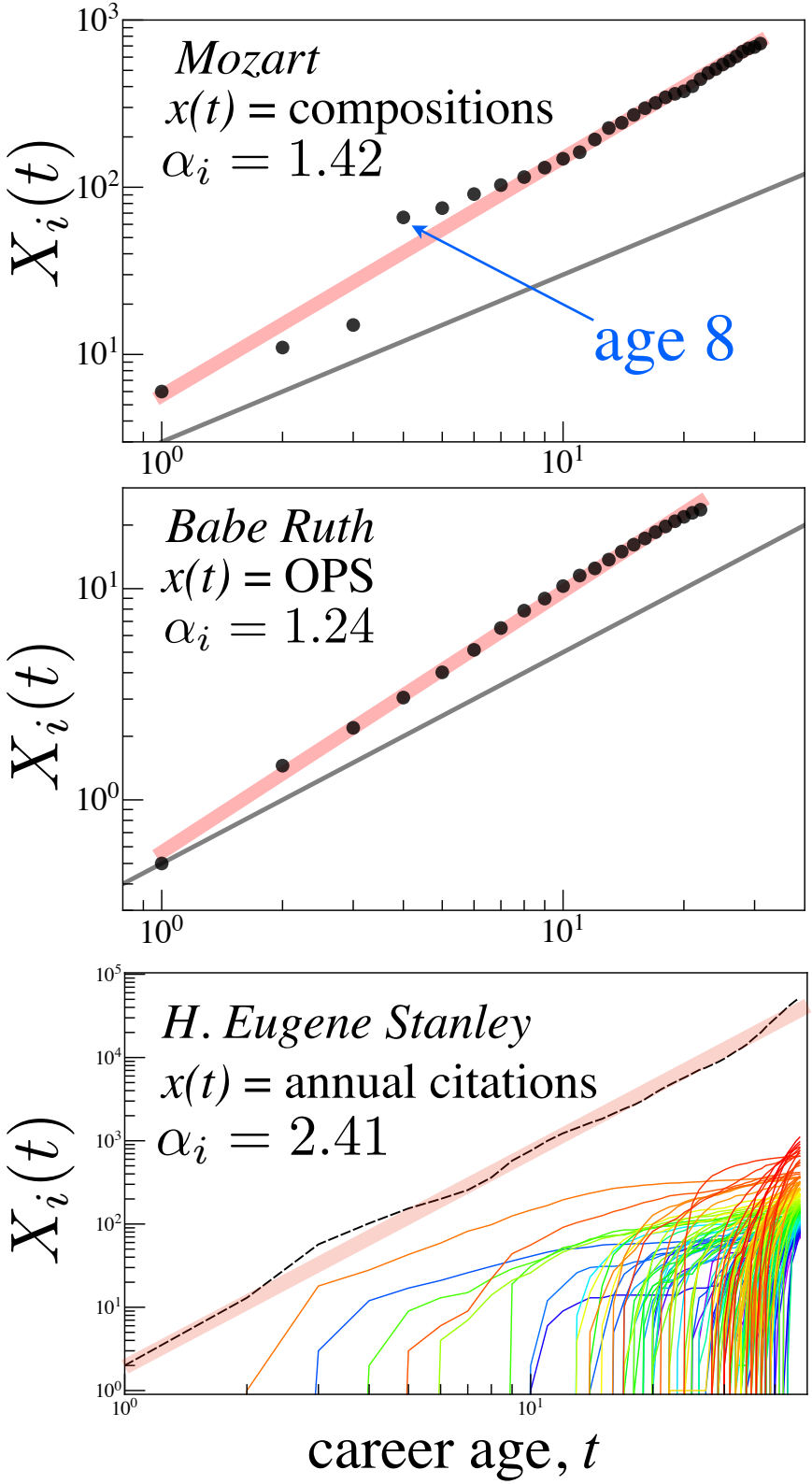
Ascent in competitive arenas: From Fenway Park to Mass Ave

The Science of Success: *Measurements and Predictions*

June 17th Harvard University

Alexander M. Petersen

IMT Institute for Advanced Studies, Lucca Italy



stellar career growth is
a non-linear process

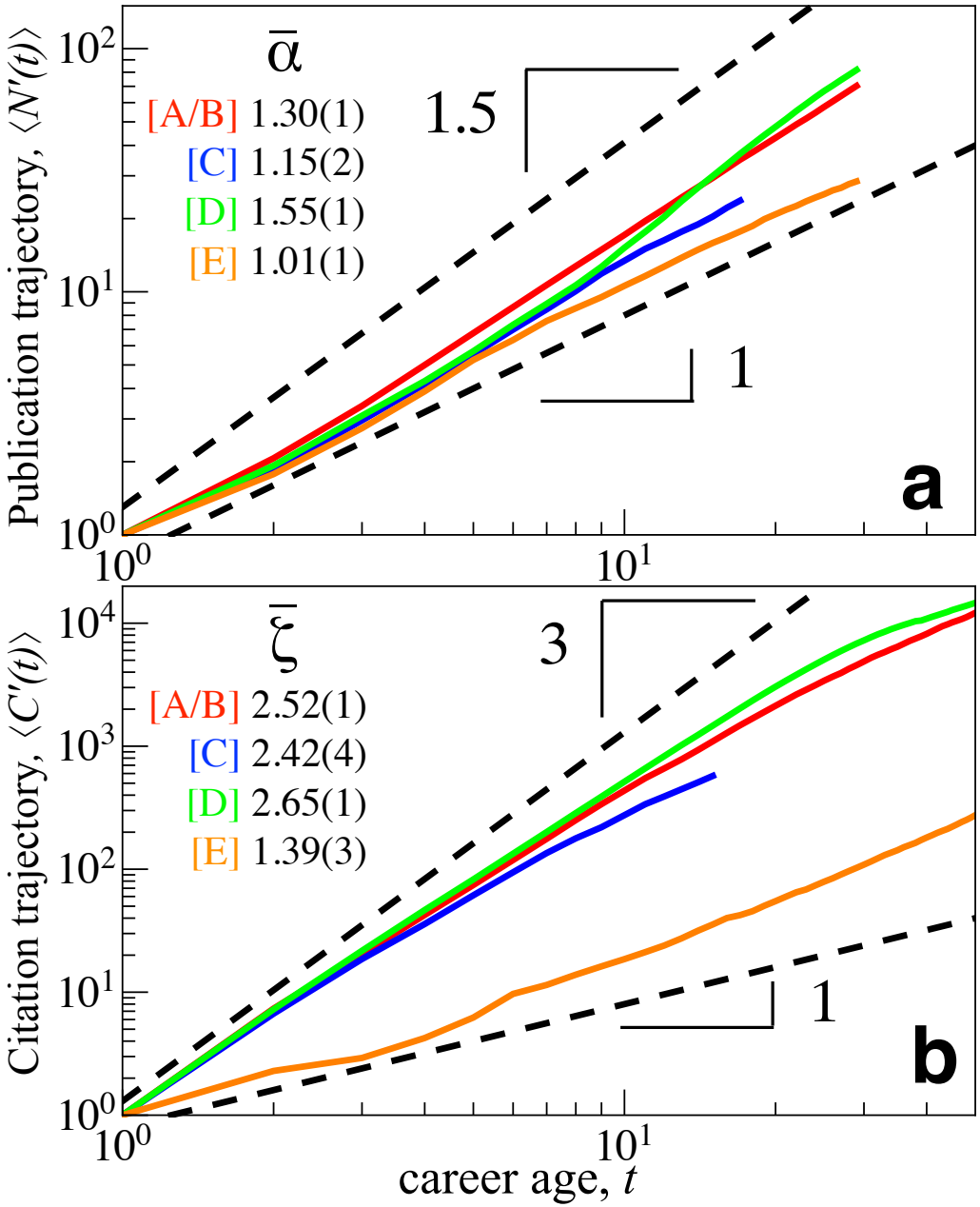
cumulative output ~ reputation

$$X_i(t) \equiv \sum_{\tau=1}^{t} x_i(\tau) \sim t^{\alpha_i}$$

$\alpha_i \approx 1 \Rightarrow$ *constant output, no growth*

$\alpha_i > 1 \Rightarrow$ *super-linear growth*

Growth patterns in "superstar" academic careers



Cumulative reputation measures:

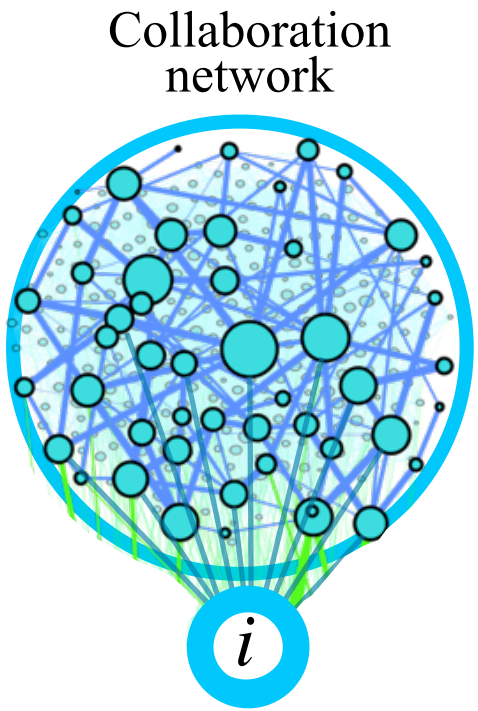
(normalized) publication trajectory $\langle N'(t) \rangle \sim t^{\bar{\alpha}}$

(normalized) cumulative citation trajectory $\langle C'(t) \rangle \sim t^{\bar{\zeta}}$

$\zeta > \alpha > 1 \Rightarrow$ *super-linear growth*

Cumulative advantage ~
careers become "attractors" of new opportunities instead of "pursuers"

What makes science special (complex)?



