Different Strategies in Scientific Publishing (How to be (i)relevant)

Carlos Baquero

DEI, FEUP, Universidade do Porto

Talk at IMDEA Software Institute, Madrid , June 25th 2025

Scientific Publishing

Why do people engage in Scientific Publishing?

- To Communicate findings
- To get Feedback, Validate research
- (Repeat, build, improve)

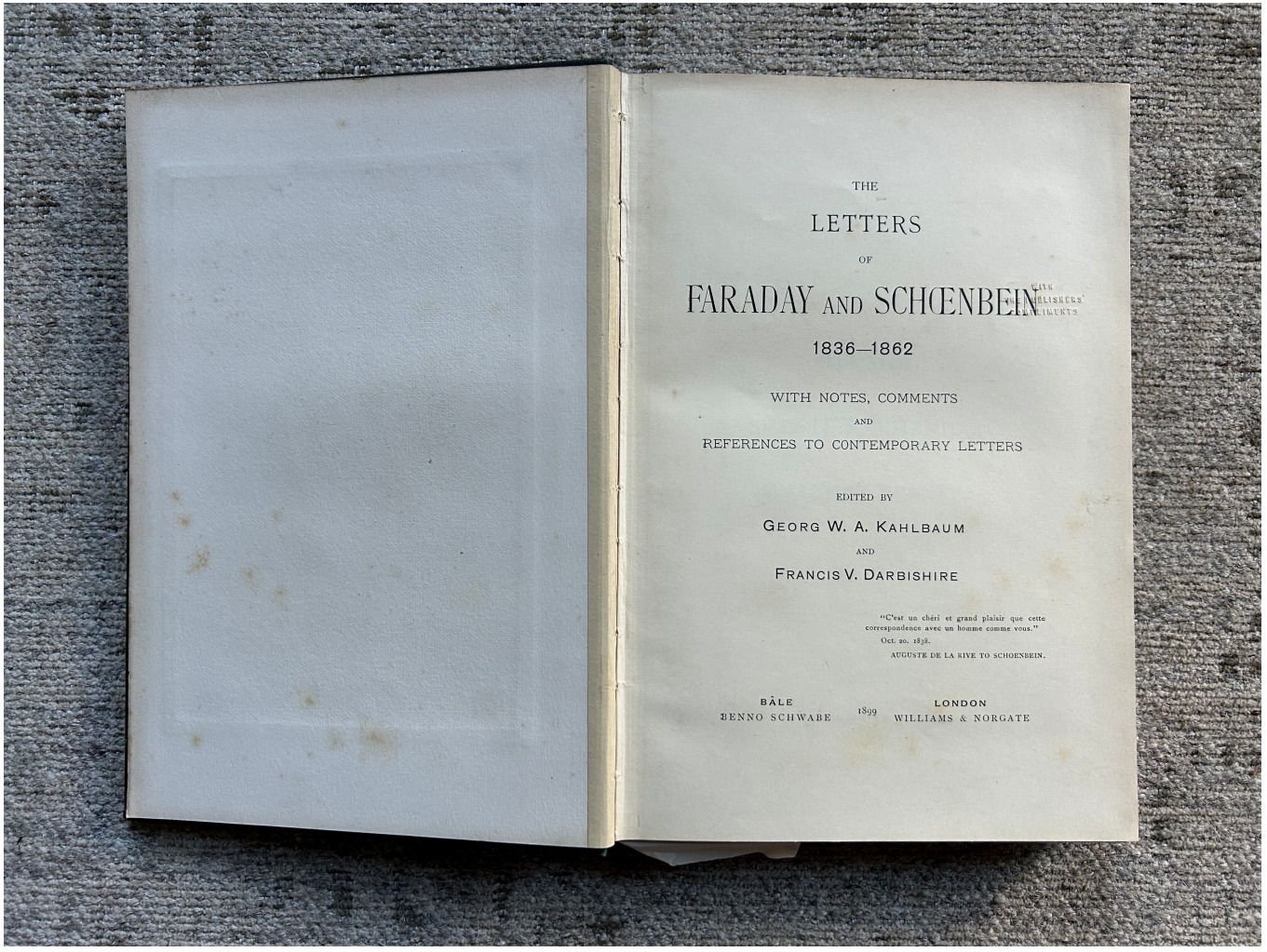
Scientific Publishing

Why do people engage in Scientific Publishing?

- To Communicate findings
- To get Feedback, Validate research
- (Repeat, build, improve)

Scientific Publishing

Communication used to be simple



THE

LETTERS

OF

FARADAY AND SCHŒNBEIN

1836—1862

WITH NOTES, COMMENTS

AND

REFERENCES TO CONTEMPORARY LETTERS

EDITED BY

GEORG W. A. KAHLBAUM

AND

FRANCIS V. DARBISHIRE

"C'est un chéri et grand plaisir que cette correspondence avec un homme comme vous."
Oct. 20. 1838.
AUGUSTE DE LA RIVE TO SCHOENBEIN.

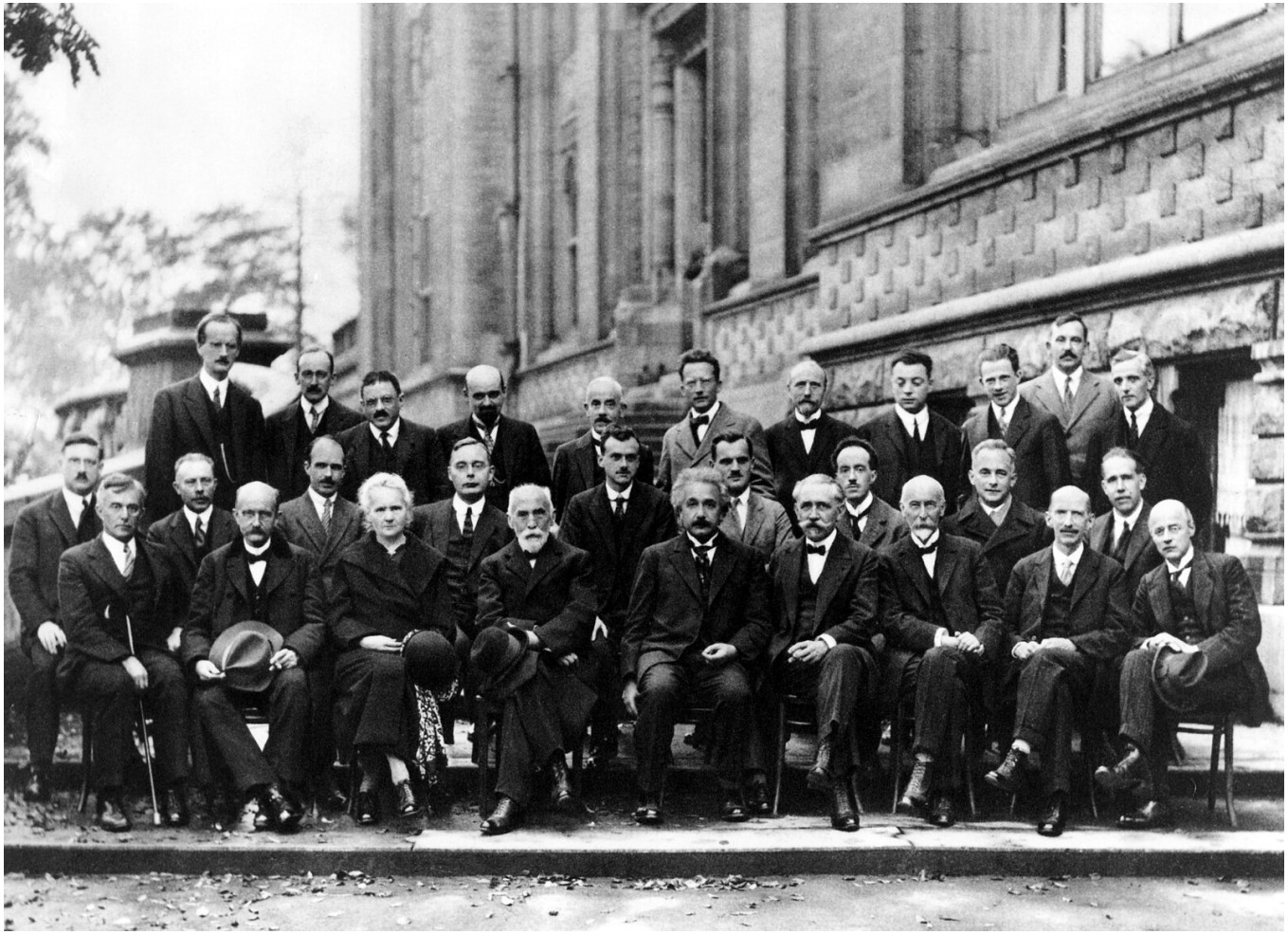
BÂLE
BENNO SCHWABE

1899

LONDON
WILLIAMS & NORGATE

Scientific Publishing

Feedback used to be simple



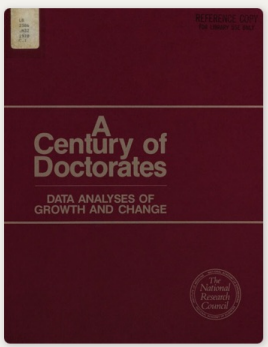
Solvay Conference, Brussels, 1927

Scientific Publishing

Feedback is still simple, in niche areas

Dare Summer School, Brussels, 2023

The Growth of PHDs



VIEW LARGER COVER

Century of Doctorates

Data Analyses of Growth and Change : U.S. PhD'S--Their Numbers, Origins, Characteristics, and the Institutions From Which They Come : a Report to the National Science Foundation, to the National Endowment for the Humanities, and to

