Incentives in Science

Carolyn Stein

Economic Analyses of Science Workshop
Oxford University, May 2024

Science as a non-market incentive

Academic freedom

The power of credit

The consequences of a credit-driven system
- Credit leads to intense competition
- Credit requires scientists to defend their work

The question

What motivates scientists to do science?

Basic vs. applied science

- **Basic science** seeks to expand human knowledge, but not to create or invent something. There is no obvious commercial value to the result of basic research
- **Applied science** seeks to solve practical problems and often yields something that is commercially valuable
- Basic research is important: "People cannot foresee the future well enough to predict what's going to develop from basic research. If we only did applied research, we would still be making better spears." – George Smoot
- But how do we incentivize people to produce it?

Basic vs. applied science: an example

Basic science

Applied science

Francisco Mojica studied bacteria in Spanish salt flats in the 80s and 90s. He noticed odd bits of repeated DNA in these bacteria...which paved the way for CRISPR

Long stretches of short tandem repeats are present in the largest replicons of the Archaea *Haloferax mediterranei* and *Haloferax volcanii* and could be involved in replicon partitioning

F.J.M. Mojica, C. Ferrer, G. Juez, F. Rodríguez-Valera

First published: July 1995 | https://doi.org/10.1111/j.1365-2958.1995.mmi_17010085.x | Citations: 205

Vertex pharmaceuticals has developed a CRISPR gene editing treatment to cure patients with sickle-cell anemia, which is currently under FDA review

Profit motivates applied science

- We have robust evidence that innovation responds to demand:
  - Acemoglu and Linn (2004) studies drug development in response to demographic shifts
  - Finkelstein (2004) studies vaccine development in response to policy changes
  - Moscona and Sastry (2023) studies heat-tolerant seeds in response to climate change
- We also have evidence that innovation response to increased IP protection
  - Budish, Williams, and Roin (2016) shows that cancer drug innovation responds to the (effective) patent term
  - Giorcelli and Moser (2020) shows that IP led to the development of more operas in the 19th century
  - Moscona (2022) shows that patent protection led to increased development of seed varieties

The incentives governing basic science are murkier

- What are scientists trying to maximize?
- How does this govern their behavior?
- What are the implications for the production of science?

Mertonian norms

Throughout this talk, it can be helpful to think about four norms (or perhaps ideals) of science developed by the sociologist Robert K. Merton:

1. **Communality:** ideas ought to be freely shared to promote collaboration; secrecy is the opposite of this norm
2. **Universalism:** scientific validity is independent of the status or attributes of its participants
3. **Disinterestedness:** scientists should act for the benefit of the scientific enterprise – not for specific outcomes or personal gain
4. **Organized skepticism:** scientific claims should be exposed to scrutiny before being accepted

When do these norms break down?

Science as a non-market incentive

Academic freedom

The power of credit

The consequences of a credit-driven system

- Credit leads to intense competition
- Credit requires scientists to defend their work

Academic freedom can be a strong motivator

- Stern (2004) "Do Scientists Pay to Be Scientists?" studies whether researchers take a pay cut to be given more scientific freedom
- Surveys biology post-docs with multiple job offers and collects characteristics of the jobs
  - Salary
  - Measures of scientific freedom (allowed to publish discoveries, allowed to continue postdoc projects, whether there are incentives to publish)
- Argues that all job offers should be roughly similarly attractive (formal offers only issued if candidate is serious)
- Can the run a hedonic regression with individual fixed effects

Researchers do value academic freedom

Results suggest postdocs accept a 20% pay cut in exchange for more academic freedom

Table 3 Hedonic Wage Regression: Overall Sample Dependent Variable = LN(SALARY), # of Observations = 121

	Permission to publish			Combination model	Science index model	
	(3–1)	(3–2)	(3–3)	(3–4)	(3–5)	(3–6)
	Baseline (NO FE)	Baseline (w/FE)	Full model (w/FE)	Full model (w/FE)	Full Model (w/FE)	Full Model (w/FE)
PERMIT_PUB	0.027 (0.186)	**−0.266** **(0.114)**	**−0.191** **(0.105)**	−0.089 (0.103)		
CONTINUE RESEARCH				**−0.134** **(0.060)**		
INCENT_PUB				−0.036 (0.028)		
SCIENCE INDEX					**−0.114** **(0.053)**	−0.078 (0.057)
EQUIPMENT				**0.063** **(0.033)**	**0.057** **(0.030)**	**0.053** **(0.031)**
CONTROLS						
PROMOTION			**0.041** **(0.025)**	**0.046** **(0.021)**	**0.042** **(0.021)**	0.031 (0.023)
STOCK_DUMMY			**0.196** **(0.085)**	**0.234** **(0.074)**	**0.260** **(0.067)**	**0.190** **(0.077)**
ACCEPTED JOB			−0.013 (0.040)	0.002 (0.043)	−0.0001 (0.043)	−0.002 (0.044)
JOBTYPE CONTROLS	no	no	yes (5; Sig.)	no	no	yes (5)
Individual fixed effects	no	yes (52; Sig.)	yes (52; Sig.)	yes (52; Sig.)	yes (52; Sig.)	yes (52; Sig.)
R-squared	0.001	0.915	0.955	0.958	0.954	0.958

Notes. Only persons with multiple job offers are included.

