

All You Need is QaL

The Crisis in Academic Publishing: Symptoms, Diagnoses, and a Potential Solution Concept



Karl T. Ulrich

(Joint with Gérard Cachon and Christian Terwiesch)

July 9, 2026

OID Department Seminar

Published Articles are the Coin of the Realm

- Unit of "work" in academia
- Vary widely in embodied effort
- "Refereed" journal an important categorization
- "A" journal is the most important categorization
- "Quality" matters, but assessing quality is tricky

PROSPECT THEORY: AN ANALYSIS OF DECISION UNDER RISK

DANIEL KAHNEMAN; AMOS TVERSKY

Econometrica (pre-1986); Mar 1979; 47, 2; ABI/INFORM Global
pg. 263

ECONOMETRICA

VOLUME 47

MARCH, 1979

NUMBER 2

PROSPECT THEORY: AN ANALYSIS OF DECISION UNDER RISK

BY DANIEL KAHNEMAN AND AMOS TVERSKY¹

This paper presents a critique of expected utility theory as a descriptive model of decision making under risk, and develops an alternative model, called prospect theory. Choices among risky prospects exhibit several pervasive effects that are inconsistent with the basic tenets of utility theory. In particular, people underweight outcomes that are merely probable in comparison with outcomes that are obtained with certainty. This tendency, called the certainty effect, contributes to risk aversion in choices involving sure gains and to risk seeking in choices involving sure losses. In addition, people generally discard components that are shared by all prospects under consideration. This tendency, called the isolation effect, leads to inconsistent preferences when the same choice is presented in different forms. An alternative theory of choice is developed, in which value is assigned to gains and losses rather than to final assets and in which probabilities are replaced by decision weights. The value function is normally concave for gains, commonly convex for losses, and is generally steeper for losses than for gains. Decision weights are generally lower than the corresponding probabilities, except in the range of low probabilities. Overweighting of low probabilities may contribute to the attractiveness of both insurance and gambling.

1. INTRODUCTION

EXPECTED UTILITY THEORY has dominated the analysis of decision making under risk. It has been generally accepted as a normative model of rational choice [24], and widely applied as a descriptive model of economic behavior, e.g. [15, 4]. Thus, it is assumed that all reasonable people would wish to obey the axioms of the theory [47, 36], and that most people actually do, most of the time.

The present paper describes several classes of choice problems in which preferences systematically violate the axioms of expected utility theory. In the light of these observations we argue that utility theory, as it is commonly interpreted and applied, is not an adequate descriptive model and we propose an alternative account of choice under risk.

What are the stakeholder needs for an academic publishing system?

Credibly **certify correctness**.

Provide reliable **evidence of importance** and impact.

} “Quality”

Minimize the **calendar time** from creation of work to "refereed publication" CV status.

Minimize **unproductive author and community effort**.

Engage the community in **constructive feedback**.

Embrace revision over time with a transparent history and serving the latest version.

Exhibit grace towards low-citation papers and authors (the vast majority of the field).

Enable discovery, search, and retrieval.

Operate in an **economically sustainable** way.

A cursory tour of academic forums (Hacker News, r/AskAcademia, r/professors) turns up every one of these, in far more colorful form. On what the community wants: Mulligan, Hall & Raphael (2013), JASIST 64(1):132–161; Ithaka S+R US Faculty Survey (2021).

What are the stakeholder needs for an academic publishing system?

- B+ Credibly **certify correctness**.
- D Provide reliable **evidence of importance** and impact.
- D- Minimize the **calendar time** from creation of work to "refereed publication" CV status.
- C Minimize **unproductive author and community effort**.

- D Engage the community in **constructive feedback**.
- F **Embrace revision** over time with a transparent history and serving the latest version.
- B- **Exhibit grace** towards low-citation papers and authors (the vast majority of the field).
- C **Enable discovery**, search, and retrieval.
- B+ Operate in an **economically sustainable** way.



Grading Existing Journal
System in OIDD Fields

A cursory tour of academic forums (Hacker News, r/AskAcademia, r/professors) turns up every one of these, in far more colorful form. On what the community wants: Mulligan, Hall & Raphael (2013), JASIST 64(1):132–161; Ithaka S+R US Faculty Survey (2021).

What We Mean by Quality

- Many dimensions of quality: interestingness, importance, impact, novelty, elegance, humor, clarity, rigor, popularity, methodological purity, and so on.
- For simplicity, we collapse them to a single scalar Q , the eventual recognized value of the work; correctness and importance are inputs to Q , not separately certified axes.
- Ex ante uncertainty about Q is high.
- Review effort narrows that uncertainty only modestly, and possibly mostly on the dimension of correctness; time reveals most of it, particularly on the importance dimension.

Cachon, Girotra & Netessine (2020), "Interesting, Important, and Impactful Operations Management," M&SOM 22(1)

nature

[Explore content](#) ▾ [About the journal](#) ▾ [Publish with us](#) ▾ [Subscribe](#)

[nature](#) > [letters](#) > article

Letter | Published: 04 June 1998

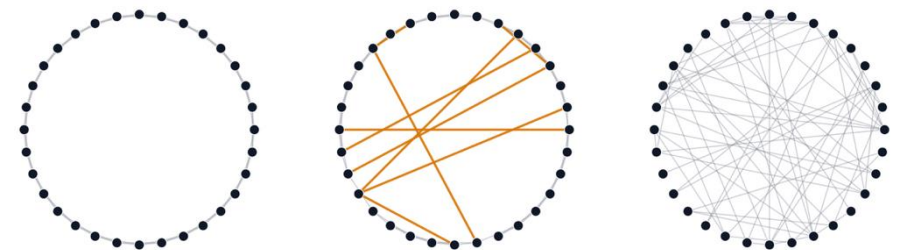
Collective dynamics of 'small-world' networks

[Duncan J. Watts](#) ✉ & [Steven H. Strogatz](#)

[Nature](#) **393**, 440–442 (1998) | [Cite this article](#)

[Save article](#)

225k Accesses | 36k Citations | 643 Altmetric | [Metrics](#)



A Metric for Q: Authority-Weighted Citations (essentially "PageRank")

False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant

[JP Simmons](#), [LD Nelson](#)... - Psychological ..., 2011 - journals.sagepub.com