(1978)

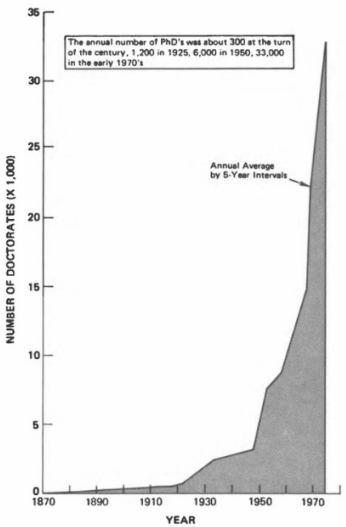
Download Free PDF | Read Free Online

Contributor(s): National Research Council; Commission on Human Resources; Board on Human-Resource Data and Analyses; Lindsey R. Harmon

The Growth of PHDs

Page 5

6



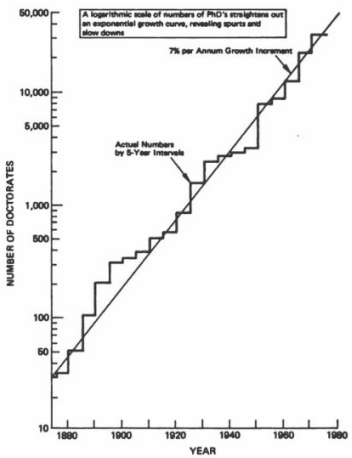
SOURCE: NRC, Commission on Human Resources

FIGURE 1 Doctorates granted annually.

While dramatic, this graph has a number of drawbacks from the standpoint of interpretation. The data cover a period in which the annual number of degrees increased a thousandfold. It is easier to visualize such an exponential growth process by plotting the data on a semi-logarithmic scale. This is done in Figure 2, which shows the average number of degrees granted per year for each 5-year period from 1875 through 1974. A straight line drawn through the "stair steps" of the graph depicts a steady 7 percent annual growth rate over this century. The deviations from this steady growth are informative, but one must allow for a greater degree of uncertainty of the data and the effects of small numbers in the years prior to the twentieth century. A slowing down is apparent for 15 years after 1895, and the year-by-year data of Table 1 show a particularly sharp decline during World War I. A growth spurt follows in the 1920's, then a slowing down during the years of the economic depression of the 1930's. Again, year-by-year data show a very sharp drop in PhD's granted during World War II and an upswing later that is even more dramatic than the huge step in Figure 2 at the beginning of the 1950's. Another slowing down appears after 1950; the growth of the "GI period" (about 1945-1950) was obviously not sustainable, and a secondary effect of World War II appeared in the late 1950's. This was a lean period due to the interruption and postponement of undergraduate education by the war; the gap moved on to the PhD level about 1957. Following this there is a steady increase through the 1960's, which experienced the highest sustained growth in PhD output since the beginning of graduate education. The early 1970's show a sharp break in the growth curve.

The output of PhD's, depicted graphically in Figures 1 and 2, is shown numerically in Table 1, which provides both annual data and 5-year summaries. As noted earlier, the data prior to 1920 are from the USOE, except for the years 1917 and 1919, which had to be filled in from NRC sources, since the USOE data became biennial after 1916.

A third way of looking at PhD growth is shown in Figure 3, which depicts the 5-year summaries in PhD graduation numbers as successive tree rings, each ring adding to the previous number of doctorates granted. In Figure 3, the *area* of each new ring is proportional to the number of new degrees granted in the 5-year



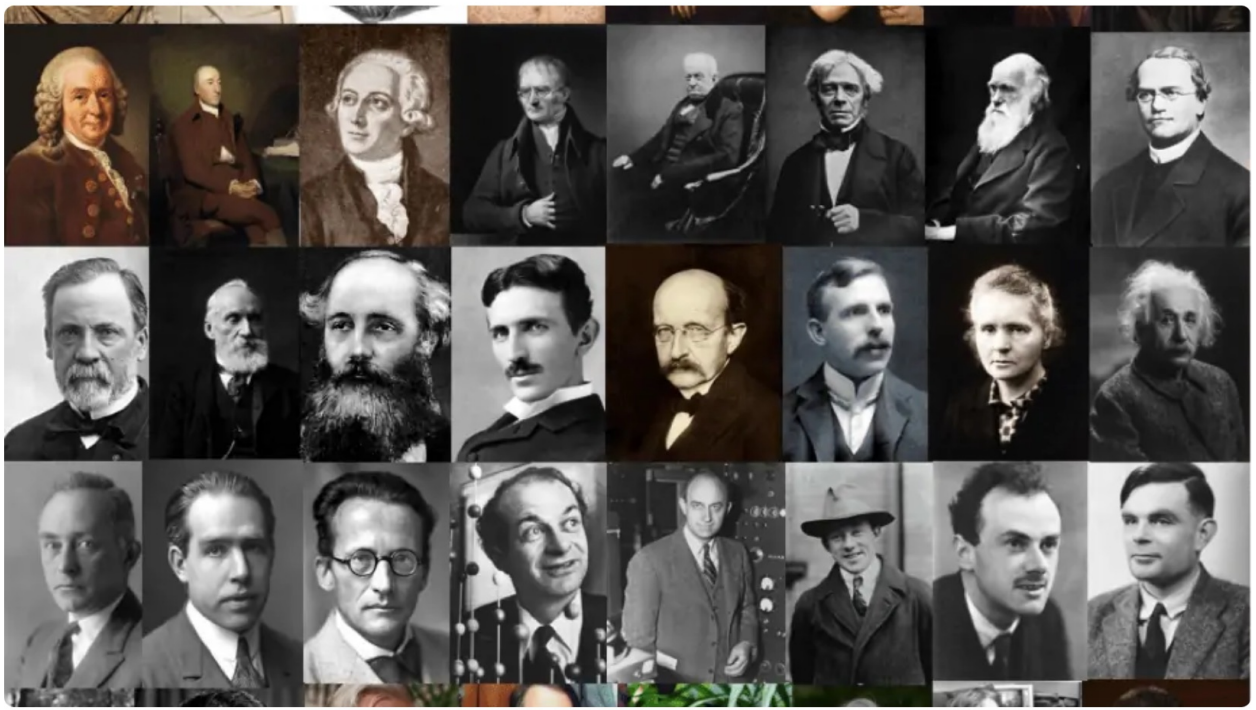
SOURCE: NRC, Commission on Human Resources

FIGURE 2 Doctorates granted annually (logarithmic scale).

The Growth of PHDs

90% of All the Scientists That Ever Lived Are Alive Today

Published: November 5, 2015 Author: Eric Gastfriend



Are we making excellent science? Yes, and No

Standing Out

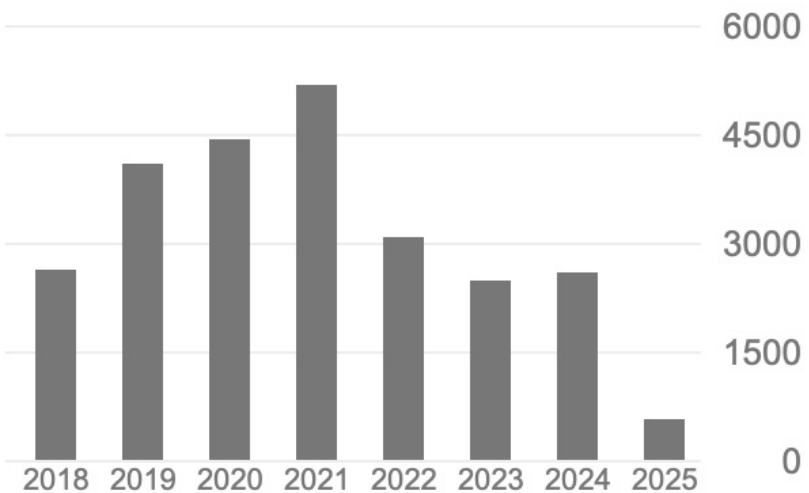
When there is a lot of competition, nature finds a way to stand out

(By Jatin Sindhu - Own work, CC BY-SA 4.0)

Standing Out
Gaming the citation indexes

When there is a lot of competition, nature finds a way to stand out

Cited by		VIEW ALL
	All	Since 2020
Citations	39651	18417
h-index	118	67
i10-index	554	350



Standing Out
Gaming the citation indexes

These impressive statistics were obtained by citation gaming. For instance when giving a talk (like this one) the scientist in question often managed to write a short summary and inject 200 citations. Half to itself and half to its new journal.