Science as a non-market incentive

Academic freedom

The power of credit

The consequences of a credit-driven system
- Credit leads to intense competition
- Credit requires scientists to defend their work

How else do we encourage basic research?

Many scientific norms can be viewed through the lens of providing incentives to engage in basic research:

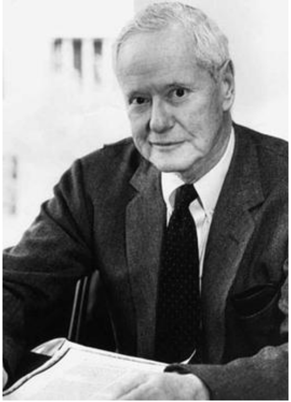
- Grants
- Prizes
- Eponymy

All designed to compensate researchers. Not with profits, but with credit, acclaim, etc.

"My love of natural science...has been much aided by the ambition to be esteemed by my fellow naturalists" – Charles Darwin

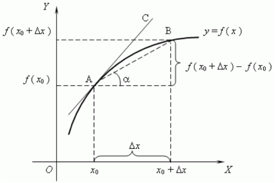
The importance of credit and recognition

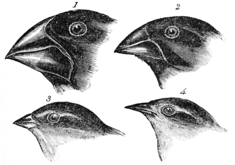
"In short, property rights in science become whittled down to just this one: the **recognition by others** of the scientist's distinctive part in having brought the result into being."
- Robert K. Merton (1957)

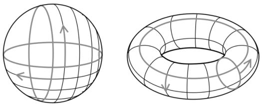


Priority in scientific discovery

- **Priority**: Credit given to the individual who *first* makes a scientific discovery.
- If being first yields more credit, not surprising that there are often fierce disputes over priority
- Notable scientific races and priority disputes:
  - Newton versus Leibniz - Calculus
  - Darwin versus Wallace - Natural Selection and Evolution
  - Perelman versus Yau, Zhu, and Cao - Proof of the Poincaré Conjecture
- Merton (1961) assembles 264 cases of "multiple discovery"







Scooped! Estimating rewards for priority in science (Hill and Stein, 2023)

1. What is the causal effect of getting scooped?
   - Short-run effect on project: Publication, journal placement, and citations
   - Long-run effect on career: Future productivity of scientists

Scooped! Estimating rewards for priority in science (Hill and Stein, 2023)

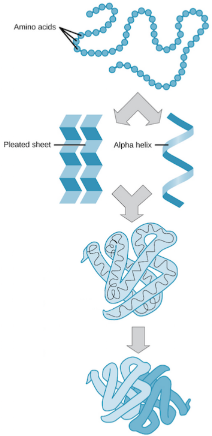
1. What is the causal effect of getting scooped?
    - Short-run effect on project: Publication, journal placement, and citations
    - Long-run effect on career: Future productivity of scientists
2. Does the priority reward system reinforce inequality in science? (Matthew Effect)
    - What drives citations: being first or being famous?

Key empirical challenges

1. Need a setting with well-defined problems and “one right answer.”
2. Need an objective measure of scientific proximity.
3. Need a view of potential abandonments prior to publication.

What is structural biology?

- Structural biologists determine the molecular structure of proteins, DNA, and RNA.
- Proteins carry out most of the functions within cells, and often "form determines function."
- Structures are solved by X-ray crystallography. Successful experiments result in diffraction data and a model that describes the protein shape.

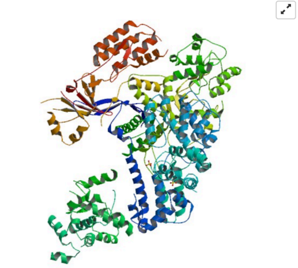


The Protein Data Bank

- The Protein Data Bank (PDB) contains structural data of 100,000+ proteins and meta-data about projects.
- Major scientific journals require scientists to submit their structure data to the PDB before publication.
- All structures are deposited confidentially a few months before article publication.
- Bioinformatics algorithm links projects with identical biological features.

PDB example: Cas-9

Biological Assembly 1



Macromolecule Content

- Total Structure Weight: 318476.84
- Atom Count: 18888
- Residue Count: 2744
- Unique protein chains: 1

Display Files | Download Files

4CMP unique structure ID

Crystal structure of S. pyogenes Cas9

DOI: 10.2210/pdb4CMP/pdb

Classification: HYDROLASE
Organism(s): Streptococcus pyogenes serotype M1
Expression System: Escherichia coli BL21(DE3)

key dates

Deposited: 2014-01-16 **Released:** 2014-02-12
Deposition Author(s): Jinek, M., Jiang, F., Taylor, D.W., Sternberg, S.H., Kaya, E., Ma, E., Anders, C., Hauer, M., Zhou, K., Lin, S., Kaplan, M., Iavarone, A.T., Charpentier, E., Nogales, E., Doudna, J.A.