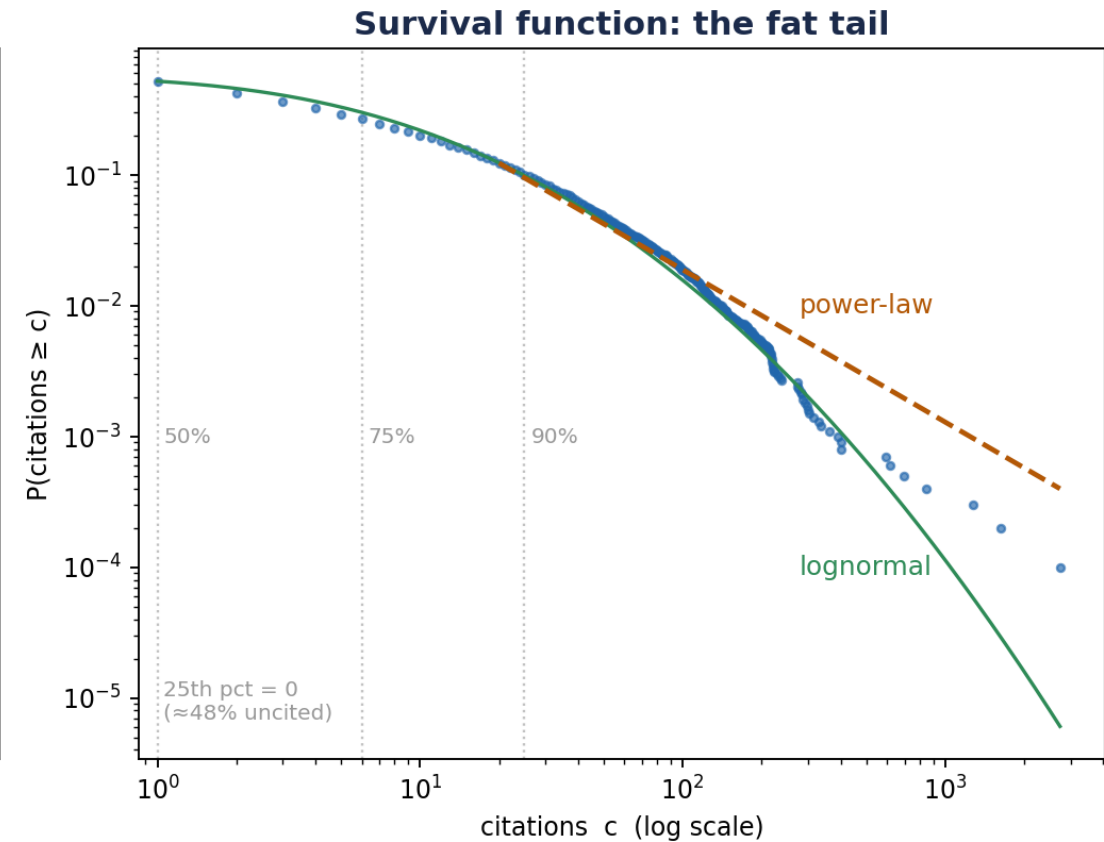
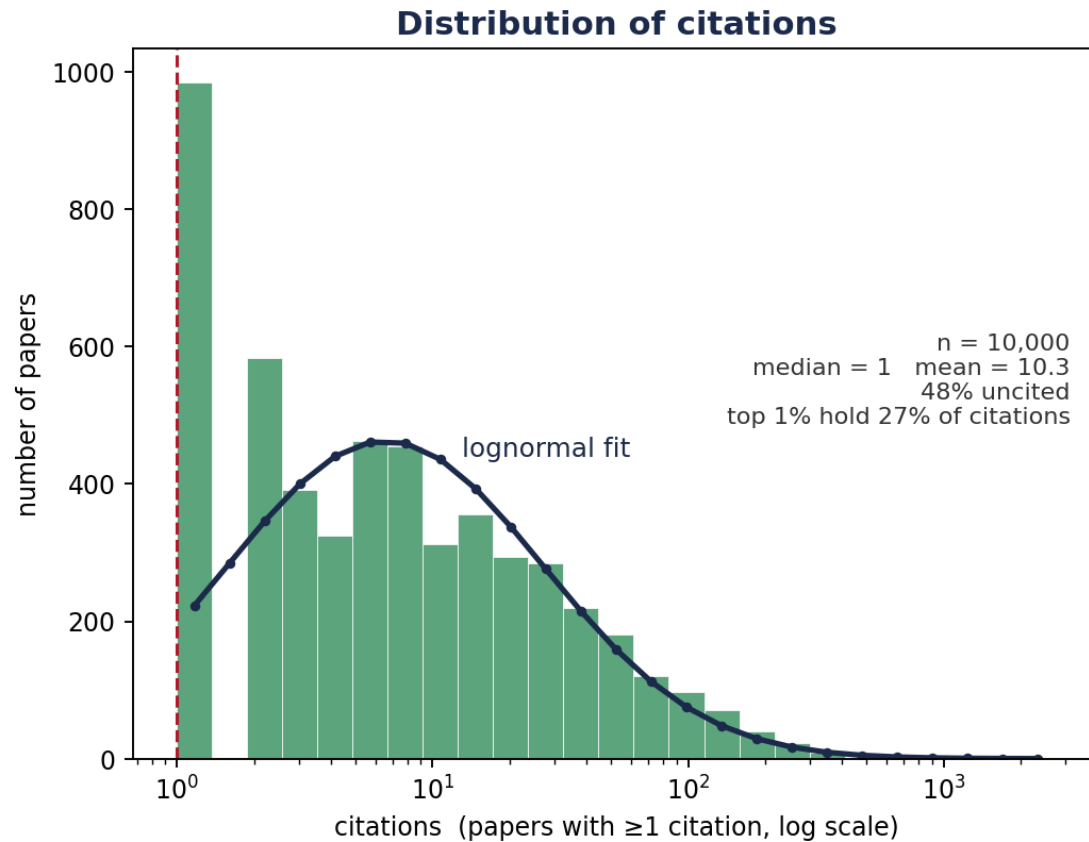
... Finally, a field known for publishing **false positives** ... **false-positive** rate of 5% (ie, $p \leq .05$), current standards for disclosing details of data collection and analyses make **false positives** ...

☆ Save  Cite **Cited by 9374** [Related articles](#) [All 50 versions](#)

- In theory, could estimate Q from a wealth of independent signals, which accrue over time, including views, citations, mentions, downloads.
- The more independent the signals, the less the estimate of Q can be gamed.
- *Citations* are well understood and reasonably well behaved statistically; authority weighting is a fairly easy extension.
- In the long run, hard to imagine a paper being important and having impact if not cited.
- Citations already among most widely used metrics for Q in hiring and promotion
- (Of course there are edge cases where citations do not reflect Q, e.g., a paper with a widely cited methodological blunder.)

The Distribution of Q Is Heavy-Tailed (and the Mode is 0)

Citation distribution, Management Science & Operations Research (OpenAlex, all journals · published 2015, cited through 2026 (~11-yr window) · n=10,000 sample)



Lognormal body with a power-law upper tail (cf. Radicchi et al. 2008; Brzezinski 2015; Thelwall 2016).

Quality Q, proxied by citations.

Management Science & OR, OpenAlex (all journals), published 2015, cited through 2026; n=10,000. ~48% uncited, median 1, yet the top 1% hold 27% of all citations. Psychology is essentially identical (median 0, ~53% uncited, $\sigma \approx 1.5$, top 1% ≈ 26%): citation distributions are universal across fields.

Radicchi, Fortunato & Castellano (2008), PNAS 105(45):17268–17272; Brzezinski (2015), Scientometrics 103(1):213–228; Thelwall (2016), Journal of Informetrics 10(2):336–346.

Reporting Q as a Percentile Class

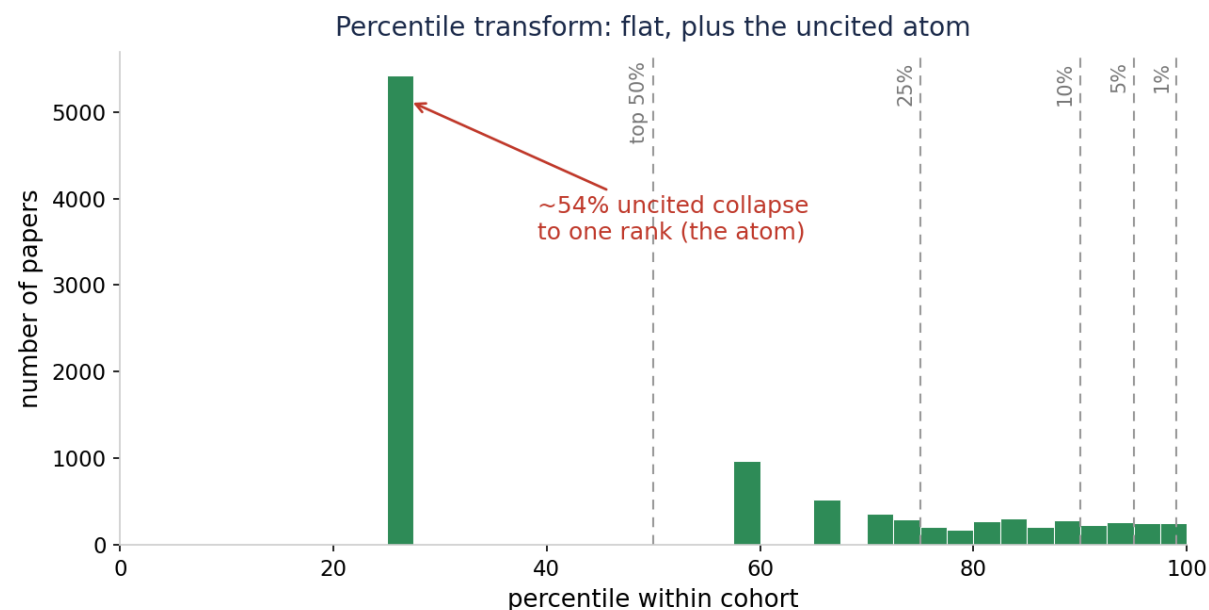
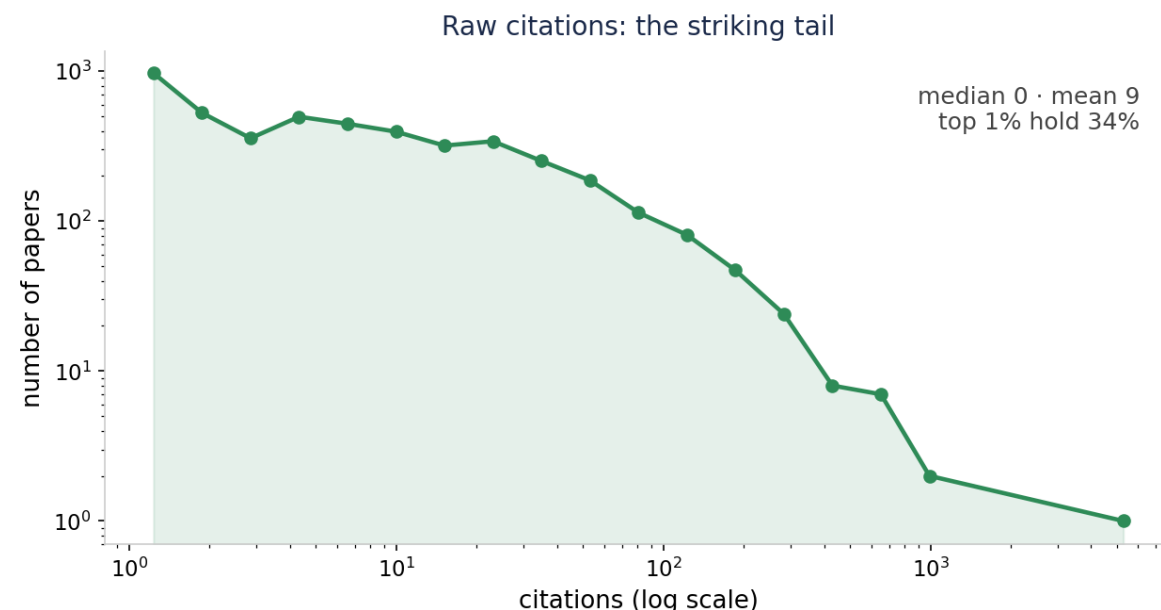
- Raw citation counts are heavy-tailed and hard to compare across fields and years.
- Can report each paper in terms of percentile within its own (sub)field and publication year.
- Report membership in classes: top 1%, 5%, 10%, 25%, 50%.
- These are the NSF / Leiden percentile classes, a standard in research evaluation.
- About half of all papers sit at or near zero citations: a mass at the bottom.
- A percentile is intuitive, field-fair, and robust to the tail.

