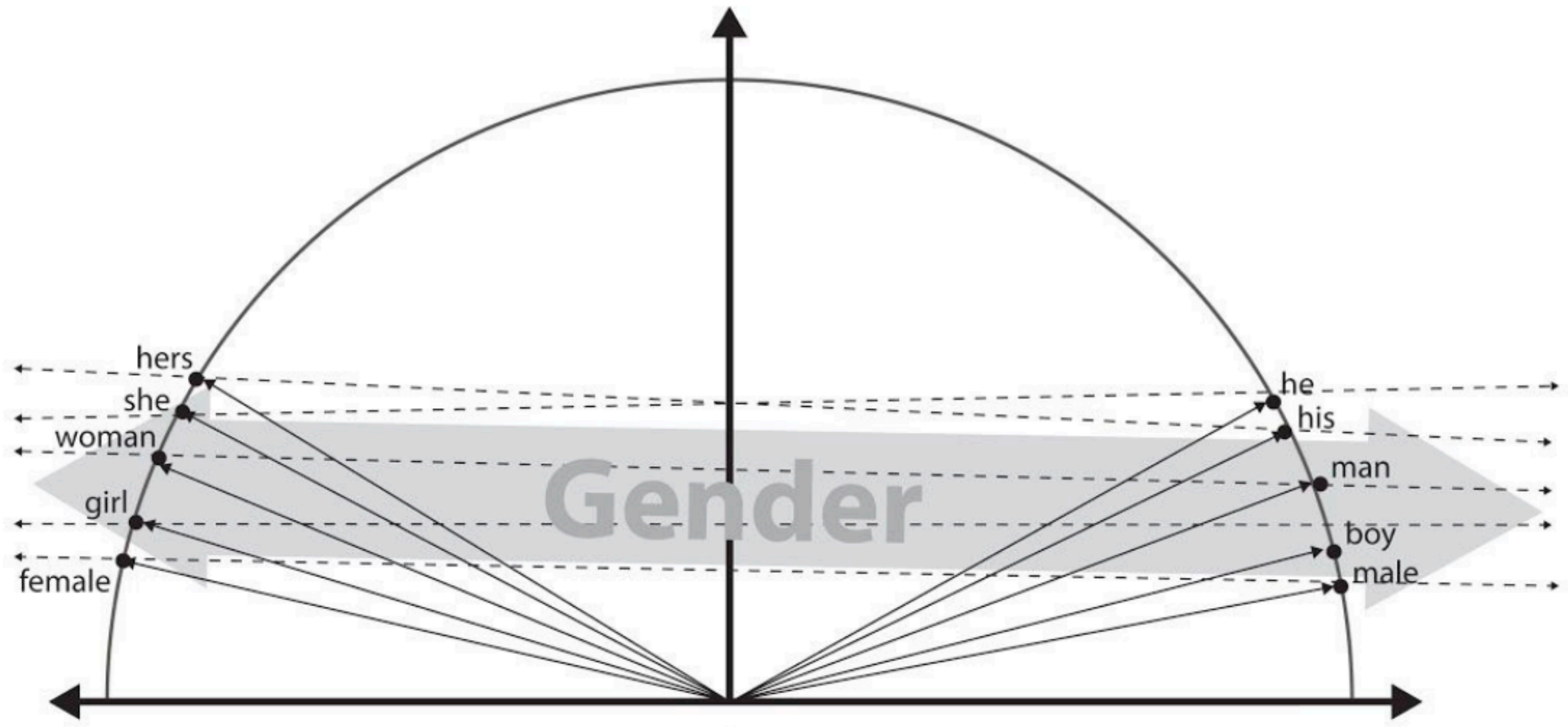


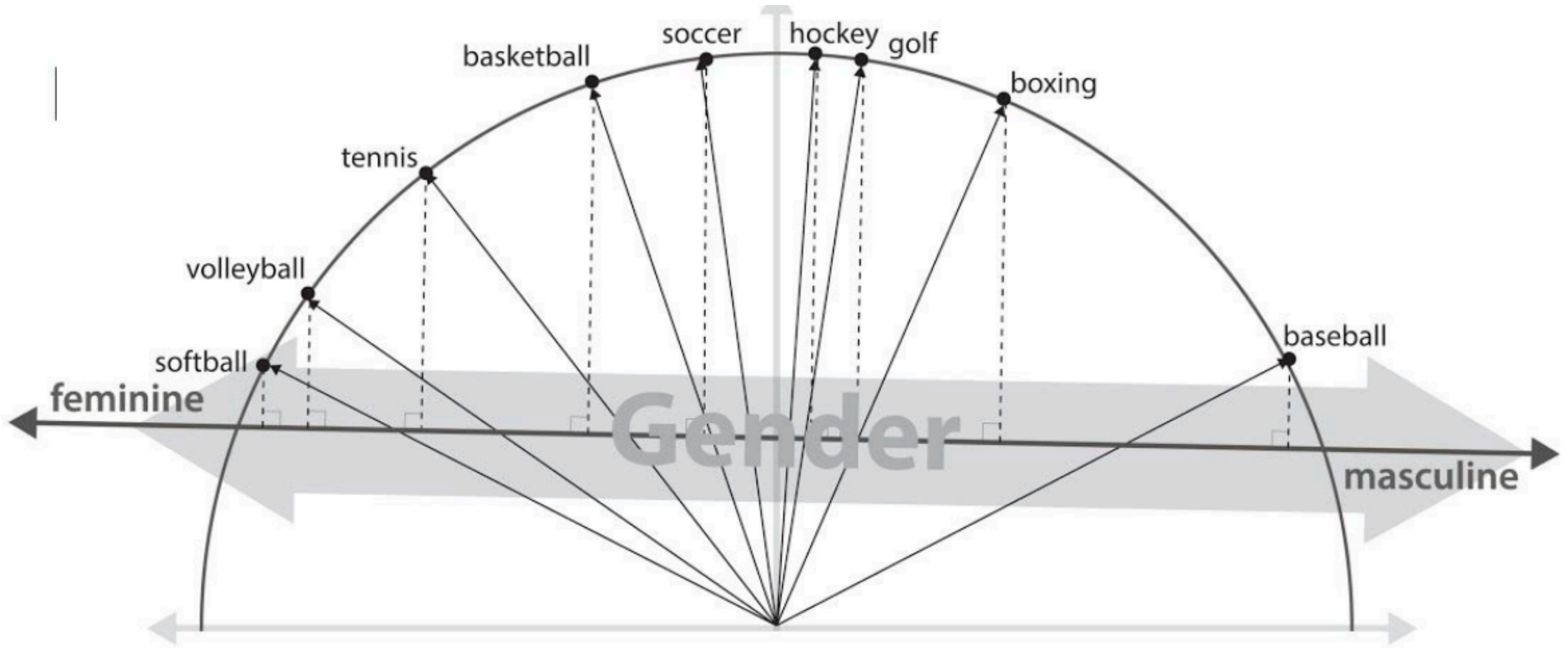
Step 2. Locate words on the semantic axis