No citation shall go wasted :)

For more details


EL PAÍS

Science

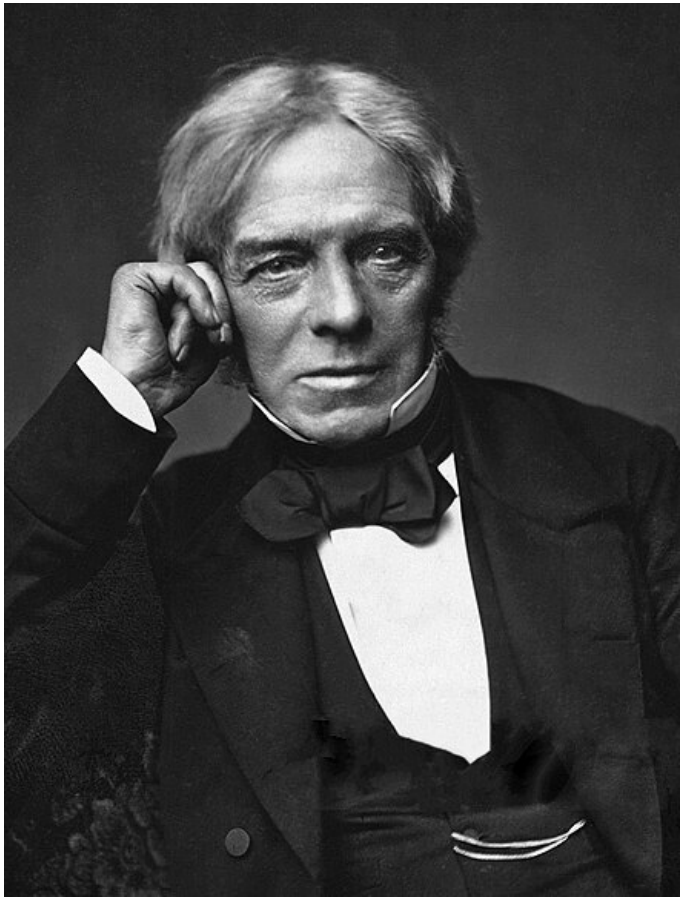
SILICON VALLEY · YOUTUBE · GOOGLE · LATEST NEWS

ETHICS IN SCIENCE >

The seven lies of the AI expert who cited himself thousands of times on scientific papers

Scientific Communication

Communication used to be simple



Faraday and Schonbein, didn't care much with citations

Scientific Communication

Communication can be informal (Joe Hellerstein Blog, UCB)

{ async stream }

asynchronous data · distributed systems · stream of consciousness

HOME ABOUT CONTACT

CRDTs #4: Convergence, Determinism, Lower Bounds and Inflation

May 29, 2025 • 16 min read

The CRDT literature sometimes leaves room for mathematical ambiguity. Maybe because the bulk of the work tends to be targeted at systems researchers and developers, like a lot of work on eventual consistency.

The discussion below untangles three subtle but important ideas in CRDT design, which turn out to be interrelated:

1. Determinism vs Convergence guarantees
2. Early `read`s: lower bounds or not?
3. Algebraic property requirements for `update` functions.

⚠ The CRDT guarantee of *strong eventual consistency does **not** guarantee determinism!* If you want your CRDTs to be deterministic, or you want to treat CRDT `read`s as lower bounds, then your `update` functions must be **inflationary**.

> *This is the 4th post in a series I'm doing on CRDTs. This one is a bit more technical and narrow than my previous CRDT posts, but practitioners should still know about the main conclusions. This post also contains a small formal result that seems novel -- Strong Eventual Consistency does not guarantee determinism! I'd be curious to hear about prior work that makes this point.*

Scientific Communication

Communication can be informal (Joe Hellerstein Blog, UCB)

pssalmeida 17 days ago

Nice to see someone finally noticing the monotonic vs inflation issue! (Carlos Baquero drew my attention to this post.) Some remarks:

About non-determinism. The issue is not determinism, it is about ensured progress without losing updates. The kind of "non-determinism" that happens if we do not have inflations is that some updates will "randomly" disappear as if they had not been issued. If we start from state A, issue some updates and end up in some smaller state B, state B will not subsume A but will be overwritten by A. So, a message carrying an older state (e.g., a duplicate) will cause operations to be discarded. That is unacceptable: we want a newer state to win over an older state. So, It is not a form of benign non-determinism, but goes against one main goal of CRDTs: avoiding losing updates. So we always want state mutators to be inflations. That, together with the join, is what allows the weak dissemination guarantees from the network in state-based CRDTs, including message duplication, unlike op-based CRDTs.

The decrement in the CSUR paper was an example for the standard Int lattice. If we reverse the order, then increment is not an inflation. The point is that although both increment and decrement are monotonic, we cannot have both as inflations in the same Int lattice. This forces a PN-Counter to have two components.

About lower bounds. Again, we want inflations not for the reason of not "spoiling lower bounds". We want to always have inflations for the same reason: not to lose updates, regardless of whether the domain in the object being modelled is a lattice and, if that is the case, whether all operations are inflations in that lattice. A counter object with inc and dec operations does not provide "lower bounds" but is perfectly fine for many purposes. So, even if we do not have "lower bounds to spoil", we want the state mutators corresponding to inc and dec to be both inflations in the CRDT state lattice, i.e., we want a PN-Counter CRDT to implement it.

↑ 1 👍 1 5 replies

jhellerstein 16 days ago Owner edited

Lots of good thoughts to discuss!

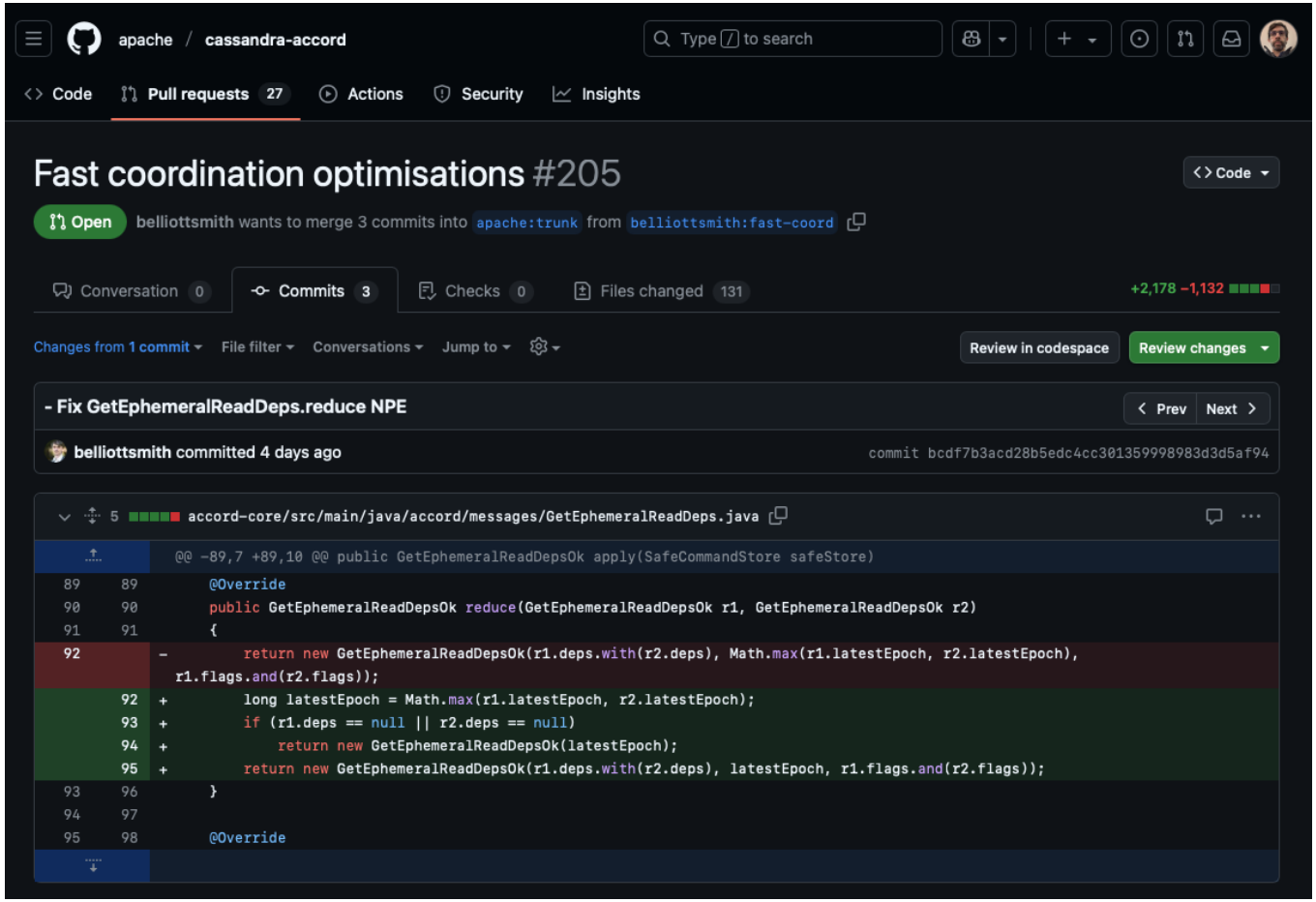
We don't disagree on your first point: it's both about non-deterministic reads as I said, and it's due to non-deterministically "lost" (or overwritten) updates as you said. These are not conflicting points, they just reflect a different focus: mine on downstream usage ("read"s and other functions that expose properties of the transient state), yours on guarantees for updates.