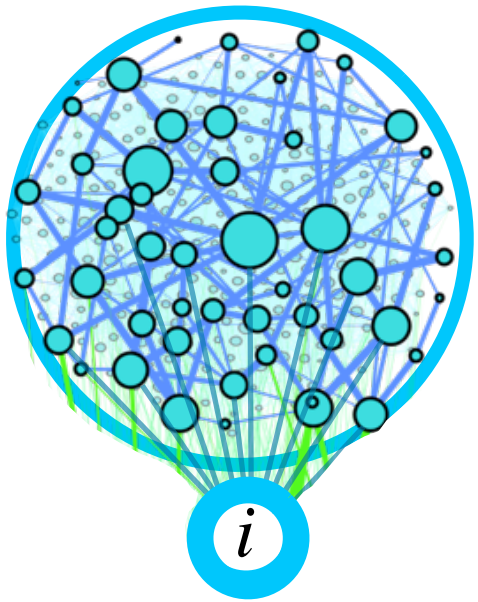
principal
investigator

Interactions mediated by social "forces":

- Collaboration (attractive)
- Competition for priority (repulsive)
- Knowledge (an "exchange particle")

Diverse collaboration strategies

Collaboration network



Interactions mediated by social "forces":

- Collaboration (attractive)
- Competition for priority (repulsive)
- Knowledge (an "exchange particle")

diverse collaboration strategies even within the same field!

Watson-Crick strategy:

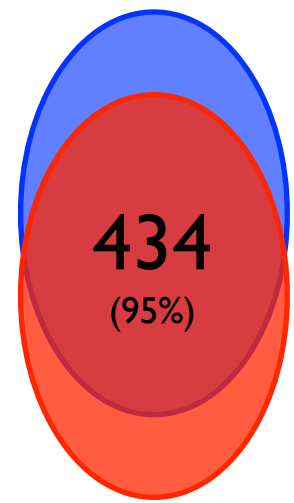
* **Michael Stuart Brown**
* **Joseph L. Goldstein**

Recipients of the 1985 Nobel Prize in Physiology or Medicine for describing the regulation of cholesterol metabolism.

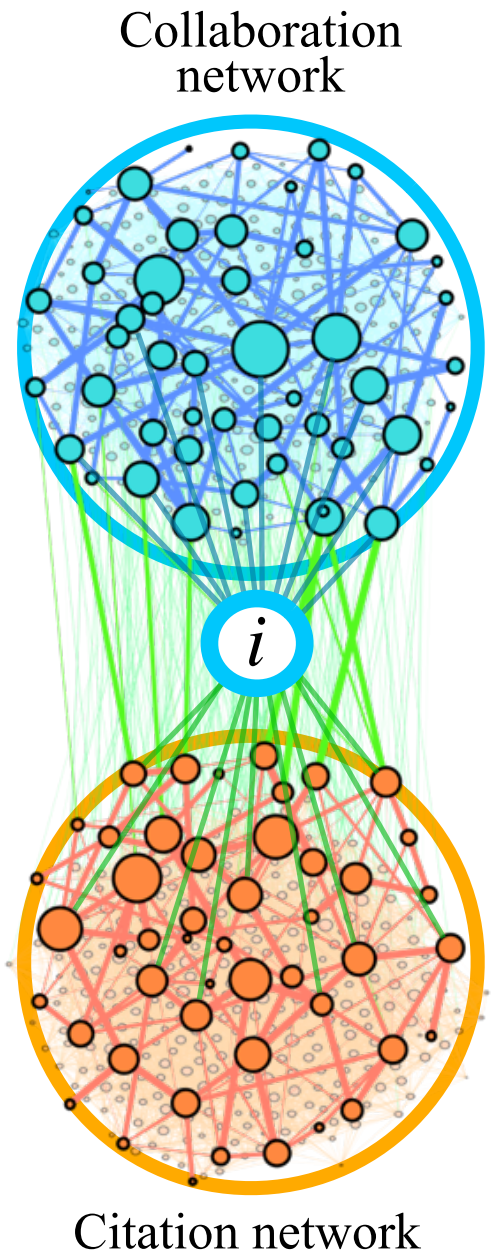
Solo-artist strategy:

* **Marilyn Kozak**

$N = 70, N_{solo} = 59$ (84%)



Co-evolving network of networks



Complexity

- coevolutionary system
- behavioral components
- embedded social processes
  - reputation
  - economic incentives

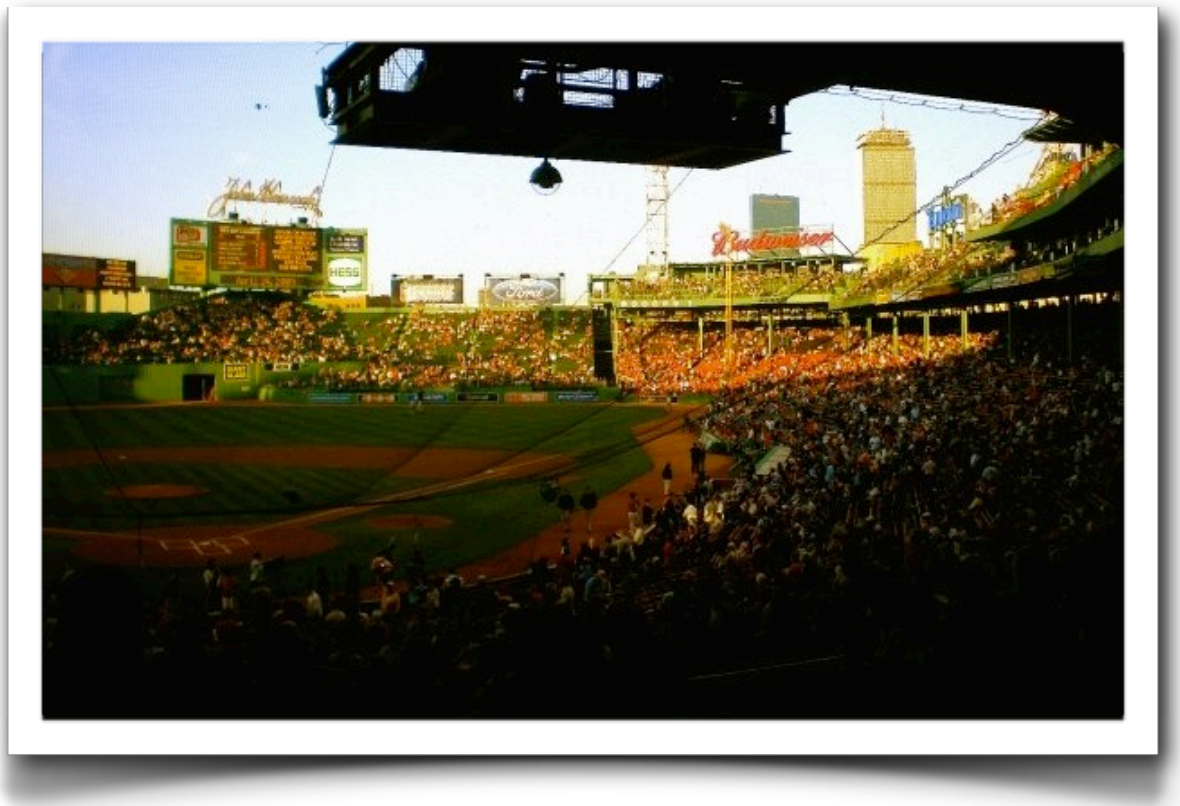
Reputation and Impact in Academic Careers, ArXiv: *1303.7274*
A. M. Petersen, S. Fortunato, R. K. Pan, K. Kaski, O. Penner,
M. Riccaboni, H. E. Stanley, F. Pammolli

Institutions Markets Technologies
IMT
INSTITUTE FOR ADVANCED STUDIES LUCCA

Competitive arenas

Physical Review Letters
moving physics forward

Science
AAAS



nature

Cell

The NEW ENGLAND JOURNAL of MEDICINE

PNAS
Proceedings of the National Academy of Sciences of the United States of America

Competitive arenas

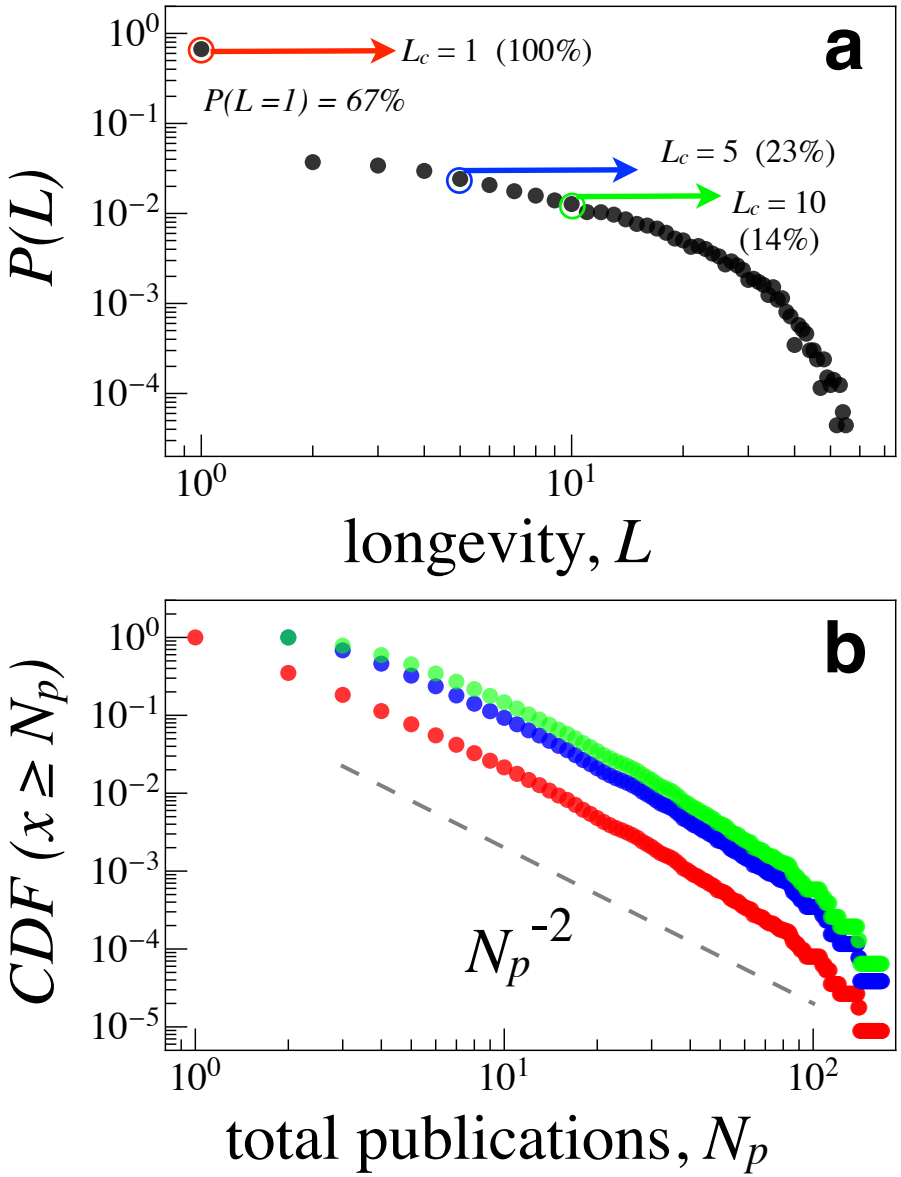
Disambiguation strategy: Analyzed subset of "rare" surnames