Compute a cosine similarity between a word vector and axis vector

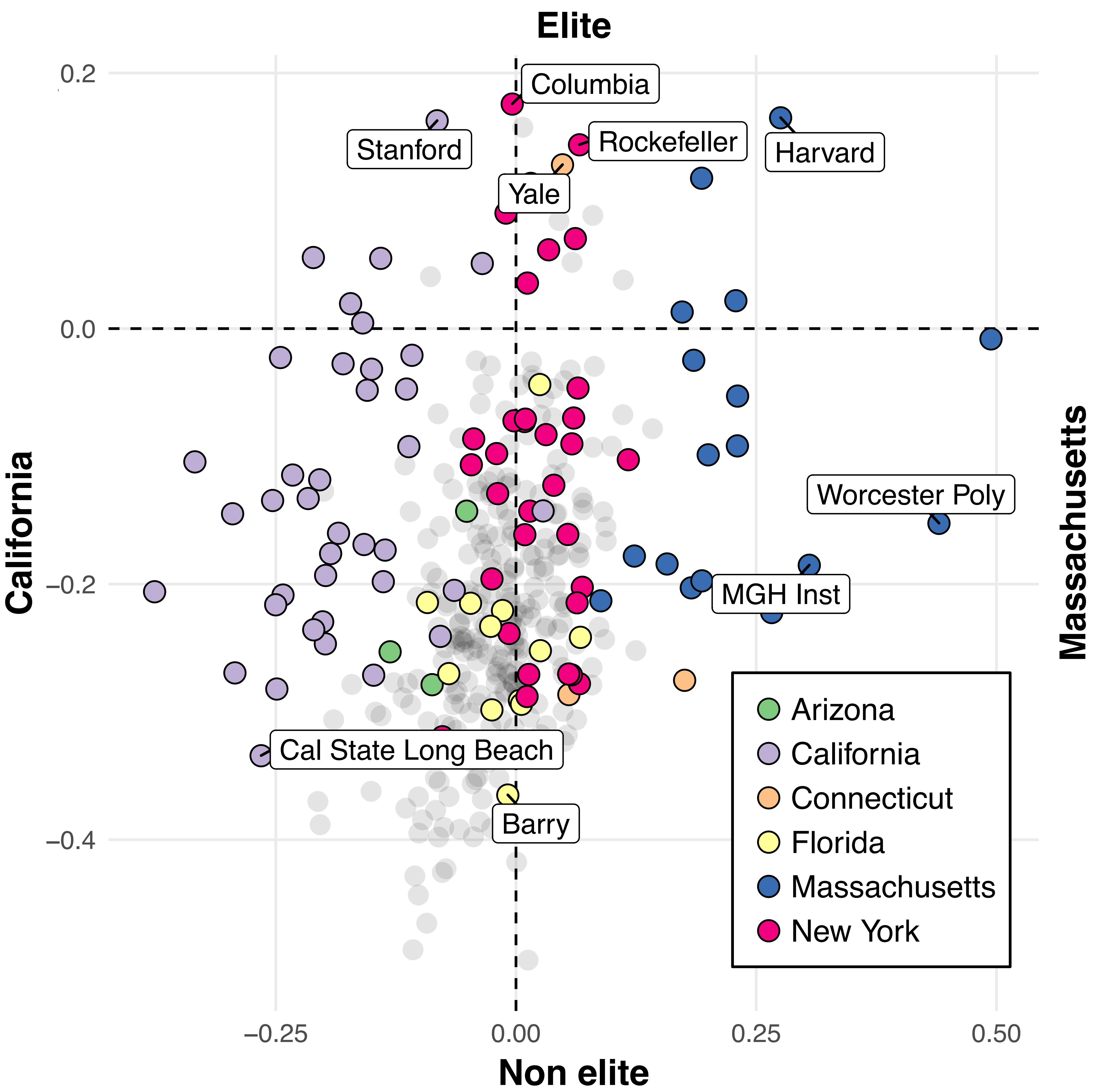
An, J., Kwak, H., & Ahn, Y.-Y. (2018). SemAxis: A Lightweight Framework to Characterize Domain-Specific Word Semantics Beyond Sentiment. *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 2450–2461.