Percentile classes: National Science Board, Science & Engineering Indicators (NSB-2025-7); CWTS Leiden Ranking.

Parameter-free alternative: Glänzel & Schubert (1988), "Characteristic Scores and Scales in Assessing Citation Impact," Journal of Information Science 14(2):123–127.

The transform is flat only where the data is continuous. The heavy ties at low citation counts (0, 1, 2, 3 ...) are the exception — each is a lump. This isn't a plotting artifact; it's the honest behavior, and it's the same reason the QaL gives all uncited papers one percentile and low-cited papers cluster: papers you truly can't tell apart get the same rank.

Same data, two views — Management Science & OR, 2015 cohort (n = 10,000)



Ranking within a (sub)field-and-year cohort turns the heavy tail into a uniform target; the only lump is the uncited atom. Bands are the NSF percentile classes (top 50 / 25 / 10 / 5 / 1%).

Academic Publishing - The Job to be Done

- **Administration:** time stamp, copy edit, typeset, archive, distribute
- **Development:** peer feedback and guidance for revision
- **Promotion:** communicating existence to the relevant audience
- **Certification of quality:** branding the work as a signal of quality

- Historically, *journals* performed all of these functions
- Most functions (except development) were completed nearly simultaneously, around the time of "publication"
- Today the functions are increasingly unbundled (SSRN, arXiv, personal sites, social media, DIY publication-quality document prep, LLM copy editing)
- Administration is now largely irrelevant (PDF downloads, SSRN, open access)

The functions of scholarly communication were canonically identified as registration, certification, awareness, and archiving by Roosendaal & Geurts (1997), "Forces and Functions in Scientific Communication"; the case for unbundling (decoupling) them into independent services was made by Priem & Hemminger (2012), *Frontiers in Computational Neuroscience* 6:19.

SPRINGER NATURE

Search

Authors

Publish an article

Publish a book

Publish conference proceedings

< Publish your research

Publish an article

Join thousands of researchers worldwide that have published their work in one of our 3,000+ Springer Nature journals.

Step-by-step guide to article publishing

- > 1. Prepare your article
- > 2. Write a cover letter
- > 3. Submit your manuscript
- > 4. Technical check
- > 5. Editor and peer review

The one job journals claim to still do (certification of quality) can't really be done, for two reasons

1. The instrument is noisy

Inter-rater reliability is low (~ 0.34):
2–3 reviewers give a verdict little better than a guess.

Reliability rises only slowly with more judges; you would need >10 to dependably classify a marginal paper.

Fixable in principle: add judges (Spearman–Brown).

2. The target is mostly exogenous

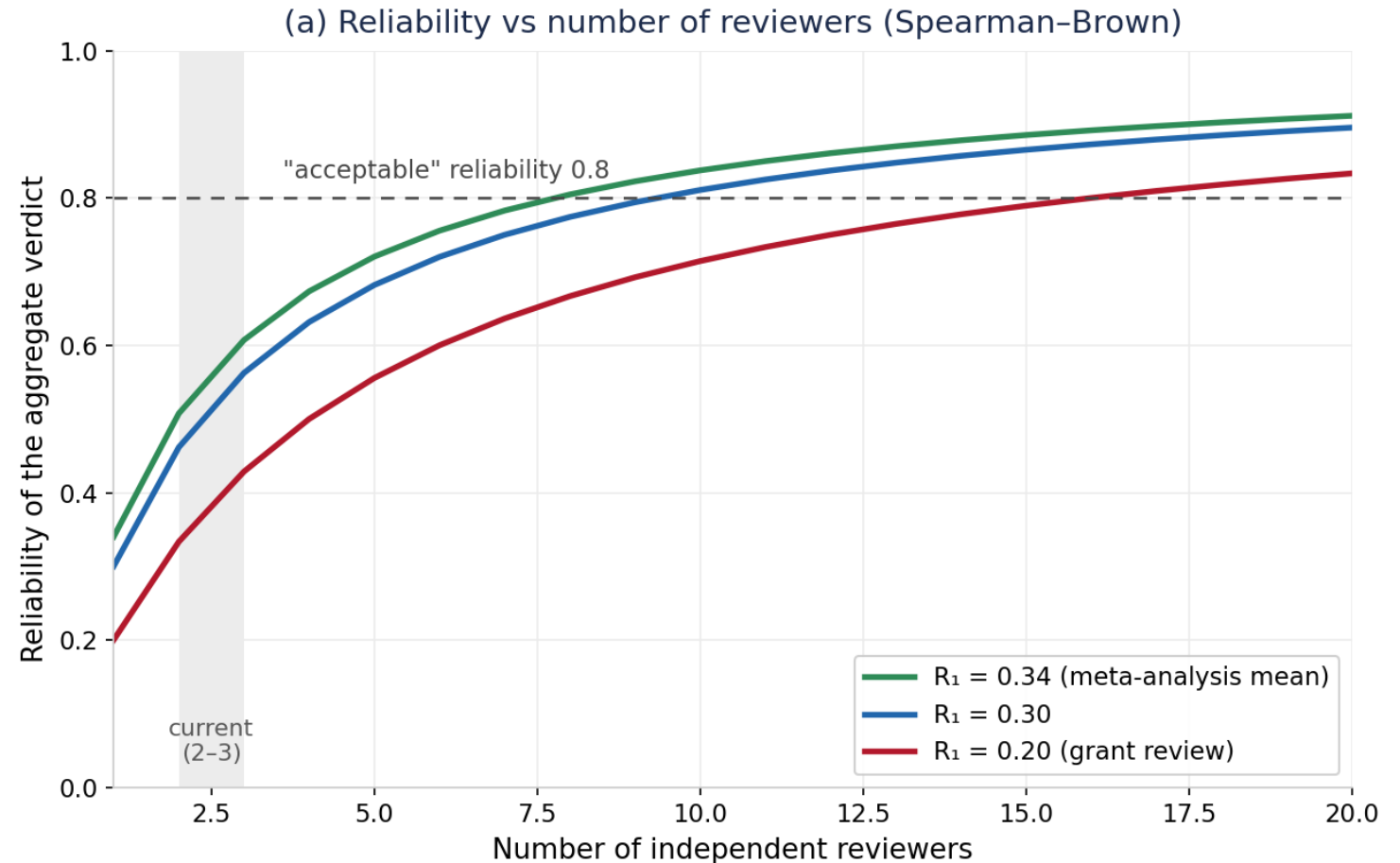
Even a perfect read on a paper's knowable quality predicts little of its eventual Q.

Most of what determines eventual Q (timing, attention, who cites whom) is realized after the decision.

Not fixable by any number of judges.