Literature Download Primary Citation

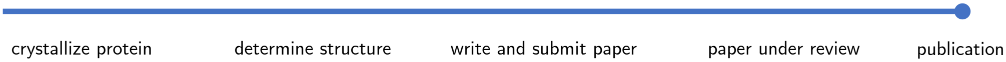
Structures of Cas9 Endonucleases Reveal RNA-Mediated Conformational Activation.

Jinek, M., Jiang, F., Taylor, D.W., Sternberg, S.H., Kaya, E., Ma, E., Anders, C., Hauer, M., Zhou, K., Lin, S., Kaplan, M., Iavarone, A.T., Charpentier, E., Nogales, E., Doudna, J.A.
(2014) Science **343**: 47997

PubMed: 24505130 Search on PubMed Search on PubMed Central
DOI: 10.1126/science.1247997
Primary Citation of Related Structures:
4OGE, 4OGC, 4CMQ

PubMed Abstract:
Type II CRISPR (clustered regularly interspaced short palindromic repeats)-Cas (CRISPR-associated) systems use an RNA-guided DNA endonuclease, Cas9, to generate double-strand breaks in invasive DNA during an adaptive bacterial immune response. Cas9 h ...

Project timeline



Project timeline

PDB dates:

Deposit Date:
Team uploads project details to the PDB database in secret

crystallize protein — determine structure — write and submit paper — paper under review — publication

Project timeline

PDB deposit hidden from public

PDB dates:

Deposit Date:
Team uploads project details to the PDB database in secret

Release Date:
Project is released at time of publication* for public view

crystallize protein

determine structure

write and submit paper

paper under review

publication

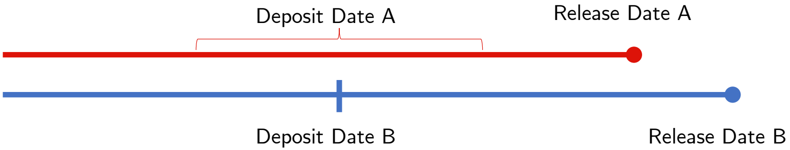
Mean = 6.5 months
Median = 5.1 months

*If project goes unpublished, data is released publicly after one year

Scoop definition

Rules:
1. Take two projects that have identical sequence, different authors.
2. Assert that both projects are deposited before the first project is released.
3. Call the first to release the winner, call the second project "scooped."

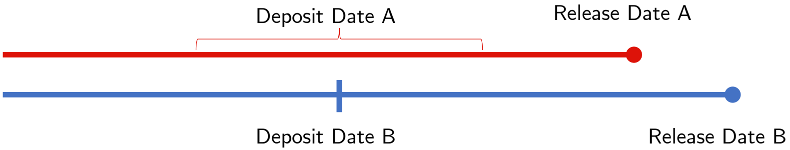
Scenario 1: Project A scoops Project B



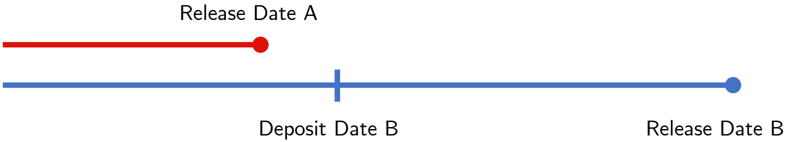
Scoop definition

Rules:
1. Take two projects that have identical sequence, different authors.
2. Assert that both projects are deposited before the first project is released.
3. Call the first to release the winner, call the second project "scooped."

Scenario 1: Project A scoops Project B

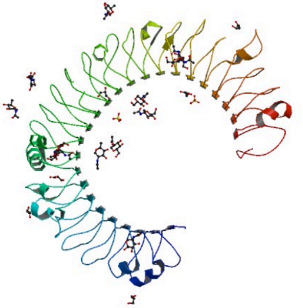


Scenario 2: Project A and Project B are excluded from racing sample



Example race: Toll-like receptor 3

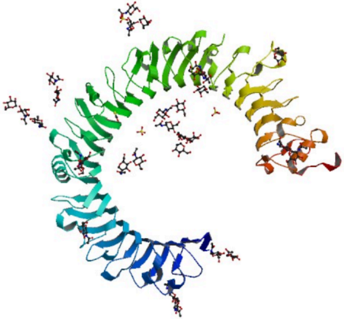
Winning Deposit: 1ZIW



Affiliation: Scripps Research Institute
Deposit Date: April 27, 2005
Release Date: June 28, 2005

Journal: *Science*
Journal Impact Factor: 30.9
5-year Citations: 196

Scooped Deposit: 2A0Z



Affiliation: National Institutes of Health
Deposit Date: June 27, 2005
Release Date: August 2, 2005