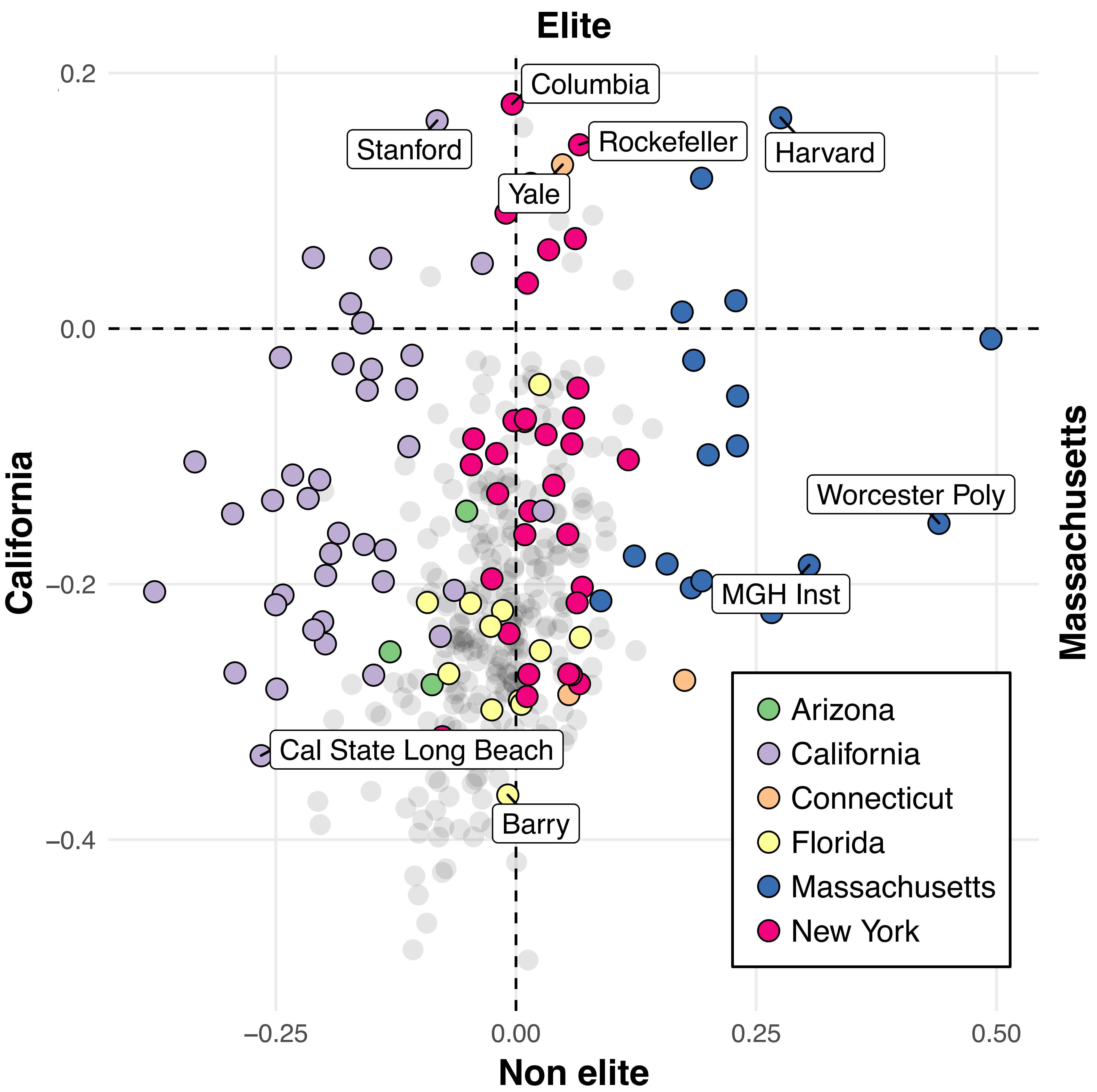


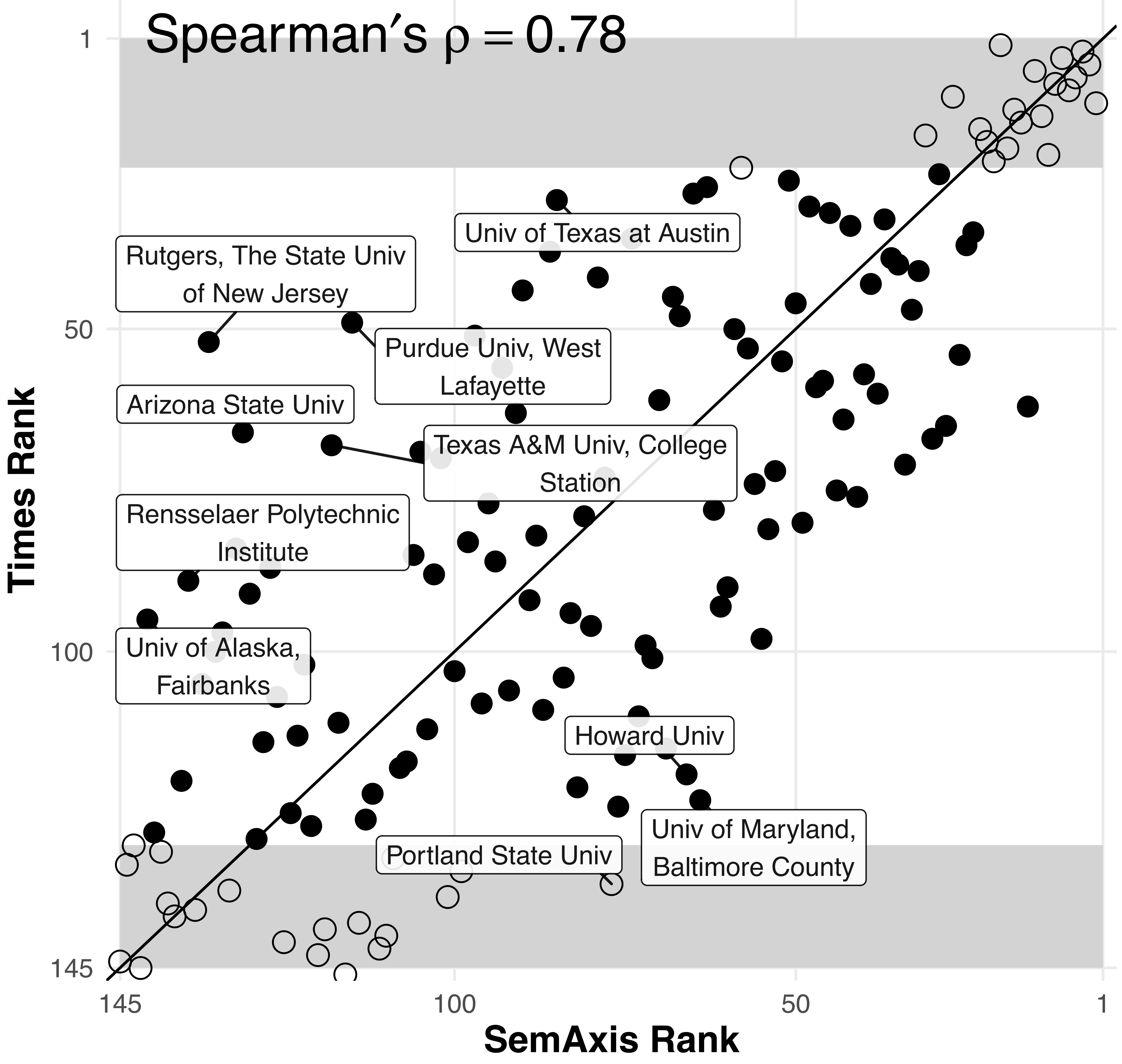
Kozlowski, A. C., Taddy, M., & Evans, J. A. (2019). The Geometry of Culture: Analyzing the Meanings of Class through