Reason 1 limits any small panel.

Reason 2 limits any ex-ante process.



A Simple Model, with Evidence for Each Parameter

Eventual quality decomposes into a knowable component plus an exogenous shock; each reviewer sees the knowable component through noise:

$$Q = q_0 + u \cdot s_i = q_0 + e_i$$

Parameter	Value	Evidence
Reviewer noise (reliability of one review)	ICC $\approx$ 0.34	Bornmann, Mutz & Daniel (2010), <i>PLoS ONE</i> 5(12):e14331 (48 studies, 19,443 manuscripts); Cicchetti (1991), <i>Behavioral & Brain Sciences</i> 14(1):119–186; NeurlPS revisit, ~50% of scoring subjective — Cortes & Lawrence (2021), arXiv:2109.09774
Validity ceiling (knowable quality $\rightarrow$ eventual Q)	R ² $\approx$ 0.06	Success in cumulative-advantage markets is intrinsically unpredictable — Salganik, Dodds & Watts (2006), <i>Science</i> 311:854–856; review scores do not predict citations of accepted papers — Cortes & Lawrence (2021); impact often realized years later ("sleeping beauties") — Ke, Ferrara, Radicchi & Flammini (2015), <i>PNAS</i> 112(24):7426–7431

q_0 = the knowable (ex ante) component of quality; u = the exogenous shock realized after the decision, so eventual quality $Q = q_0 + u$; e_i = reviewer error. A review recovers at best q_0 ; even perfectly, it explains only about 6% of Q . Outcomes are binned into the NSF percentile classes; the validity ceiling is sourced from the publishing literature, not from our own idea-evaluation work.

The Result: More Reviewers Barely Help

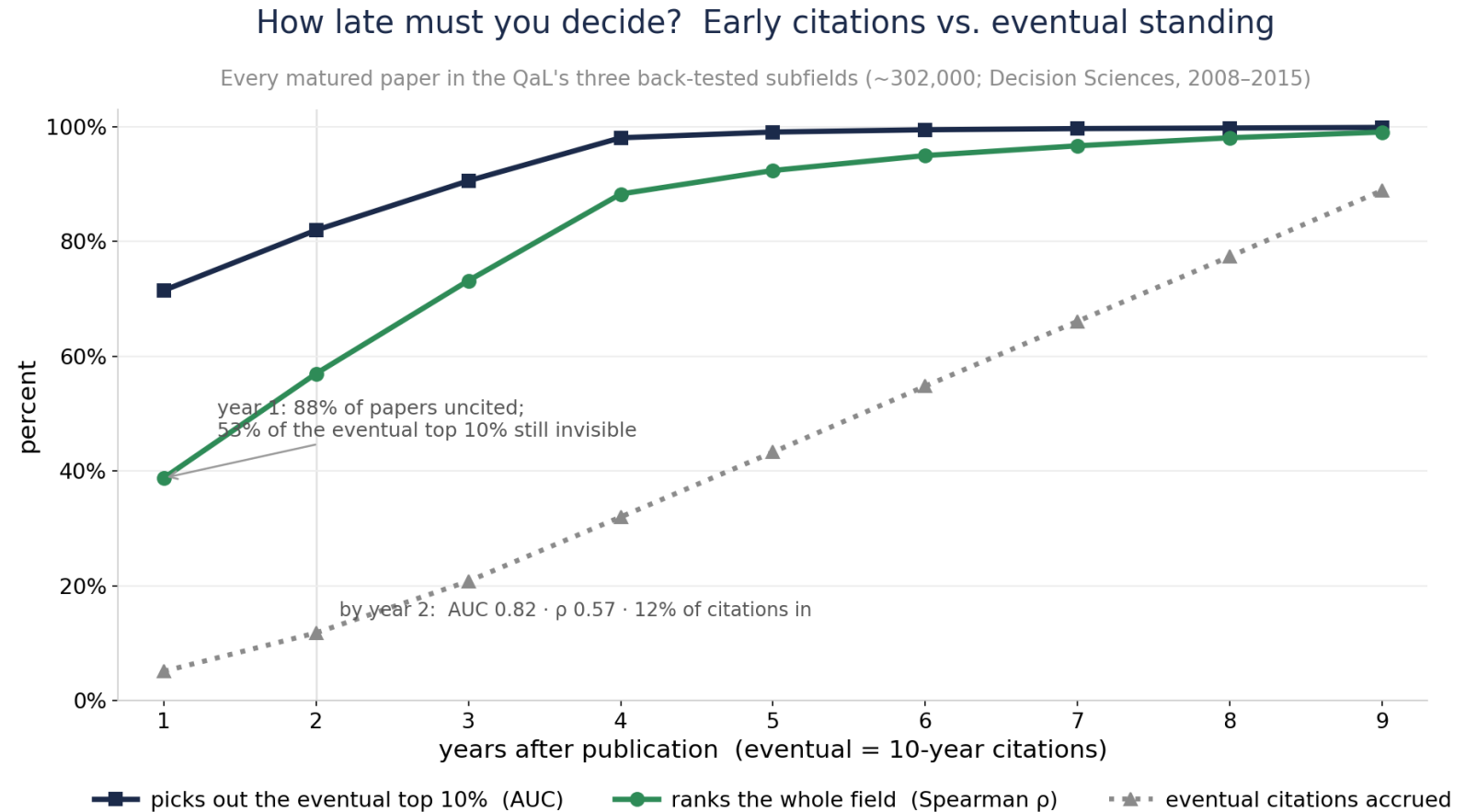
Probability of correctly identifying	By chance	3 reviewers	10 reviewers	Perfect read of q_0
a top 1% paper	1%	3%	4%	4%
a top 10% paper	10%	17%	18%	19%
a top 25% paper	25%	33%	35%	36%
a top 50% paper	50%	56%	57%	58%



Probability a paper that truly belongs in a tier is placed there, under the model above (ICC 0.34; validity ceiling R^2 0.06). The last column is a perfect (noise-free) read of q_0 — the knowable ex-ante quality defined on the previous slide ($Q = q_0 + u$) — not foreknowledge of the outcome; even then a top-1% paper is tagged only ~4%, because ~94% of impact is exogenous. Going from 3 to 10 reviewers barely moves the numbers: the validity ceiling, not reviewer count, is the binding constraint.

Early Citations Predict Eventual Rank

- Sample of 302k papers (three Decision-Sciences subfields, 2008–2015; eventual = the 10-year percentile)
- Top 10% in year N as a predictor of eventual top 10%: AUC 0.82 by year 2, 0.98 by year 4. *Even 2-year citations likely much better than 3 reviewers ex ante.*
- The atom at zero (88% uncited at year 1) flatters a plain rank correlation (Spearman ρ 0.39) and floors the tie-corrected one (Kendall τ -b 0.20); AUC (area under curve for binary classifier) ignores the uncited mass and is the best calibration metric.
- Number of citations matures slowly: only 12% of eventual citations are in by year 2, 43% by year 5.



An Alternative Approach: The Academic Ledger

- A neutral, non-profit system of record applies a light integrity screen immediately and enters the work in a ledger, without judging quality.
- Evidence of Q accumulates over time, and the branding of quality occurs with some delay, when confidence in the estimate of Q is high.
- Legacy journals co-exist and may remain important for branding and curation around topics, but the Ledger is the source of truth for Q.
- Alternatives to conventional journals may emerge, built on top of the Ledger to deliver other benefits (e.g., engagement with a community, improvement of the work, focused curation around topics).

aL

academicLedger

BETA

Explore

Author

About

For authors

Talk

Signed in as Karl Ulrich · ORCID verified

Sign out

Explore the Ledger

Search and rank records by QaL within a stated reference class and vintage. Leaderboards are just this view with a fixed filter and a QaL sort.

SEARCH TITLE OR AUTHOR

FIELD

VINTAGE (SINCE)

e.g. bandit, Cachon, Terwiesch

All fields

Any

☐ Calibrated only

LEADERBOARDS

Top QaL · MS&OR

Top QaL · Info Systems

Top QaL · Decision Sciences

SORT

QaL · synthetic ★

QaL · field

Most cited

Newest

Both reference classes are shown: QaL · field (the single OpenAlex subfield & vintage) and QaL · synthetic ★ (the official class — the paper ranked against a recency-weighted blend of its true intellectual community's cohorts; computed for the prefilled set, "—" until computed). Click any column to re-sort; the official number stays the synthetic field. Divergence between the two flags a field-sensitive paper and is itself information.