Journal	Years	Articles	Authors, N^j
CELL	1974-2012	12,349	19,491 (1,753)
Nat./PNAS/Sci.	1958-2012	219,656	112,777 (14,478)
NEJM	1958-2012	18,347	33,149 (2,897)
PRL	1958-2012	98,739	55,827 (10,206)

TABLE I: Summary of journal datsets. N^j is the number of unique surnames we were able to identify in each journal j over the denoted period. The N^j value in parentheses denotes the number of careers with $L_i \geq 5$.

Peering inside the high-impact arena

Nature/PNAS/Science



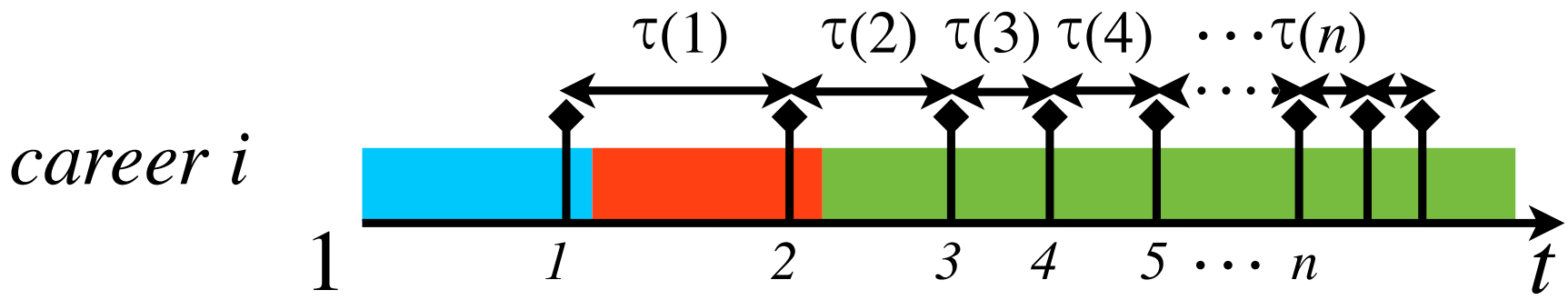
Basic measures for survival and achievement

Longevity $L_i^j \equiv t_{i,f}^j - t_{i,0}^j + 1$ *in a given journal set* is extremely right-skewed, in agreement with the quantitative predictions of a rich-get-richer career progress model

Likewise, since production is highly correlated with longevity, the distribution of cumulative publications is also extremely right-skewed (Lotka's law)

"Cumulative advantage"

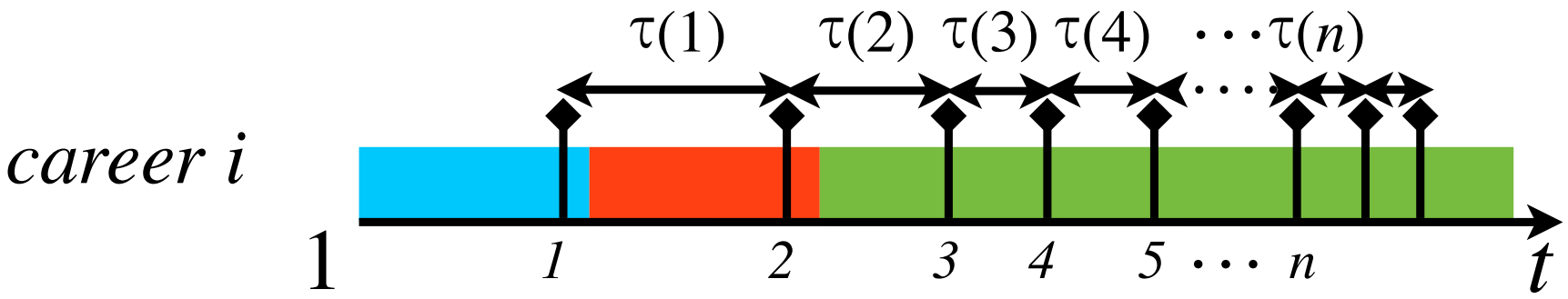
For each career *i* we track his/her longitudinal publication rate by aggregating over publications in a *specific set* of high-impact journals

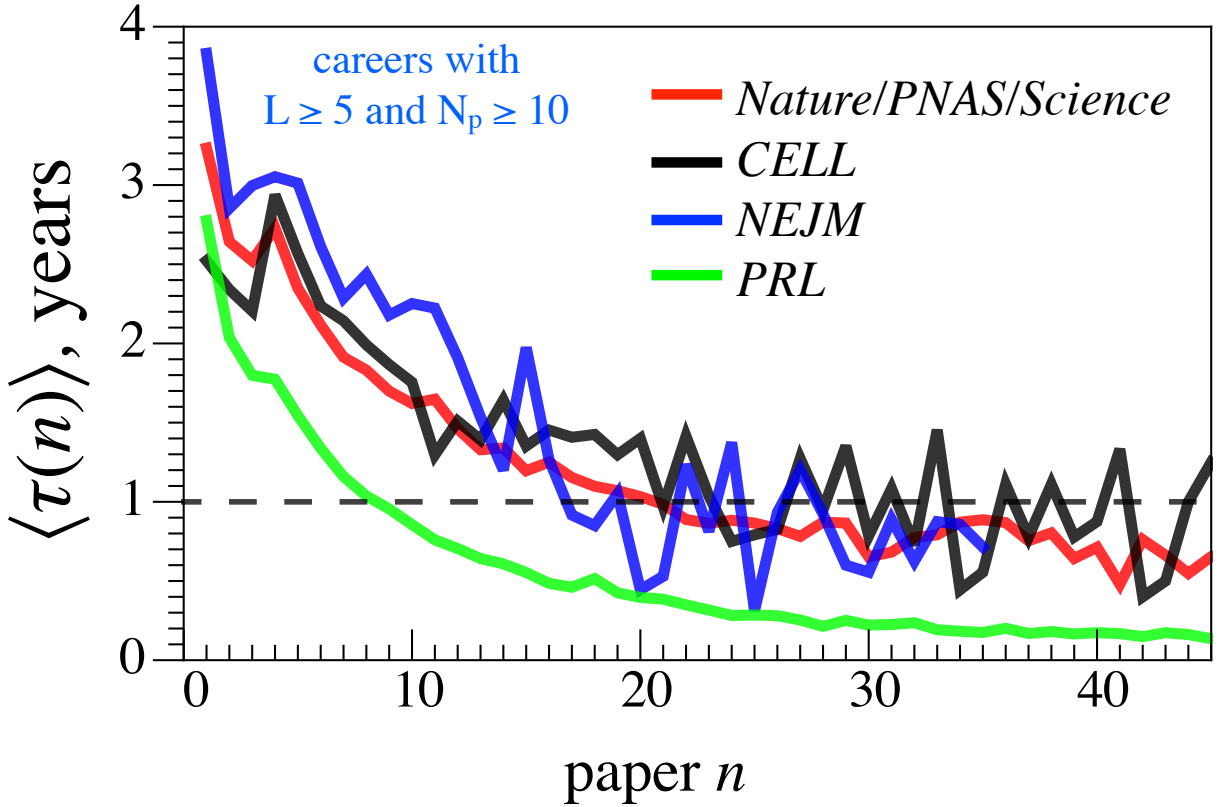


Q: What is the characteristic waiting time $\tau_i(n)$ between an author's n^{th} paper and $(n+1)^{th}$ paper?

"Cumulative advantage"

For each career *i* we track his/her longitudinal publication rate by aggregating over publications in a *specific set* of high-impact journals