Word Embeddings. *American Sociological Review*, *84*(5), 905–949.

SemAxis using Geography and Prestige



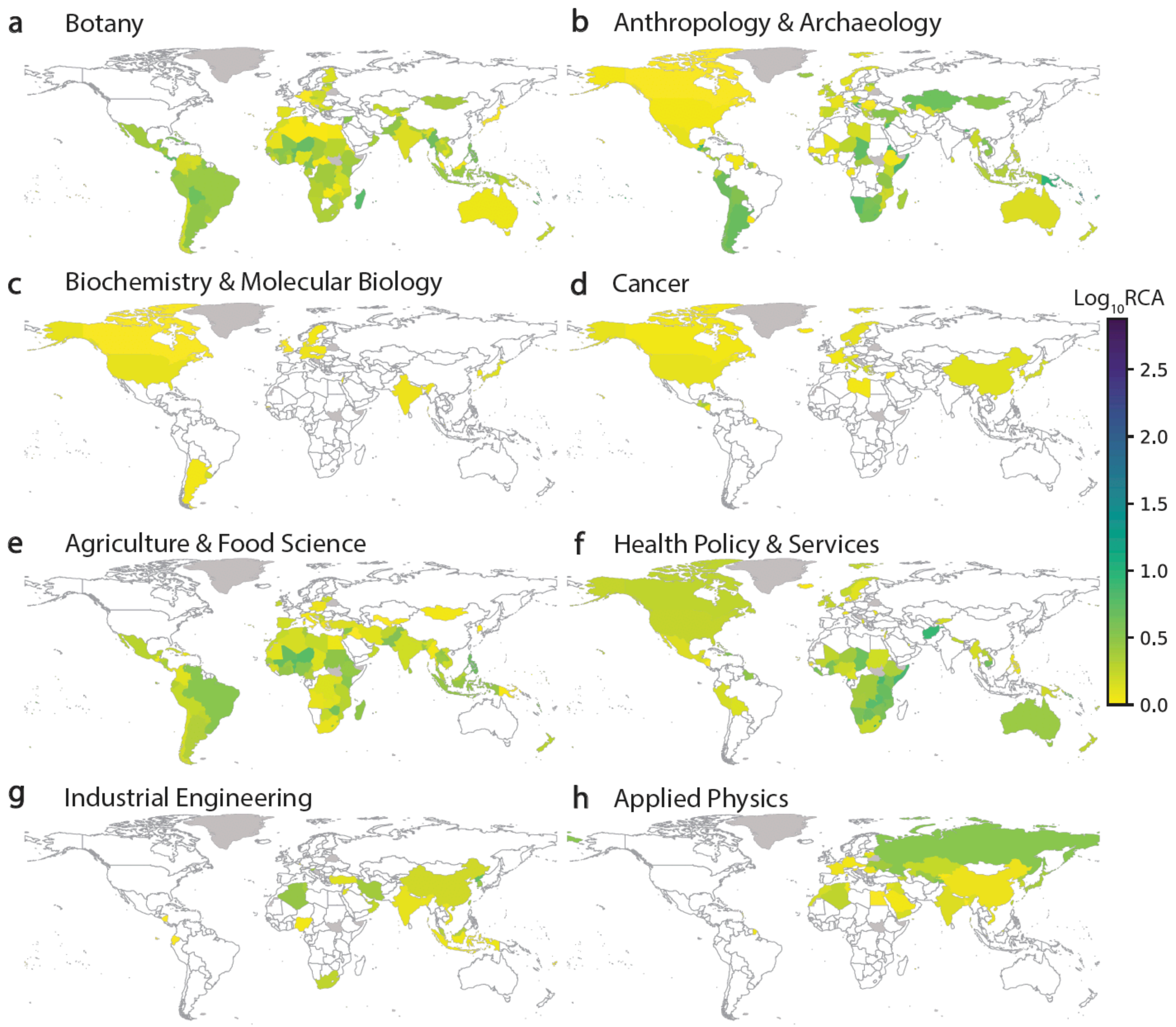
SemAxis reconstructs university prestige