WRT whether non-inflationary updates create benign or malignant non-determinism, I have no opinion. I won't try to assume what CRDT designs are *supposed* to do, but non-inflationary updates *do* match the spec of SEC.

Similarly, lower bounds for reads or lost updates... we don't disagree, it's the same issue of what lens we choose: read guarantees (visibility) vs write guarantees. My feeling is that if "reads" (broadly defined, including functions that expose properties of the state) are ill-defined, then there's little point proving things about updates.

Scientific Communication

Communication can be practical (Cassandra Accord Github)

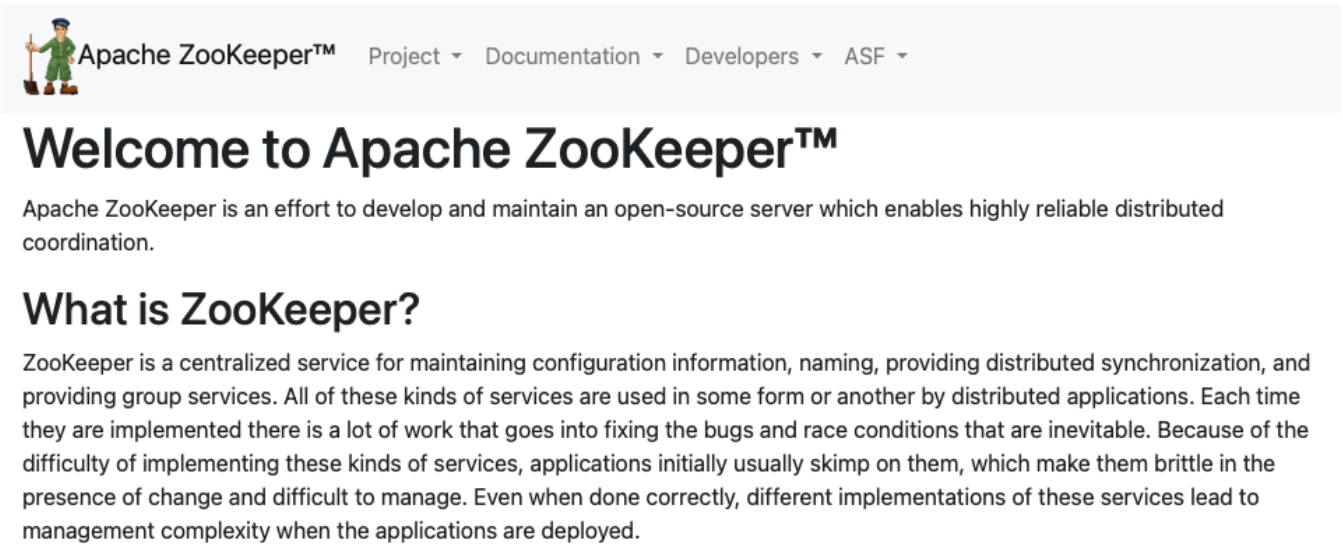


This is actual applied science, used in Apple products.

Applied Science

The two faces of Applied Science

(a) Practical tools with a solid background



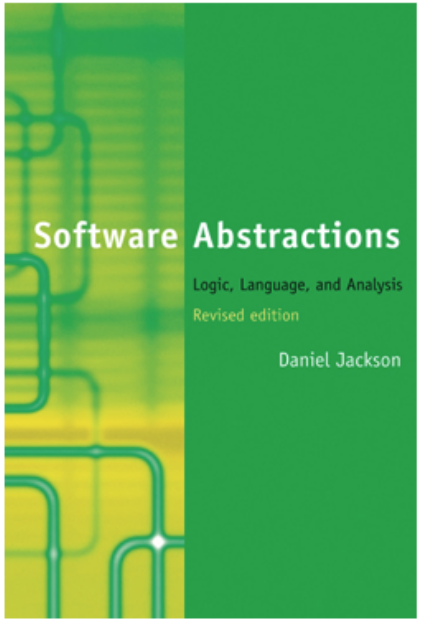
Applied Science

The two faces of Applied Science

(a) Practical tools with a solid background

Alloy

Alloy is an open source language and analyzer for software modeling. It has been used in a wide range of applications, from finding holes in security mechanisms to designing telephone switching networks. This site provides language documentation, tool downloads, and a repository of links to case studies and applications. As the open source community grows, this site will also provide access to extensions of the Alloy Analyzer, and tools built on top of it and on top of Kodkod, its model finding engine.



Last release

The last release to date (2025/01/09) is Alloy 6.2.0 .

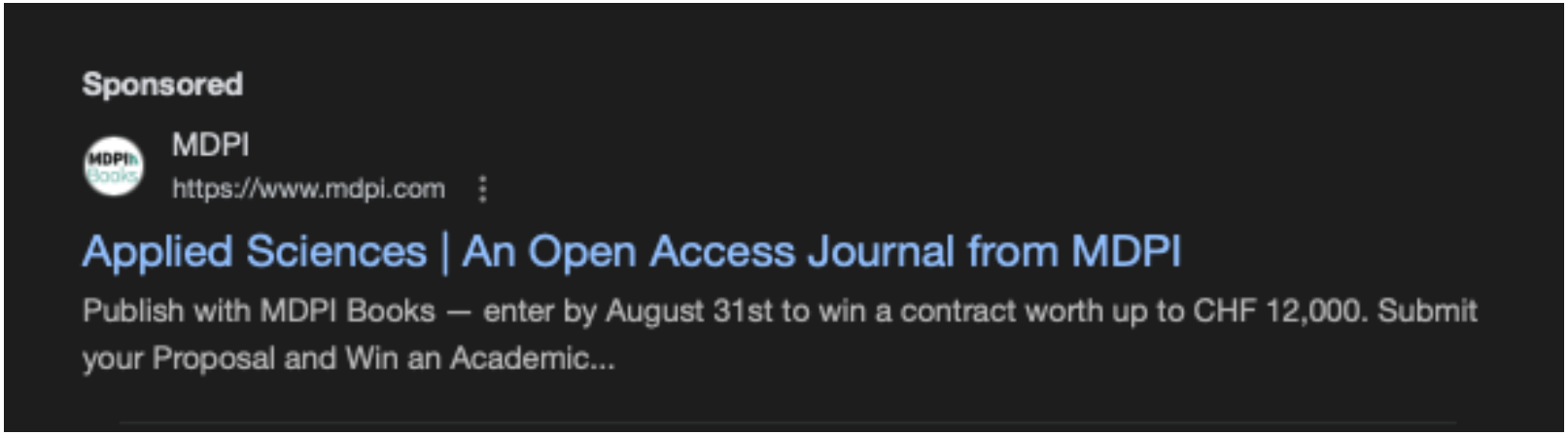
Alloy 6

Alloy 6 is a major revision w.r.t Alloy 4, that adds mutable state, a temporal logic and accompanying solvers as well as an improved Visualizer. Specifying the behavior of systems gets easier in many cases.

Applied Science

The two faces of Applied Science

(b) Serial publishing of standard techniques over diverse settings



(1st result from googling “Applied Science”)

Goodhart's Adage

Goodhart's law

"when a measure becomes a target, it ceases to be a good measure"

- Less focus on indexes
- Less focus o ranks
- Explain the results and their impact
- More science and more peer evaluation
- (AI is not the solution, yet)

DORA initiative

https://sfdora.org/read/

San Francisco Declaration on Research Assessment

General Recommendation

1. Do not use journal-based metrics, such as Journal Impact Factors, as a surrogate measure of the quality of individual research articles, to assess an individual scientist's contributions, or in hiring, promotion, or funding decisions.