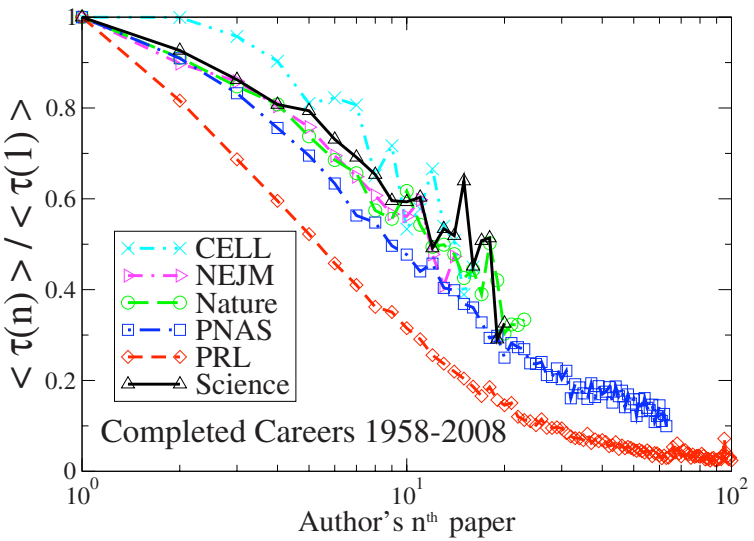


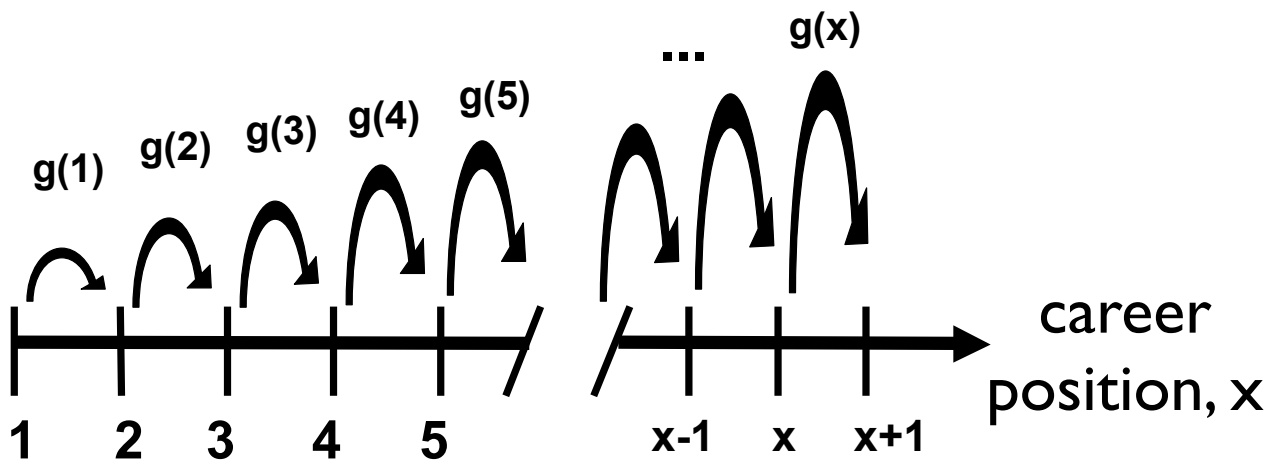
Q: What is the characteristic waiting time $\tau_i(n)$ between an author's n^{th} paper and $(n+1)^{th}$ paper?

By the 10th paper, the waiting time between publications has decreased by ~ factor of 2!

Modeling the "Rich-get-richer" effect

- Forward progress follows a stochastic "progress rate" $g(x)$
- Cumulative advantage: $g(x)$ increases with career position x





$$g(x) = 1 / \langle \tau(x) \rangle$$

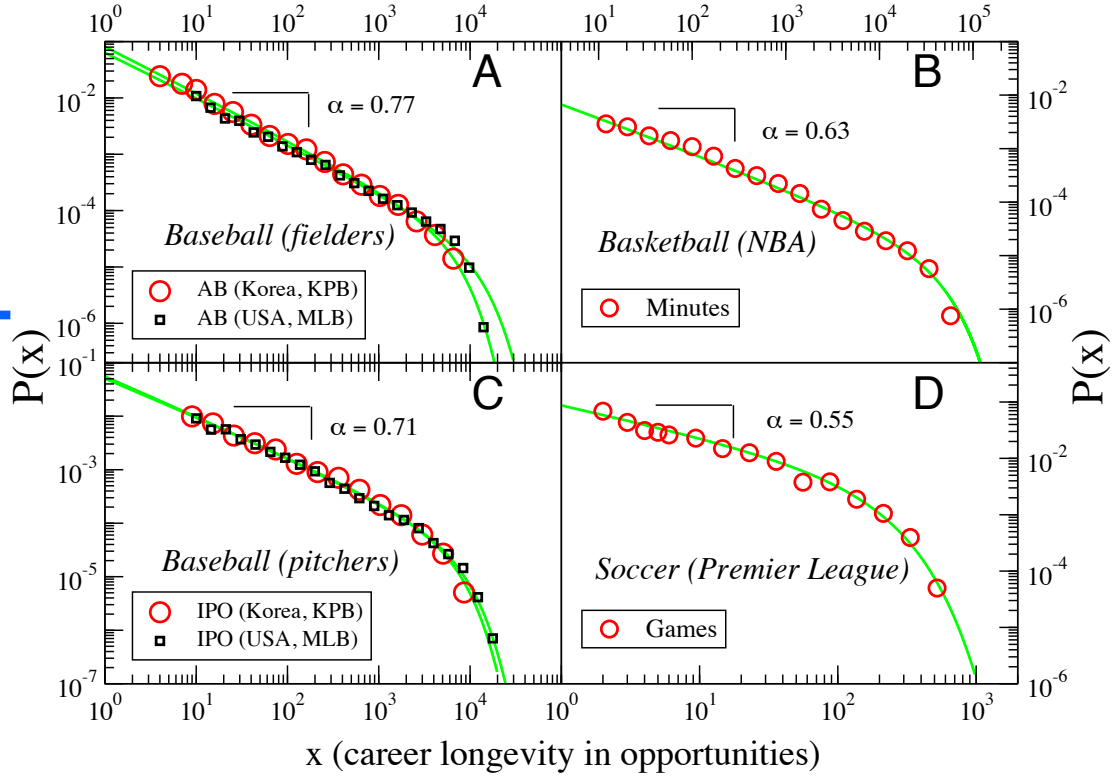
The progress probability g *is the inverse of the mean waiting time* τ

Methods for measuring the citations and productivity of scientists across time and discipline, A. M. Petersen, F. Wang, H. E. Stanley. Phys. Rev. E 81, 036114 (2010).

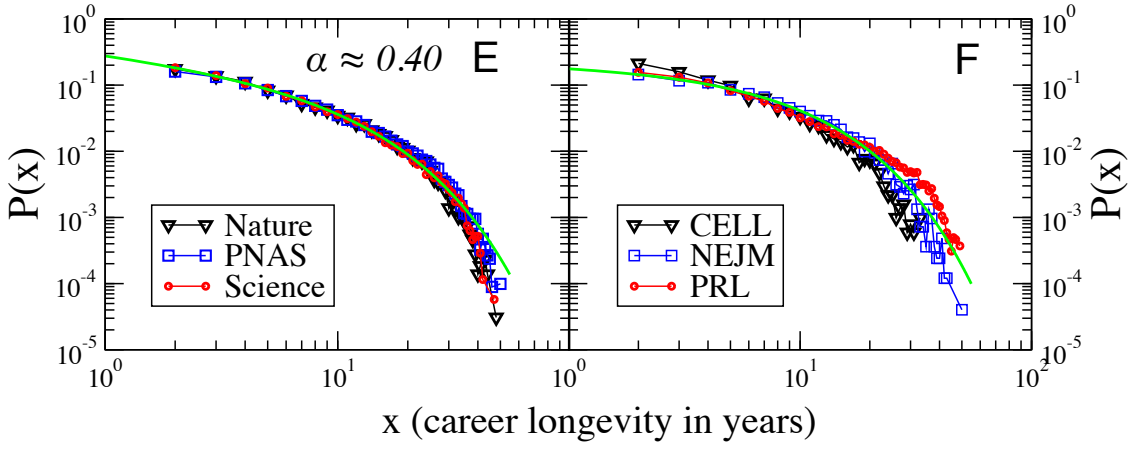
Quantitative and empirical demonstration of the Matthew effect in a study of career longevity. A. M. Petersen, W.-S. Jung, J.-S. Yang, H. E. Stanley. Proc. Natl. Acad. Sci. USA 108, 18-23 (2011).

Statistical regularities in the career longevity distribution

Pro Sports



Academia



opportunities ~ time duration

Major League Baseball

- 130+ years of player statistics, ~ 15,000 careers

``One-hit wonders"

- 3% of all fielders finish their career with ONE at-bat!
- 3% of all pitchers finish their career with less than one inning pitched!

``Iron horses"

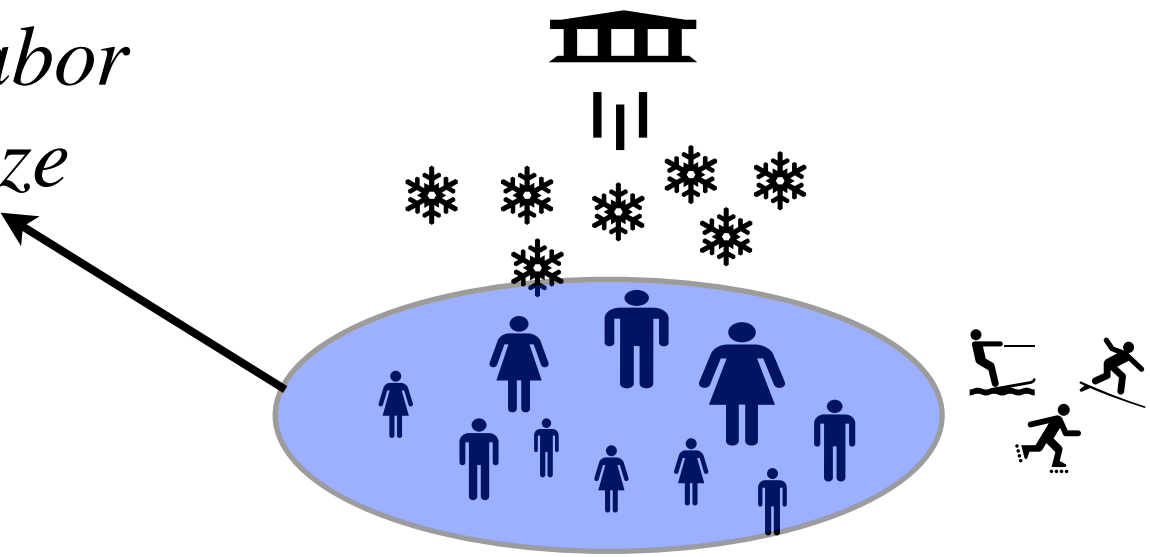
- Lou Gehrig (the Iron Horse): NY Yankees (1923-1939)
- Played in 2,130 consecutive games in 15 seasons! 8001 career at-bats!
- Career & life stunted by the fatal neuromuscular disease, amyotrophic lateral sclerosis (ALS), aka Lou Gehrig's Disease

Modeling competition

Agent-based model of competition with achievement appraisal

Achievement measured by $n_i(t)$, the number of opportunities (ex. publications) captured in time period t

I = *finite labor force size*



Persistence and Uncertainty in the Academic Career,
A. M. Petersen, M. Riccaboni, H. E. Stanley, F. Pammolli.
Proc. Natl. Acad. Sci. USA 109, 5213-5218 (2012).