Countries do different kinds of research

Miao, L., **Murray, D.**, Jung, W., Larivière, V., Sugimoto, C. R., Ahn, Y., The scientific development of nations. (In Preparation)



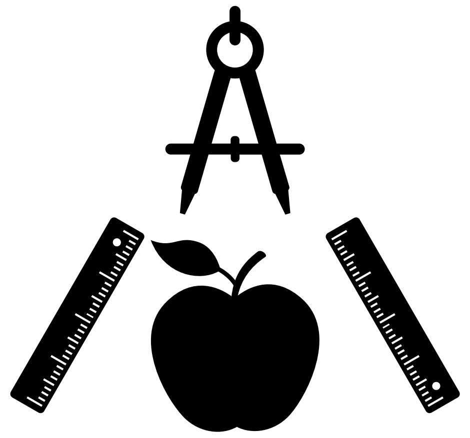
Summary of main findings

Peer review at *eLife*

Disparity and bias in peer review outcomes at eLife

Homophily matters in outcomes!

Homophily matters in reviewer team formation!



Student-teacher evaluations

Demographic factors are related to teacher ratings

Gender, ethnicity, attractiveness, more

Disagreement in science

Disagreement roughly follows the hierarchy of sciences

But heterogeneous based on the particular characteristics of fields

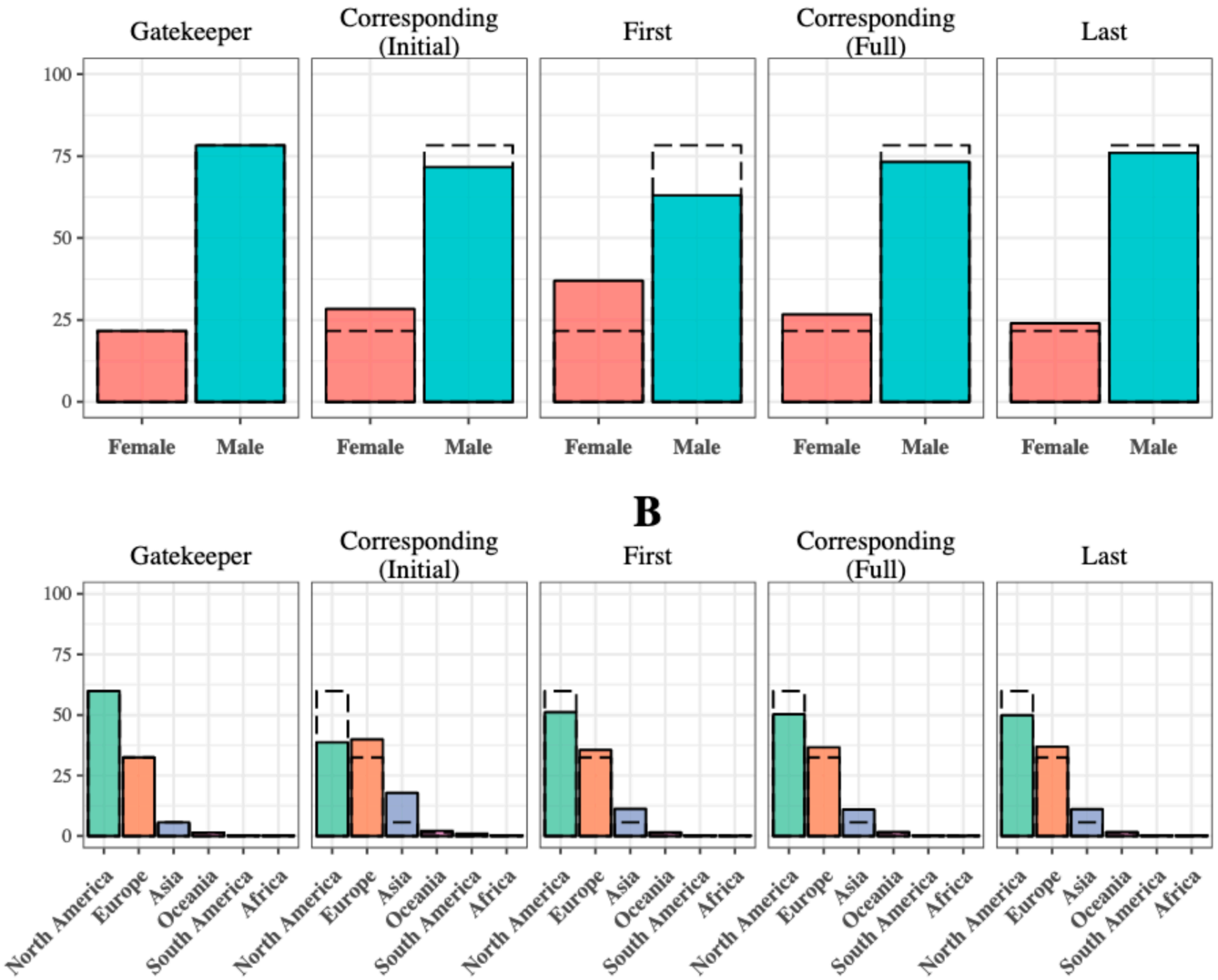
Global scientific mobility

Mobility is constrained by geography, language, history, prestige, etc.

