# Open Science as an Economic Institution

*Pierre Azoulay*
*MIT Sloan School of Management*

*NBER Innovation Boot Camp – July 15th, 2023*

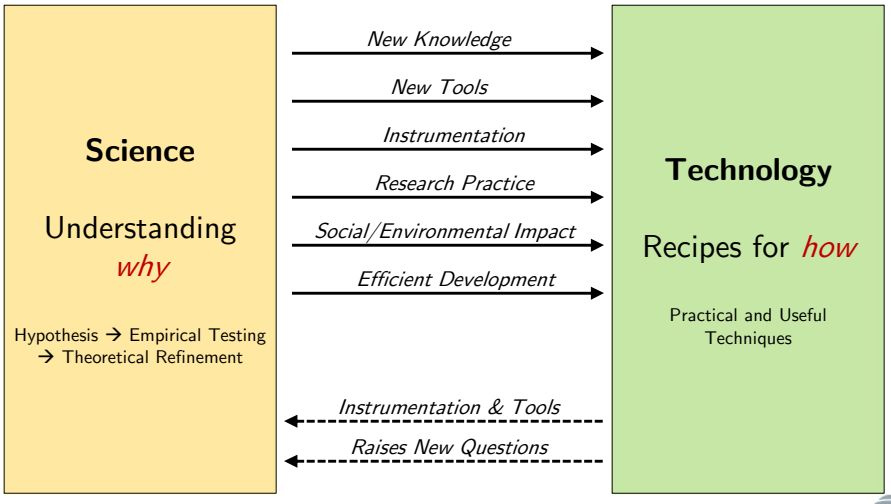

*@pierre_azoulay*

## Outline

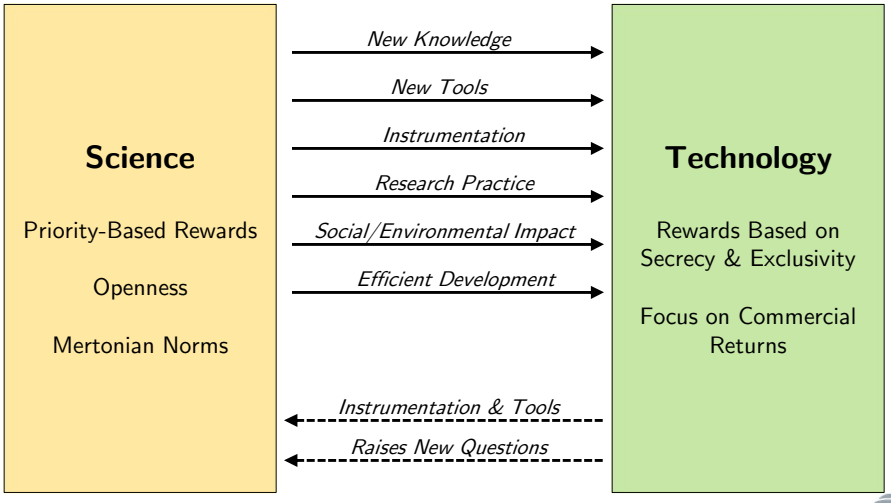
- ***What is science?***
  - *Science as a social institution*
  - *The relationship between science and technology*
- ***The direction of science***
- ***Scientific competition and its consequences***
- ***Science and its institutions***
- ***What are the "returns" to science? (wait for Kyle)***

## WHAT IS "SCIENCE"?

## The relationship between science and technology




## *Science as a distinctive incentive system (Dasgupta and David 1994)*




## *Merton's CUDOS: The Normative Structure of Science [1942]*

- ***Communalism** – the common ownership of scientific discoveries, according to which scientists give up intellectual property rights in exchange for recognition and esteem*
- ***Universalism** – according to which claims to truth are evaluated in terms of universal or impersonal criteria, and not on the basis of race, class, gender, religion, or nationality*
- ***Disinterestedness** – according to which scientists are rewarded for acting in ways that outwardly appear to be selfless*
- ***Originality** – the ultimate scientific reward is the "thin" intellectual property right of credit for having made a particular discovery*
- ***Skepticism** – all ideas must be tested and are subject to rigorous, structured community scrutiny*

## *Science as a social institution*

## *The Matthew Effect (Merton 1965)*

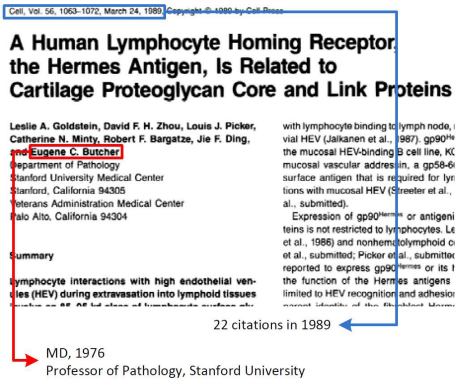
- ***Seemingly high importance of early luck and resources in shaping the skewed distribution of research productivity and scientific status***
  - *"if I have not seen as far as others, it is because giants were standing upon my shoulders" – Hal Abelson*

***"Rayleigh's name was either omitted or accidentally detached*** **[from a manuscript]** ***and the Committee turned it down as the work of one of those curious persons called paradoxers. However, when the authorship was discovered, the paper was found to have merits after all."***

## *Matthew: Effect or Fable? Azoulay, Stuart, & Wang 2014*

- ***Distinguish between producers (scientists) and products (articles)***
- ***Focus on the impact of a discrete change in producer status, i.e., a "status shock:" HHMI Appointment***
- ***Restrict the set of products to those that first appeared before the shock***
- ***Measure the status premium (or discount) by examining changes in deference patterns after the shock, relative to before***

## *Treated and control articles*



Cell, Vol. 56, 1063–1072, March 24, 1989, Copyright © 1989 by Cell Press

**A Human Lymphocyte Homing Receptor: the Hermes Antigen, Is Related to Cartilage Proteoglycan Core and Link Proteins**

Leslie A. Goldstein, David F. H. Zhou, Louis J. Picker, Catherine N. Minty, Robert F. Bargatze, Jie F. Ding, and Eugene C. Butcher

Department of Pathology
Stanford University Medical Center
Stanford, California 94305
Veterans Administration Medical Center
Palo Alto, California 94304

Summary

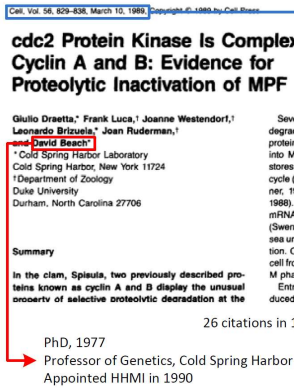
Lymphocyte interactions with high endothelial venules (HEV) during extravasation into lymphoid tissues

with lymphocyte binding to lymph node, vial HEV (Jalkanen et al., 1987). gp90 the mucosal HEV-binding B cell line, KO mucosal vascular addressin, a gp58-6 surface antigen that is required for lym tions with mucosal HEV (Streeter et al., al., submitted).

