Ascent in competitive arenas: From Fenway Park to Mass Ave

The Science of Success: Measurements and Predictions

June 17th Harvard University

Science of Success Symposia

Alexander M. Petersen
IMT Institute for Advanced Studies, Lucca Italy
stellar career growth is a non-linear process

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

\( \zeta > \alpha > 1 \Rightarrow \text{super-linear growth} \)

Cumulative advantage \( \sim \) careers become “attractors” of new opportunities instead of “pursuers”
What makes science special (complex)?

Interactions mediated by social “forces”:
- Collaboration (attractive)
- Competition for priority (repulsive)
- Knowledge (an “exchange particle”)

Interactions mediated by social “forces”:
Diverse collaboration strategies

Interactions mediated by social “forces”:
- Collaboration (attractive)
- Competition for priority (repulsive)
- Knowledge (an “exchange particle”)

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
Competitive arenas
Competitive arenas

Disambiguation strategy: Analyzed subset of “rare” surnames

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

TABLE I: Summary of journal datasets. $N^j$ 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 \(\sim\) 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 \( x \)

\[
g(x) = \frac{1}{\langle \tau(x) \rangle}
\]

The progress probability \( g \) is the inverse of the mean waiting time \( \tau \)


Statistical regularities in the career longevity distribution

opportunities \sim \text{time duration}

Major League Baseball
- 130+ years of player statistics, \sim 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 \( t \)

\[ I = \text{finite labor force size} \]

Persistence and Uncertainty in the Academic Career,
A. M. Petersen, M. Riccaboni, H. E. Stanley, F. Pammolli.
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

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

**Appraisal timescale $1/c$**

- $c \rightarrow 0$: appraisal over all lifetime achievements ($\sim$ 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.
Institutional trends in Science

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

200+ years


• 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

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?

Nature/PNAS/Science

# papers

Ave # coauthors per paper

annual growth rate = 0.004

x 1.5

Nature/PNAS/Science

PLoS One

Open Access Journals

PLoS One:

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

⇒ x 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.
...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:

<table>
<thead>
<tr>
<th>Rank</th>
<th>Name</th>
<th>Final Season (L)</th>
<th>Career Metric</th>
<th>Rank* (Rank)</th>
<th>Name</th>
<th>Final Season (L)</th>
<th>Career Metric</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Barry Bonds</td>
<td>2007 (22)</td>
<td>762</td>
<td>1(3)</td>
<td>Babe Ruth</td>
<td>1935 (22)</td>
<td>1215</td>
</tr>
<tr>
<td>2</td>
<td>Hank Aaron</td>
<td>1976 (23)</td>
<td>755</td>
<td>2(23)</td>
<td>Mel Ott</td>
<td>1947 (22)</td>
<td>637</td>
</tr>
<tr>
<td>3</td>
<td>Babe Ruth</td>
<td>1935 (22)</td>
<td>794</td>
<td>3(26)</td>
<td>Lou Gehrig</td>
<td>1939 (17)</td>
<td>635</td>
</tr>
<tr>
<td>4</td>
<td>Willie Mays</td>
<td>1973 (22)</td>
<td>660</td>
<td>3(17)</td>
<td>Jimmie Foxx</td>
<td>1945 (20)</td>
<td>635</td>
</tr>
<tr>
<td>5</td>
<td>Ken Griffey Jr.</td>
<td>2009 (21)</td>
<td>630</td>
<td>5(2)</td>
<td>Hank Aaron</td>
<td>1976 (23)</td>
<td>582</td>
</tr>
<tr>
<td>6</td>
<td>Sammy Sosa</td>
<td>2007 (18)</td>
<td>609</td>
<td>6(124)</td>
<td>Rogers Hornsby</td>
<td>1937 (23)</td>
<td>528</td>
</tr>
<tr>
<td>7</td>
<td>Frank Robinson</td>
<td>1976 (21)</td>
<td>586</td>
<td>7(192)</td>
<td>Cy Williams</td>
<td>1930 (19)</td>
<td>527</td>
</tr>
<tr>
<td>8</td>
<td>Alex Rodriguez</td>
<td>2009 (16)</td>
<td>583</td>
<td>8(1)</td>
<td>Barry Bonds</td>
<td>2007 (22)</td>
<td>502</td>
</tr>
<tr>
<td>9</td>
<td>Mark McGwire</td>
<td>2001 (16)</td>
<td>583</td>
<td>9(4)</td>
<td>Willie Mays</td>
<td>1973 (22)</td>
<td>490</td>
</tr>
<tr>
<td>10</td>
<td>Harmon Killebrew</td>
<td>1975 (22)</td>
<td>573</td>
<td>10(18)</td>
<td>Ted Williams</td>
<td>1960 (19)</td>
<td>482</td>
</tr>
<tr>
<td>11</td>
<td>Rafael Palmeiro</td>
<td>2005 (20)</td>
<td>569</td>
<td>11(13)</td>
<td>Reggie Jackson</td>
<td>1987 (21)</td>
<td>478</td>
</tr>
<tr>
<td>12</td>
<td>Jim Thome</td>
<td>2009 (19)</td>
<td>564</td>
<td>12(14)</td>
<td>Mike Schmidt</td>
<td>1989 (18)</td>
<td>463</td>
</tr>
<tr>
<td>14</td>
<td>Mike Schmidt</td>
<td>1989 (18)</td>
<td>548</td>
<td>14(10)</td>
<td>Harmon Killebrew</td>
<td>1975 (22)</td>
<td>437</td>
</tr>
<tr>
<td>15</td>
<td>Manny Ramirez</td>
<td>2009 (17)</td>
<td>546</td>
<td>15(577)</td>
<td>Gavvy Cravath</td>
<td>1920 (11)</td>
<td>433</td>
</tr>
<tr>
<td>16</td>
<td>Mickey Mantle</td>
<td>1968 (18)</td>
<td>536</td>
<td>16(718)</td>
<td>Honus Wagner</td>
<td>1917 (21)</td>
<td>420</td>
</tr>
<tr>
<td>17</td>
<td>Jimmie Foxx</td>
<td>1945 (20)</td>
<td>534</td>
<td>17(18)</td>
<td>Willie McGovey</td>
<td>1980 (22)</td>
<td>417</td>
</tr>
<tr>
<td>18</td>
<td>Ted Williams</td>
<td>1960 (19)</td>
<td>521</td>
<td>18(557)</td>
<td>Harry Stovey</td>
<td>1893 (14)</td>
<td>413</td>
</tr>
<tr>
<td>20</td>
<td>Willie McGovey</td>
<td>1980 (22)</td>
<td>521</td>
<td>20(28)</td>
<td>Stan Musial</td>
<td>1963 (22)</td>
<td>410</td>
</tr>
</tbody>
</table>

Methods for detrending success metrics to account for inflationary and deflationary factors
A. M. Petersen, O. Penner, H. E. Stanley.

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.
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 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

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.

NATURE|Vol 456|11 December 2008

“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...”

“...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.”
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
• **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.

---


---

**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/
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.