Their effects vary at different scales of analysis

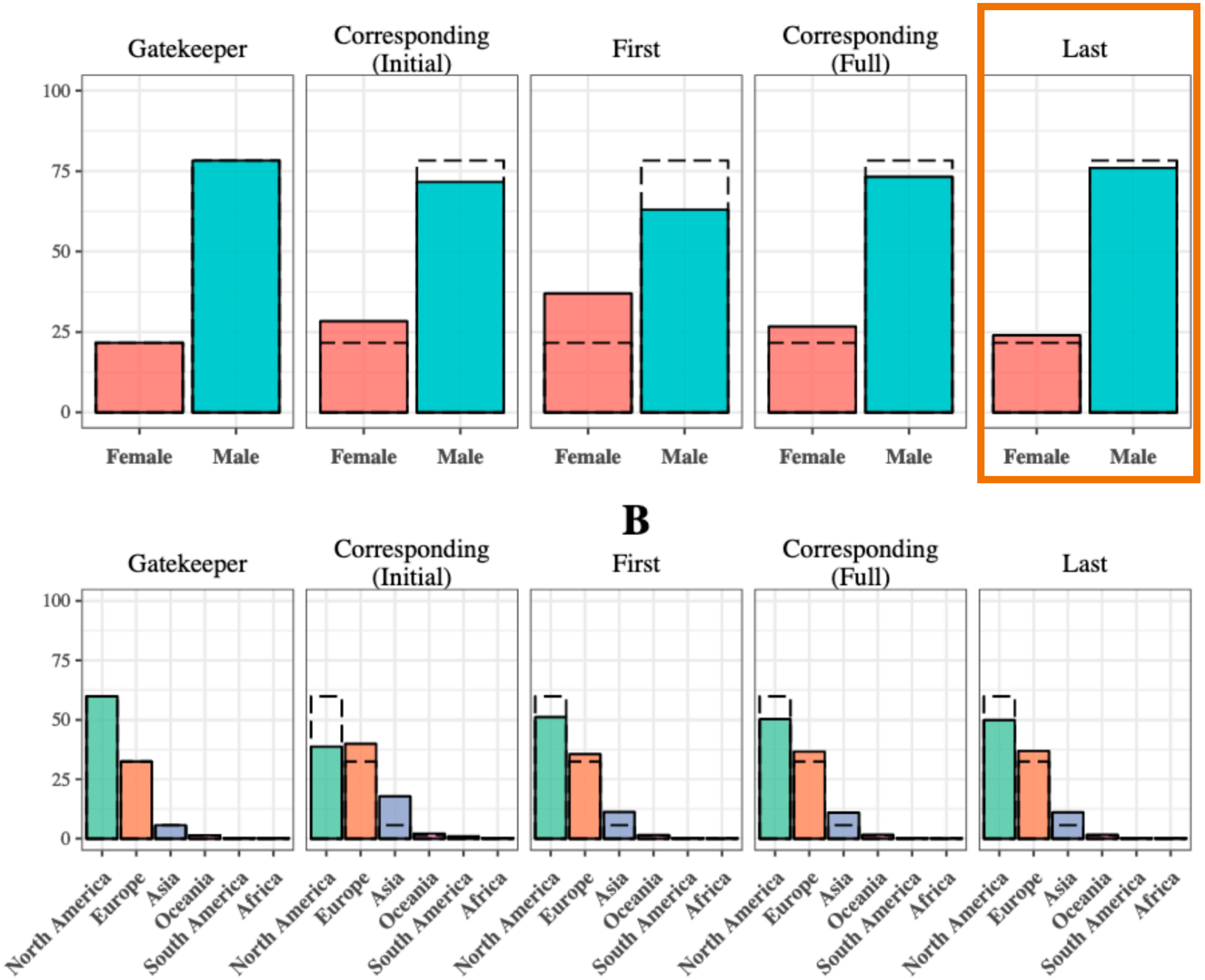
Representation of gatekeepers

Gender roughly-representative of senior author, but less so for country



Representation of gatekeepers

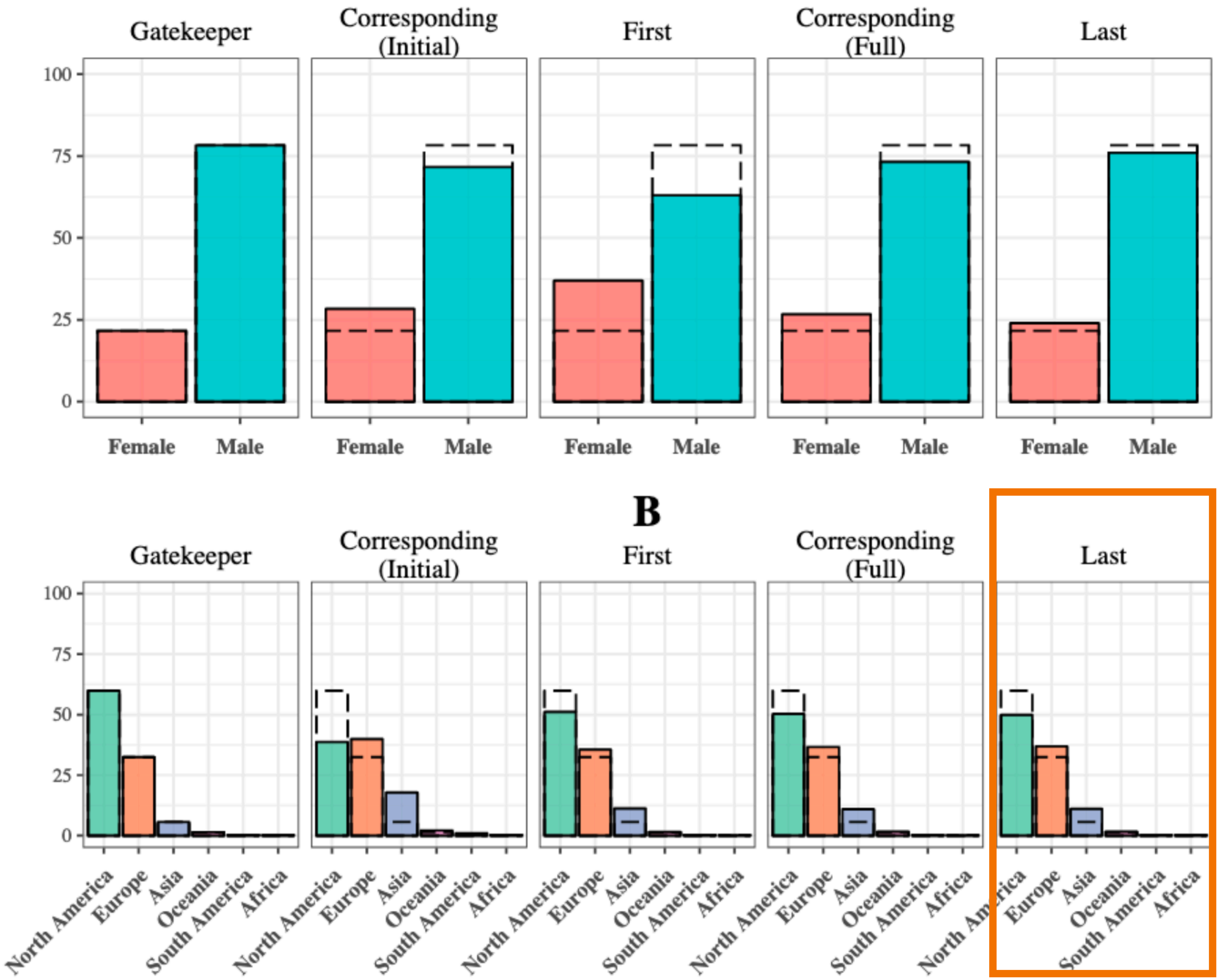
Gender roughly-representative of senior author, but less so for country