Journal: *PNAS*
Journal Impact Factor: 10.2
5-year Citations: 129

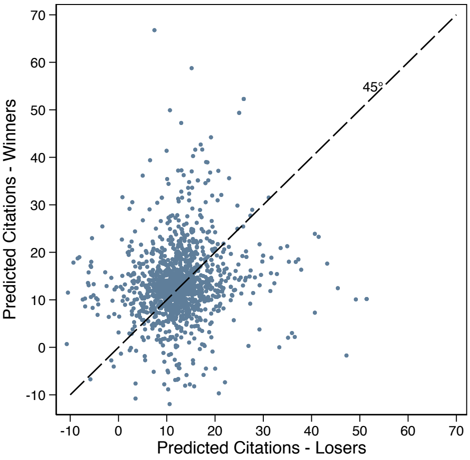
Predicted citation balance

Race winners are not randomly assigned, but seem highly unpredictable.

Lasso model of predicted citations:

- Team size and age
- Past deposits and publications
- University rank and location

Difference in predicted citations: 0.66 (p-value = 0.076)



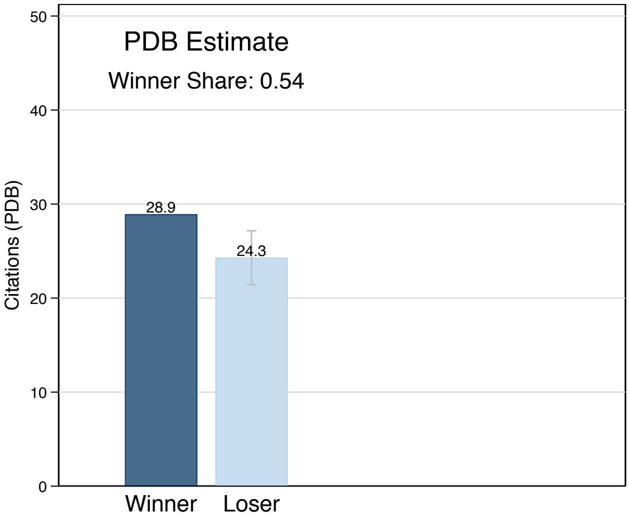
Estimating the scoop penalty

- Basic specification: For deposit i of protein (race) p:

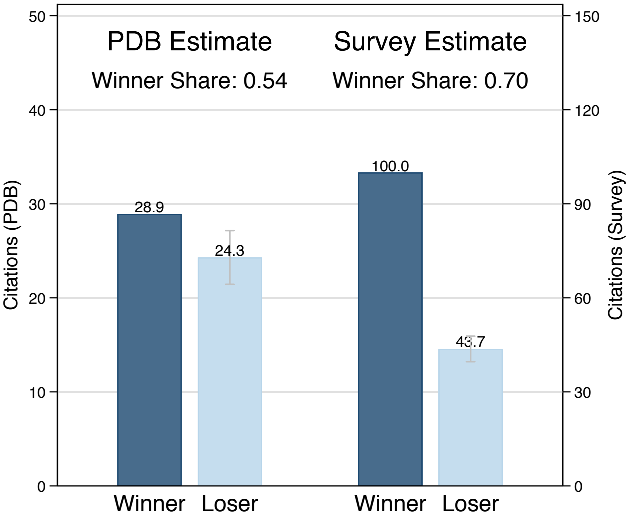
$$Y_{ip} = \alpha + \beta Scooped_i + X_i'\delta + \gamma_p + \epsilon_{ip}$$

  where
  - $Scooped_i$ is a dummy for losing priority race.
  - γ_p is the coefficient on a protein (i.e. race) fixed effect.
  - X_i is a vector of individual and lab controls selected by PDS-Lasso method (Belloni et al. 2014).

Citation penalty



Citation penalty



Scoop penalty: alternative outcomes

Dependent variable	Published (1)	Std. journal impact factor (2)	Top-ten journal (3)	Five-year citations (4)	Top-10% five year citations (5)
Panel A. No controls					
Scooped	-0.025	-0.192***	-0.066***	-0.245***	-0.037***
	(0.015)	(0.044)	(0.020)	(0.071)	(0.014)
Panel B. Base controls					
Scooped	-0.026**	-0.183***	-0.064***	-0.216***	-0.028**
	(0.013)	(0.044)	(0.021)	(0.063)	(0.014)
Panel C. PDS-Lasso selected controls					
Scooped	-0.026***	-0.186***	-0.063***	-0.208***	-0.036***
	(0.010)	(0.032)	(0.015)	(0.045)	(0.010)
Winner Y mean	0.879	-0.027	0.320	28.830	0.149
Observations	3,279	3,279	3,279	2,514	2,514

Notes: This table presents regression estimates of the scoop penalty, following equation 2 in the text. Each regression contains protein (i.e., race) fixed effects. Observations are at the structure level. Each coefficient is from a separate regression. Panel A presents results from a specification with no controls. Panel B adds the base set of controls as listed in Table 3. Panel C uses controls selected by the PDS-Lasso method. Standard errors are in parentheses, and are clustered at the race level. Column 4 regression uses asinh(five-year citations) as the dependent variable, but Winner Y Mean is reported in levels for ease of interpretation.
*$p<0.1$, **$p<0.05$, ***$p<0.01$.