500+ records · sorted by QaL · field ↓

#	TITLE	FIELDS	YR	CITES	QaL · FIELD ▼	QaL · SYNTHETIC ★
1	Grit: Perseverance and passion for long-term goals. Angela Duckworth, Christopher Peterson, Michael D. Matthews, Dennis R. Kelly · Journal of Personality and Social Psychology · source	Clin Psych 23% · +7 more	2007	6,770	100 [99–100]	100 [99–100]
2	Choices, Values, and Frames Daniel Kahneman, Amos Tversky · World Scientific handbook in financial economic series · source	Dec Sci 45% · +7 more	2013	5,410	100 [99–100]	100 [98.9–100]
3	Random graphs with arbitrary degree distributions and their applications M. E. J. Newman, Steven H. Strogatz, Duncan J. Watts · Physical review. E, Statistical physics, plasmas, fluids, and related interdisciplinary topics · OA · source	Stat & Nonl Phys 35% · +7 more	2001	3,721	100 [97.5–100]	100 [99–100]
4	Beyond Computation: Information Technology, Organizational Transformation and Business Performance Erik Brynjolfsson, Lorin M. Hitt · The Journal of Economic Perspectives · OA · source	Econ 50% · +7 more	2000	3,221	100 [97.5–100]	100 [99–100]

Wharton

Ulrich

14

Beta Build

- Built using OpenAlex as the source data.
- Estimates QaL for virtually any paper in any field (N ~ 500M)
- Basic approach:
 - Construct a synthetic subfield for any paper (using hierarchy of methods depending on paper age)
 - Locate the paper in the synthetic subfield in percentile rank relative to subfield-year cohort.
 - Percentile rank in each subfield is exact...an observation. Combined by weighting subfields.
 - Provide 90 percent confidence interval of eventual QaL (calibrated by observing realized outcomes of similar vintage-subfield cohorts)

Serving Democracy: Evidence of Voting Resource Disparity in Florida

Gérard P. Cachon, Dawson Kaaua

Management Science · PoliSci & IR 50% · +7 more · Posted 2022 · 4 years on the record

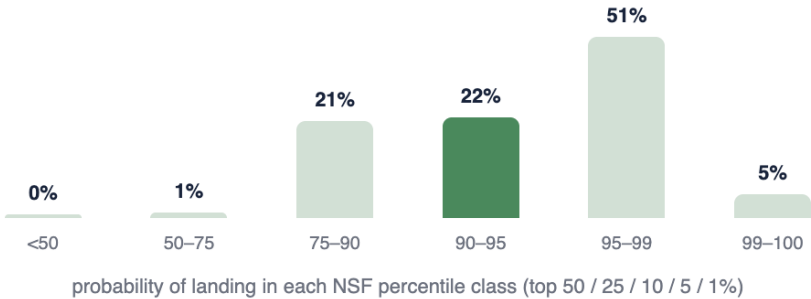
QUALITY ON THE ACADEMIC LEDGER

QaL

89 **94** 98

eventual percentile · 90% interval · preliminary — this community isn't back-tested yet

Reference class: synthetic field · official headline recency-weighted blend of its community's cohorts · preliminary · 34% back-tested · field pct 97.79



Decided late: the interval is wide early and narrows as evidence accrues.

SYNTHETIC FIELD · WEIGHTED BY RECENT REFERENCES

Political Science and International Relations (50%) · Management Information Systems (15%) · Artificial Intelligence (10%) · Sociology and Political Science (8%) · Public Administration (6%) · Transportation (5%) · Economics and Econometrics (4%) · +1 more (1%)

The official reference class is this recency-weighted blend of the communities the paper actually draws on (QaL_spec \$5), not OpenAlex's single label ("Social Sciences"). Divergence is the signal.

EVIDENCE BEHIND THE ESTIMATE — pulled from OpenAlex

13 Citations (total)	94.6% Observed percentile (synthetic field)	None Retractions / corrections
4 yr Age since posting	Social Sciences Detected field (OpenAlex)	

Computing QaL: A Worked Example

Subfield in the synthetic field	Weight	Pctile	Contribution
Political Science & International Relations	37.7%	97.8	36.9
Management Information Systems	15.3%	91.0	13.9
Economics & Econometrics	10.9%	93.7	10.2
Artificial Intelligence	8.2%	86.7	7.1
Sociology & Political Science	8.0%	95.2	7.6
Strategy & Management	7.3%	90.8	6.6
Public Administration	5.2%	93.0	4.8
Transportation	4.4%	88.6	3.9
Marketing	3.0%	86.2	2.6
Total	100%		93.7

Contribution = weight × percentile; the paper is ranked within each subfield's 2022 cohort. The subfields disagree (86th–98th) — averaging them is what pulls the blended percentile down to 93.7.

Serving Democracy — voting-resource disparity in Florida, Cachon and Kaaua (2022); 13 citations.

OpenAlex tags it **one** subfield (*Political Science & International Relations*). QaL instead ranks it in the **synthetic field** — the 9 subfields its references and co-citations actually span.

Observed = Σ (weight × percentile) = **93.7**

Extrapolate to eventual (its percentile at the 10-year horizon) — it's only 4 years old, so each calibrated subfield maps observed → eventual, blended over the 95% of weight that's calibrated:

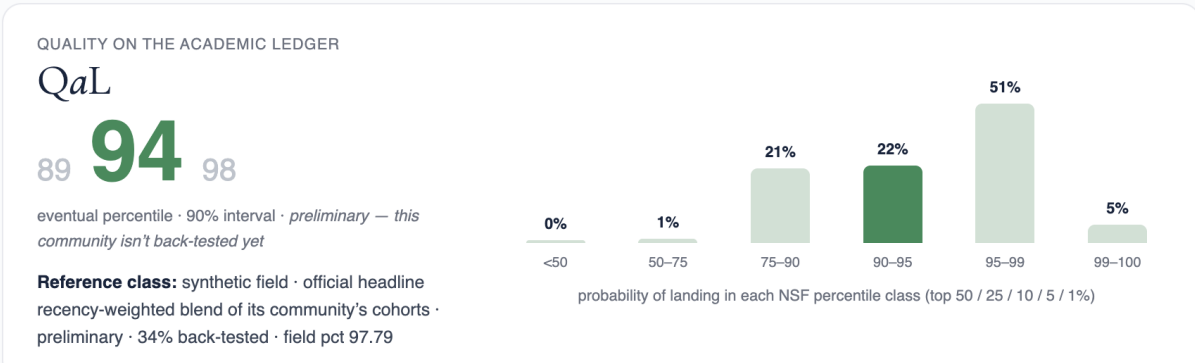
QaL 93 · 90% CI [88, 97]

a likely top-10% paper — the eventual standing is a genuine **forecast**, with an interval that **narrows as the paper matures**, not just today's citation count.

