Preprints, Inequality, and Trust: Designing Scientific Communication for Humans and AI

TMI, Faculty of Engineering, the University of Tokyo
Chiaki Miura

UTokyo

Chiaki Miura

Study of
Peer Review /
Preprint

Study of Preprints and peer review

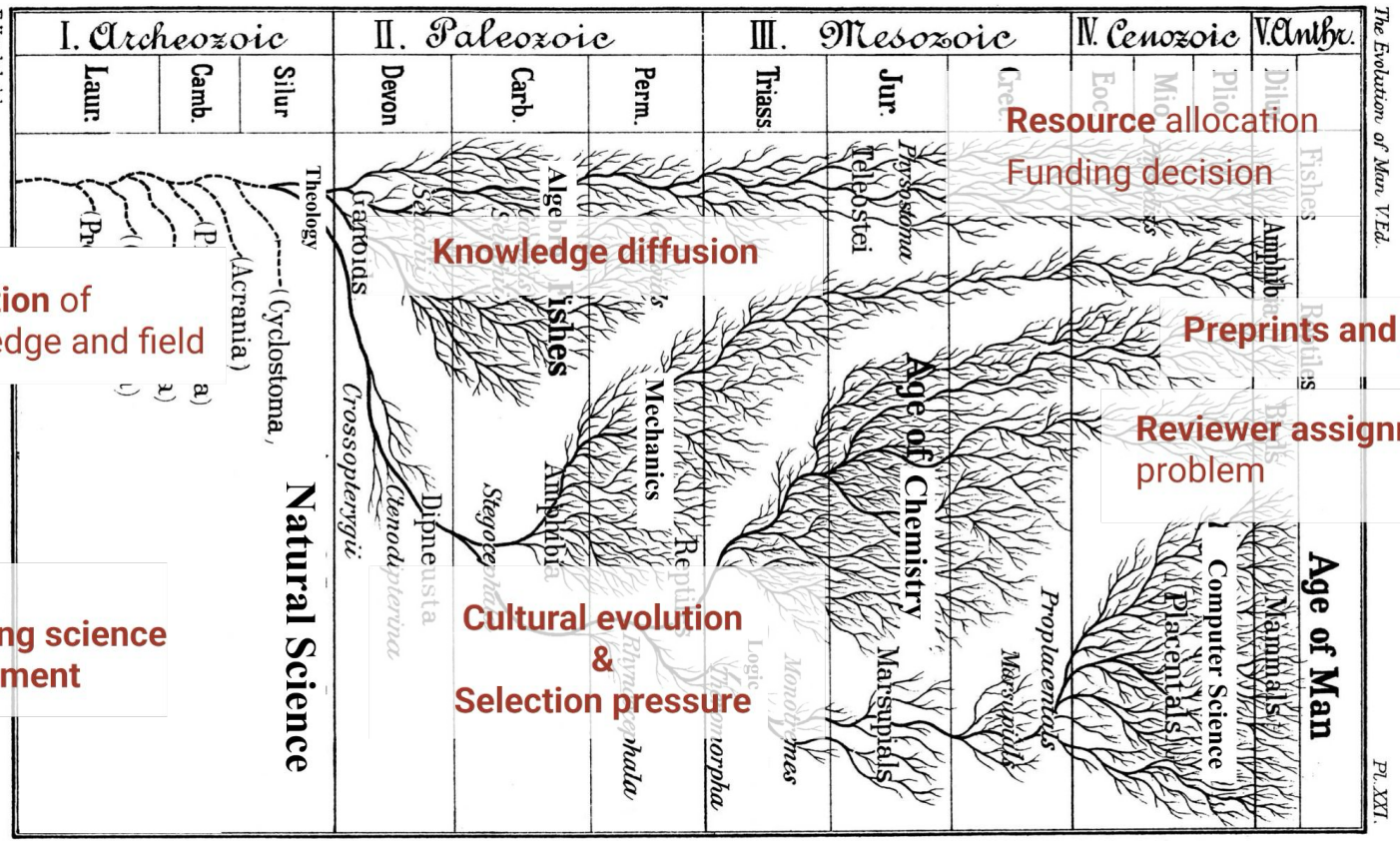
- **Preprint**: the emerging scholarly channel
  - it made the “letters” between authors **citable**
  - the best source to study **peer review**

Why does it matter?

How future science navigate itself through epistemic landscape?

- Tension between **openness** and **reliability** in preprints
- 900 papers posted by single author per month / LLM recommendation rely on OA
- Peer review is **selection pressure** on scientific claims
- Designing reward signals in the forthcoming age of AI-generated knowledge

Peer Review



Resource allocation
Funding decision

Knowledge diffusion

Preprints and journals,

Reviewer assignment problem

Extinction of knowledge and field

Cultural evolution & Selection pressure

Measuring science advancement

Peer Review

Journal

Preprint

Citation to Preprint: an age-old yet increasingly relevant phenomenon

- The first year of 21st c, Dec. An article on Nature
- Citing preprint without "this is not peer reviewed"
- [illegible] for the matter

news and views feature

Quantum theory's last challenge

Giovanni Amelino-Camelia

Quantum theory is 100 years old and still going strong. Combining general relativity with quantum mechanics is the last hurdle to be overcome [illegible]

This year we celebrate 100 years of quantum theory, and in particular the anniversary of an announcement made by Max Planck at a meeting of the German on 14 December 1900. sted in the nature of radia- ot [illegible] ribed [illegible] dence, but that required a pt: energy is not emitted or uously, but in discrete uanta. At the time, Planck he profound consequences radually physicists realized uantum concepts to under-

How frequent is the citation to preprints?