Expression of gp90 or antigeni teins is not restricted to lymphocytes. Le et al., 1986) and nonhematolymphoid c et al., submitted; Picker et al., submitte reported to express gp90 or its the function of the Hermes antigens limited to HEV recognition and adhesio

22 citations in 1989

MD, 1976
Professor of Pathology, Stanford University



Cell, Vol. 56, 829–838, March 10, 1989, Copyright © 1989 by Cell Press

**cdc2 Protein Kinase Is Complex Cyclin A and B: Evidence for Proteolytic Inactivation of MPF**

Giulio Draetta,* Frank Luca,† Joanne Westendorf,† Leonardo Brizuela,* Joan Ruderman,† and David Beach*

* Cold Spring Harbor Laboratory
Cold Spring Harbor, New York 11724
† Department of Zoology
Duke University
Durham, North Carolina 27706

Summary

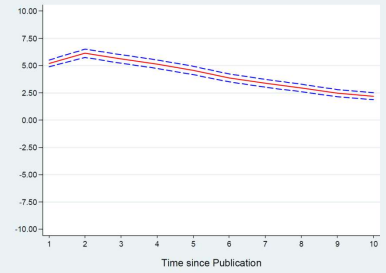
In the clam, Spisula, two previously described proteins known as cyclin A and B display the unusual property of selective proteolytic degradation at the

26 citations in 1

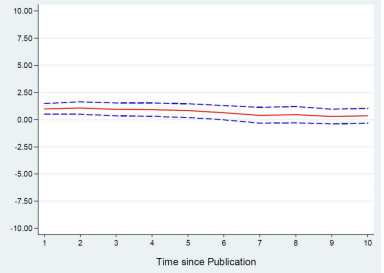
PhD, 1977
Professor of Genetics, Cold Spring Harbor
Appointed HHMI in 1990

## *Effects of HHMI appointment on citation rates [Post-Appointment Articles]*

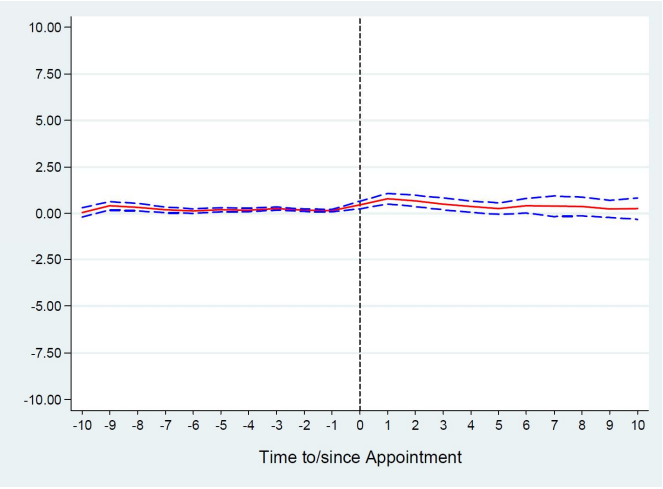
*w/o journal match*



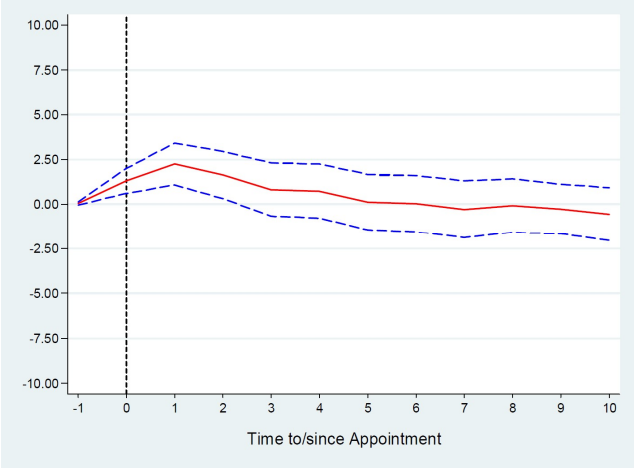

*w/ journal match*




## *Effects of HHMI appointment on citation rates [Pre-Appointment Articles]*




## *Effects of HHMI appointment on citation rates [Pre-Appointment Articles]*




## *Do scientists pay to be scientists? [Stern 2004]*

- ***A Preference Effect (a "taste" for science)***
  - *Researchers (even those in the private sector) may value participation in open science, and thus firms may earn a compensating differential by allowing participation in science in exchange for lower wages*
  - *Intrinsic preferences (Feynman, Kuhn)*
  - *Career concerns (cf. Lerner and Tirole on Open Source)*
- ***A Productivity Effect (a "ticket of admission")***
  - *Firms may benefit from access to scientific knowledge; understanding scientific discoveries (and perhaps learning about them earlier) can only be realized by firms who themselves "spill" some knowledge through participation in open science*
  - *Direct spillovers*
  - *Indirect spillovers*

## Evaluating the wage-science relationship

- ***Cross-sectional relationship between wages and science likely will reflect unobserved differences in ability***
  - *Long tradition in labor economics associated with not being able to control for unobserved heterogeneity (Rosen, 1986)*
  - *Prior work has examined job switchers (Brown, 1980) which are unfortunately subject to their own biases (Gibbons & Katz, 1992)*

- ***Prior to accepting any offer, researchers (and many professionals) receive multiple job offers***
  - *Suggests methodology for "controlling" for individual effects*
  - *Regress wage on organizational practices at the job offer level j, with a fixed effect for each individual worker i*

$$\ln(w_{ij}) = \gamma_i + \beta \cdot SCIENCE_j + \varepsilon_{ij}$$

## Hedonic wage regression (i=52, j=121)

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

## *Economic and strategic implications*

- ***Relative to a system of proprietary knowledge production, the incentives and norms of open science seem to be consistent with the objective of maximizing the rate of production of knowledge in a cumulative manner***
- ***However, the nature of the scientific priority system likely results in distortionary strategic behavior***
  - *Inefficient "herding" on hot topics or big discoveries*
  - *Complicated and costly disputes over scientific priority itself*
  - *Potential for collusion*
  - *Inefficient strategic exclusivity over data, tools, or other resources*
- ***Open science also induces a high potential for spillovers from public knowledge to applications governed by technology***

## *The relationship between science & technology*

## The linear model at work

- ***Brock's unlikely bacteria:***
  - *1967: Thomas Brock discovers Thermus Aquaticus in Yellowstone National Park geysers, classified as an extremophile*
  - *Deposited in the American Type Culture Collection*
  - *1983: Kary Mullis from Cetus conceives of a recipe — a DNA replication scheme requiring DNA polymerase that can resist extreme temperature variation*
  - *After initial attempts locally, identification of TaQ at ATCC*
- ***PCR is the foundational technology for DNA replication in all of modern molecular biology & biotechnology***
  - *1989: Thermus Aquaticus, Molecule of the Year; 1993 Nobel Prize for Mullis*
  - *The patent on PCR (held by Cetus) was sold on the "market for ideas," valued at approximately $500M.*
- ***The usefulness of extremophiles was very hard to anticipate ex ante***
- ***The "application" of the material with Mullis' insight was both a technological breakthrough and a spur for further scientific research***