Appraising prior achievement

Achievement measured by $n_i(t)$, the number of opportunities captured in time period t

The cohort of I agents compete for a fixed number of opportunities in each period over a lifespan of $t = 1 \ldots T$ periods.

In each period, the capture rate of a given individual i is calculated by an appraisal of the achievement history

capture rate

$$w_i(t) \equiv \sum_{\Delta t=1}^{t-1} n_i(t-\Delta t) e^{-c\Delta t}$$

exponential discount factor

Appraisal timescale $1/c$

$c \to 0$: appraisal over all lifetime achievements (~ tenure system)

$c > 1$: appraisal over only recent achievements (short-term contract system)

Crowding out by "kingpins"

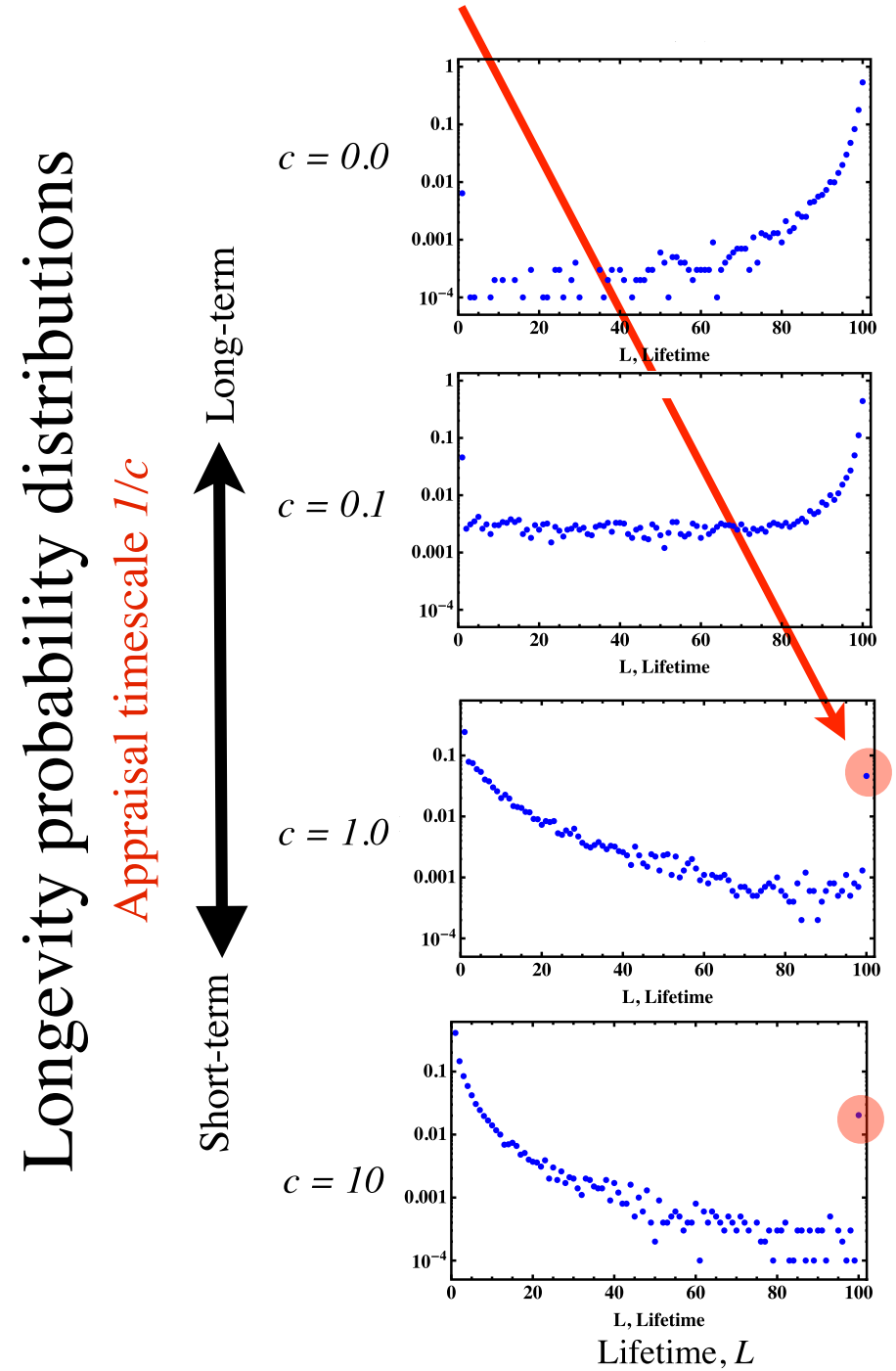
Our theoretical model suggests that

short-term appraisal systems:

* can amplify the effects of competition and uncertainty making careers more vulnerable to early termination, not necessarily due to lack of individual talent and persistence, but because of random negative production shocks.

* effectively discount the cumulative achievements of the individual.

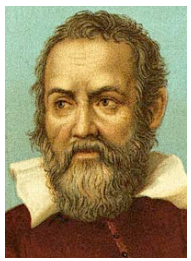
* may reduce the incentives for a young scientist to invest in human and social capital accumulation.

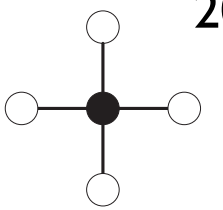


Institutions Markets Technologies
IMT
INSTITUTE FOR ADVANCED STUDIES LUCCA

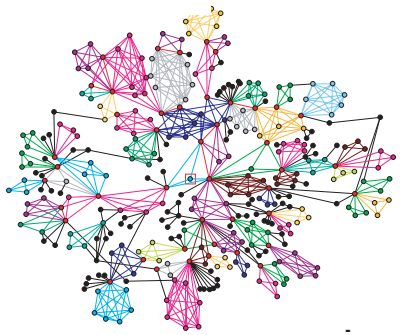
Institutional trends in Science

- emergence of small-world time-dependent collaboration networks with the increasing role of team-work in science





200+ years



G. Palla, A.-L. Barabasi, T. Vicsek. Quantifying social group evolution. Nature 446, 664-667 (2007)

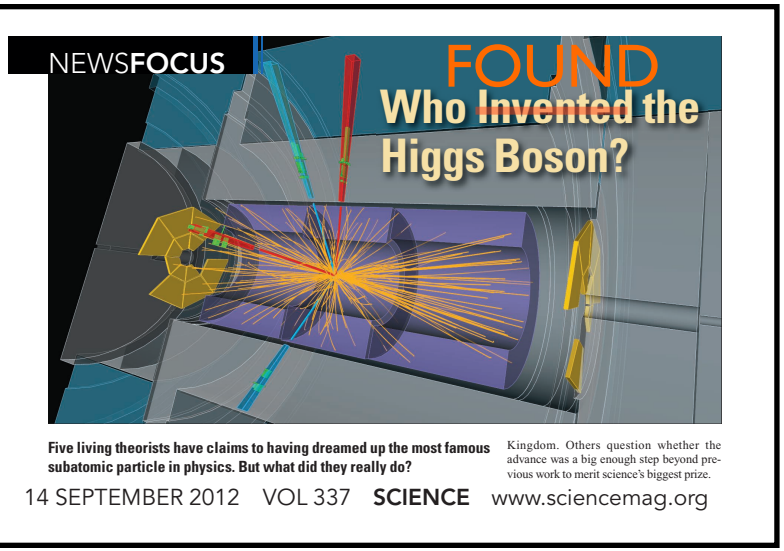
S. Wuchty, B. F. Jones, B. Uzzi. The increasing dominance of teams in production of knowledge. Science 316, 1036-9 (2007)

P. Stephan. How Economics Shapes Science. (Harvard Univ. Press, 2012)

Paul A. David. *The Historical Origins of 'Open Science': An essay on patronage, reputation, and common agency contracting in the scientific revolution.* Capitalism and Society 3(2): Article 5 (2008).

- organizational shifts in the business structure of research universities
- shifts away from tenure towards shorter-term contracts + bottle neck in the number of tenure-track positions available
- redefining the role of teaching -vs- research faculty
- shifts in the competitive aspects of science, universities, and scientists: reputation tournaments in omnipresent (online) competition arenas

Institutional context: *Increasing team size & changing incentive system*



"50-way tie for the Nobel Prize"

www.sciencemag.org **SCIENCE** VOL 336 6 APRIL 2012

Published by AAAS

CITATION IMPACT 9 DECEMBER 2011 VOL 334 **SCIENCE**

Saudi Universities Offer Cash In Exchange for Academic Prestige

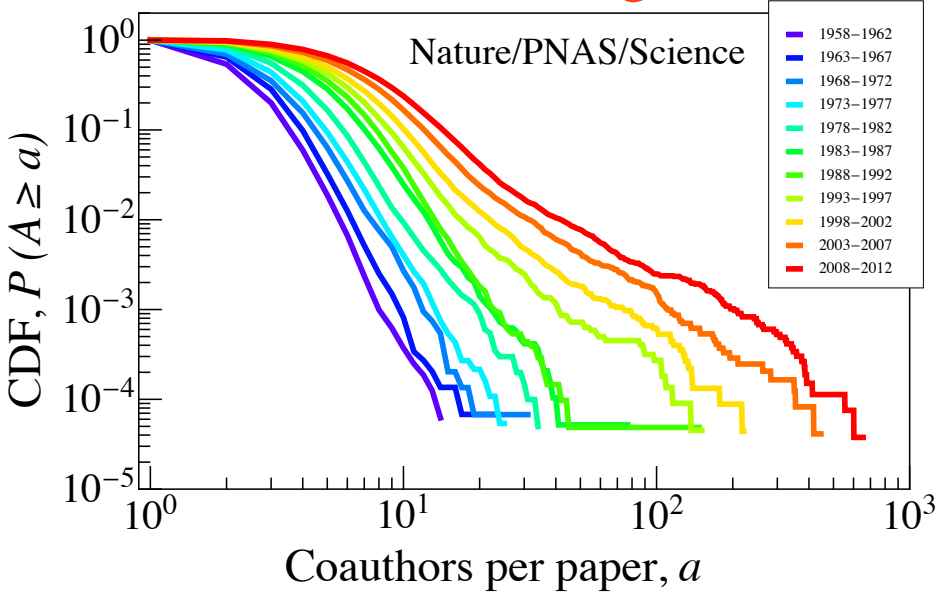
Two Saudi institutions are aggressively acquiring the affiliations of overseas scientists with an eye to gaining visibility in research journals