Is it occurring outside of preprint friendly field?

e-mail: amelino@roma1.infn.it

1. [illegible]
2. Stachel, J. In *Black Holes, Gravitational Radiation and the Universe: Essays in Honour of C. V. Vishveshwara* (eds Iyer, B. R. & Bhawal, B.) 525–534 (Kluwer Academic, Dordrecht, 1999).
3. Rovelli, C. *J. Math. Phys.* **41,** 3776–3800 (2000).
4. Amelino-Camelia, G. http://xxx.lanl.gov/abs/gr-qc/9910089
5. 't Hooft, G. *Quantum Grav.* **16,** 3263–3279 (1999).
6. Ashtekar, A., Rovelli, C. & Smolin, L. *Phys. Rev. Lett.* **69,** 237–240 (1992).
7. Green, M. B., Schwarz, J. H. & Witten, E. *Superstring Theory* (Cambridge Univ. Press, Cambridge, 1987).
8. Polchinski, J. *String theory* (Cambridge Univ. Press, Cambridge,

Citation to Preprint: an age-old yet increasingly relevant phenomenon

- Citation to preprint exist, and likely increasing
- Confounding factors: increasing size of preprints, quality of preprints, editorial policy climate, dominance of preprint friendly field

Fixed Journal Subset

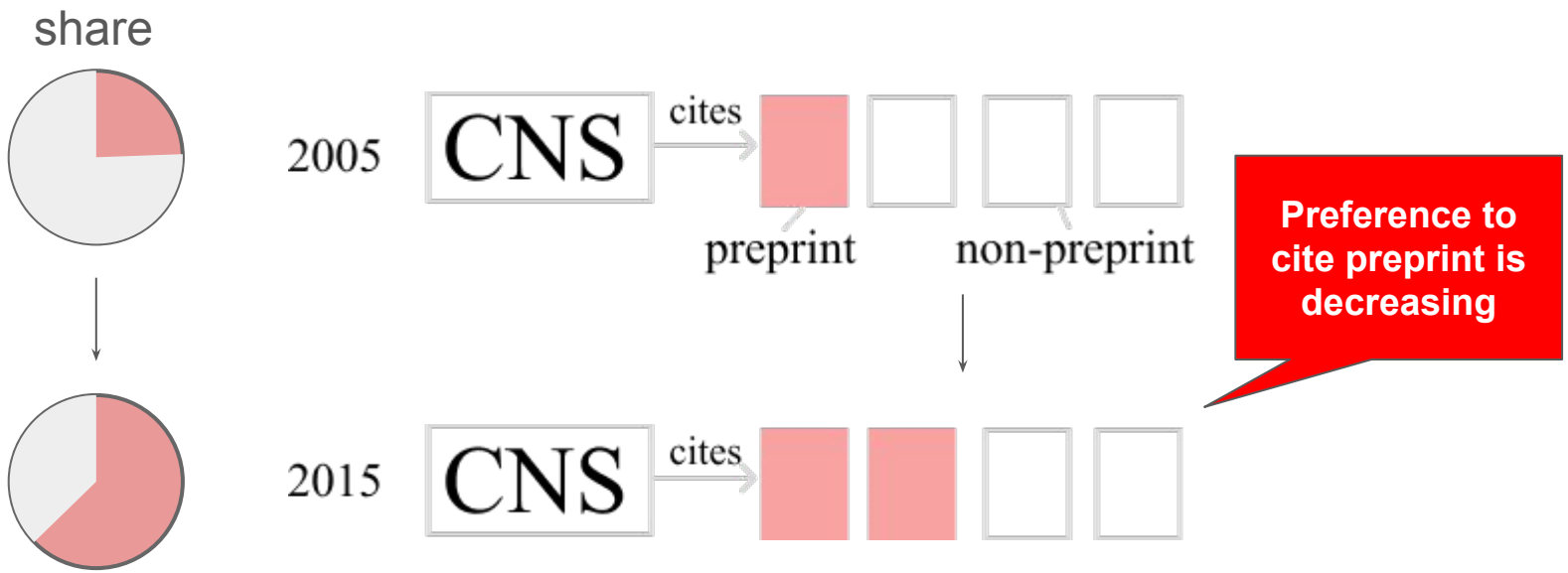
Inhouse consistent editorial standard (Nature, Science, Cell)
Journal inception < 2000

Null Model

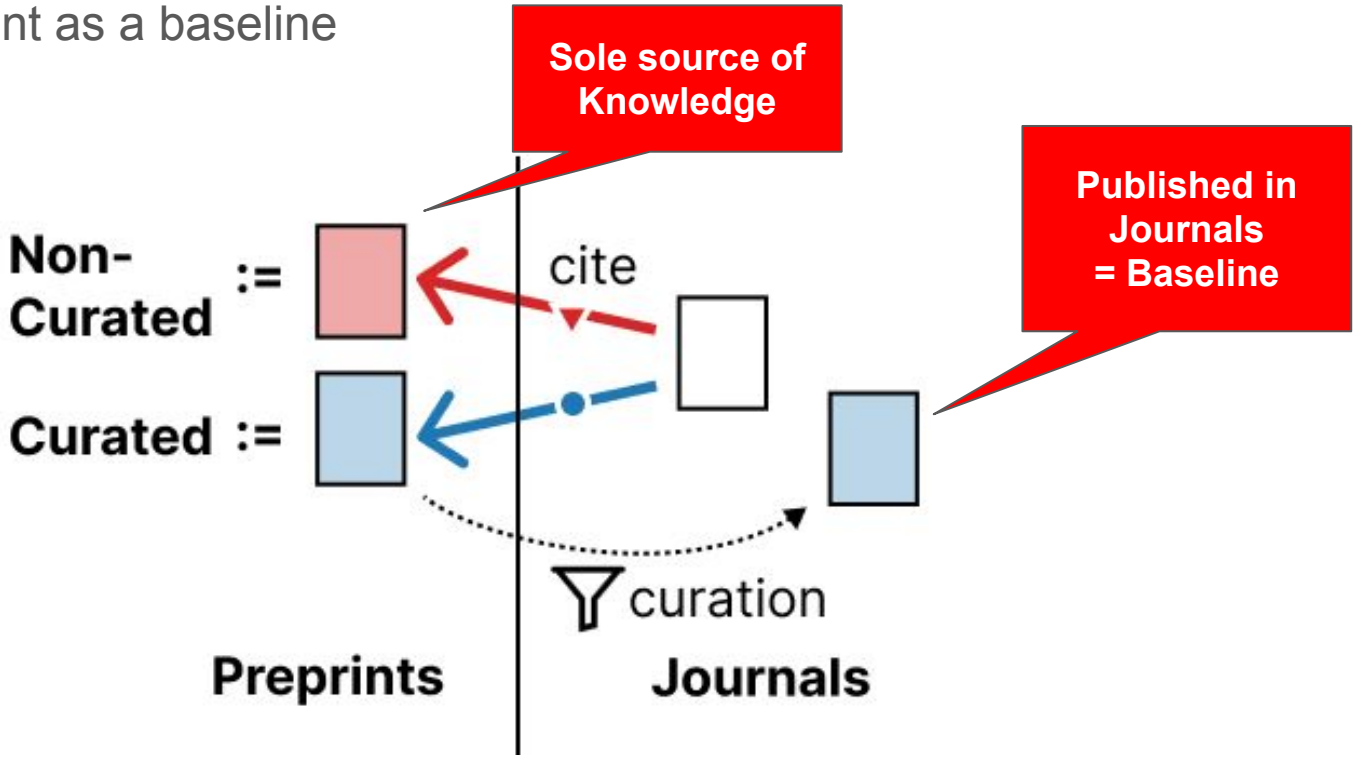
1. article share citation likelihood ratio (control for **quantity**)
2. comparison by curation (control for **quality**)

article share citation likelihood ratio

- preprints' share in all citable items vs aggregated count of citation to preprint