Serving Democracy: Evidence of Voting Resource Disparity in Florida

Gérard P. Cachon, Dawson Kaaua

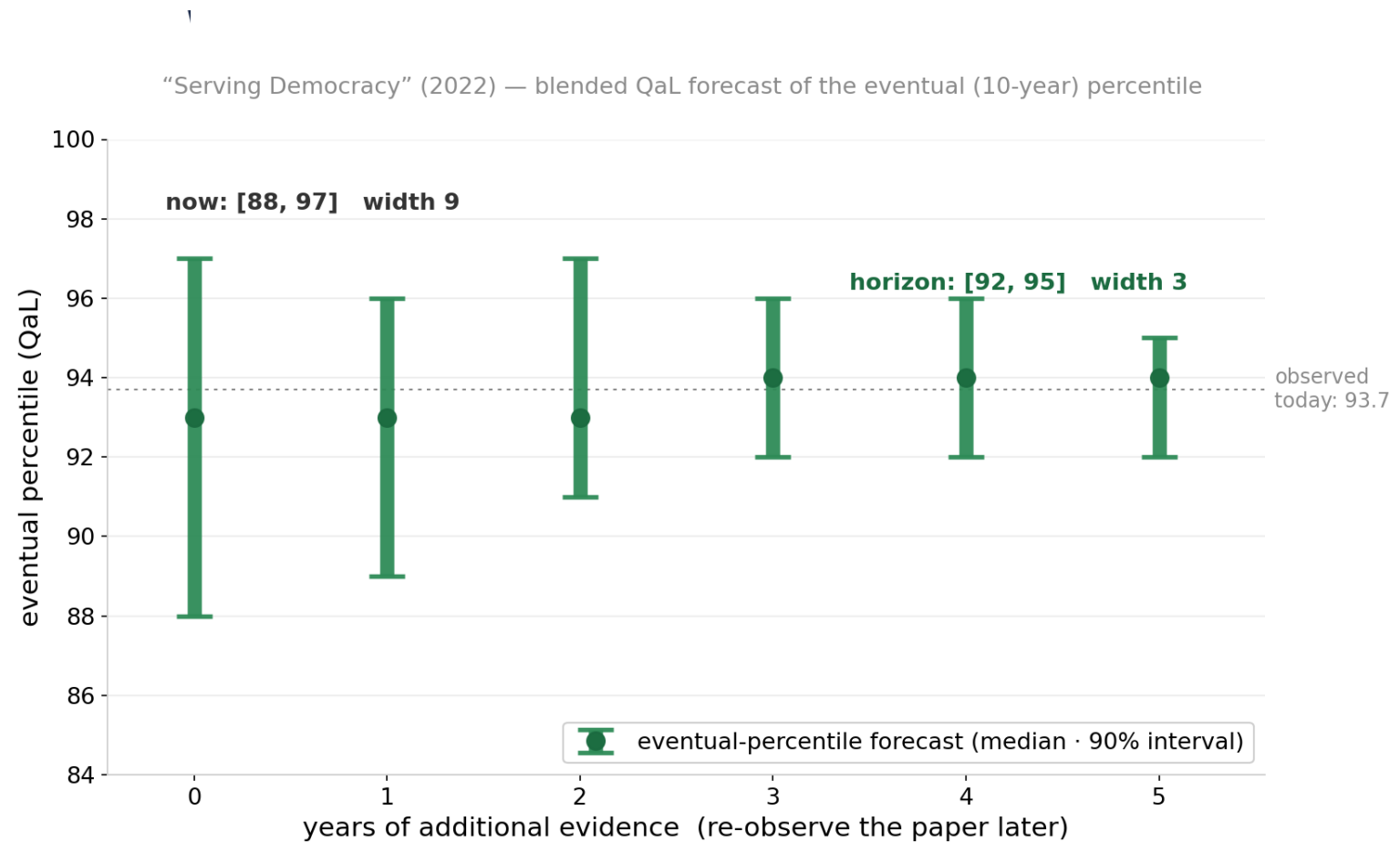
Management Science · PoliSci & IR 50% · +7 more · Posted 2022 · 4 years on the record



OpenAlex taxonomy:
<https://help.openalex.org/hc/en-us/articles/24736129405719-Topics>

Forecasting from Now to the Horizon

- A young paper's cohort isn't mature; citations are still accruing, so today's rank isn't its final one.
- Calibration learned from communities we have watched to maturity maps observed percentile + age $\rightarrow$ a distribution over the eventual percentile.
- The 90% interval narrows as the paper ages toward the 10-year horizon, $[88, 97] \rightarrow [92, 95]$.
- That shrinking uncertainty is the case for delaying assessment of Q: the eventual standing is a forecast you can sharpen by waiting.
- Still, the estimate is pretty good after a few citations.



Ideas are Dimes a Dozen: Large Language Models for Idea Generation in Innovation

Karan Girotra, Lennart Meincke, Christian Terwiesch, Karl T. Ulrich

SSRN Electronic Journal · Exp & Cog Psych 21% · +7 more · Posted 2023 · 3 years on the record

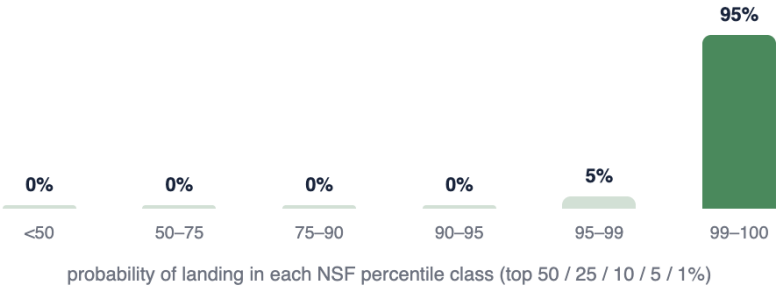
QUALITY ON THE ACADEMIC LEDGER

QaL

97.3 **99.8** 100

eventual percentile · point estimate with 90% interval · illustrative

Reference class: synthetic field · official headline
recency-weighted blend of its community's cohorts ·
74% back-tested · field pct 99.89



Decided late: the interval is wide early and narrows as evidence accrues.

SYNTHETIC FIELD · WEIGHTED BY RECENT REFERENCES

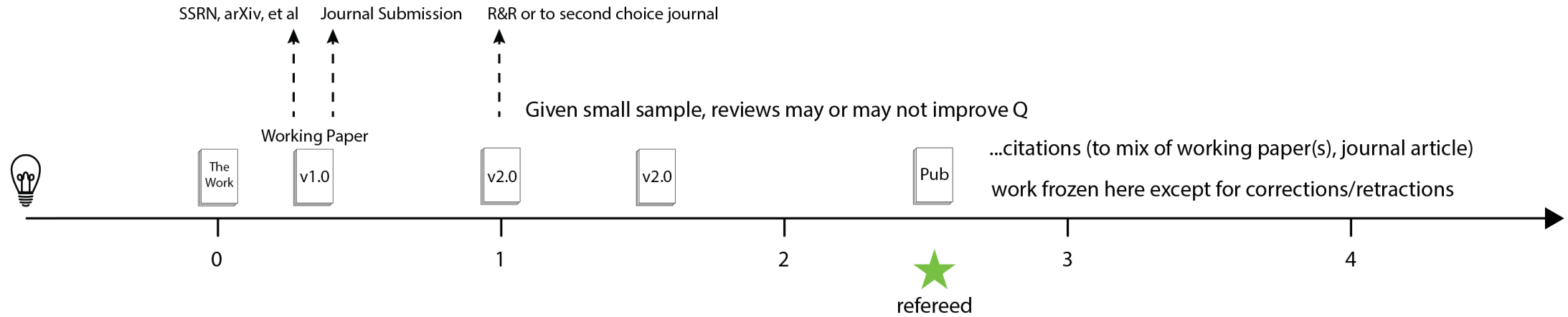
Experimental and Cognitive Psychology (21%) · Artificial Intelligence (17%) · Sociology and Political Science (11%) · Health Informatics (10%) · Computer Science Applications (5%) · Economics and Econometrics (5%) · Management Information Systems (4%) · Cognitive Neuroscience (4%)

The official reference class is this recency-weighted blend of the communities the paper actually draws on (QaL_spec §5), not OpenAlex's single label ("Computer Science"). Divergence is the signal.

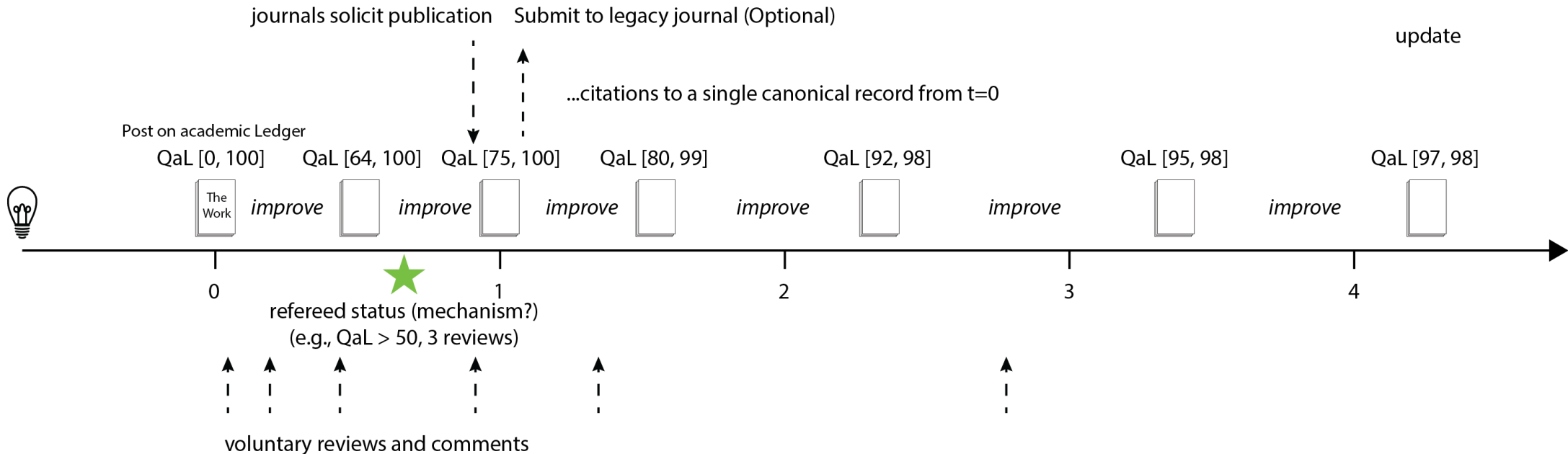
EVIDENCE BEHIND THE ESTIMATE — pulled from OpenAlex

168 Citations (total)	99.51% Observed percentile (synthetic field)	None Retractions / corrections
3 yr Age since posting	Computer Science Detected field (OpenAlex)	

Existing Timeline - Legacy Journals



Possible Timeline - academic Ledger



How Might we Provide "Refereed" Status with Minimal Friction

- "Incidental" voluntary reviews on the ledger.
e.g., signed comments from readers who are already interested organically in the work:
 - (a) what did I most like about this paper?
 - (b) how could the author make this paper better?)
- Existing journals could solicit articles from work posted on the Ledger.
- Emergent virtual journals with volunteer editorial boards for a very specific topic. (I might actually want to be a reviewer for a *Journal of Innovation Processes*.)
- Automated LLM feedback using best current practices, posted with information on the prompt and the model.