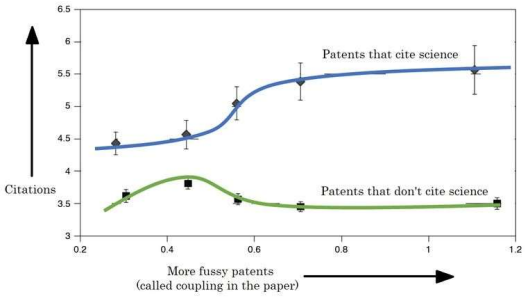
## Science as a map of unfamiliar terrain

- ***More science leads to more technological progress, but only a minority of new technologies directly rely on science***
- ***Science can provide an imperfect map of this unknown terrain, helping inventors step wisely***
- ***Science can obviously benefit unexplored regions of the technological landscape***
  - *But there may also be regions that are well-trod, but treacherous*
  - *Sorenson & Fleming (2004) provide evidence that science is especially useful for this type of "fussy" technologies*

## Sorenson and Fleming (2004)

- ***Relies of USPTO patent classification***
- ***Key measure based on how well a class seems to "play nice" with other technologies***
  - *If a class is frequently attached to a patent alongside a wide range of other classifications, then coupling is loose*
  - *if a class is only ever assigned to a patent with one other classification, then coupling is tight*
- ***Survey of inventors to show this measure is correlated with inventors self-assessments of how sensitive their own inventions are to small changes***
  - *Not merely picking up how novel the technology is*

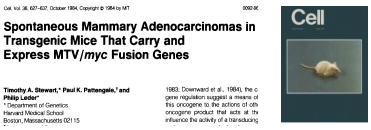
## Main result




- ***Patents primarily composed of "fussy" technologies seem to disproportionately benefit from science, measured by citation of a scientific article***

## *When a demonstration of why is also an example of how: The Harvard OncoMouse*

- ***1984: Leder & Stewart, from the Harvard Medical School, develop the "Oncomouse"***
  - *First mouse with genes inserted to predispose mouse to cancer*
  - *A significant advance along two dimensions:*
    - *Advancing basic research into the role of genes in cancer*
    - *An input into applied research focused on cancer therapies*
- ***Oncomouse is a "dual" discovery and serves as foundation for:***
  - *On-going scientific discovery*
  - *Translation, innovation & economic growth*
- ***Leder publishes a seminal article in Cell, and Harvard (and its licensee DuPont) are granted a US patent in 1988***
  - *Distribution comes with controversial licensing restrictions on use (e.g., reach-through rights and article review)*



Cell, Vol. 38, 627-637, October 1984, Copyright © 1984 by MIT 0092-86

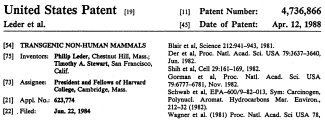
**Spontaneous Mammary Adenocarcinomas in Transgenic Mice That Carry and Express MTV/*myc* Fusion Genes**

**Timothy A. Stewart,* Paul K. Pattengale,† and Philip Leder***

* Department of Genetics
Harvard Medical School
Boston, Massachusetts 02115

1982, Downward et al., 1984), the c
gene regulation suggest a means of
this oncogene to the actions of oth
oncogene product that acts at its
influence the activity of a transducing

Cell



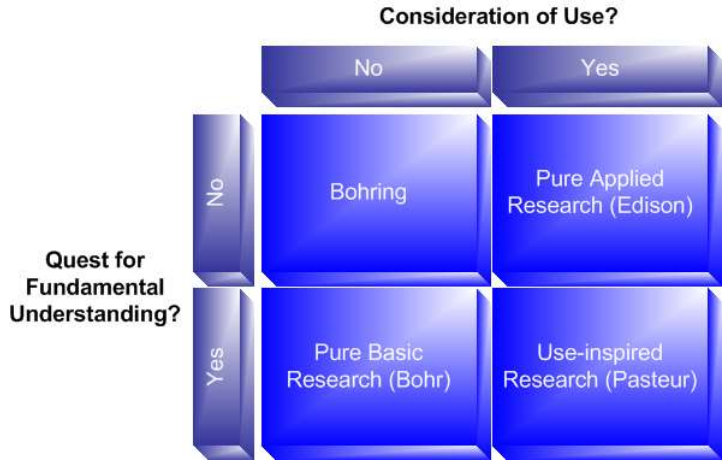
**United States Patent** [19]
**Leder et al.**

[11] **Patent Number: 4,736,866**
[45] **Date of Patent: Apr. 12, 1988**

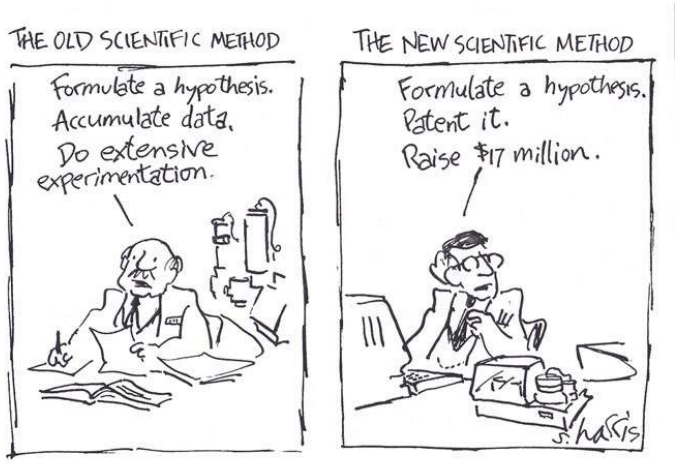
[54] **TRANSGENIC NON-HUMAN MAMMALS**
[75] Inventors: **Philip Leder**, Chestnut Hill, Mass.; **Timothy A. Stewart**, San Francisco, Calif.
[73] Assignee: **President and Fellows of Harvard College**, Cambridge, Mass.
[21] Appl. No.: **623,774**
[22] Filed: **Jun. 22, 1984**

Blair et al, Science 212:941-943, 1981.
Der et al, Proc. Natl. Acad. Sci. USA 79:3637-3640, Jun. 1982.
Shih et al, Cell 29:161-169, 1982.
Gorman et al., Proc. Natl. Acad. Sci. USA 79:6777-6781, Nov. 1982.
Subwah et al, EPA-460/9-82-013, Sym: Carcinogen, Polynucl. Aromat. Hydrocarbons Mar. Environ., 212-32 (1982).
Wagner et al. (1981) Proc. Natl. Acad. Sci USA 78,

## *Pasteur's Quadrant*




## *A more cynical view of Pasteur's quadrant*




## *The Challenge of Pasteur's Quadrant*

- ***What are the key economic implications of the fact that scientists seem to be governed by a distinctive set of "values" that are somewhat independent of pure monetary gain?***
- ***How do the norms and institutions of open science cohere with the nature of the incentive contracting problem between researchers (who may have preferences to participate in open science) and research funders?***