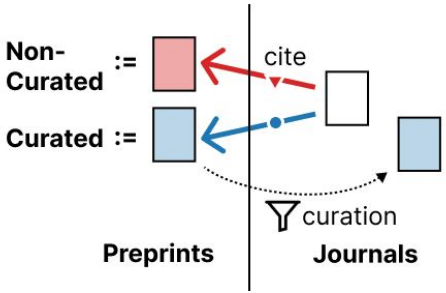


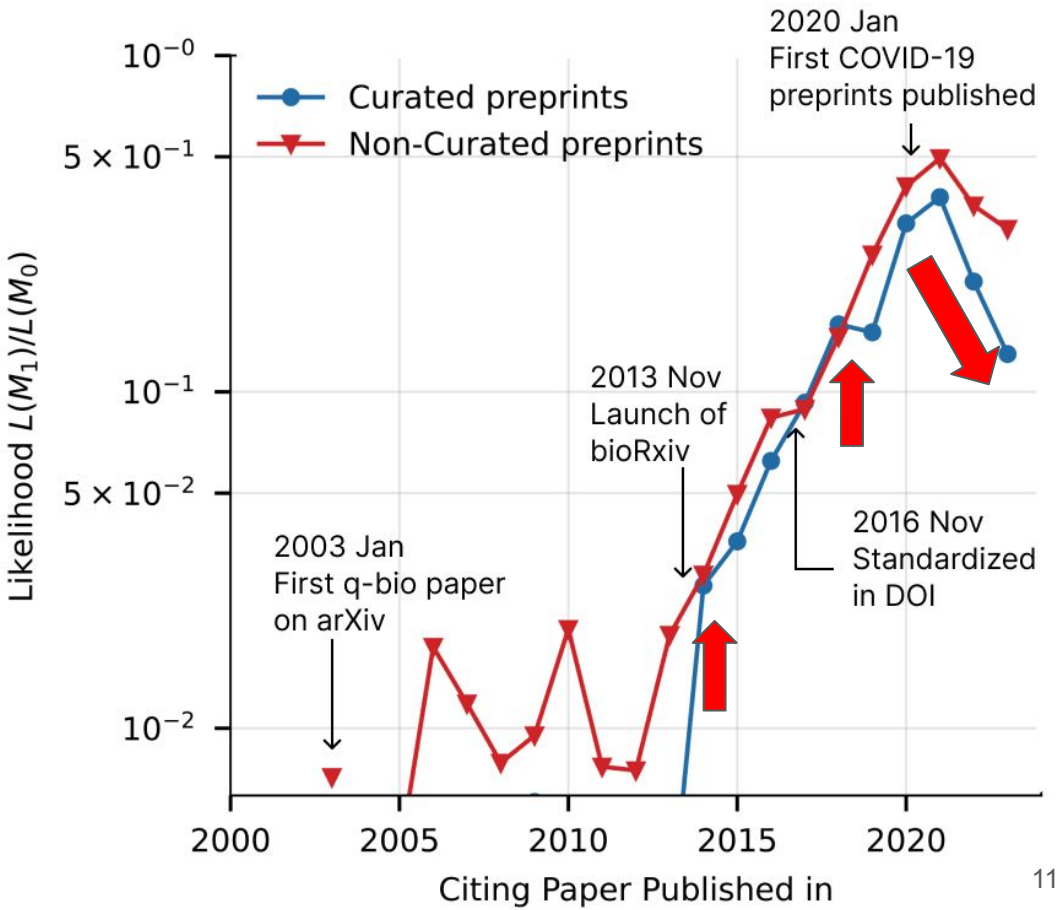
- Curated preprint as a baseline



Less than expected by chance yet increasingly preferred

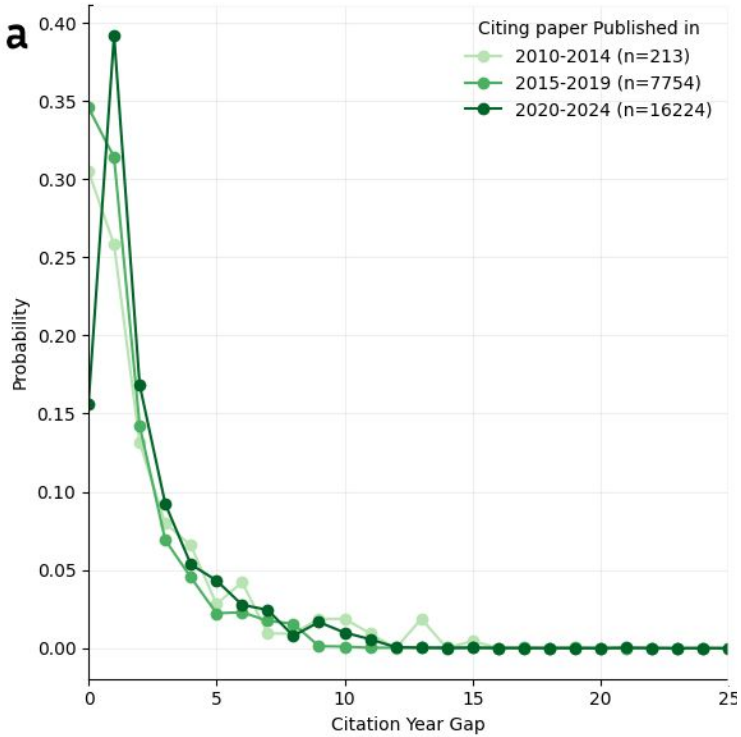
- Non-curated Preprints are increasingly cited by journals
- Preference to cite preprints…
  - is strengthened by institutionalization
  - began to erode after COVID