IE: Research Evaluation

https://www.informatics-europe.org/

The Strain on Publishing

Volume 5, Issue 4

Fall 2024

November 01 2024

The strain on scientific publishing

Mark A. Hanson, Pablo Gómez Barreiro, Paolo Crosetto, Dan Brockington

Check for updates

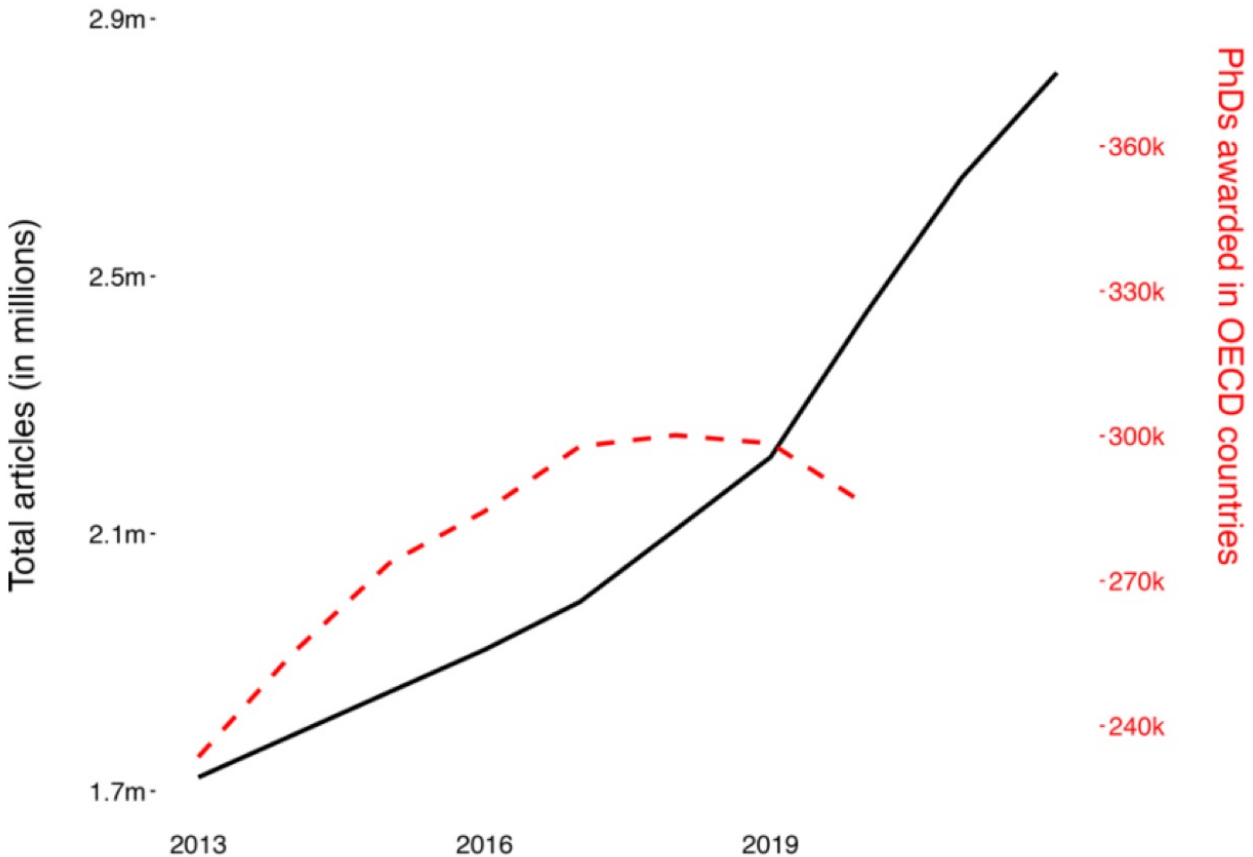
Author and Article Information

Quantitative Science Studies (2024) 5 (4): 823–843.

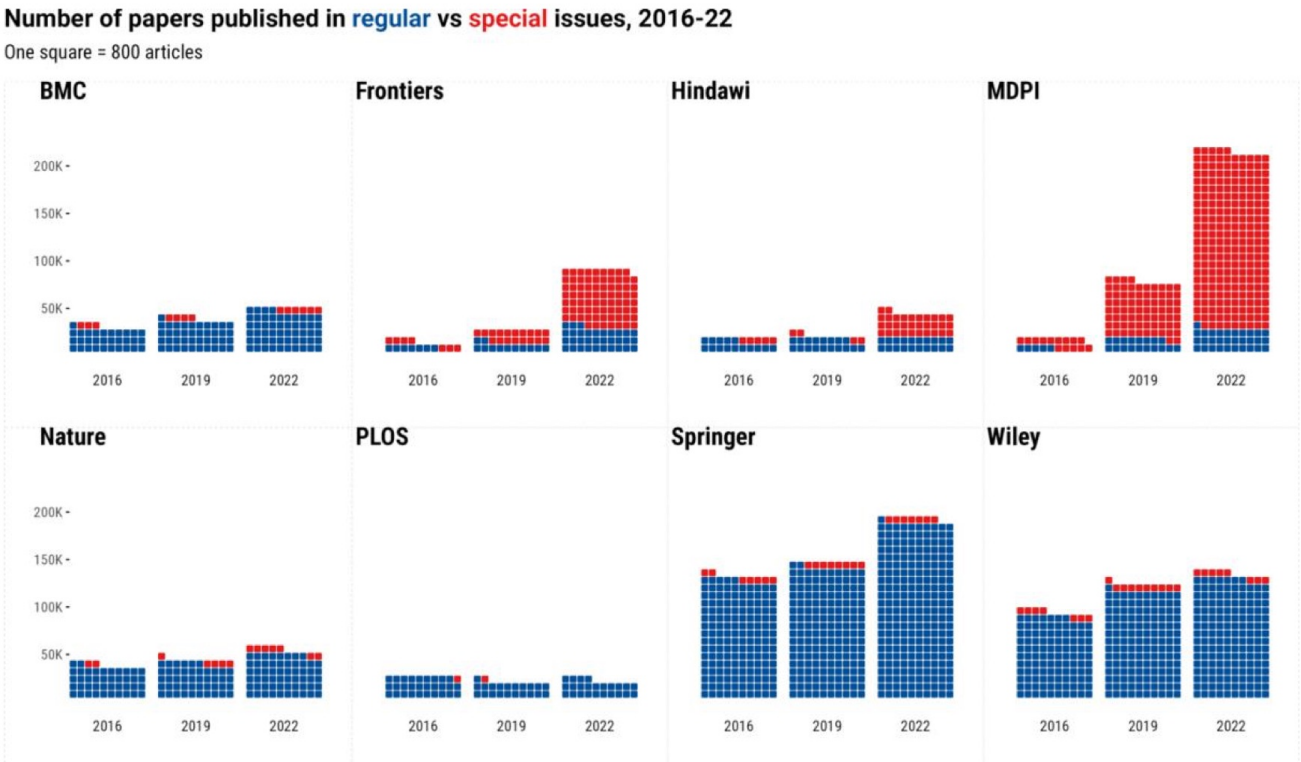
https://doi.org/10.1162/qss_a_00327 **Article history**

Cite **PDF** **Permissions** **Share** **Views**

The Strain on Publishing



The Strain on Publishing



Hyperprolific Authors

nature

View all journals | Search | Log in

Explore content | About the journal | Publish with us | Subscribe | Sign up for alerts | RSS feed

nature › comment › article

COMMENT | 12 September 2018

Thousands of scientists publish a paper every five days

To highlight uncertain norms in authorship, John P. A. Ioannidis, Richard Klavans and Kevin W. Boyack identified the most prolific scientists of recent years.