## *Academic Freedom, Private-Sector Focus, & the Process of Innovation (Aghion, Dewatripont, and Stein, 2008)*

- ***Why does academia exist? Usual answer includes imperfect IPRs combined with knowledge spillovers***
  - *But recall Pasteur's quadrant: the connection between the "basicness" of a line of research and the degree of appropriability of the resulting output is ambiguous*
  - *Even if we need basic research to be subsidized (because of limited appropriability), why does this need to happen in academia?*
- ***ADS 2008 develop a model that***
  - *clarifies the respective advantages and disadvantages of academic and private-sector research*
  - *allows one to say when—in the process of developing an idea from its very earliest stages to a finished commercial product—it is normatively optimal to make the transition from academia to the private sector*
- ***At the heart of the model is a decision right:***
  - *Academia boils down to a commitment mechanism that ensures scientists can choose the projects they work on*
  - *In private-sector research, the decision rights inevitably resides with the owner/manager of the firm, who can (and will) largely dictate project choice and methods to the individual scientists who work for the firm*

## *Academic Freedom, Private-Sector Focus, & the Process of Innovation (Aghion, Dewatripont, and Stein, 2008)*

- ***A simple model of the impact of science/academia as a method for organizing privately funded research***
- ***Consider a k-stage research process, in which financial returns V are only realized when the firm successfully completes all stages***
- ***Model "science" or "academia" as an organizational design choice, in which the firm cedes control rights over research direction to researchers (i.e., this is a model of "freedom")***
  - *Ignore the issue of appropriability*
  - *With probability $\alpha$, researcher has preferences for research direction which advances commercialization, and is successful (conditional on choosing that direction) with probability p; note that with $1-\alpha$, research gets utility z from an alternative direction and interests are misaligned*
- ***Firms can either retain control rights for themselves (enhancing the potential for commercialization) or cede control to researchers and benefit from a lower wage structure***

## *Basic intuition*

- ***Consider a case where commercialization involves two steps***
- ***In the last stage, firm chooses to retain control rights if the gains to ensuring that the right final "step" is taken outweighs the wage benefit from ceding control to the researcher (i.e., pV > z)***
- ***However, in the first stage, firm only chooses to retain control rights if the gains to ensuring that all steps outweighs the wage benefit, (i.e., $pE(\Pi_1) > z$)***
- ***Key insight: "academic freedom" is most attractive at the "earliest" stages of the research process and is associated with exploration***

## *Exploration incentives*

- ***ADS consider the possibility of research lines "branching out"***
- ***Suppose that there are two potentially legitimate research projects inside the firm:***
  - *An "applied" project that is only two stages away from a commercial payoff*
  - *A more "basic" project that is five stages away from any payoff*
- ***Which organizational form is more likely to explore?***
  - *It is possible that the ultimate payoff on the more basic project is sufficiently high that, evaluated at academic-sector wages, it is not only positive net present value (NPV), but of greater NPV than the applied project.*
  - *It is also possible that, evaluated at private-sector wages, the basic project is negative NPV.*
    - *If this is the case, then when a private-sector firm has the decision rights, it will allocate all of its scientists to the applied project, and completely ignore the basic project*
    - *By contrast, if the ideas were left freely available to academic scientists, there would naturally tend to be some progress on both projects, as individual scientists followed their own interests.*
- ***It is possible that the returns to freedom are higher when researchers are able to exercise openness, since the benefits from control are more salient when one is able to publicly reveal the information in the scientific literature***

## *Evidence on the Benefits of Openness Of Mice and Academics (Murray et al. 2016)*

- ***What role does scientific openness play in scientific research?***
- ***What types of research are promoted by openness?***
- ***Control rights approach suggests two effects of openness:***
  - *Vertical exploitation – downstream exploitation increases*
  - *Horizontal exploration – entirely new, diverse lines of basic research increase*
- ***The paper exploits the natural experiment created by the shift in openness from NIH agreements and traces out the impact on citations to articles impacted by the agreement***

## *The mouse revolution as a research setting*

- ***Over the past twenty years, a "revolution" in the use of genetically engineered research mice as a tool for life sciences progress***
  - *Mice could now be "engineered" to have a particular gene inserted or removed to mimic a disease e.g., cancer or diabetes*
  - *Over 13,000 specialized mice published in scientific literature*
- ***2007 Nobel Prize in Medicine to Mario R. Capecchi, Martin J. Evans and Oliver Smithies for "gene modification in mice"***
- ***Openness: While the development of genetically modified mice has tremendous for potential application in both basic and applied research, the ability to initiate research "lines" based on new mice require gaining access to those specific mice***
  - *Mice are costly to make and require specialized techniques including embryo manipulation, stem cell adaptation, and molecular biology*
  - *Many mice are also covered by intellectual property rights and so require a license contract with upstream researchers*

## Natural experiment in openness

- ***1990s: Openness crisis***
  - *scientists demand openness to DuPont's OncoMice*
- ***1999: Harold Varmus at NIH intervenes and signs MoU with DuPont to make OncoMice subject to a "simple" license with no reach-through***
  - *An unexpected shift in the openness of mouse genetics research*

## Data sources

- ***Data Sources***
  - *Mouse Genome Informatics database catalogs over 13,000 mice & links each mouse to an original publication in a scientific journal (mouse-articles)*
  - *PubMed for information about mouse-articles & ISI Web of Science SCI for citations*
- ***Sampling Strategy***
  - *Identify universe of MGI mouse-articles published 1983-1998 sample on four types of mouse-articles (2,638 unique mice in 2,223 mouse-articles)*
  - *Cre-Lox (52), Oncomouse (160), Knock-Out (2171), Spontaneous (255)*
- ***For each mouse-article collect information about the forward citations***
  - *525,865 total citations (from pub year thru 2006)*
  - *Aggregated up into 27,442 citation-years*
- ***For each citing article code key article/author characteristics***

## Results: Vertical Exploitation

Dep. var. = Annual citations
[Incidence rate ratios reported in square brackets]
Estimated coefficients in second line
(Block bootstrapped SEs reported in parentheses)

| | OLS | Negative binomial | | | |
|---|---|---|---|---|---|
| | (4-1) Baseline model, DV = log Annual citations | (4-2) Baseline model | (4-3) Baseline model with treatment effect dynamics | (4-4) Treatment effects by Cre-lox and Onco | (4-5) Baseline model, citations from high quality journals only[d] |
| Post-NIH | [1.229]*** 0.206 (0.052) | [1.302]*** 0.264 (0.062) | | | [1.409]*** 0.343 (0.080) |
| Post-NIH, Short-term[b] | | | [1.220]*** 0.199 (0.064) | | |
| Post-NIH, Long-term[c] | | | [1.429]*** 0.357 (0.074) | | |
| Post-Cre-lox | | | | [1.467]*** 0.383 (0.115) | |
| Post-Onco | | | | [1.267]*** 0.236 (0.060) | |
| Age FEs | Yes | Yes | Yes | Yes | Yes |
| Year FEs | Yes | Yes | Yes | Yes | Yes |
| Article FEs | Yes | Yes | Yes | Yes | Yes |
| log-likelihood | — | −55,919.8 | −55,906.1 | −55,912.4 | −34,112.8 |
| Observations | 22,265 | 22,265 | 22,265 | 22,265 | 21,574 |