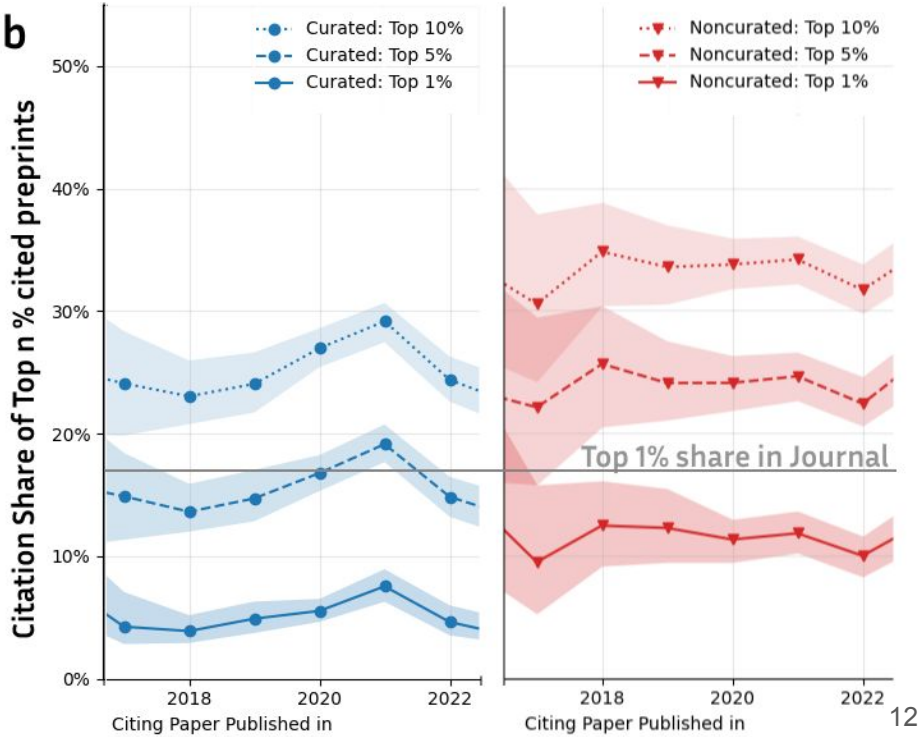


Robust to other confounding factors

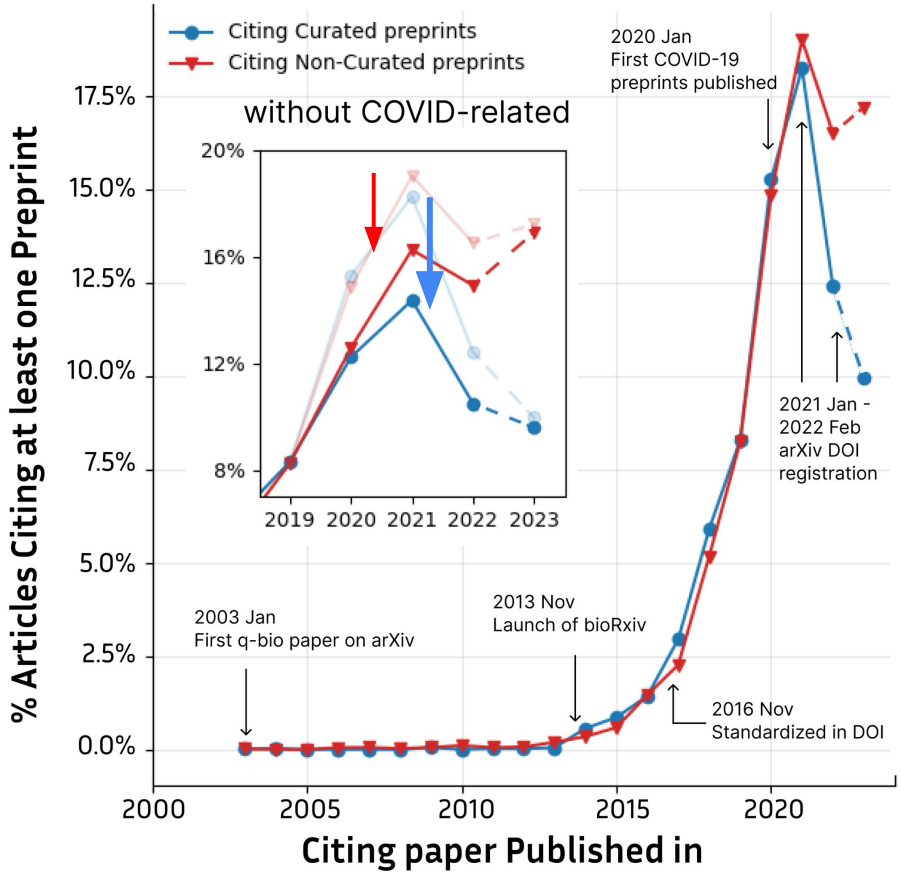
a) Reference age is stable



b) Top share is stable



Robust to other confounding factors



COVID-19 shock affected curated preprint more

Robust to other confounding factors

Table 2: The characteristic terms of titles in non-curated preprints. Top unigrams with significantly high log likelihood ratio ($\lambda > 6.635$, $p < 0.01$) for non-curated preprints with respect to curated preprints. Both frequent (+) and missing (-) terms are shown. The terms are sorted from highest λ to the lowest. We omit individual λ values for simplicity.

+ (significantly more frequent than curated)	- (significantly less frequent than curated)
learn, network, convolut, deep, adversari, neural, machin, train, reinforc, languag, model, task, algorithm, approach, supervis, graph, [illegible], understand, imag, gener, recurr, use, rang, probabilist, object, alzheim, bayesian, s, infer, explain, ai, person, attent, multimod, interpret, segment, al, u, imput, inform	cov, sar, covid, patient, basi, serolog, infect, cryo, contact, bacteri, antibodi, parasit, assembl, loop, protein, structur, bnt, pandem, genom, [illegible], receptor, rapid, termin, decis, european, impact, crispr, divers