SCIENCE POLICY 5 AUGUST 2011 VOL 333 **SCIENCE**

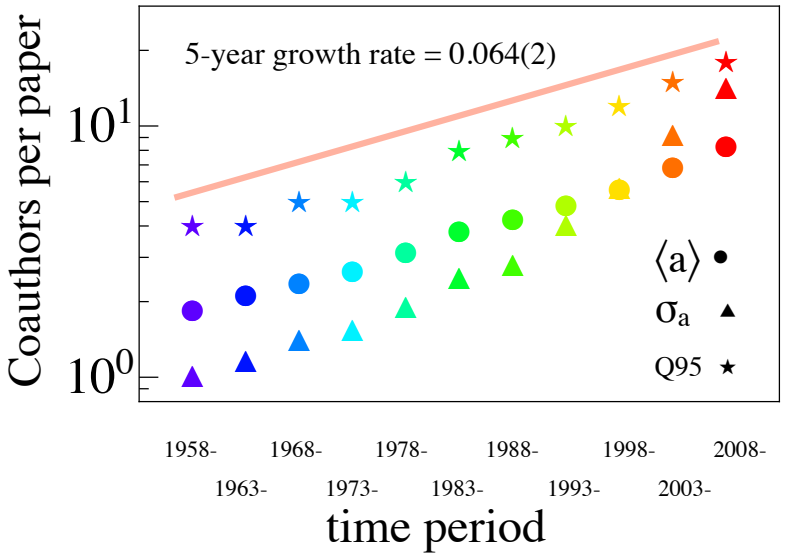
Changing Incentives to Publish

Chiara Franzoni,[1] Giuseppe Scellato,[2,3] Paula Stephan[4,5,6]*

genome.gov
National Human Genome Research Institute
National Institutes of Health



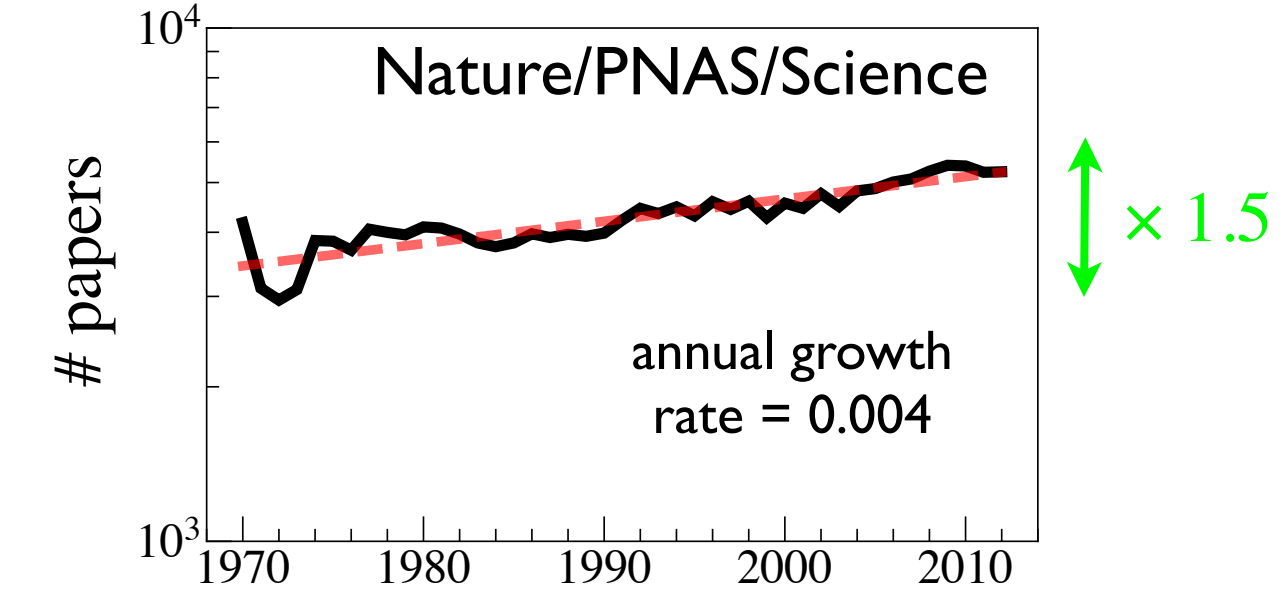
ENCODE



Q: how to "fairly" distribute credit in a system dominated by teams?

Scientific output inflation

what is the relative impact/visibility of a publication today -vs- Y years ago?

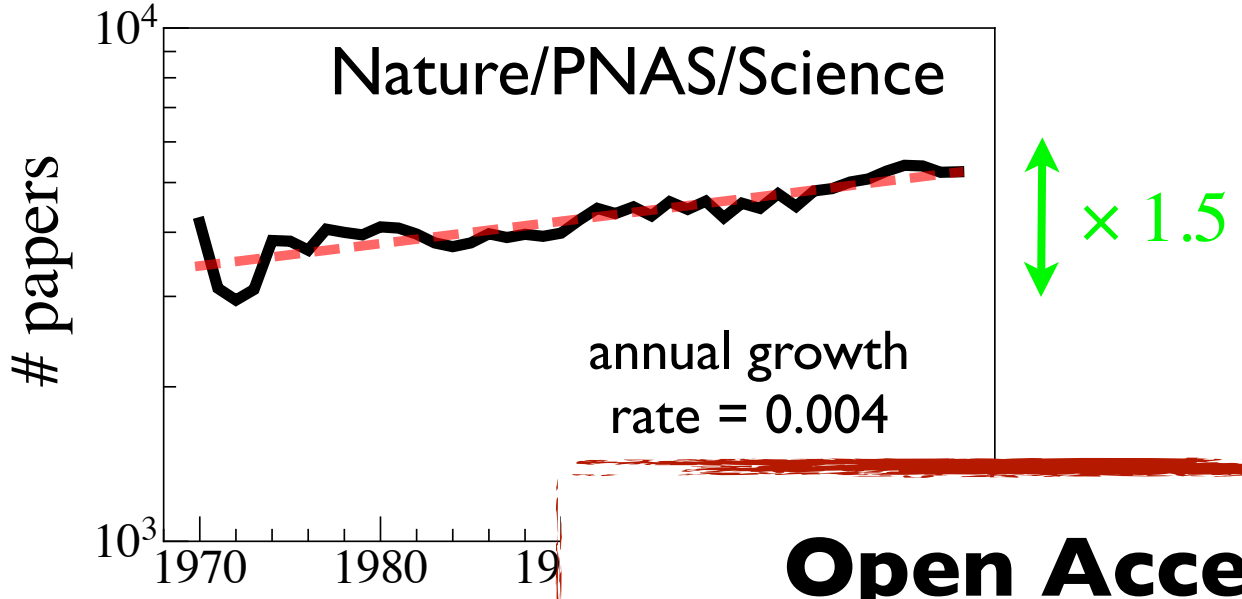


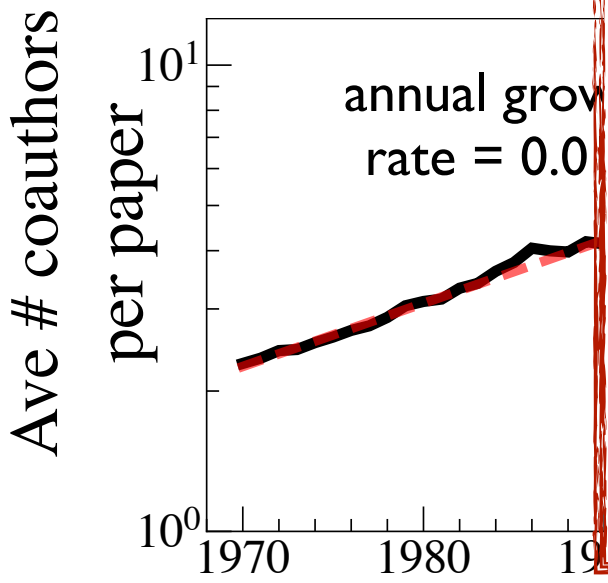
Scientific output increase due to technological factors, population growth, and "output inflation"

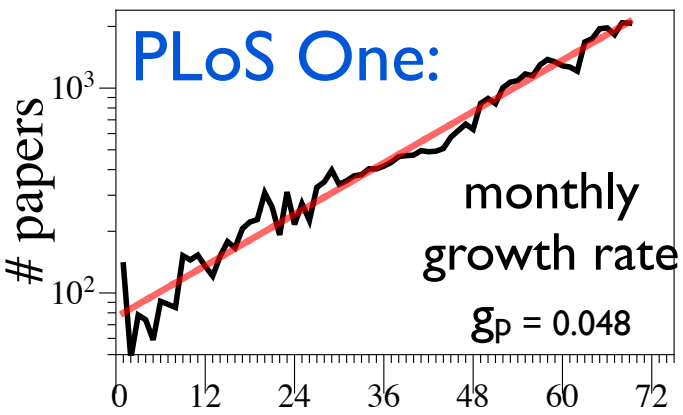
growth of team science

Scientific output inflation

what is the relative impact/visibility of a publication today -vs- Y years ago?