## Results: Horizontal Exploration

| Negative Binomial | Keywords | | Journals | |
|---|---|---|---|---|
| | Annual Citations with New keywords | Annual Citations with Old keywords | Annual Citations in New Journals | Annual Citations in Old Journals |
| Post Shock | 1.260*** | 0.925 | 1.381*** | 1.201* |
| Conditional Fixed Effects for Article, Margin-Age and Margin-Calendar Year, Window Effects | | | | |

## Key Findings

- ***A significant increase in the rate of follow-on citations for "mouse-articles" impacted by the NIH agreements***
- ***This boost in follow-on research is driven by***
  - *Contributions by "new" authors or institutions (reprint authors or institutions that had not previously cited the original mouse-article)*
  - *More diverse types of research (articles using previously unused keywords or published in journals that had not previously cited the original mouse-article)*
  - *No detectable reduction in the flow of new mouse creation.*
- ***Results highlight a neglected impact of IP: reductions in the diversity of experimentation arising from a single idea***

## Frictions at the academia/industry interface

1977: Discovery (purification) of EPO by Eugene Goldwasser (U of Chicago)

1977-1981
For 5 years, Goldwasser tries desperately to interest firms to produce EPO
Rejected by: University of Chicago; Parke-Davis, Abbott Labs

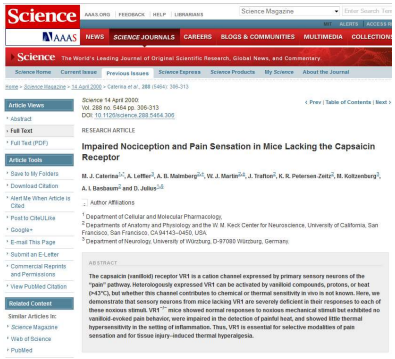
1984: Amgen sequences EPO gene

1987: Amgen produces recombinant EPO

- ***In ADS 2008, the hand-off from academia to the private sector might not happen at the optimal time, but it is essentially frictionless***
- ***Bikard (2018) provides evidence of under-utilization of knowledge coming out of universities***
  - *Under what circumstances is a piece of scientific knowledge translated into a new technology?*
  - *Specifically, does it matter if the discovery took place in a university vs. a private firm?*
- ***Key empirical lever: scientific twins stemming from simultaneous discoveries***

## *In the winter of 1999*

***Vanilloid receptor-1 contributes to chemical and thermal sensitivity in mice***



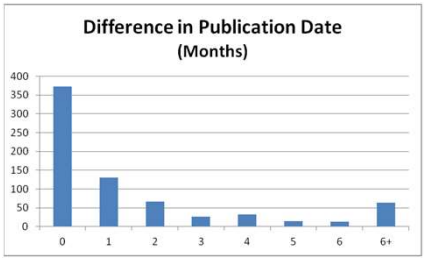

Submitted: 18 January 2000; Published: 14 April 2000
Address: UCSF in San Francisco (CA)

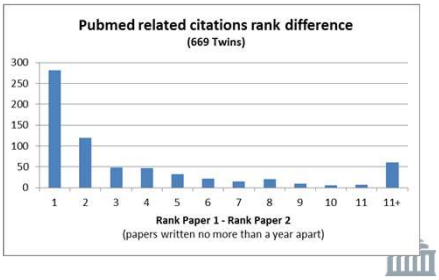

Submitted: 20 December 1999; Published: 11 May 2000
Address: Smithkline Beecham in Harlow (UK)

***"Paper twins": same knowledge simultaneously emerges in two distinct environments***

## *Bikard's (2020) Twin Identification Algorithm*

- ***Most pairs were published in the exact same month***
  - *Avg. difference in months is 1.8*
- ***267 twins were published in the same issue of the same journal***
- ***PubMed Related Citations Algorithm (based on word similarity) rank them next to each other 42% of the time.***
  - *Rank difference <10 for 90% of the twins*
- ***Out 10 interviewees, 9 told me about the twin paper without me asking***
  - *One got really upset when I mentioned the twin*




## *Main result:*
## *Academic twin 20-30% less cited in private-sector patents, relative to the native corporate result*

| Variable | Dependent variable = reference (1/0) | | |
|---|---|---|---|
| | Conditional logit; controls only (Model 5-1) | Conditional logit; main effect (Model 5-2) | LPM; main effect (Model 5-3) |
| *Academic origin* | | −0.673*** | −0.238*** |
| | | (0.25) | (0.07) |
| *Paper is more detailed* | 0.623* | 0.755** | 0.184** |
| | (0.36) | (0.30) | (0.08) |
| *Paper has richer theory* | −1.073 | −1.096 | −0.287 |
| | (0.87) | (0.72) | (0.18) |
| *Paper is more sophisticated* | −1.177 | −0.99 | −0.249 |
| | (0.92) | (0.80) | (0.19) |
| *Paper has more practical emphasis* | 0.593** | 0.467 | 0.0843 |
| | (0.29) | (0.32) | (0.09) |
| *Paper is clearer* | 14.55*** | 14.58*** | 0.696*** |
| | (1.41) | (1.35) | (0.19) |
| *U.S. paper* | 1.544** | 1.701*** | 0.408*** |
| | (0.67) | (0.66) | (0.15) |
| *Journal impact factor* | 0.0148 | 0.00264 | 0.00118 |
| | (0.03) | (0.03) | (0.01) |
| *Patent-paper pair* | 0.204 | −0.0079 | −0.0478 |
| | (0.51) | (0.45) | (0.10) |
| *Number of authors* | 0.0745 | 0.0773 | −0.0588 |
| | (0.49) | (0.39) | (0.11) |
| *Authors' publication stock* | 0.221 | 0.349 | 0.124 |
| | (0.34) | (0.32) | (0.08) |
| *Authors' patent stock* | −0.0401 | −0.117 | −0.0367 |
| | (0.12) | (0.17) | (0.04) |
| *Time lag* | 0.313 | 0.506 | 0.0641 |
| | (0.43) | (0.41) | (0.09) |
| *Geographic distance* | −0.115 | −0.0949 | −0.029 |
| | (0.10) | (0.10) | (0.03) |
| *Same country* | −0.426 | −0.402 | −0.124 |
| | (0.53) | (0.55) | (0.14) |
| *Constant* | | | 0.145 |
| | | | (0.36) |
| Observations | 523 | 523 | 924 |
| No. simultaneous discovery/ patent dyads | 225 | 225 | 480 |
| Pseudo-$R^2$ | 0.119 | 0.153 | 0.149 |
| Log-likelihood | −163.3 | −157.1 | −310.1 |
| Simultaneous discovery/ patent fixed effects | Yes | Yes | Yes |

*Notes.* Observations are paper-patent dyads. Column 3 contains more observations as LPM does not drop cases in which the patent cites all the paper twins. Fixed effects are at the set-of-paper-twins/patent dyad level. Standard errors are clustered throughout at the level of the set of paper twins.
*$p < 0.1$; **$p < 0.0.05$; ***$p < 0.01$.