Gender representation of gatekeepers roughly similar to last authors submitting to *eLife*

Representation of gatekeepers

Gender roughly-representative of senior author, but less so for country

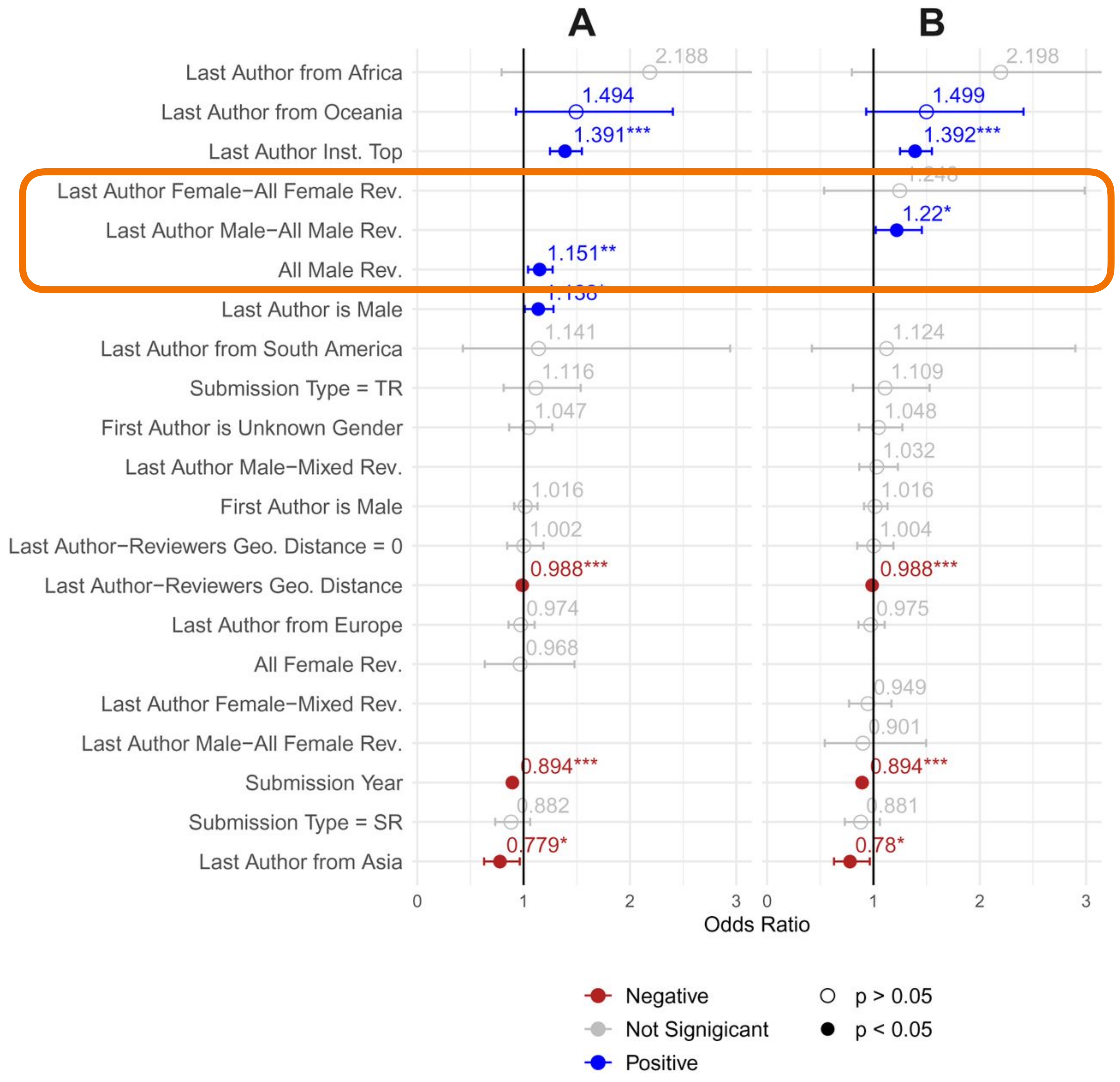


Gender representation of gatekeepers roughly similar to last authors submitting to *eLife*