Open Access Journals

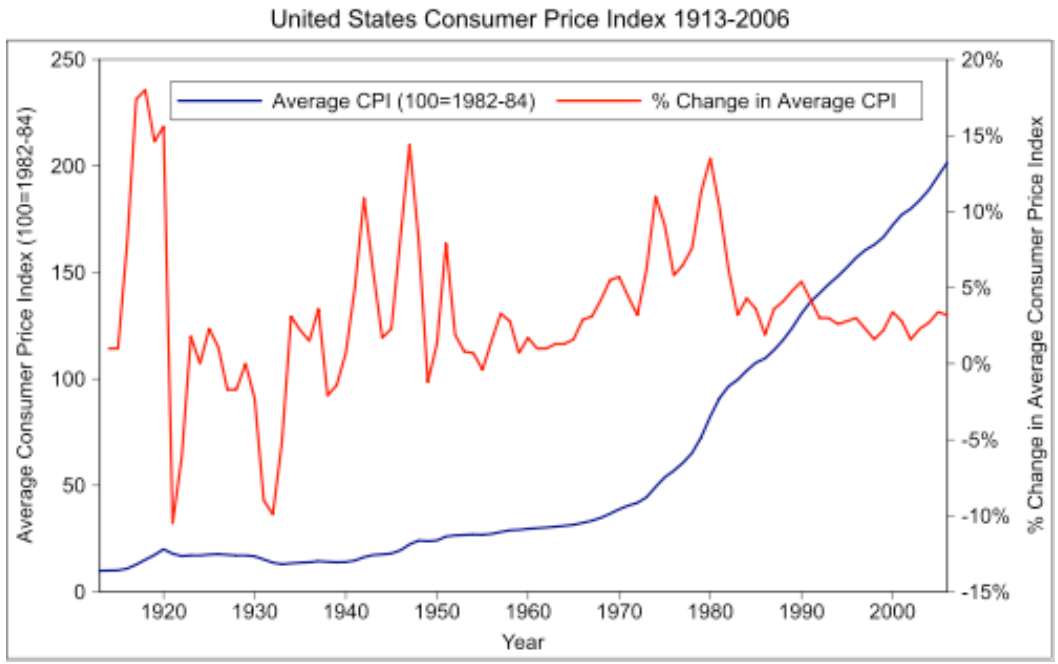
PLoS One:

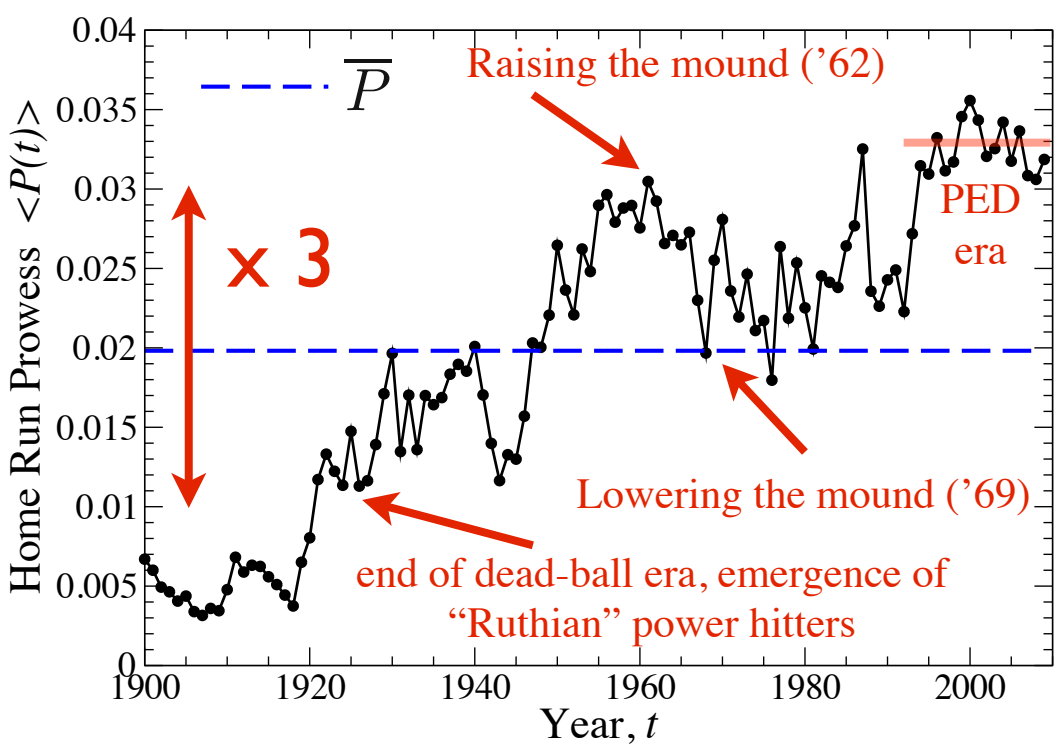
~ 6,700 articles in 2010 and ~ 14,000 in 2011

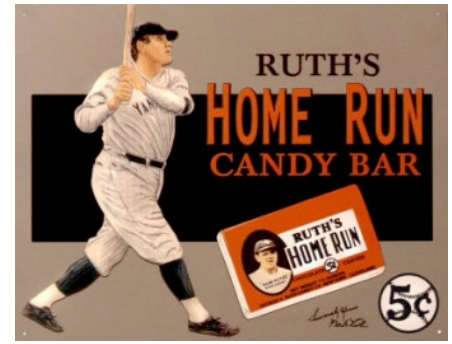
⇒ × 2 growth in one year alone!

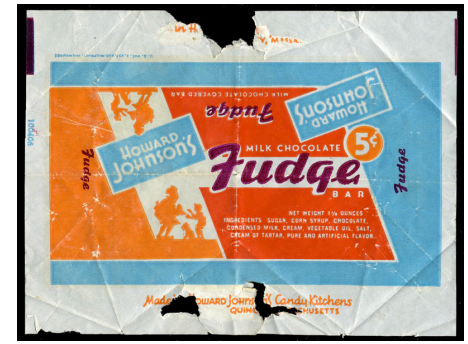
... who is reading/refereeing all these papers??

Accounting for Inflation



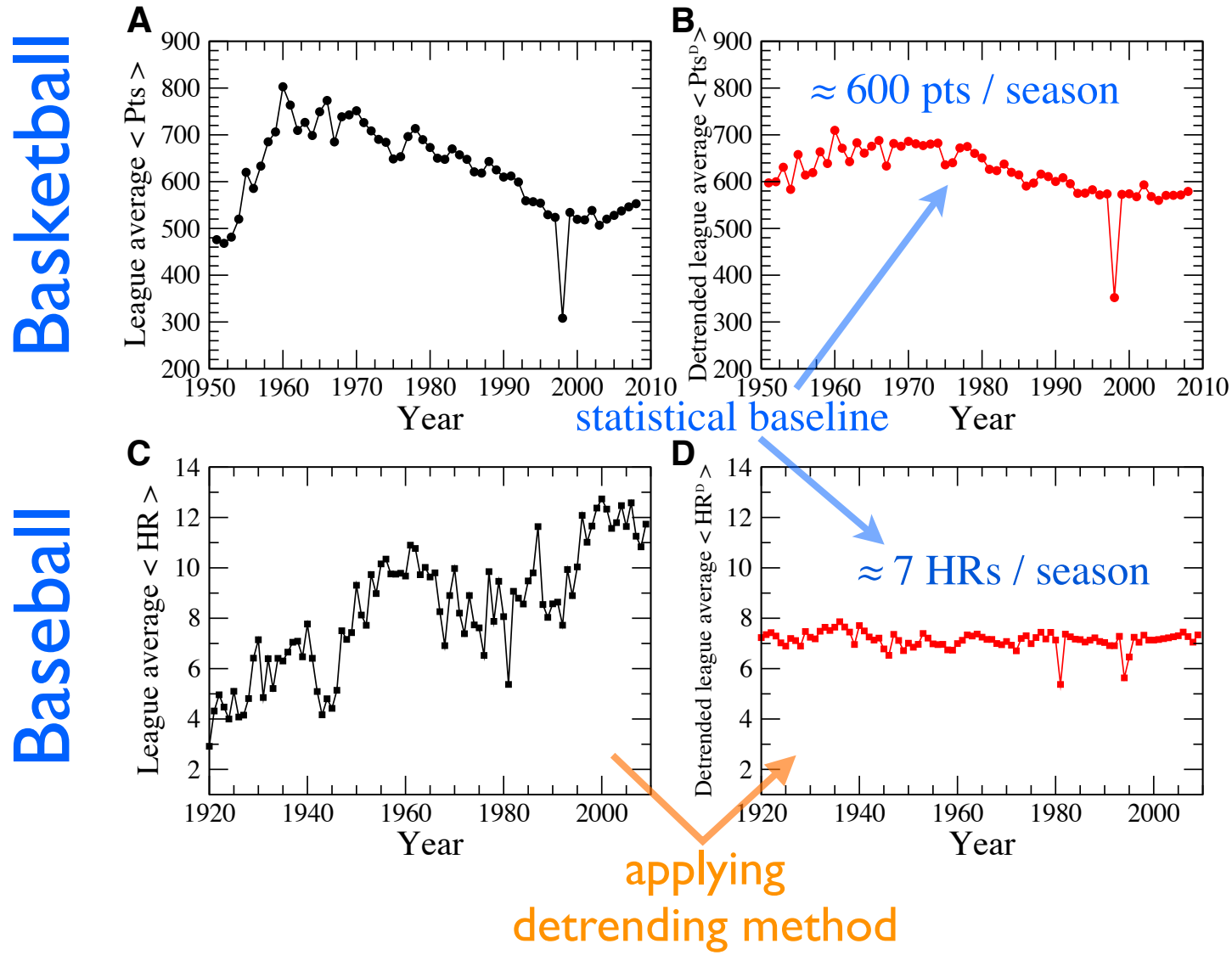






Just as the price of a candy bar has increased by a factor of ~ 20 over the last 100 years (roughly 3% inflation rate), the home run hitting ability of players has also increased by a significant factor over the same period

Accounting for socio-technological factors that underly achievement



Quantitative measures for success are important for comparing both individual and group accomplishments, often achieved in different time periods.

However, the evolutionary nature of competition results in a non-stationary rate of success, can make comparing accomplishments across time statistically biased.

the big debate...Career Home Runs....