# *Challenges to the "linear model"*

- ***Probably a good first-order description, but:***
  - *What about feedback? (Rosenberg on chemical engineering, Mokyr)*
  - *Pasteur's Quadrant: What does basic and applied mean?*
    - *Results harder to appropriate? No.*
    - *Results closer to ultimate commercial payoff?*
    - *Results that provides broader shoulders, for more follow-on innovators, to stand on?*
- ***How does the transmission from academia to the private sector happen?***
  - *Who selects ideas for innovation?*
- ***Why do universities patent?***
  - *Because they hold on too long in the linear model?*
  - *Because they do research located in Pasteur's Quadrant?*

## *Summary*

- ***Is science more than a type of knowledge? YES***
- ***Over the past decade, a mixture of theoretical and empirical research suggests that open science is a distinctive economic institution***
  - *An incentive system that overcomes the "paradox" of directly paying for ideas that seems so central to endogenous technical progress*
- ***Increasing amount of scholarly effort devoted to examining the impact of specific institutions and potential for strategic behavior undermining this objective***

## *THE DIRECTION OF SCIENCE*

## *Do scientists sort themselves "efficiently" across research areas?*

- ***Well, probably not…but how might one think about distortions in a systematic way?***
  - *WWSPD? WFK…*
- ***Very different from thinking about bias in the direction of technical change (à la Daron)***
  - *How does the social planner internalize researchers' intrinsic innovation, if at all?*
  - *Diversity objectives in terms of approaches?*
  - *Connections to technology and consumer preferences?*
  - *What do we assume the social planner understand, and what eludes even her?*
- ***Never forget one thing: No one is (really) in charge!***
  - *No one is even supposed to be in charge!*

## *Empirical work on institutions and the direction of scientific effort*

- ***Azoulay et al. (2019) investigate whether superstars can skew the agenda of their fields to follow a specific trajectory***
- ***Myers (2020) on the "elasticity" of science investigates how much scientists need to be paid to switch areas***

## *Azoulay et al. (2019) Does Science Advance One Funeral at a Time?*


**Planck's Principle:**
"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"

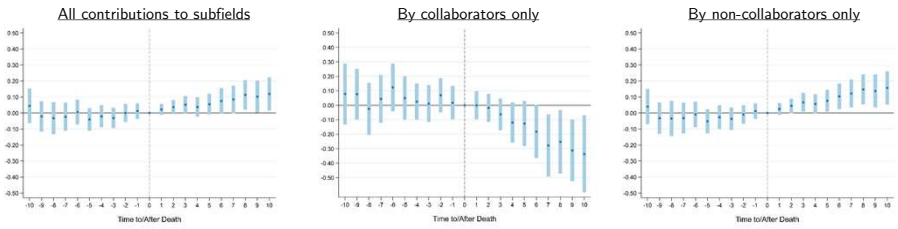
## *Research approach*

- ***Focus on "normal science" (i.e., the regular work of scientists experimenting within a settled paradigm)***
  - *Not Kuhnian "paradigm shifts" (which are very rare)*
- ***An actual delineation of biomedical research subfields***
  - *PubMed Related Citations Algorithm [PMRA]*
- ***A lever to tease out the impact of eminent scientists***
  - *Premature death*

## Findings

- ***Massive negative impact of superstar extinction on publication flows for collaborators in the subfield***
- ***Offset by positive effect on publication flows for non-collaborators***
  - *Outsiders, not competitors, drive the effect*
- ***"Angular velocity": renewal of intellectual sources the research draws upon***
- ***Gatekeeping***
  - *increase in entry more pronounced when the departing stars leave a larger "hole" to fill or are particularly prominent*
  - *increase in entry less pronounced when the subfield is intellectually or socially "coherent" or when the star leaves behind a praetorian guard to manage his/her legacy*

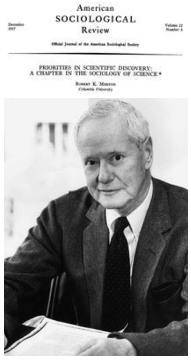
## Event study pictures




## *SCIENTIFIC COMPETITION*

## *How are scientists rewarded for novel discoveries?*

*"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)**


American
SOCIOLOGICAL
Review

PRIORITIES IN SCIENTIFIC DISCOVERY:
A CHAPTER IN THE SOCIOLOGY OF SCIENCE*

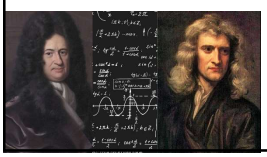
ROBERT K. MERTON
Columbia University

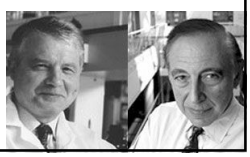
## The nature of scientific priority

- ***An odd type of property right***
  - *Not a direct monetary reward*
  - *Not a control right*
  - *Simply a "thin" intellectual property right – "the recognition by others of the scientist's distinctive part into having the result brought into being."*
- ***Adjudicated itself by the scientific community***
  - *Eponymy rather than anonymity*
- ***The potential for mischief***
  - *Fraud*
  - *Plagiarism*

## Priorities in Scientific Discovery [1957]

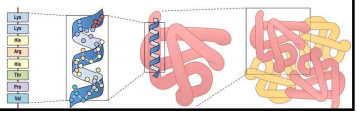
- ***The history of science is replete with intense (and intensely complicated) disputes over scientific priority – who was the particular person to make a particular discovery***
  - *This is not simply a matter of egotism – many disputes are fought by supposedly independent parties, and, in many cases, the subjects of the dispute stay "above the fray"*
  - *Indeed, in some (but not all) cases, researchers undertake steps to share credit or recognize others contributions (e.g., Darwin and Wallace)*
- ***The norms and behaviors to accord scientific priority reflects the fundamental interest in providing a reward for originality***
  - *But balanced against the competing norm of humility*





## Priority races: Empirical evidence

- ***In a pair of papers, Ryan Hill and Carolyn Stein provide the first systematic empirical look at priority races***
- ***"Scooped! Estimating Rewards for Priority in Science"***
  - *What is the causal effect of losing a priority race on project and scientist outcomes?*
- ***"Race to the Bottom: Competition and Quality in Science"***
  - *The dark side of competition: scientists may cut corners and reduce quality in their pursuit to publish first*
- ***Both papers use leverage the same data and setting: structural biology and the Protein Data Bank (PDB)***



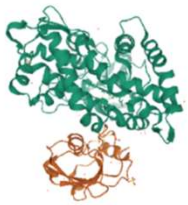
## Scooped!