But gatekeepers are much more North American than authorship

Even when controlling for other factors…

- Year of submission
- Submission type
- Author demographics
- Reviewer composition
- Distance between authors + reviewer
- Combination term
- Male last author + Male reviewers > Female last authors + all male reviewers). No such effect for mixed reviewer teams. **Reviewers matter!**



Neural embedding

"The quick brown fox …"

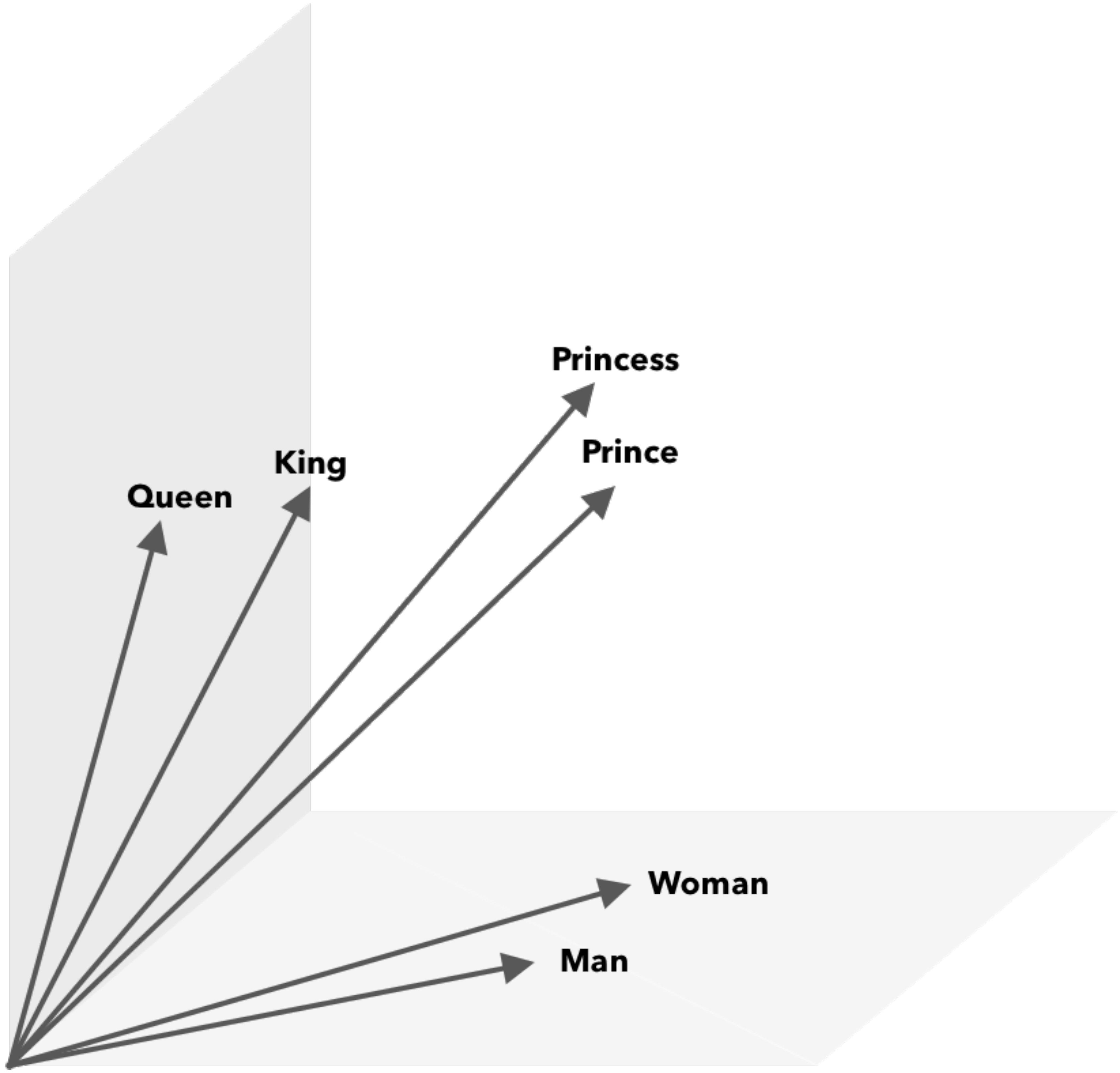
Word2vec
(or other word embedding methods)

"quick": [0.51, 0.12, 0.69, …]
"brown": [0.11, 0.92, 0.29, …]

Dense vectors

The geometry of the vector space encodes semantic relationships

Using cosine distance



Murray, D., Yoon, J., Kojaku, S., Costas, R., Jung, W.-S., Milojević, S., & Ahn, Y.-Y. (2020). Unsupervised embedding of trajectories captures the latent structure of mobility. *(In revision)*

They don’t have to be “real” words or sentences

Any "sentences" — a sequence of elements from a finite vocabulary — work! We can use **trajectories of scientists** as **sentences** and **organizations** as **words**.