Machine learning

Clinical, Pathological

Key Takeaways

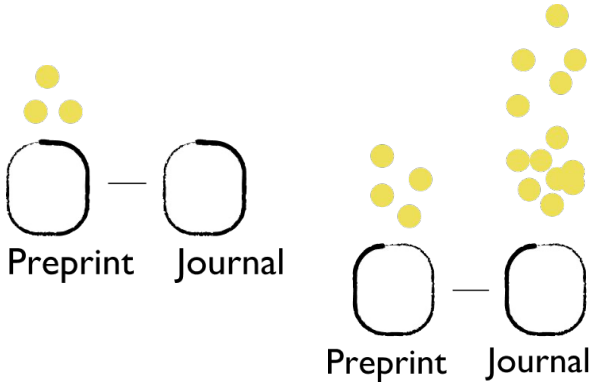
- ➔ **Non-curated preprints are increasingly cited** by top journals
- ➔ The increase is not explainable by **heightened research cycle nor a few influential preprints**
- ➔ **Perceived legitimacy is the key** to maintain citability of preprint

Preprint has its own citation economies

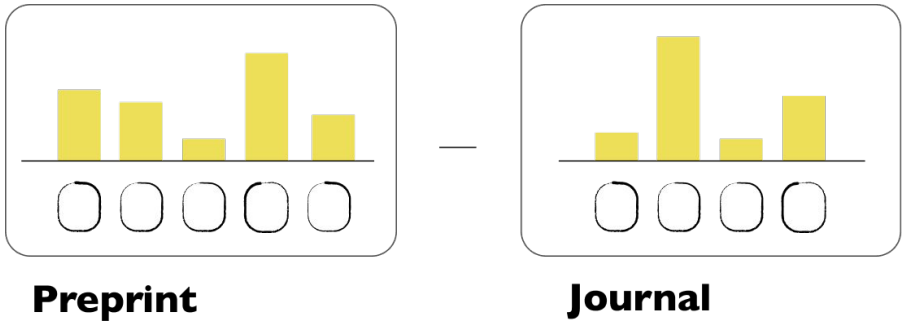
How can we understand peer review through the lens of citation?

Focusing on the distribution of overall citation counts

- The citation count of individual papers is influenced by various factors:
  - stochastic factors (preferential attachment)
  - publication cycle
- How does researchers' citation behavior change with or without peer review?

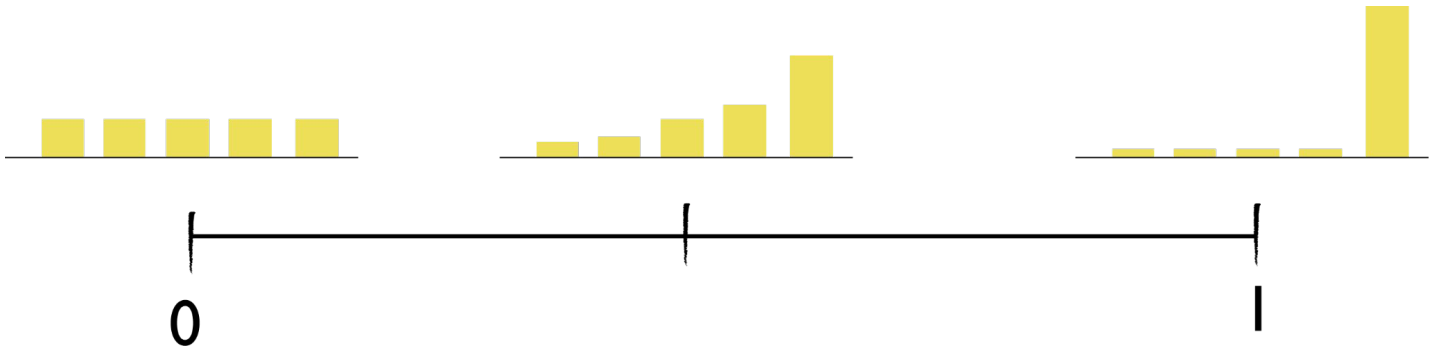


- Focus on the distribution, controlling for average impact and venue age [Nielsen & Andersen 2021].



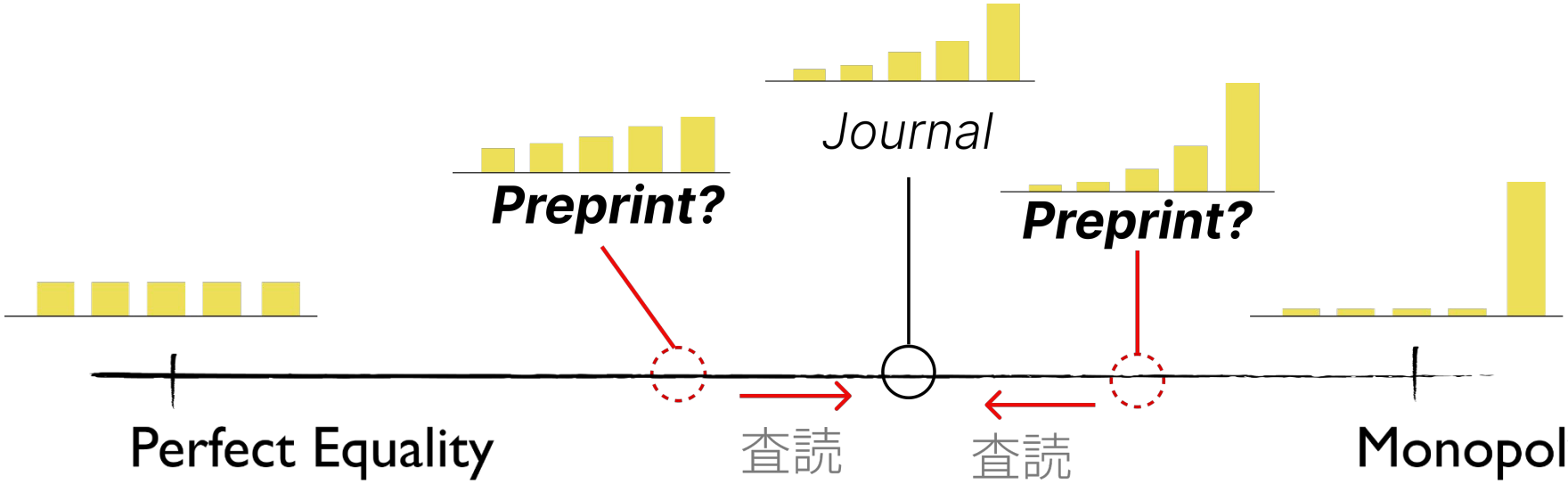
Inequality in Attention Distribution

- Attention inequality is measured by Gini coefficient of a venue
  - If every paper in the venue is cited at equal pace, if on average higher or lower, inequality is 0.
  - Inequality equal 1 means one paper accumulates all the citation that comes to the venue.