By John P. A. Ioannidis, Richard Klavans & Kevin W. Boyack

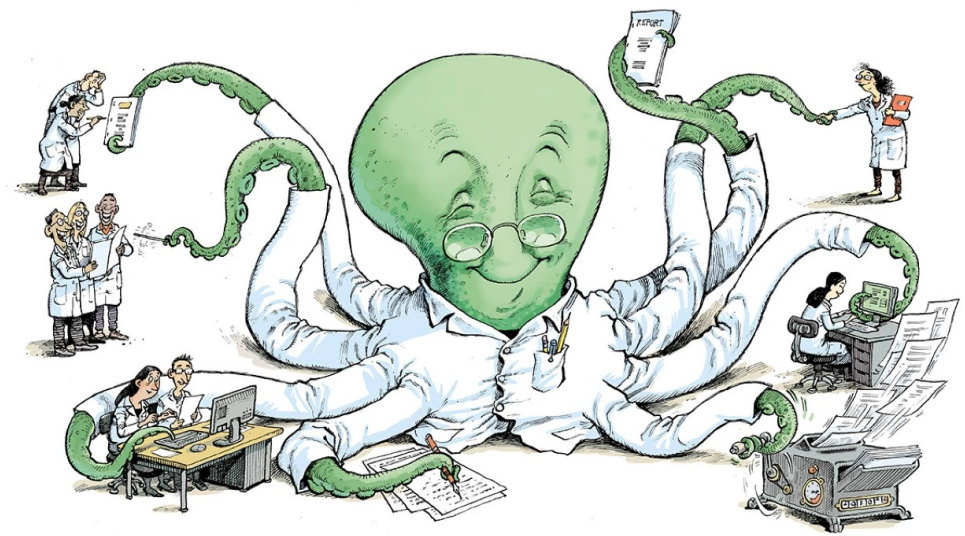
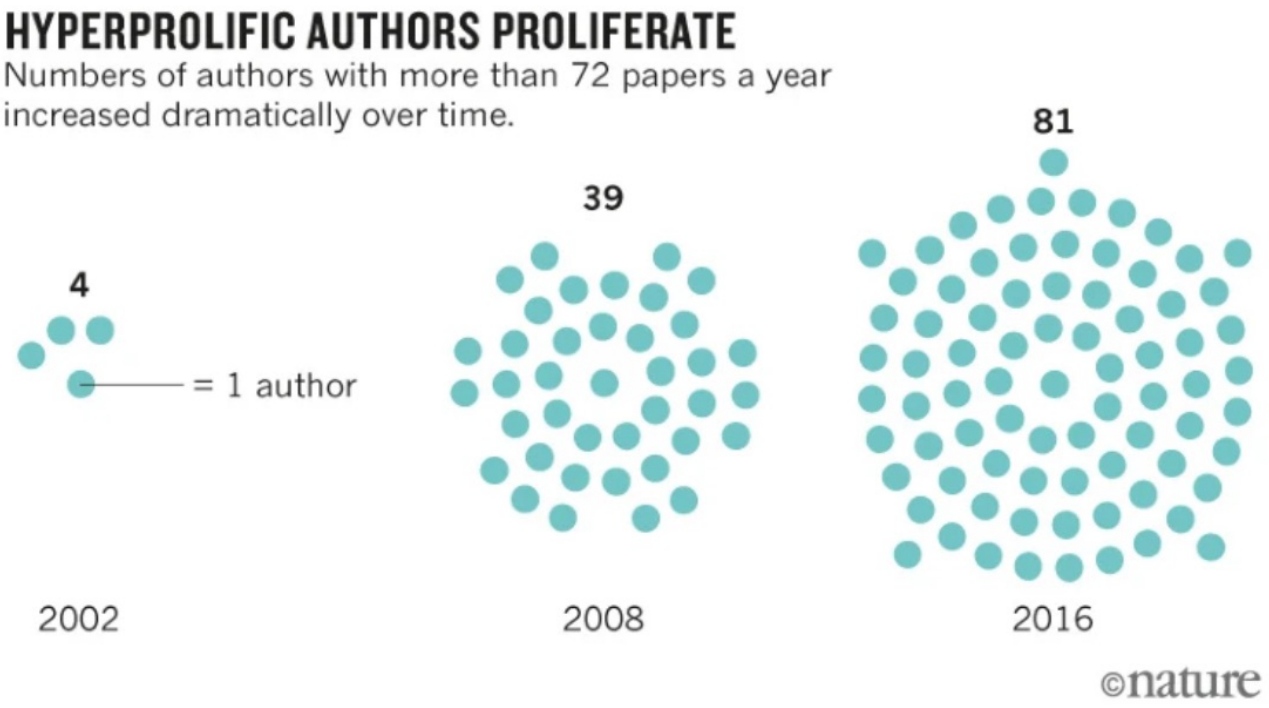


Illustration by David Parkins

Hyperprolific Authors



Source: J. P. A. Ioannidis, R. Klavans & K. W. Boyack

Picking Publication Targets

Overview of the text in

BLOG@CACM

Picking Publication Targets

By Carlos Baquero

Communications of the ACM, March 2022, Vol. 65 No. 3, Pages 10-11

10.1145/3510545

The Context

Publication targets depend on:

- **The quality** and length of the work. Each paper is different
- **The moment**. Deadlines matter. Venue quality changes
- **The career/project** phase. Persons and grants need outputs

The Review Process

What can we expect of the review process?

NeurIPS experiment

Acceptance decisions are highly inconsistent among reviewer subsets . . . the randomness in the decision increases as acceptance rates decrease. Reviewers are more consistent in detecting weaker papers than in agreeing on the best ones. Paper clarity was the best predictor of future paper impact.

The Review Process

What can we expect of the review process?

NeurIPS experiment

Acceptance decisions are highly inconsistent among reviewer subsets . . . the randomness in the decision increases as acceptance rates decrease. Reviewers are more consistent in detecting weaker papers than in agreeing on the best ones. Paper clarity was the best predictor of future paper impact.

The Review Process

What can we expect of the review process?

NeurIPS experiment

Acceptance decisions are highly inconsistent among reviewer subsets . . . the randomness in the decision increases as acceptance rates decrease. Reviewers are more consistent in detecting weaker papers than in agreeing on the best ones. Paper clarity was the best predictor of future paper impact.

The Review Process

What can we expect of the review process?

NeurIPS experiment

Acceptance decisions are highly inconsistent among reviewer subsets . . . the randomness in the decision increases as acceptance rates decrease. Reviewers are more consistent in detecting weaker papers than in agreeing on the best ones. Paper clarity was the best predictor of future paper impact.

The Review Process

What can we expect of the review process?

Conferences

- Close to 50% of the decision is random (NeurIPS case)
- In light of this, resubmission with small changes is a strategy
- Chasing reviewer whims is not efficient

The Review Process

What can we expect of the review process?

Conferences

- Close to 50% of the decision is random (NeurIPS case)
- In light of this, resubmission with small changes is a strategy
- Chasing reviewer whims is not efficient

The Review Process

What can we expect of the review process?

Conferences

- Close to 50% of the decision is random (NeurIPS case)
- In light of this, resubmission with small changes is a strategy
- Chasing reviewer whims is not efficient

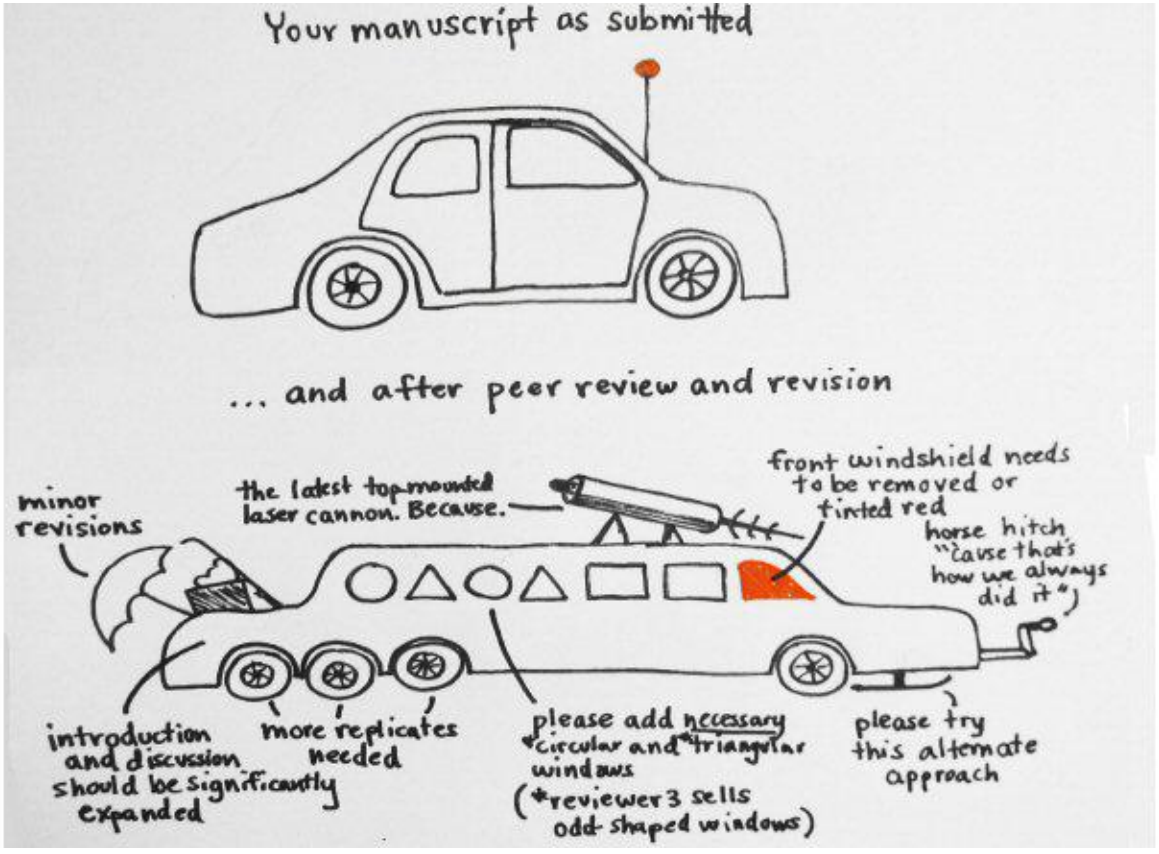
The Review Process

What can we expect of the review process?

Conferences

- Close to 50% of the decision is random (NeurIPS case)
- In light of this, resubmission with small changes is a strategy
- Chasing reviewer whims is not efficient

The Whims



- Better venues bring better quality feedback
- Across conference re-submissions reviewers keep changing

The Review Process

How can we stop the reviewers from changing?

Journals

- **Hybrids**: VLDB conference/journal
- **Classic**: IEEE, ACM, Elsevier, . . . (OA as option)
- **Diamond Open Access**: *The Programming Journal*
- ———————
- **Borderline**: Plos One
- **Professional "Predatory"**: MDPI, Frontiers
- **Classic Predatory**: WSEAS et all (funny english team)

The Review Process

How can we stop the reviewers from changing?