The long-run consequences of being scooped

- Long run outcomes (excluding winning/scooped paper):
  - Active in PDB five years later
  - Total publications five years later
  - Total citations five years later
- Estimate for scientist s, deposit i, for protein (race) p:

$$Y_{isp} = \alpha + \beta Scooped_{is} + X_{is}'\delta + \gamma_p + \varepsilon_{isp}$$

- Estimate separately for novices (<1 year of PDB experience) and veterans.

Long-run results

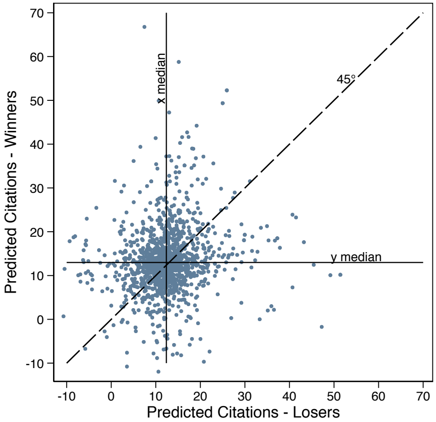
Dependent variable	Any PubMed within five years (1)	Any PDB within five years (2)	Total count within five years after race: PubMed publications (3)	PDB publications (4)	Top-ten publications (5)	Citation-weighted publications (6)	Top-10% cited publications (7)
Panel A. All scientists							
Scooped	-0.018***	-0.042***	-1.165	-0.085	-0.114	-0.172***	-0.414**
	(0.006)	(0.010)	(1.051)	(0.220)	(0.100)	(0.044)	(0.180)
Winner Y mean	0.841	0.702	45.869	7.154	3.610	497.203	7.741
Observations	8,624	8,624	8,624	8,624	8,624	6,484	6,484
Panel B. Novices							
Scooped	-0.057***	-0.040**	-0.021	0.003	0.104	-0.321***	-0.102
	(0.018)	(0.019)	(0.276)	(0.168)	(0.068)	(0.103)	(0.109)
Winner Y mean	0.469	0.356	4.243	1.890	0.616	75.691	1.165
Observations	2,033	2,033	2,033	2,033	2,033	1,529	1,529
Panel C. Veterans							
Scooped	-0.006*	-0.040***	-1.200	-0.176	-0.197	-0.130***	-0.568**
	(0.003)	(0.012)	(1.556)	(0.308)	(0.144)	(0.043)	(0.252)
Winner Y mean	0.990	0.839	61.681	9.261	4.787	667.421	10.388
Observations	5,821	5,821	5,821	5,821	5,821	4,378	4,378

Priority and inequality

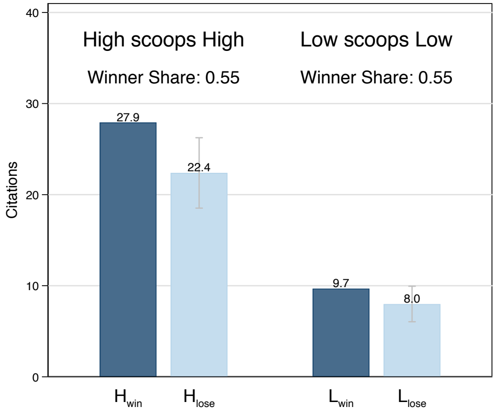
- Merton proposes two key drivers of academic attention:
  - Priority
  - Matthew Effect
- We test which of these effects dominates by comparing citations in races between high- and low-reputation teams
- See the statistical discrimination model in the paper

Defining reputation

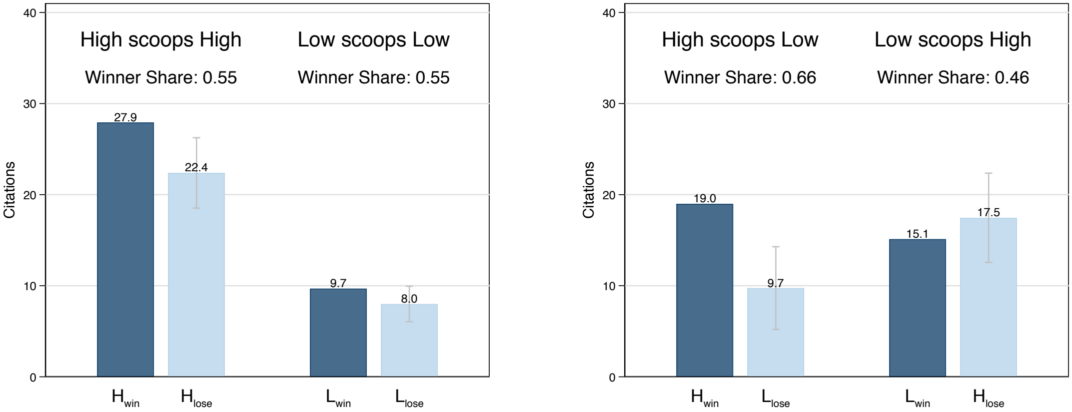
- Define pre-existing reputation using LASSO-generated predicted citations
- Define H teams as those with above median predicted citations and L teams as those with below median



Evenly-matched and mismatched races



Evenly-matched and mismatched races



Conclusion

Getting scooped lowers citations, but rewards are more evenly distributed than previously thought.