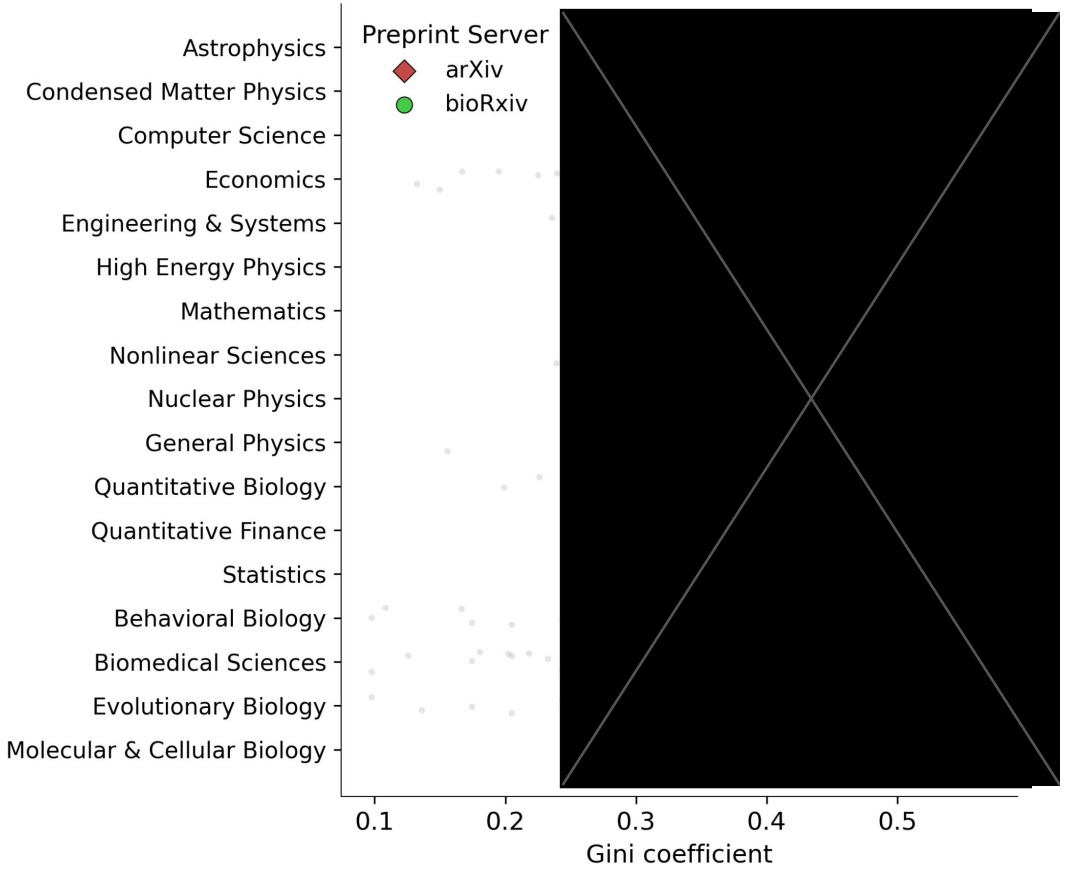
Inequality in Attention Distribution

- Will the attention inequality be Journal < Preprint, or the reverse?
  - Inequality exists within venues: some papers are cited more than others
  - Literatures pointed to Peer review bias, which may widen the attention
  - We measure the effect of peer review in decreasing attention inequality



Confidential

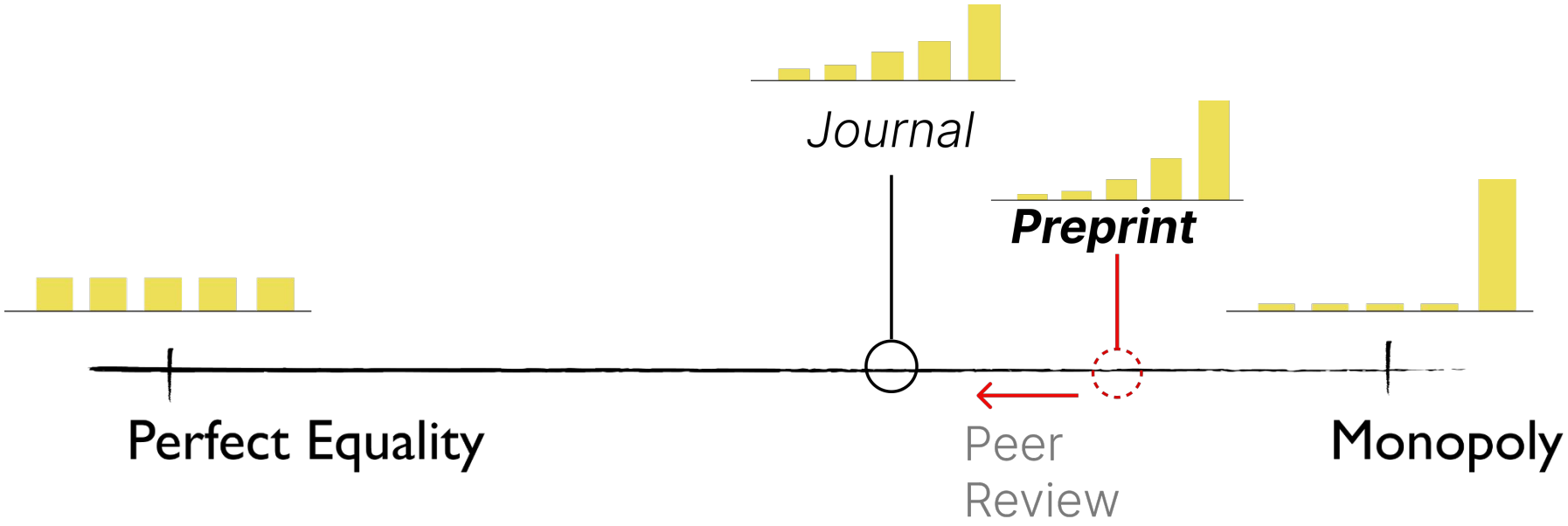
Inequality is larger in preprints than journals