Articles in Refereed Journals

- [13] Karl T. Ulrich, "Digital Darwinism: Steering the Evolution of Artificial Life in Socio-Technical Systems," *AI and Ethics* (2026) 6:268. <https://doi.org/10.1007/s43681-026-01057-8>
- [14] Karl T. Ulrich, "Re-Consider the Lobster: Animal Lives in Protein Supply Chains," *Sustainability*, 17, 2025, p. 7034. <https://doi.org/10.3390/su17157034>
- [15] Karl T. Ulrich, "Minimum Spatial Housing Requirements for Human Flourishing," *Buildings*, 15, 2025, p. 2623. <https://doi.org/10.3390/buildings15152623>



Working Papers including Papers under Review or Revision

- [48] Ulrich, Karl T. "The Satoshi Overhang: Why the Bear Case is Bounded." *under review*. <https://doi.org/10.48550/arXiv.2604.27694>
- [49] Ulrich, Karl T. "Why is Skiing Fun? A Compound Superstimulus Theory of Risky Effortful Movement," *under review*.
- [50] Girotra, Karan and Meincke, Lennart and Terwiesch, Christian and Ulrich, Karl T., Ideas are Dimes a Dozen: Large Language Models for Idea Generation in Innovation. Report – Mack Center for Technological Innovation. University of Pennsylvania (July 10, 2023). Available at SSRN: <https://ssrn.com/abstract=4526071>
- [51] Kornish, Laura J. and Karl T. Ulrich, "Assessing the Quality of Selection Processes," under revision for resubmission, March 2017.
- [52] Wooten, J. and K. Ulrich, "The Impact of Visibility in Innovation Tournaments: Evidence from Field Experiments," under revision for resubmission, March 2017.
- [53] Terwiesch, Christian and Karl T. Ulrich, "Will Video Kill the Classroom Star? The Threat and Opportunity of Massively Open On-Line Courses for Full-Time MBA Programs," Report – Mack Center for Technological Innovation. University of Pennsylvania, July 2014.
- [54] Girotra, K. and K. Ulrich, "Empirical Evidence for Domain Name Performance," Working Paper, The Wharton School, May 2011.
- [55] Ulrich, Karl T., "The Environmental Paradox of Bicycling," Working Paper, The Wharton School, July 2006.

Certification of a Paper could Evolve with Time

certified

Cachon, G., C. Terwiesch, and K. Ulrich. 2026. The Crisis in Academic Publishing. *Academic Ledger* [0, 100]. <https://doi.org/10.59312/aldg.2026.0427>

refereed

Cachon, G., C. Terwiesch, and K. Ulrich. 2026. The Crisis in Academic Publishing. *Academic Ledger* [75, 100]. <https://doi.org/10.59312/aldg.2026.0427>

canon

Cachon, G., C. Terwiesch, and K. Ulrich. 2026. The Crisis in Academic Publishing. *Academic Ledger* [95, 100]. <https://doi.org/10.59312/aldg.2026.0427>

P

canon

Cachon, G., C. Terwiesch, and K. Ulrich. 2026. The Crisis in Academic Publishing. *Academic Ledger*. (also appears in *Quantitative Science Studies* 7(2):512–540, 2026). <https://doi.org/10.59312/aldg.2026.0427>

One record, one DOI, one act of authorship: the line is upgraded in place; the circled P flags the journal-published version, with its citation appended.

arXiv:1606.06565v2 [cs.AI] 25 Jul 2016

Concrete Problems in AI Safety

Dario Amodei* Google Brain
Chris Olah* Google Brain
Jacob Steinhardt Stanford University
Paul Christiano UC Berkeley
John Schulman OpenAI
Dan Mané Google Brain

Abstract

Rapid progress in machine learning and artificial intelligence (AI) has brought increasing attention to the potential impacts of AI technologies on society. In this paper we discuss one such potential impact: the problem of accidents in machine learning systems, defined as unintended and harmful behavior that may emerge from poor design of real-world AI systems. We present a list of five practical research problems related to accident risk, categorized according to whether the problem originates from having the wrong objective function ("avoiding side effects" and "avoiding reward hacking"), an objective function that is too expensive to evaluate frequently ("scalable supervision"), or undesirable behavior during the learning process ("safe exploration" and "distributional shift"). We review previous work in these areas as well as suggesting research directions with a focus on relevance to cutting-edge AI systems. Finally, we consider the high-level question of how to think most productively about the safety of forward-looking applications of AI.

1 Introduction

The last few years have seen rapid progress on long-standing, difficult problems in machine learning and artificial intelligence (AI), in areas as diverse as computer vision [82], video game playing [102], autonomous vehicles [86], and Go [140]. These advances have brought excitement about the positive potential for AI to transform medicine [126], science [59], and transportation [86], along with concerns about the privacy [76], security [115], fairness [3], economic [32], and military [16] implications of autonomous systems, as well as concerns about the longer-term implications of powerful AI [27, 167].

The authors believe that AI technologies are likely to be overwhelmingly beneficial for humanity, but we also believe that it is worth giving serious thought to potential challenges and risks. We strongly support work on privacy, security, fairness, economics, and policy, but in this document we discuss another class of problem which we believe is also relevant to the societal impacts of AI: the problem of accidents in machine learning systems. We define accidents as unintended and harmful behavior that may emerge from machine learning systems when we specify the wrong objective function, are

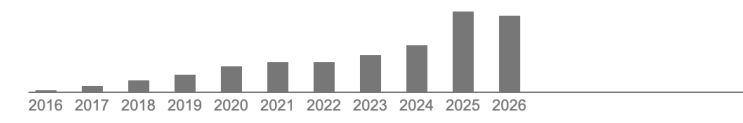
*These authors contributed equally.

Concrete problems in AI safety

Authors Dario Amodei, Chris Olah, Jacob Steinhardt, Paul Christiano, John Schulman, Dan Mané
Publication date 2016/6/21
Journal arXiv preprint arXiv:1606.06565

Description Rapid progress in machine learning and artificial intelligence (AI) has brought increasing attention to the potential impacts of AI technologies on society. In this paper we discuss one such potential impact: the problem of accidents in machine learning systems, defined as unintended and harmful behavior that may emerge from poor design of real-world AI systems. We present a list of five practical research problems related to accident risk, categorized according to whether the problem originates from having the wrong objective function ("avoiding side effects" and "avoiding reward hacking"), an objective function that is too expensive to evaluate frequently ("scalable supervision"), or undesirable behavior during the learning process ("safe exploration" and "distributional shift"). We review previous work in these areas as well as suggesting research directions with a focus on relevance to cutting-edge AI systems. Finally, we consider the high-level question of how to think most productively about the safety of forward-looking applications of AI.

Total citations Cited by 5586



Scholar articles Concrete problems in AI safety
D Amodei, C Olah, J Steinhardt, P Christiano... - arXiv preprint arXiv:1606.06565, 2016
Cited by 5586 Related articles All 6 versions

23

“App Store” for academic Ledger

- A lot of useful tools can be built on the ledger.
 - Reference checker (see demo).
 - Reference recommender.
 - Reviewer finder.
 - Rising stars.
 - Hot spots in subfields.
 - Topic notification service.

← → ↺ 🏠 🔍 academic-ledger.org/check-references ☆ 📄 ⓘ 🔔 📱

aL academicLedger BETA

Explore Check refs Author About For authors Talk

Check my references

Paste a paper's reference list. We resolve each reference against the scholarly record and show its citation impact on the Ledger. Anything we can't confidently match is flagged **check for validity** — a fast scan for fabricated, garbled, or dead citations.