Journals

- **Hybrids**: VLDB conference/journal
- **Classic**: IEEE, ACM, Elsevier, . . . (OA as option)
- **Diamond Open Access**: *The Programming Journal*
- ———————
- **Borderline**: Plos One
- **Professional "Predatory"**: MDPI, Frontiers
- **Classic Predatory**: WSEAS et all (funny english team)

The Review Process

How can we stop the reviewers from changing?

Journals

- **Hybrids**: VLDB conference/journal
- **Classic**: IEEE, ACM, Elsevier, . . . (OA as option)
- **Diamond Open Access**: *The Programming Journal*
- ———————
- **Borderline**: Plos One
- **Professional "Predatory"**: MDPI, Frontiers
- **Classic Predatory**: WSEAS et all (funny english team)

Opportunity cost

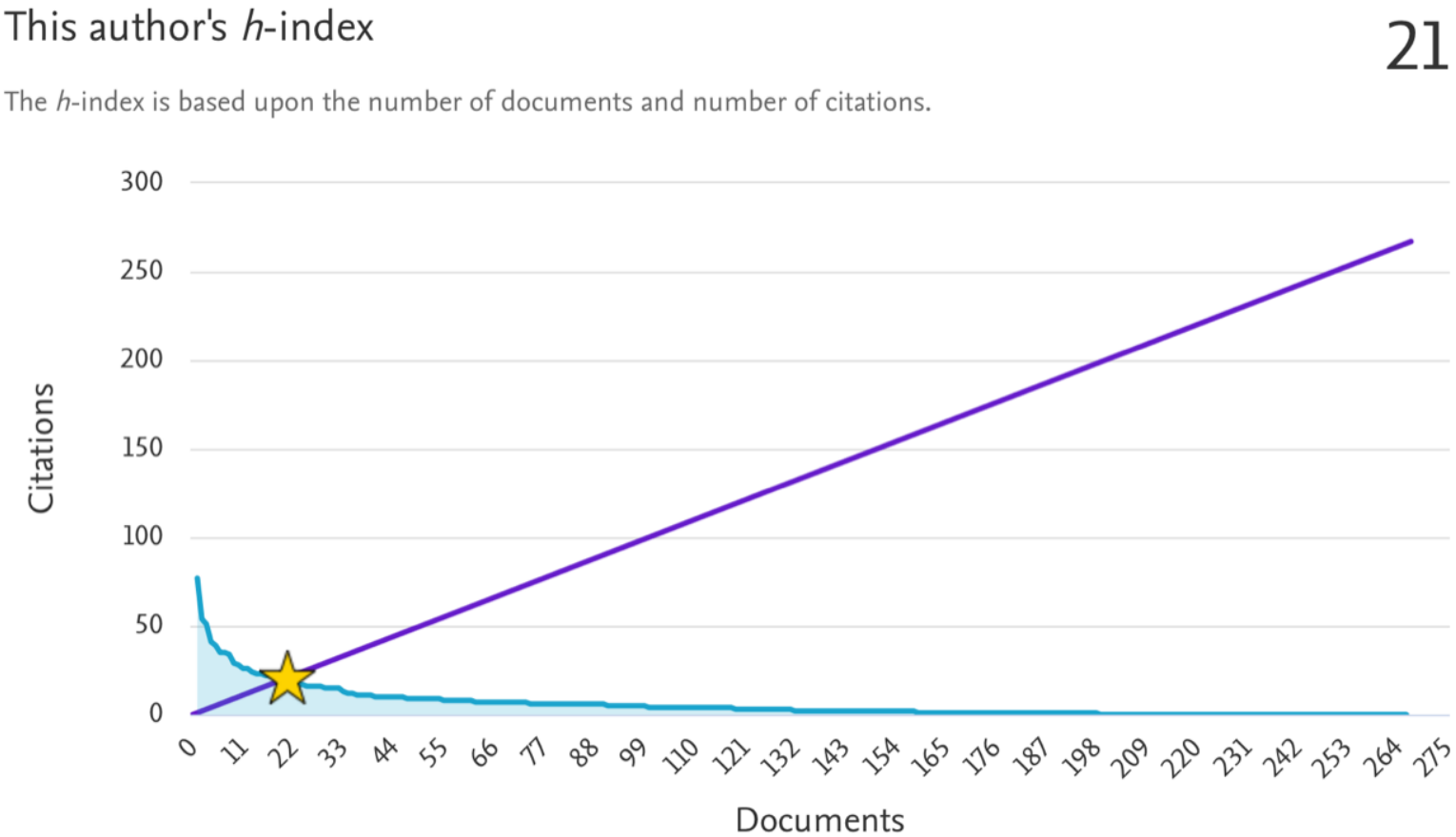
Google("Opportunity cost"):

> *"What Is Opportunity Cost? Opportunity costs represent the potential benefits that an individual, investor, or business misses out on when choosing one alternative over another. Because opportunity costs are unseen by definition, they can be easily overlooked."*

Our time for trying to make good impactful papers is finite

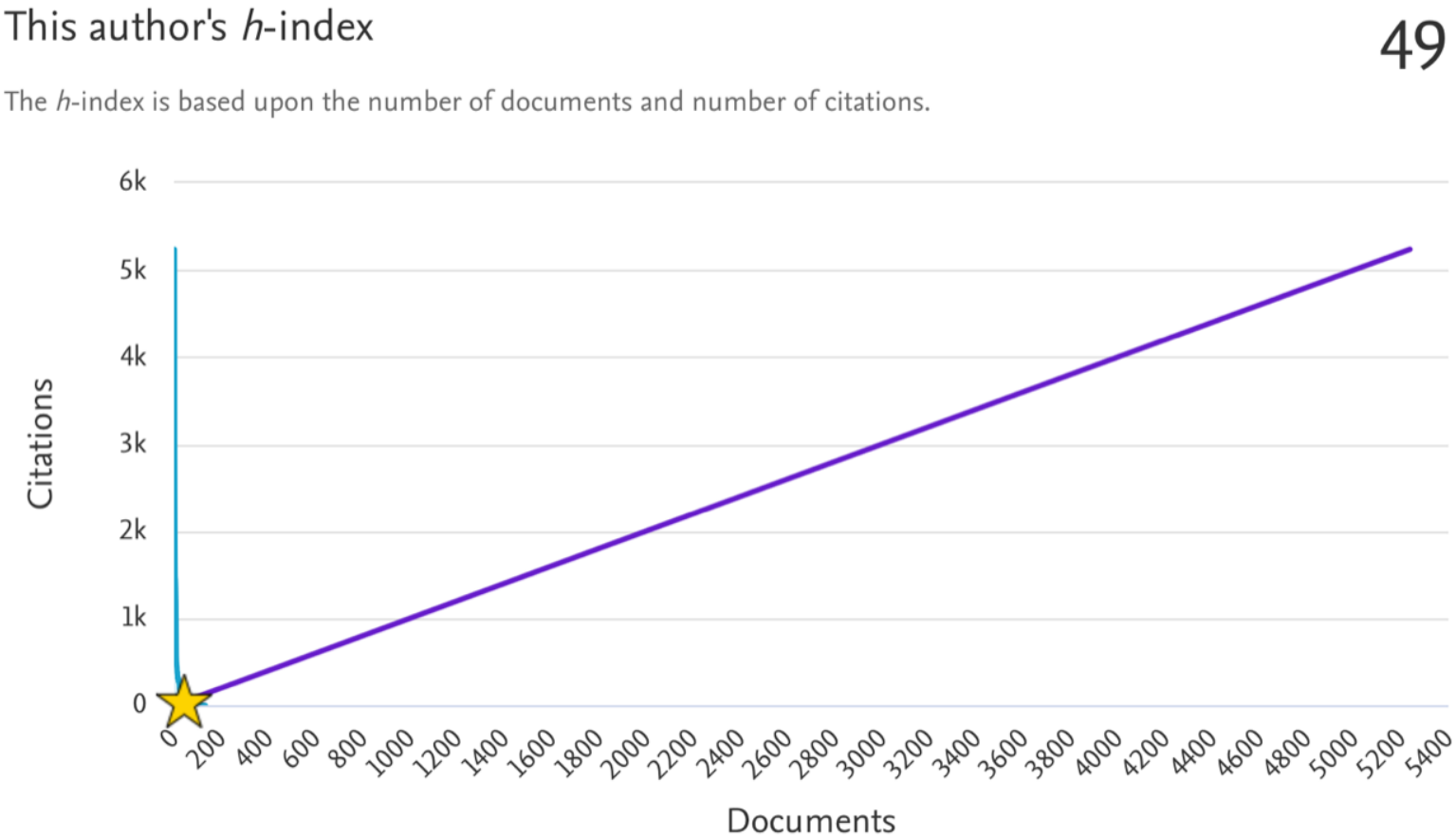
The profiles
The Maximalist Strategy

A maximalist strategy would focus on producing as many papers as possible with minimal investment in each result, and full use of conferences and journals with high acceptance rates.