Redacted

Matched Journal < Preprint

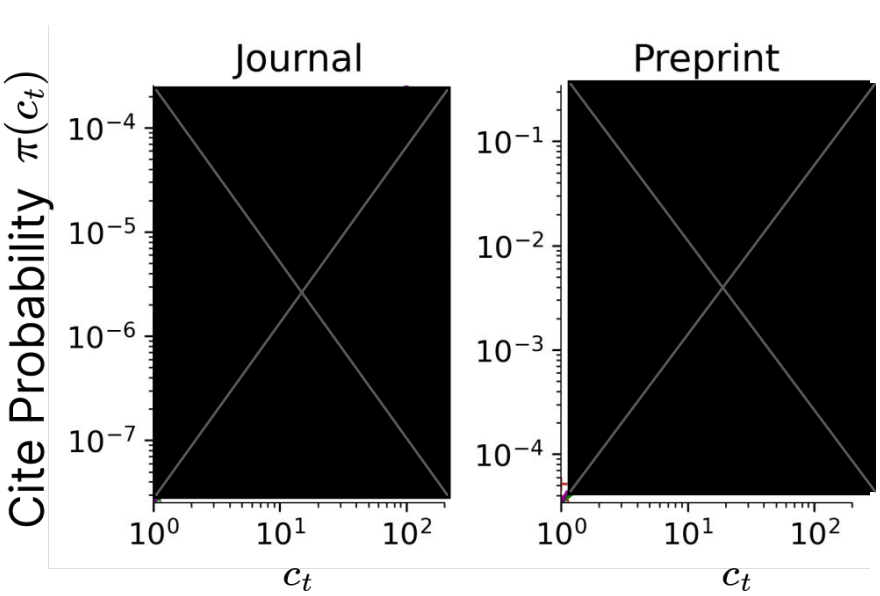
Inequality is larger in preprints than journals



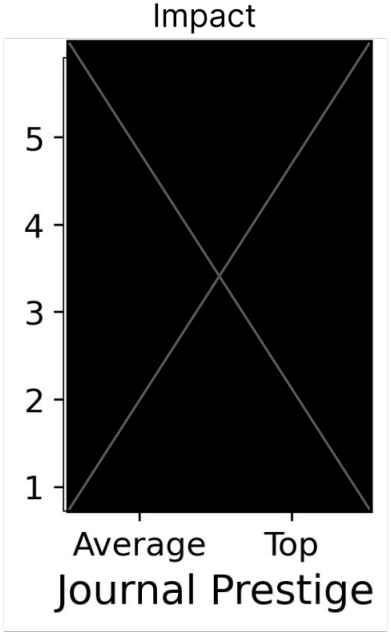
Mechanism behind the inequality

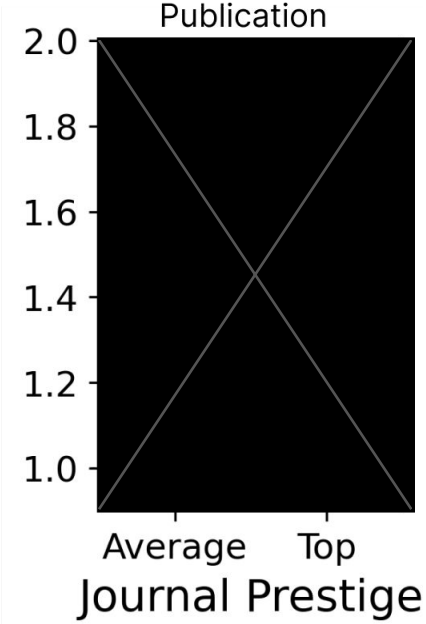
Redacted

❌ Prestige of paper



✅ Prestige of Author





Author signals the citability

Key takeaway

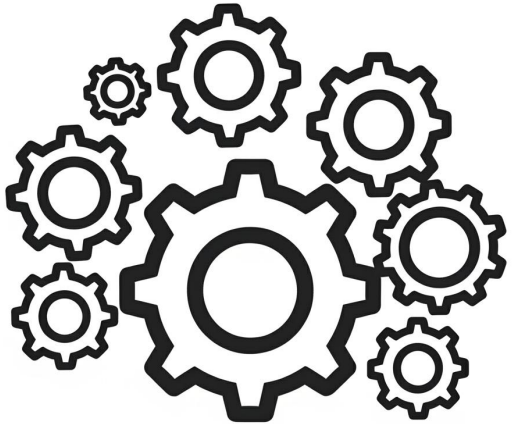
- ➔ Offers a perspective on peer review through **community attention**
- ➔ Attention **inequality within preprints is even greater** than in journals
- ➔ Suggests that **emerging researchers are discovered through peer review**.

Science systems studies

> a field of study that investigate how **scientific systems** can be organized to **maximize scientific progress.**

Science Systems Studies

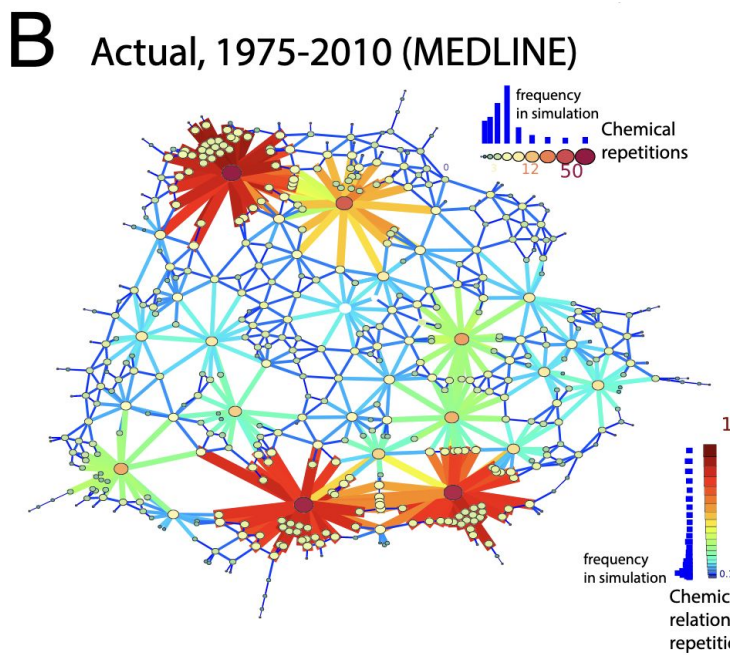
- Science is a knowledge-producing system
  - input: existing knowledge (literature)
  - output: new understanding
- Science systems
  - funding, peer review, journal, research institute, education & training, tenure track, …
  - Essentially two components (later)
- Related fields
  - Economics
  - Applied Linguistics
  - Physics
  - Machine Learning
  - Network Science