Normative implications: Is the premium for priority too large or too small?

- Priority may incentivize effort and timely disclosure.
- Racing may incentivize speed at the expense of quality and transparency.

Science as a non-market incentive

Academic freedom

The power of credit

The consequences of a credit-driven system

- Credit leads to intense competition
- Credit requires scientists to defend their work

How does competition for credit affect science?

- Might motivate researchers to work harder, in order to finish first
- Researchers might be more secretive; less collaborative
- Researchers might cut corners and do lower-quality work to finish first
- What else?

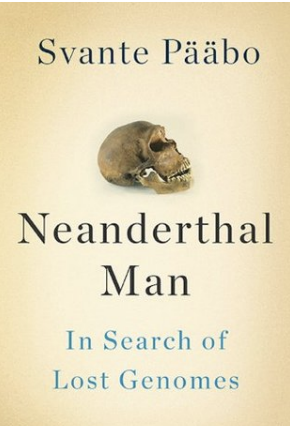
Competition and quality in science (Hill and Stein, 2024)

- This paper focuses on one potential effect of competition: lower quality work
- Again we focus on the field of structural biology for data reasons
- We find strong evidence that scientists complete competitive projects more quickly but with lower quality
- More important projects tend to be more competitive, so key implication: important projects are done more poorly

Example: sequencing the Neanderthal genome

"Hendrik's paper also illustrated a dilemma in science: doing all the analyses and experiments necessary to tell the complete story leaves you vulnerable to being beaten to the press...Even when you publish a better paper, you are seen as mopping up the details after someone who made the real breakthrough"

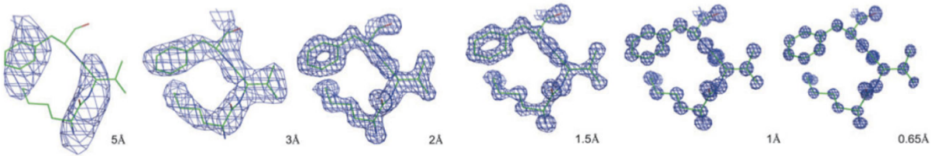
– Svante Pääbo, *Neanderthal Man: In Search of Lost Genomes*



We focus on structural biology because project quality is well measured

A unique feature of structural biology is the objective measures of project quality:

1. Refinement resolution: similar to resolution of a photograph



2. R-free: model fit, estimated on a holdout sample of the experimental data
3. Outlier share: errors in the model based on chemical properties

Combine these outcomes into a standardized quality index (higher is better)

Key results

- Consistent with our theory, more important projects are done faster and more poorly
- These more important projects do not appear to be more difficult or complex
- In addition, researchers who are in less competitive types of positions exhibit a less negative relationship between project importance and project quality. This further suggests competition is the key channel

Science as a non-market incentive

Academic freedom

The power of credit

The consequences of a credit-driven system

- Credit leads to intense competition
- Credit requires scientists to defend their work

Science is full of *nasty* debates

What drove the dinosaurs extinct?

THE NASTIEST FEUD IN SCIENCE

A Princeton geologist has endured decades of ridicule for arguing that the fifth extinction was caused not by an asteroid but by a series of colossal volcanic eruptions. But she's reopened that debate.

By Bianca Bosker

Science is full of *nasty* debates

What is the effect of immigration on native wages?

THE IMPACT OF THE MARIEL BOATLIFT ON THE MIAMI LABOR MARKET

DAVID CARD*

Using data from the Current Population Survey, this paper describes the effect of the Mariel Boatlift of 1980 on the Miami labor market. The Mariel immigrants increased the Miami labor force by 7%, and the percentage increase in labor supply to less-skilled occupations and industries was even greater because most of the immigrants were relatively unskilled. Nevertheless, the Mariel influx appears to have had virtually no effect on the wages or unemployment rates of less-skilled workers, even among Cubans who had immigrated earlier. The author suggests that the ability of Miami's labor market to rapidly absorb the Mariel immigrants was largely owing to its adjustment to other large waves of immigrants in the two decades before the Mariel Boatlift.

THE WAGE IMPACT OF THE *MARIELITOS*: A REAPPRAISAL

GEORGE J. BORJAS*

This article brings a new perspective to the analysis of the wage effects of the Mariel boatlift crisis, in which an estimated 125,000 Cuban refugees migrated to Florida between April and October, 1980. The author revisits the question of wage impacts from such a supply shock, drawing on the cumulative insights of research on the economic impact of immigration. That literature shows that the wage impact must be measured by carefully matching the skills of the immigrants with those of the incumbent workforce. Given that at least 60% of the *Marielitos* were high school dropouts, this article specifically examines the wage impact for this low-skill group. This analysis overturns the prior finding that the Mariel boatlift did not affect Miami's wage structure. The wage of high school dropouts in Miami dropped dramatically, by 10 to 30%, suggesting an elasticity of wages with respect to the number of workers between −0.5 and −1.5.