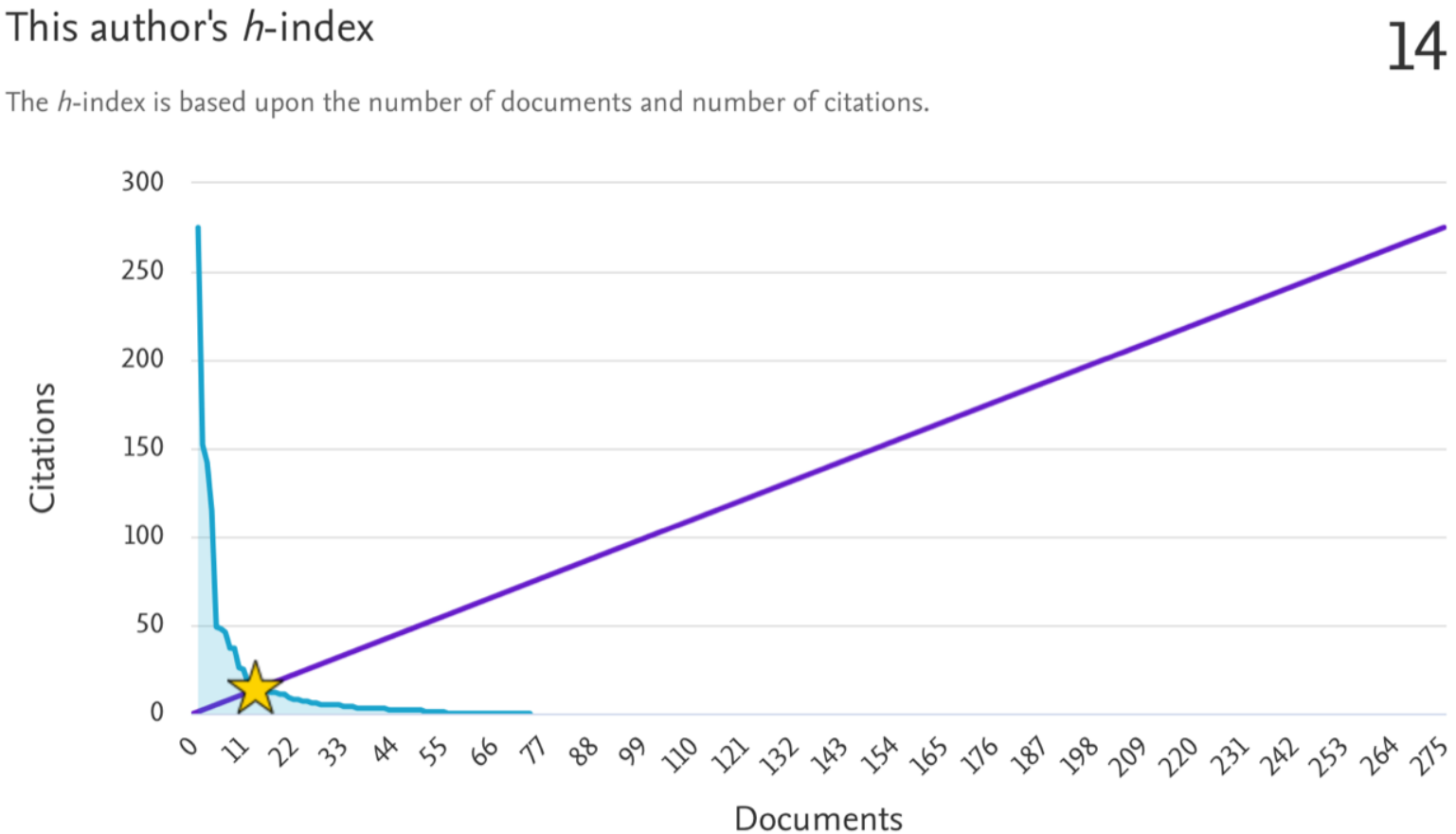
The profiles
The Perfectionist Strategy

A perfectionist strategy would strive to achieve the perfect scientific result and only publish high-potential results in the most demanding venues. (Our Ronaldo is Lamport. There are few stars)

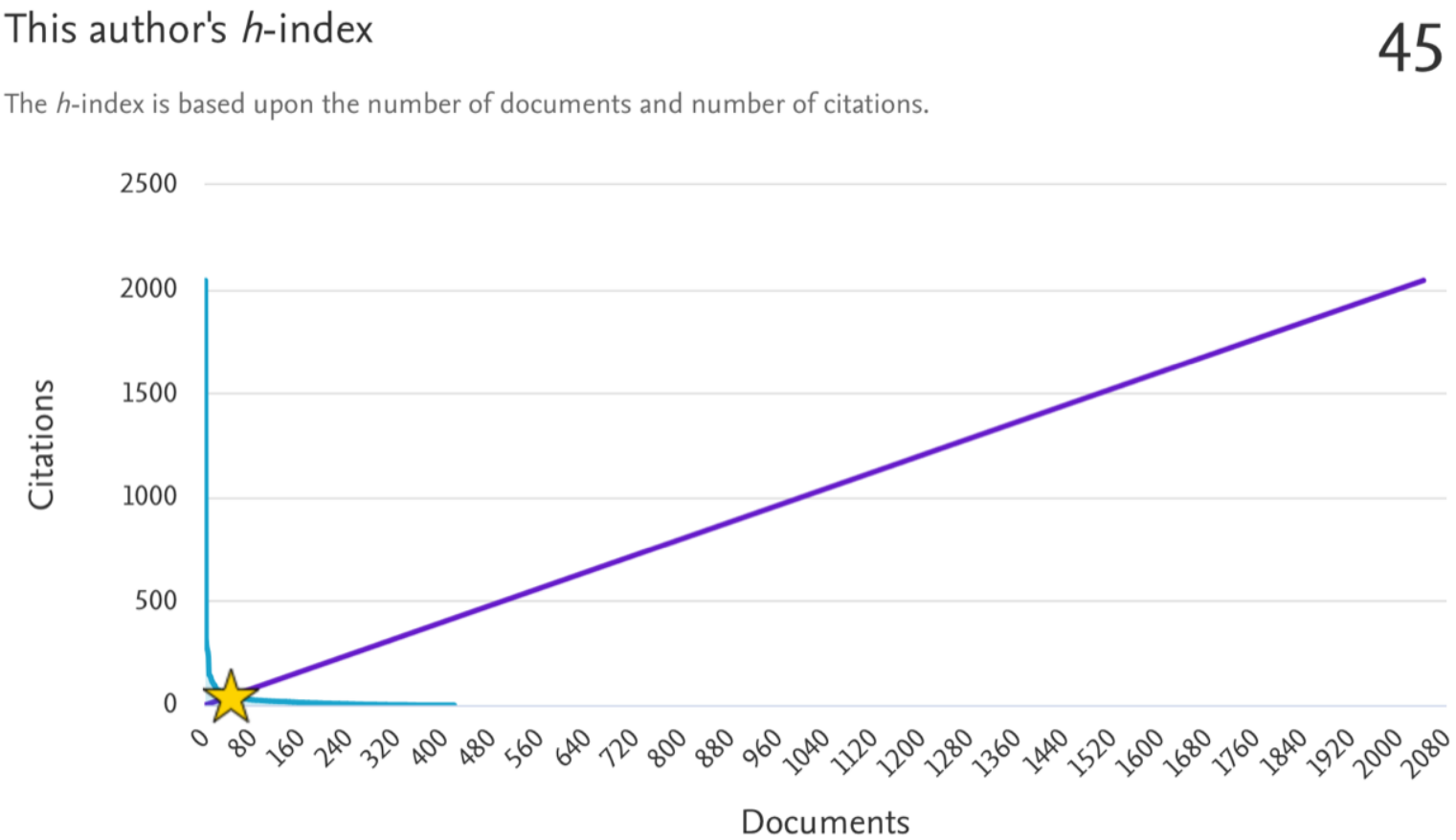


The profiles

An early stage balanced profile



The profiles
A late stage balanced profile



Keeping a balance

Randomness in outcomes and reviewer/evaluator perception

Bet in different horses, but good horses: Technical Reports, Newsletters, Good Workshops and Conferences, Good Journals, Good Implementations, Good Talks. **Luck comes from exposure**

Portfolio – Quality is better than size

- Know your best 5/10 papers.
  - How did they impact science and society?
- Keep trying to replace them by even better ones. Don't burnout, do it for the joy of learning new things

The Social Network

- We converge to the average of those around us
- Its good strategy to be the less bright in the group
- But some nice groups are only by invitation

Groucho Marx quote comes to mind:

I don't want to belong to any club that will accept me as a member

The Social Network

- We converge to the average of those around us
- Its good strategy to be the less bright in the group
- But some nice groups are only by invitation

Groucho Marx quote comes to mind:

I don't want to belong to any club that will accept me as a member

The Social Network

Bootstrapping (or how to find the cool kids)

- Phase: pre-PhD
  - PhD programme: Joint hardship makes bonds
  - Summer schools: Bertinoro, Marktoberdorf
  - Lab visits: *Pick good teams or top places*
- Phase: First PhD decade
  - Advisor role: Invest in students
  - Venues: Prioritize good venues, even if workshops
  - Co-authorship: Diversify and internationalize
- Phase: More senior years
  - Keep nurturing your network
  - Help the next generation