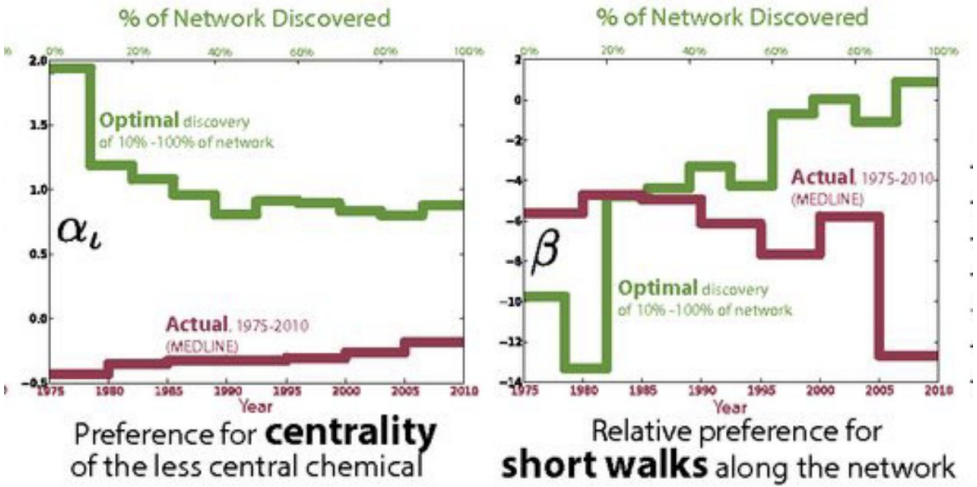


❑ Does preprint accelerate scientific progress?

- \+ accelerate publishing cycle
- \- bounded individual productivity
- \+ increase visibility
- \- increase noise
- \- curation cost

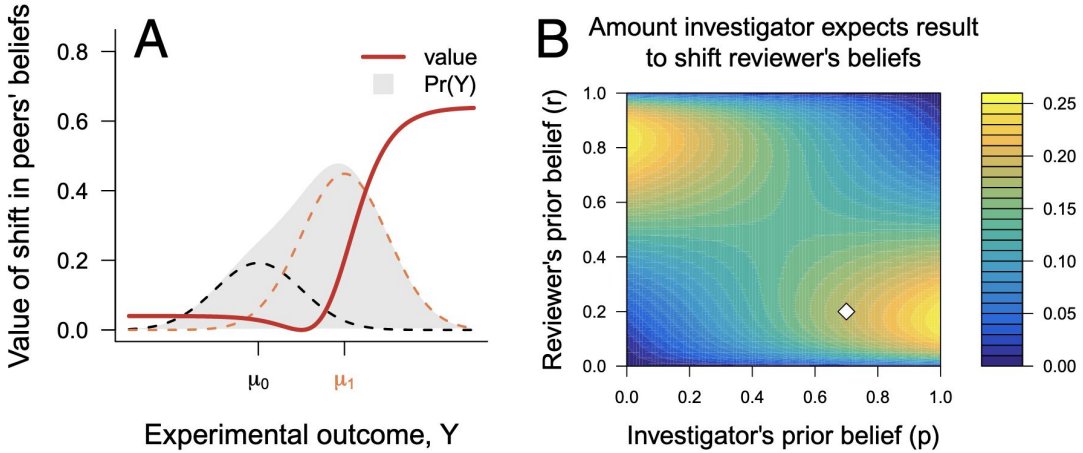
- **How researchers collaborate/coordinate/cooperate to optimize search?**





Rzhetsky 2015

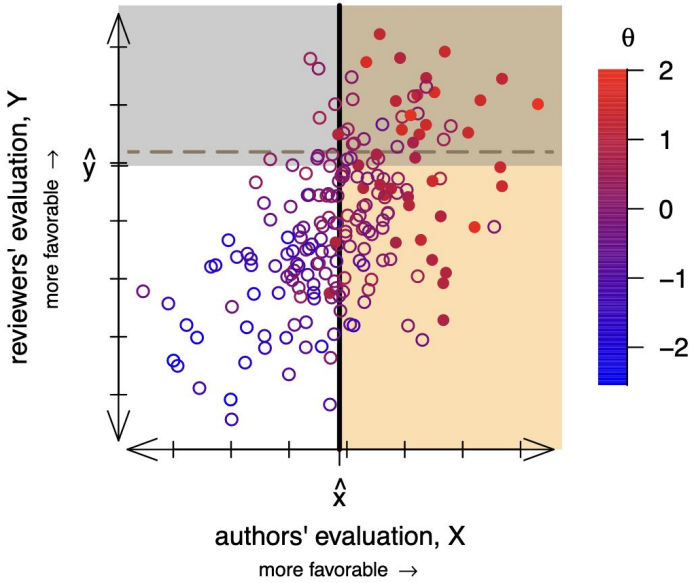
❑ How to weaken the incentives for scientific misconducts

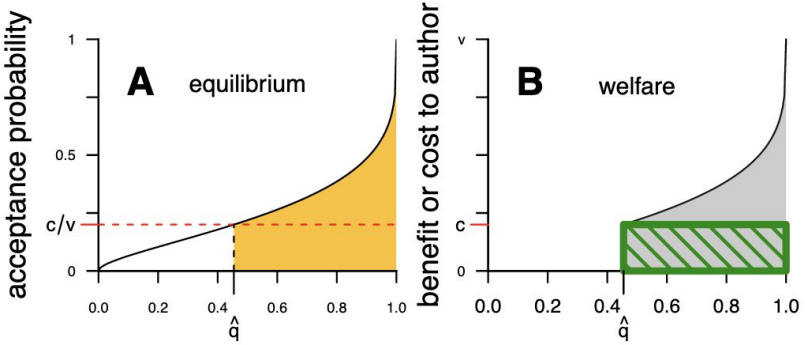


science of file drawer problem

Bergstrom&Gross 2023

- **Why reviewers disagree, and what is the best *reviewer assignment*?**





Bergstrom&Gross 2025

Thanks for listening!

Chiaki Miura

t.miura@gnt.place

Peer Review
Research Evaluation
Science Systems Studies

Blog Post and CV (Social Media)