Amodei D, Olah C, Steinhardt J, Christiano P, Schulman J, Mané D (2016) Concrete problems in AI safety. arXiv 1606.06565. <https://doi.org/10.48550/arXiv.1606.06565>

Bedau MA (2003) Artificial life: organization, adaptation, and complexity from the bottom up. Trends in Cognitive Sciences 7(11):505–512. <https://doi.org/10.1016/j.tics.2003.09.012>

Bostrom N (2014) Superintelligence: Paths, Dangers, Strategies. Oxford University Press, Oxford

Boyd R, Richerson PJ (1985) Culture and the Evolutionary Process. University of Chicago Press, Chicago

Brynjolfsson E, McAfee A (2014) The Second Machine Age. W. W. Norton, New York

21

Bryson JJ (2020) The artificial intelligence of the ethics of artificial intelligence: an introductory overview for law and regulation. In: Dubber MD, Pasquale F, Das S (eds) The Oxford Handbook of Ethics of AI. Oxford University Press, Oxford. pp 1–35

Check references

try a sample

Leigh EG (2010) The evolution of mutualism. Journal of Evolutionary Biology 23(12):2507–2528. <https://doi.org/10.1111/j.1420-9101.2010.02114.x>

The evolution of mutualism

Egbert Giles Leigh · 2010 · 225 citations · Journal of Evolutionary Biology · [view on the Ledger →](#)

MacKenzie D (2018) How algorithms interact: financial models and the markets they shape. Theory, Culture & Society 35(6):39–66. <https://doi.org/10.1177/0263276418757017>

Performing Theory?

Donald MacKenzie · 2006 · 4 citations · The MIT Press eBooks · [view on the Ledger →](#)

Maeda H, Quan-Haase A (2024) Chatting alone together: parasocial bonds with conversational agents. New Media & Society 26(1):45–64. <https://doi.org/10.1177/14614448221102900>

⚠ cited DOI belongs to a different paper — likely fabricated (resolves to: "Parasocial interactions with real and virtual influencers: The role of perceived similarity and human-likeness")

MAS (2024) Project Guardian: Policy Learnings from Tokenised-Assets Pilots. Monetary Authority of Singapore, Singapore

Monetary Policy and Reserve Management

Monetary Authority of Singapore · 2021 · 0 citations · [view on the Ledger →](#)

Margulis L (1998) Symbiotic Planet. Basic Books, New York

⚠ no result found — check for validity (closest, rejected: "Ernst Mayr: What evolution is")

Maynard Smith J, Szathmáry E (1995) The Major Transitions in Evolution. Oxford University Press, Oxford

The Major Transitions in Evolution

John Maynard Smith, Eörs Szathmáry · 1997 · 2,598 citations · Oxford University Press eBooks · [view on the Ledger →](#)

Mendoza J, Sun G, Barrett M (2023) Strategic foresight for socio-technical systems. Technological Forecasting and Social Change 191:122515. <https://doi.org/10.1016/j.techfore.2023.122515>

⚠ cited DOI belongs to a different paper — likely fabricated (resolves to: "Data sharing for business model innovation in platform ecosystems: From private data to public good")

Ulrich

24

The Legacy Journal Perspective

- The Ledger is consistent with the policies of most academic journals in social science.
- Journals could encourage authors to first publish on the Ledger, reducing their volume of low-quality submissions.
- The alternative to a hard reject could become "develop the work on the Ledger."
- Editors could proactively scan the Ledger for work to invite, evaluating fit and quality based on evidence.



Image: M. Eric Johnson

Utopia

- Administration is smooth and fast: register, screen integrity, archive, timestamp, and disseminate within days.
- True signals of quality emerge smoothly and continuously as the field reads, uses, cites, and endorses; those signals are used for critical decisions in universities.
- Authors can respond, adapt, and improve the work, with the revision history on the record.
- Incentive is to create interesting and impactful work not to fit the norms of a community.
- Journals become nice-to-have curation; the revealed evidence is what matters for reputation and for evaluating scholarship.
- A more predictable, less capricious, less arbitrary world for emerging scholars.
- Senior scholars with no patience for journals may opt out of the legacy system and still "publish."



Could existing efforts and institutions achieve most of the aims of the Ledger?

- **Google Scholar.** “could” but probably not a Google priority. May not remain neutral.
- **OpenAlex.** “could” but may not have sufficient academic standing. Resources thin. (Still, probably the best prospect for doing so.)
- **arXiv.** “could” but resources stretched; focus on CS, QFin, AI, math.
- **SSRN.** “could” but Elsevier is the owner and is conflicted.
- **Clarivate** et al. Too conflicted and sluggish.



Articles

About 7,220,000 results (0.41 sec)

Any time

Since 2026

Since 2025

Since 2022

Custom range...

Sort by relevance

Sort by date

Any type

Review articles

☐ include patents

☒ include citations

☒ Create alert

An overview of innovation

SJ Kline, N Rosenberg - The positive sum strategy: Harnessing ..., 1986 - World Scientific

Models that depict **innovation** as a smooth, well-behaved linear process badly misspecify the nature and direction of the causal factors at work. **Innovation** is complex, uncertain, ...

☆ Save ↀ Cite Cited by 9042 Related articles All 12 versions

[HTML] Understanding innovation

KB Kahn - Business horizons, 2018 - Elsevier

... **Innovation** as an outcome emphasizes what output is sought, including product **innovation**, process **innovation**, marketing **innovation**, ... to major radical **innovations**; **innovation** is not a ...

☆ Save ↀ Cite Cited by 1356 Related articles All 6 versions

Managing innovation

J Tidd - ... and Engineering Management Society Body of ..., 2023 - Wiley Online Library

... **innovation**. The degree of **innovation** includes decisions regarding the balance between incremental and more radical **innovation**, and the direction of **innovation** influences the types of ...

☆ Save ↀ Cite Cited by 17003 Related articles All 19 versions ↀ



Product & Services ▾ Subscribe Submit a paper

 **Stiftungen Und Soziale Innovationen (Foundations and Social Innovation)**

Number of pages: 45 • Posted: 11 Jul 2014

Steffen Bethmann

Center for Philanthropy Studies

A definition is provided and the process of social innovation is explained.

Finally, an ideal type of a foundation for social innovation is presented.

Social Innovation, Soziale Innovation, Stiftung, Foundation

 **Innovación Y Localización (Innovation and Location)**

Revista Internacional Administración & Finanzas, v. 8 (6) p. 77-93, 2015

Number of pages: 18 • Posted: 30 Jan 2016

José Gerardo Serafín Vera Muñoz, María Antonieta Monserrat Vera Muñoz, Rafaela Martínez Méndez and Brenda Vera Palacios

Benemérita Universidad Autónoma de Puebla (BUAP); Universidad Popular Autónoma del Estado de Puebla (UPAEP); Universidad Popular Autónoma del Estado de Puebla (UPAEP); Benemérita Universidad Autónoma de Puebla (BUAP)

y la naturaleza no lineal del proceso de innovación.

Las primeras investigaciones sobre innovación se enfocaron en las grandes empresas.

Innovation, National Innovation System, Rural Business

 **Innovative Capability and Financing Constraints for Innovation More Money, More Innovation?**

ZEW - Centre for European Economic Research Discussion Paper No. 09-081

Number of pages: 42 • Posted: 03 Feb 2010

Hanna Hottenrott and Bettina Peters

Technische Universität München (TUM); Center for European Economic Research (ZEW)

Third, we introduce the concept of innovative capability and how it affects financing constraints for innovation.

by innovative capability through increasing resource requirements.

Some Key Questions

- Does the academic publishing system require radical change, or could the incumbents address the pain points?
- Could an institution host the Ledger (e.g., "academic Ledger is hosted by the Wharton School of the University of Pennsylvania")?
- What is the unit of work? (papers, books, videos, other) Should we care?
- What resources are required to operate the Ledger at scale?
- What business models might work to make the Ledger sustainable?

 *academicLedger* BETA

Explore the Ledger

Search and rank records by QaL within a stated reference class and vintage. Leaderboards are just this

SEARCH TITLE OR AUTHOR

FIELD

e.g. bandit, Cachon, Terwiesch

All fields

LEADERBOARDS

Top QaL · MS&OR

Top QaL · Info Systems

Top QaL · Decision Sciences

 SORT

Both reference classes are shown: **QaL · field** (the single OpenAlex subfield & vintage) and **QaL · synthetic ★** (the **offic** cohorts; computed for the prefilled set, "—" until computed). Click any column to re-sort; the official number stays the synt

500+ records · sorted by QaL · field ↓