- ***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*
- ***Does the priority reward system reinforce inequality in science?***
  - *Is the scoop effect equal for high- and low-reputation teams?*
- ***Key empirical challenges***
  - *Need a setting with well-defined problems and "one right answer"*
  - *Need an objective measure of scientific proximity*
  - *Need a view of potential abandonments prior to publication*
- ***The paper analyzes more than 1,500 priority races in structural biology using the Protein Data Bank (PDB)***

## Research Design

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

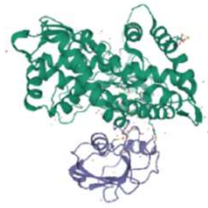
Winning Deposit: 4JWS



Affiliation: UC Irvine
Deposit Date: March 27, 2013
Release Date: June 19, 2013

Journal: *Science*
Journal Impact Factor: 31.5
5-year Citations: 52

Scooped Deposit: 3W9C



Affiliation: Leiden University
Deposit Date: April 3, 2013
Release Date: August 21, 2013

Journal: *Journal of Molecular Biology*
Journal Impact Factor: 4
5-year Citations: 39

## Results

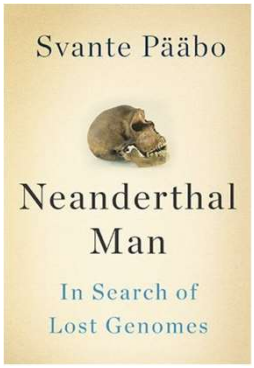
- ***Priority paper gets 54% of total citations and scooped paper gets 46%***
  - *Surveyed scientists are much more pessimistic: 74% to 26%.*
  - *Scooped projects are less likely to be published, and less likely to appear in a top-10 journal*
- ***In the next five years, scooped scientists have the same number of publications, but fewer citations***
- ***Priority system reinforces inequality:***
  - *Citation penalty is larger for low-ranked teams than it is for high-ranked teams.*

| Dependent variable | Published (1) | Std. journal impact factor (2) | Top-ten journal (3) | asinh(Five-year citations) (4) | Top-10% five year citations (5) |
|---|---|---|---|---|---|
| Scooped | -0.025*** | -0.178*** | -0.060*** | -0.197*** | -0.035*** |
| | (0.010) | (0.032) | (0.014) | (0.045) | (0.010) |
| Winner Y mean | 0.880 | -0.031 | 0.318 | 28.918 | 0.150 |
| Observations | 3,319 | 3,319 | 3,319 | 2,546 | 2,546 |

*Note:* All regressions include controls selected by PDS-Lasso as well as year fixed effects. Unpublished papers have impact factor imputed to minimum factor journal. Citation regressions restricted to papers published before 2014. Column 4 dependent variable is asinh(five-year citations) but mean citations is reported in levels.

## *Race to the bottom*

*"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"*



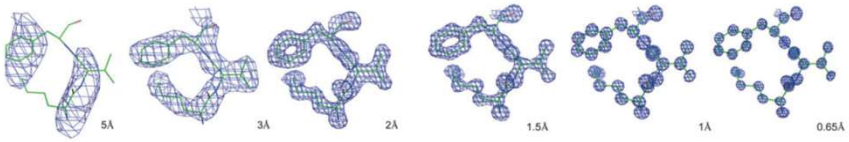

– Svante Pääbo, Neanderthal Man: In Search of Lost Genomes

## *Race to the Bottom (cont'd)*

- ***Hill & Stein build a model with the following predictions:***
  - *Most (ex-ante) important projects are more competitive, rushed, and lower quality*
- ***They find that***
  - *High-potential projects are more competitive (multiple researchers working simultaneously)*
  - *High-potential projects are completed faster and are lower quality*
  - *Follow-on work ameliorates but does not eliminate the negative relationship between potential and quality*
  - *Quality magnitudes large enough to impact usefulness of projects for drug development*

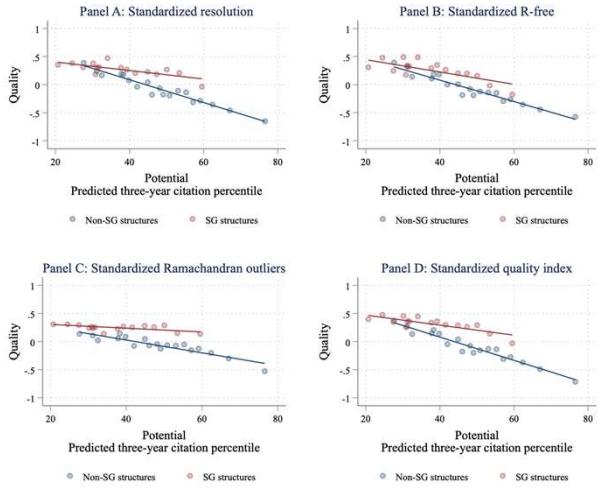
## Measurement challenge

- ***A unique feature of structural biology is the objective, ex-ante 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. *Outliers: errors in the model based on chemical properties*




- ***They combine these outcomes into a standardized quality index (higher is better)***

## Results in one picture




# *SCIENCE AND ITS INSTITUTIONS*

## *Institutions and the rate and direction of scientific advance*

- ***Recall the "ideas production function" from Jones***

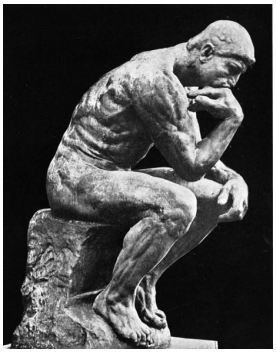
$$\dot{A} = f[A(z), H(z), K(z), z]$$

- ***Broad view of what counts as an institution***
  - *Editorial policies*
  - *Replicability rules*
  - *Funding rules and systems*
  - *Access to capital equipment and materials…*

- ***What is the impact of specific institutions on science?***

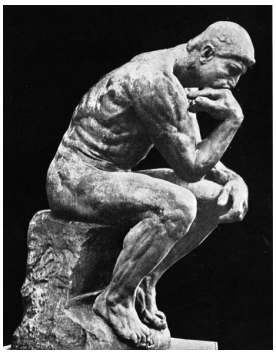
## *The investigator-initiated scientific grant: A peculiar form of contract*

- ***Scientists need $$ to do research***
- ***One way to fund research is a peculiar kind of contract: the scientific grant (Azoulay & Li 2022)***
- ***But not all grant systems are created equal***
  - *Targeted at projects or individuals?*
  - *How renewed?*
- ***Leads naturally to economic interests***
  - *How does contract design relate to the "importance" of research being undertaken?*



## *Azoulay et al. (2011) on science funding, tolerance for failure, and scientific exploration*

- ***A setting in which agents are at risk of receiving different type of grants***
  - *which embed different type of incentives*
- ***A way to measure the "quality" of ideas (i.e., the tail)***
- ***An experiment: a set of identical agents who receive only standard "exploitation" incentives***



## *Howard Hughes Medical Investigator Program*

- ***Most important private source of funding for academic biomedical research***
- ***HHMI selects about 50 "young" (i.e., cusp-of-tenure or recently tenured) life scientists from elite institutions every 3 years***
- ***Very prestigious accolade***
- ***Major source of funding for selected scientists***
- ***But more than a prize, a program:***
  - *"push the boundaries of science"*
  - *"people, not projects"*
  - *renewal every 5 years (with 2 year phase down), but first review rather lax*
  - *intensive monitoring and evaluation*