Why the acrimony?

One hypothesis:

- In a credit-driven system, one's reputation is the currency of careers
- Scientists therefore are invested in defending their work and reputation

What are the consequences?

"A new scientific truth does not triumph by convincing its opponents and making them see the light, but rather because its opponents eventually die, and a new generation grows up that is familiar with it."

– Max Planck

Science advances one funeral at a time (Azoulay et. al, 2019)

One hypothesis:

- It is difficult for science to move in new directions while the developers of old theories are defending their ideas
- Thus, we might expect that after the death of a superstar scientist, we see that field moving in new directions

Assemble 452 research-active superstar scientists who die

Start with a sample of 12,935 "superstar scientists." 452 die prematurely

Table 1: Summary Statistics — Deceased Superstar Scientists (N=452)

	Mean	Median	Std. Dev.	Min.	Max.
Year of Birth	1930.157	1930	11.011	1899	1959
Degree Year	1957.633	1957	11.426	1928	1986
Year of Death	1991.128	1992	8.055	1975	2003
Age at Death	60.971	61	9.778	34	91
Female	0.102	0	0.303	0	1
MD Degree	0.403	0	0.491	0	1
PhD Degree	0.489	0	0.500	0	1
MD/PhD Degree	0.108	0	0.311	0	1
Sudden Death	0.409	0	0.492	0	1
Nb. of Subfields	6.805	4	7.308	1	57
Career Nb. of Pubs.	138.221	112	115.704	12	1,380
Career Nb. of Citations	8,341	5,907	8,562	120	72,122
Career NIH Funding	$16,637,919	$10,899,139	$25,441,933	0	$329,968,960
Sits on NIH Study Section	0.007	0	0.081	0	1
Career Nb. of Editorials	0.131	0	0.996	0	17

Note: Sample consists of 452 superstar life scientists who died while still actively engaged in research. See Appendix A for more details on sample construction.

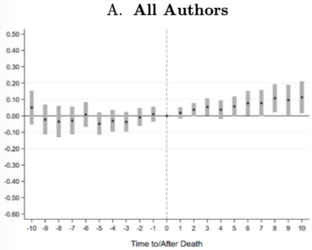
Empirical strategy

- Analysis is performed at the subfield level. Subfields are classified using the PubMed Related Citations Algorithm
  - Despite having the word *citations* in the title, the algorithm relies on the paper titles, abstracts, and MeSH keywords
- Create a set of control superstars who did not die
- Treated subfields have a superstar death; control subfields have a living superstar
  - I am omitting some details here, there is a careful matching process
- Main specification is a diff-in-diff at the subfield level (i) year (t) level:

$$y_{it} = \exp[\beta_0 + \beta_1 AfterDeath_{it} + \beta_2 AfterDeath_{it} \times Treat_i + f(Age_{it}) + \delta_t + \gamma_i + \varepsilon_{it}]$$

Fields with deaths become more productive, driven by non-collaborators

Figure 2
Effect of Star Scientist Death on Subfield Growth and Decline

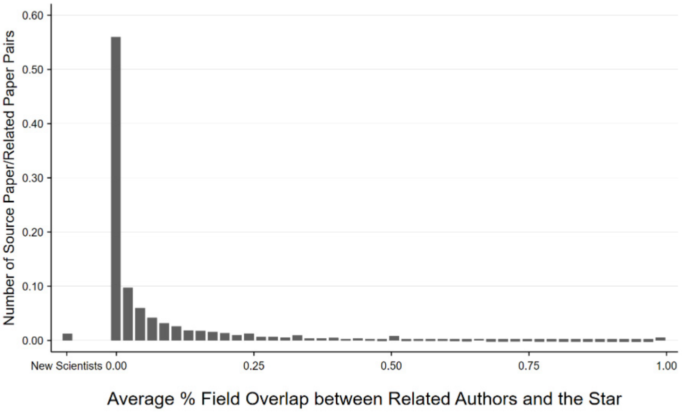


Note: The dark dots in the above plots correspond to coefficient estimates stemming from conditional (subfield) fixed effects Poisson specifications in which publication flows in subfields are regressed onto year effects, subfield age effects, as well as 20 interaction terms between treatment status and the number of years before/after the death event (the indicator variable for treatment status interacted with the year of death is omitted). The specifications also include a full set of lead and lag terms common to both the treated and control subfields to fully account for transitory trends in subfield activity around the time of the death. The 95% confidence interval (corresponding to robust standard errors, clustered at the level of the star scientist) around these estimates is plotted with vertical light grey lines; Panel A corresponds to a dynamic version of the specification in column (1) of Table 3; Panel B corresponds to a dynamic version of the specification in column (2) of Table 3; Panel C corresponds to a dynamic version of the specification in column (3) of Table 3.