Traditional Rank				Detrended Rank			
Rank	Name	Final Season (L)	Career Metric	Rank*(Rank)	Name	Final Season (L)	Career Metric
1	Barry Bonds	2007 (22)	762	1(3)	Babe Ruth	1935 (22)	1215
2	Hank Aaron	1976 (23)	755	2(23)	Mel Ott	1947 (22)	637
3	Babe Ruth	1935 (22)	714	3(26)	Lou Gehrig	1939 (17)	635
4	Willie Mays	1973 (22)	660	3(17)	Jimmie Foxx	1945 (20)	635
5	Ken Griffey Jr.	2009 (21)	630	5(2)	Hank Aaron	1976 (23)	582
6	Sammy Sosa	2007 (18)	609	6(124)	Rogers Hornsby	1937 (23)	528
7	Frank Robinson	1976 (21)	586	7(192)	Cy Williams	1930 (19)	527
8	Alex Rodriguez	2009 (16)	583	8(1)	Barry Bonds	2007 (22)	502
8	Mark McGwire	2001 (16)	583	9(4)	Willie Mays	1973 (22)	490
10	Harmon Killebrew	1975 (22)	573	10(18)	Ted Williams	1960 (19)	482
11	Rafael Palmeiro	2005 (20)	569	11(13)	Reggie Jackson	1987 (21)	478
12	Jim Thome	2009 (19)	564	12(14)	Mike Schmidt	1989 (18)	463
13	Reggie Jackson	1987 (21)	563	13(7)	Frank Robinson	1976 (21)	444
14	Mike Schmidt	1989 (18)	548	14(10)	Harmon Killebrew	1975 (22)	437
15	Manny Ramirez	2009 (17)	546	15(577)	Gavvy Cravath	1920 (11)	433
16	Mickey Mantle	1968 (18)	536	16(718)	Honus Wagner	1917 (21)	420
17	Jimmie Foxx	1945 (20)	534	17(18)	Willie McCovey	1980 (22)	417
18	Ted Williams	1960 (19)	521	18(557)	Harry Stovey	1893 (14)	413
18	Frank Thomas	2008 (19)	521	19(5)	Ken Griffey Jr.	2009 (21)	411
18	Willie McCovey	1980 (22)	521	20(28)	Stan Musial	1963 (22)	410

...for extensive top-50 tables for Hits, HR, RBI, K, W calculated for single seasons and also over entire the career consult the papers downloadable at:

Methods for detrending success metrics to account for inflationary and deflationary factors
A. M. Petersen, O. Penner, H. E. Stanley.
Eur. Phys. J. B 79, 67-78 (2011).

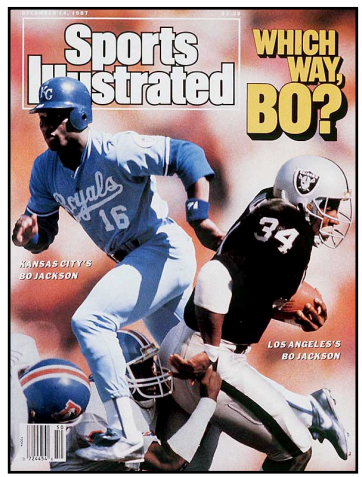
and an analogous statistical analysis of basketball career statistics:

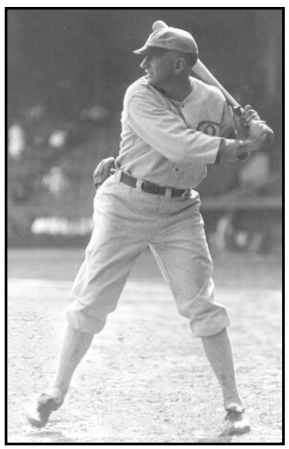
A method for the unbiased comparison of MLB and NBA career statistics across era
A. M. Petersen, O. Penner.
MIT Sloan Sports Analytics Conference 2012.

Physiological/Behavioral components of competition

High competition levels can make careers vulnerable to early career negative production shocks (ie stress, burn-out, productivity lulls, etc.)

Achievement-oriented systems: incentives for cut-throat "zero-sum" behavior, i.e. use of performance/cognitive enhancing drugs, possibly leading to blatant cheating/falsification





Ethical scandals reveal the price of success

Jan Hendrik Schön Scandal (2001)

On October 31, 2002, *Science* withdrew eight papers written by Schön

On December 20, 2002, *Physical Review* withdrew six papers

On March 5, 2003, *Nature* withdrew seven papers

Diederik Alexander Stapel Scandal (2011)

Social psychologist made up data for at least 30 publications according to preliminary investigation, which is still ongoing.

Hisashi Moriguchi Scandal (2012)

"Transplant of induced pluripotent stem cells to treat heart failure probably never happened.... He is affiliated with University of Tokyo but not with Massachusetts General Hospital nor with Harvard Medical School. The study did not receive Institutional Review Board approval." nature.com

Cognizant Enhancement Drugs (CED)

Professor's little helper

The use of cognitive-enhancing drugs by both ill and healthy individuals raises ethical questions that should not be ignored, argue **Barbara Sahakian** and **Sharon Morein-Zamir**.

NATURE|Vol 450|20/27 December 2007

" Is it cheating to use cognitive-enhancing drugs?.... How would you react if you knew your colleagues — or your students — were taking cognitive enhancers?... we know that a number of our scientific colleagues ... already use modafinil [Modiodal, Provigil] to counteract the effects of jetlag, to enhance productivity or mental energy, or to deal with demanding and important intellectual challenges..."

NATURE|Vol 452|10 April 2008

Poll results: look who's doping

In January, *Nature* launched an informal survey into readers' use of cognition-enhancing drugs. **Brendan Maher** has waded through the results and found large-scale use and a mix of attitudes towards the drugs.

"One in five respondents said they had used drugs for non-medical reasons to stimulate their focus, concentration or memory. Use did not differ greatly across age-groups..., which will surprise some. "

Towards responsible use of cognitive-enhancing drugs by the healthy

Society must respond to the growing demand for cognitive enhancement. That response must start by rejecting the idea that 'enhancement' is a dirty word, argue **Henry Greely and colleagues**.

"...one survey estimated that almost 7% of students in US universities have used prescription stimulants [Adderall and Ritalin] in this way, and that on some campuses, up to 25% of students had used them in the past year. These students are early adopters of a trend that is likely to grow, and indications suggest that they're not alone."

NATURE|Vol 456|11 December 2008

Ethics and scientific careers

- **Competition ("fairness"):**
  - strategizing / extreme behavior, e.g. scientific fraud
  - CED (cognitive enhancing drugs)
  - free-riding in team science, individual vs group: the"tragedy of the scientific commons"
- **Funding:**
  - financial incentives & who should subsidize early career risk
  - how to attribute / appraise / reward achievement, especially in the case of extremely large team projects
- **Careers:** predicting future career achievement using incomplete information and poorly understood/ designed achievement measures

Institutions Markets Technologies
IMT
INSTITUTE FOR ADVANCED STUDIES LUCCA

Food for thought

- Competition and Reward: There are many analogies between the superstars in science and the superstars in professional sports, possibly arising from the generic aspects of competition. Currently, the contract length, compensation, and appraisal timescale in these two professions are VERY different. Is science becoming more like professional sports?

- Science as an evolving institution: An institutional setting that neglects specific features of academic career trajectories (increasing returns from knowledge spillovers and cumulative advantage, collaboration factors, career uncertainty) is likely to be inefficient and unfair. But what is "fair"?

- Complex career dynamics: Knowledge, reputation, and collaboration spillovers are major factors leading to increasing returns along the scientific career trajectory. A data-centric ("big data") understanding of the production function of individual scientists can improve academic policies aimed at increasing career sustainability by decreasing career risk.

- *Quantitative and empirical demonstration of the Matthew effect in a study of career longevity, A. M. Petersen, W.-S. Jung, J.-S. Yang, H. E. Stanley. Proc. Natl. Acad. Sci. USA 108, 18-23 (2011).*
- *Persistence and Uncertainty in the Academic Career, A. M. Petersen, M. Riccaboni, H. E. Stanley, F. Pammolli. Proc. Natl. Acad. Sci. USA 109, 5213-5218 (2012).*
- *On the distribution of career longevity and the evolution of home run prowess in professional baseball, A. M. Petersen, W.-S. Jung, H. E. Stanley, Europhysics Letters 83, 50010 (2008).*
- *Methods for measuring the citations and productivity of scientists across time and discipline, A. M. Petersen, F. Wang, H. E. Stanley. Phys. Rev. E 81, 036114 (2010).*
- *Methods for detrending success metrics to account for inflationary and deflationary factors, A. M. Petersen, O. Penner, H. E. Stanley , Eur. Phys. J. B 79, 67-78 (2011).*
- *Reputation and impact in academic careers, A. M. Petersen, S. Fortunato, R. K. Pan, K. Kaski, O. Penner, M. Riccaboni, H. E. Stanley, F. Pammolli, (submitted) ArXiv:1303.7274*

Thank You!

A special thanks to my collaborators: Santo Fortunato, Woo-Sung Jung, Fabio Pammolli, Raj Pan, Orion Penner, Massimo Riccaboni, Gene Stanley, Sauro Succi, Fengzhong Wang, and Jae-Sook Yang

http://physics.bu.edu/~amp17/

Abstract:

Ascent in competitive arenas: From Fenway Park to Mass Ave

Competitive arenas are abundant in society and are characterized by at least three basic principles: limited opportunities, cumulative advantage, and the boundless ambitions of highly driven individuals. Using longitudinal career data for several hundred top-cited physicists, biologists, and mathematicians, I will show that stellar careers can be classified by common growth patterns. And while much is known about the stellar ascent of the likes of Mozart, Babe Ruth, and Einstein, little is known about their numerous out-shined competitors. Using data from six high-impact journals complemented by comprehensive career data spanning the entire history of the Major League Baseball labor force, I will further illustrate how the skewed distributions for diverse career achievement measures can be explained by simple models for career progress and competition. Context also matters, and one cannot understate the role that institutions play in establishing competitive norms and terms of fair play. As science continues to evolve towards a bigger and more interconnected system, an institutional setting which neglects the features of competition may inadvertently give rise to shifts in performance incentives and promote a "tragedy of the scientific commons" marked by the dilemma of individual versus the group. To this end, there is an increasing need to better understand the ethics of competition, as evidenced by both the frequency of research scandals and the widespread emergence of performance- (and even cognitive-) enhancing drugs in society's competitive arenas, which together highlight the risk that individuals are willing to accept in their pursuit for even the slightest competitive advantage.