## *Incubating future Nobelists...*

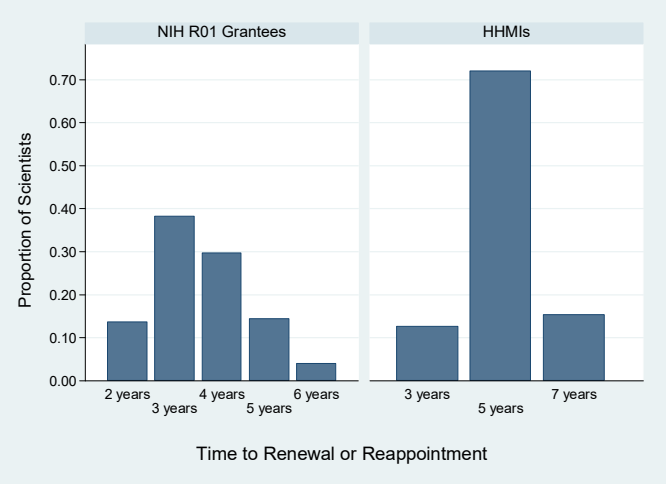
## *Why HHMI?*

- ***Program features match closely the characteristics of incentive systems that Manso (2011) claims should encourage exploration***
- ***But important to recognize that the program could have other effects as well, e.g., anointment***

## *Exploitation incentives: NIH funding*

- ***R01 grants from the NIH***
  - *support particular projects, not individuals*
  - *must be renewed every 3-5 years*
    - *No points for "trying hard"*
    - *Low-cost monitoring*
    - *Probability of renewal shrouded in uncertainty (where will the "pay line" be in 3/4/5 years?)*
  - *common criticism: provides incentives to choose less risky topics (Kolata 2009)*

## *Time horizons*




## *Comparison between the two sources of funding*

| NIH R01 Grants | HHMI Investigator Program |
|---|---|
| 3-year funding | 5-year funding |
| first review is similar to any other review | first review is rather lax |
| funds dry up upon non-renewal | two-year phase-down upon non-renewal |
| some feedback in the renewal process | feedback from renowned scientists |
| funding is for a particular project | "people, not projects" |

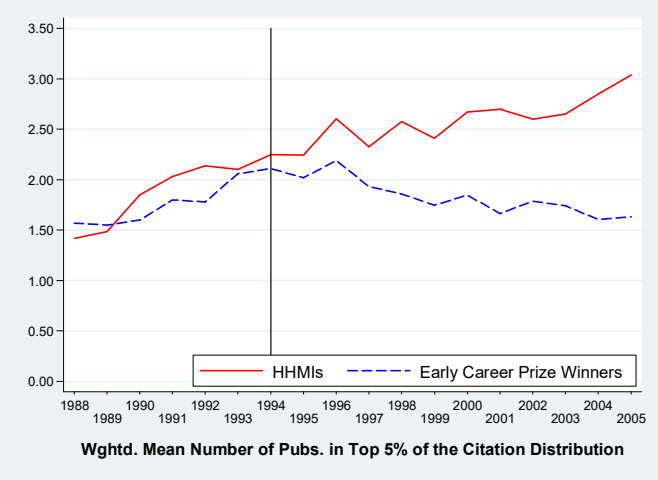
## Challenges

- ***How Does One Measure "Creativity"?***
- ***Selection vs. Treatment***
- ***Interpretation of Treatment Effect: Incentives vs. ...***

## Measuring creativity

- ***Creativity := Outcomes "in the Tail"***
  - *Number of papers in top quantiles of the citation distribution*
  - *Prizes (Nobel, Lasker, election to NAS, IoM…)*
  - *Grad Student/Postdoc Placement*
- ***Creativity := Frequency of "Strike-outs"***
  - *Number of papers in the bottom quartile of the citation distribution*
- ***Creativity as Process: Extent of "Branching Out" or Recombination (Weitzman 1998; Burt 2004; Simonton 2004)***
  - *…relative to the scientist's own pre-appointment output*
  - *…relative to the world's scientific frontier*

## Results in one picture



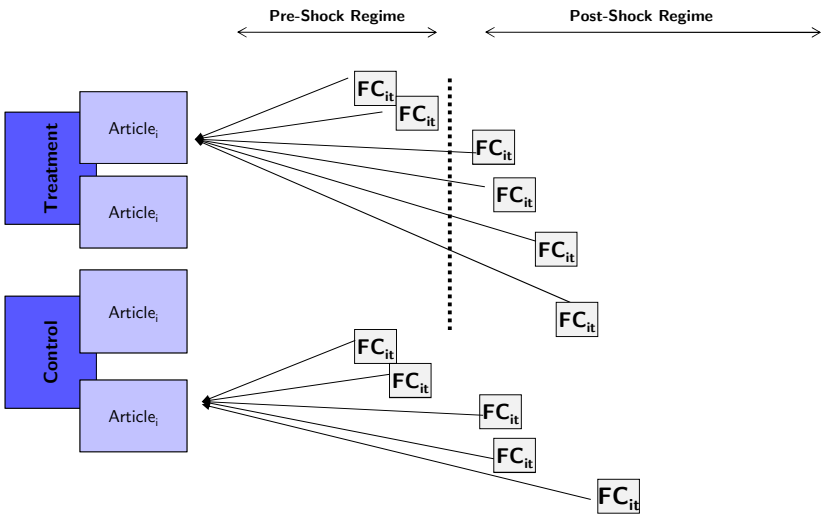

## Problems with the AGZM evidence

- ***Poor man's identification strategy***
  - *selection on observables*
  - *combined with differencing*
- ***Can't distinguish the effect of $$ from the effect of longer time horizons/rich feedback/freedom to experiment***
- ***Can't distinguish between the incentives and sorting effects of HHMI appointments (best postdocs/grad students seek out HHMIs)***
- ***Can't filter out the effect of anointment***
  - *But other evidence suggests these effects are small (Azoulay, Stuart, & Wang 2010)*
- ***Collaboration contaminates the control group***
  - *Downward bias?*

## *Furman and Stern (AER 2011)*

- ***Question:*** ***How do institutional forms influence the disclosure of knowledge, with implications for cumulativeness and the capacity to harness potential spillovers?***
- ***Identification Strategy:*** ***Differences-in-differences. Take a fixed piece of knowledge (e.g., a paper). Examine changes in citation behavior before and after some "exogenous" event (treatment). Compare to control group of similar piece of knowledge (e.g. paper with similar ex-ante citations) that does not experience treatment.***
- ***Setting:*** ***Biological resource centers (BRCs). Deposit of organisms en masse in BRC (exogenous event) allow other researchers to utilizes these organisms.***

## *Research design*




## *Figure 2: Citation effect of BRC deposit*

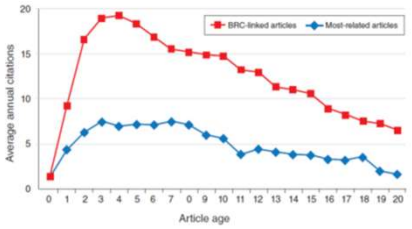



FIGURE 1. AVERAGE ANNUAL CITATIONS BY AGE, BRC VERSUS CONTROL ARTICLES

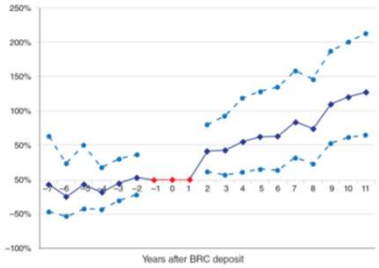



FIGURE 2. PRE- AND POSTDEPOSIT EFFECTS ON FORWARD CITATIONS