Who are these non-collaborators?

Majority have never worked in the superstar's subfield prior to death

A. Distribution of Intellectual Proximity



Note: 50.53% of the related articles by non-collaborators have authors whose past output has zero overlap with the star's subfield.
1.24% of the related articles have authors that are new scientists (and therefore also have no intellectual overlap with the star's subfield).

These new publications are disproportionately high impact

Table 4: Scientific Impact of Entry

	Vintage-specific long-run citation quantile						
	All Pubs	Bttm. Quartile	2^{nd} Quartile	3^{rd} Quartile	Btw. 75^{th} and 95^{th} pctl.	Btw. 95^{th} and 99^{th} pctl.	Above 99^{th} pctl.
After Death	0.082^{**} (0.029)	-0.028 (0.036)	0.008 (0.033)	0.031 (0.032)	0.125^{**} (0.035)	0.232^{**} (0.049)	0.320^{**} (0.081)
Nb. of Investigators	6,260	6,222	6,260	6,257	6,255	6,161	5,283
Nb. of Fields	34,218	33,714	34,206	34,212	34,210	33,207	21,852
Nb. of Field-Year Obs.	1,259,176	1,240,802	1,258,738	1,258,954	1,258,880	1,221,952	804,122
Log Likelihood	-2,768,252	-689,465	-1,125,555	-1,432,223	-1,469,096	-542,735	-156,519

Note: Estimates stem from conditional (subfield) fixed effects Poisson specifications. The dependent variable is the total number of publications by non-collaborators in a subfield in a particular year, where these publications fall in a particular quantile bin of the long-run, vintage-adjusted citation distribution for the universe of journal articles in *PubMed*. All models incorporate a full suite of year effects and subfield age effects, as well as a term common to both treated and control subfields that switches from zero to one after the death of the star. Exponentiating the coefficients and differencing from one yield numbers interpretable as elasticities. For example, the estimates in column (1), Panel A, imply that treated subfields see an increase in the number of contributions by non-collaborators after the superstar passes away—a statistically significant 100×(exp[0.082]-1)=8.55%.

Robust standard errors in parentheses, clustered at the level of the star scientist. $^{\dagger}p < 0.10$, $^{*}p < 0.05$, $^{**}p < 0.01$.

The scientific orientation of these new articles is nuanced

- Panel A suggests new articles are central to the subfield (as measured by PRMA)
- Panel B says that the references are more novel, bringing new sources of inspiration
- Panel C suggests new articles are closer to the scientific frontier

Table 5: Entry and Research Direction

Panel A	**Cardinal Measure**		**Ordinal Measure**	
	Intllct. Proximate Articles	Intllct. Distant Articles	Intllct. Proximate Articles	Intllct. Distant Articles
After Death	0.091** (0.030)	0.028 (0.035)	0.117** (0.028)	-0.024 (0.037)
Nb. of Investigators	6,228	6,099	6,260	6,017
Nb. of Fields	33,375	32,232	34,218	31,712
Nb. of Field-Year Obs.	1,228,157	1,186,589	1,259,176	1,167,423
Log Likelihood	-1,628,374	-1,816,449	-1,893,982	-1,628,170
Panel B	**In-field vs. Out-of-field References**		**Backward Citations to the Star's Bibliome**	
	w/ in-field references	w/o in-field references	w/ references to the star	w/o references to the star
After Death	-0.023 (0.041)	0.128** (0.031)	0.078* (0.036)	0.152** (0.034)
Nb. of Investigators	6,195	6,260	6,247	6,259
Nb. of Fields	32,721	34,218	34,179	34,147
Nb. of Field-Year Obs.	1,204,315	1,259,176	1,257,747	1,256,576
Log Likelihood	-792,795	-2,510,344	-1,914,448	-1,767,571
Panel C	**Vintage of Cited References**		**Vintage of 2-way MeSH Term Combinations**	
	Young	Old	Young	Old
After Death	0.071* (0.035)	-0.010 (0.034)	0.090** (0.033)	0.029 (0.036)
Nb. of Investigators	6,260	6,260	6,258	6,260
Nb. of Fields	34,218	34,214	34,206	34,210
Nb. of Field-Year Obs.	1,259,176	1,259,044	1,258,732	1,258,906
Log Likelihood	-2,124,598	-1,613,457	-1,853,062	-1,784,275

Channels

- Effect is larger when star is more eminent
- No clear evidence that stars withhold resources
  - Very few of the stars held editorial or funding positions around the time of their death
  - Some evidence that effects are larger if fewer superstar *collaborators* hold these positions *rightarrow* evidence of indirect influence

Conclusion

- Scientists are human and humans respond to incentives
- The credit-based system that science employs leads to scientists deviating from the Mertonian norms/ideals
- Can we do better? Are there alternative ways to organize and motivate scientific work?