Supporting Information Appendix:
Quantitative and empirical demonstration of the Matthew effect in a study of career longevity

Alexander M. Petersen,[1] Woo-Sung Jung,[2,1] Jae-Suk Yang,[3] and H. Eugene Stanley[1]

[1]Center for Polymer Studies and Department of Physics,
Boston University, Boston, Massachusetts 02215, USA
[2]Department of Physics and Basic Science Research Institute,
Pohang University of Science and Technology, Pohang 790-784, Republic of Korea
[3]Columbia Business School, Columbia University, New York, NY 10027, USA

I. DATA AND METHODS

The publication data analyzed in this paper was downloaded from *ISI Web of Knowledge* in May 2009. We restrict our analysis to publications termed as "Articles", which excludes reviews, letters to editor, corrections, etc. Each article summary includes a field for the author identification consisting of a last name and first and middle initial (eg. the author name John M. Doe would be stored as "Doe, J" or "Doe, JM" depending on the author's designation). From these fields, we collect the career works of individual authors within a particular journal together, and analyze metrics for career longevity and success.

For author i we combine all articles in journal j for which he/she was listed as coauthor. The total number of papers for author i in journal j over the 50-year period is n_i. Following methods from lifetime statistics [S1], we use a standard method to isolate "completed" careers from our data set which begins at year Y_0 and ends at year Y_f. For each author i, we calculate $\langle \Delta\tau_i \rangle$, the average time $\Delta\tau_i$ between successive publications in a particular journal. A career which begins with the first recorded publication in year $y_{i,0}$ and ends with the final recorded publication in year $y_{i,f}$ is considered "complete", if the following two criteria are met:

(1) $y_{i,f} \leq Y_f - \langle \Delta\tau_i \rangle$

(2) $y_{i,0} \geq Y_0 + \langle \Delta\tau_i \rangle$.

This method estimates that the career begins in year $y_{i,0} - \langle \Delta\tau_i \rangle$ and ends in year $y_{i,f} + \langle \Delta\tau_i \rangle$. If either the estimated beginning or ending year do not lie within the range of the data base, than we discount the career as incomplete to first approximation. Statistically, this means that there is a significant probability that this author published before Y_0 or will publish after Y_f. We then estimate the career length within journal j as $L_{i,j} = y_{i,f} - y_{i,0} + 1$, and do not consider careers with $y_{i,f} = y_{i,0}$. This reduces the size of the data set by approximately 25% (compare the raw data set sizes N to the pruned data set size N^* in Table S1).

There are several potential sources of systematic error in the use of this database:

(i) Degenerate names → increases career totals. Radicchi *et al.* [S2] observe that this method of concatenated author ID leads to a pdf $P(d)$ of degeneracy d which scales as $P(d) \sim d^{-3}$.

(ii) Authors using middle initials in some but not all instances of publication → decreases career totals.

(iii) A mid-career change of last name → decreases career totals.

(iv) Sampling bias due to finite time period. Recent young careers are biased toward short careers. Long careers located towards the beginning Y_0 or end Y_f of the database are biased towards short careers.

II. A ROBUST METHOD FOR CLASSIFYING CAREERS

Professional sports leagues are geared around annual championships that celebrate the accomplishments of teams over a whole season. On a player level, professional sports leagues annually induct retired players into "halls of fame" in order to celebrate and honor stellar careers. Induction immediately secures an eternal legacy for those that are chosen. However, there is no standard method for inducting players into a *Hall of Fame*, with subjective and political factors affecting the induction process. In [S5] we quantitatively normalize seasonal statistics so to remove time-dependent factors that influence success. This provides a framework for comparing career statistics across historical eras.

In this section we propose a generic and robust method for measuring careers. We find that the pdf for career longevity can be approximated by the gamma distribution,

$$Gamma(x;\alpha,x_c) = \frac{x^{-\alpha}e^{-x/x_c}}{x_c^{1-\alpha}\Gamma(1-\alpha)} \ , \tag{S1}$$

with moments $\langle x^n\rangle = x_c^n\frac{\Gamma(1-\alpha+n)}{\Gamma(1-\alpha)}$, where we restrict our considerations to the case of $\alpha \leq 1$, with $x_c >> 1$. This distribution allows us to calculate the extreme value x^* such that only f percentage of players exceed this value according to the pdf $P(x)$,

$$f = \int_{x^*}^{\infty} \frac{x^{-\alpha}e^{-x/x_c}}{x_c^{1-\alpha}\Gamma(1-\alpha)}dx = \frac{\Gamma[1-\alpha,\frac{x^*}{x_c}]}{\Gamma(1-\alpha)} = Q[1-\alpha,\frac{x^*}{x_c}] \ , \tag{S2}$$

where $\Gamma[1-\alpha,\frac{x^*}{x_c}]$ is the incomplete gamma function and $Q[1-\alpha,\frac{x^*}{x_c}]$ is the regularized gamma function. This function can be easily inverted numerically using computer packages, e.g. *Mathematica*, which results in the statistical benchmark

$$x^* = x_c\, Q^{-1}[1-\alpha, f]. \tag{S3}$$

In [S5] we use the maximum likelihood estimator (MLE) for the Gamma pdf to estimate the parameters α and x_c for each pdf. The values we obtain using MLE are systematically smaller for α values and for x_c values, but the relative differences are negligible.

In Table S2 we provide statistical benchmarks x^* corresponding to career longevity and career metrics for several sports. For the calculation of each x^* we use the parameter values α and x_c calculated from a least-squares fit to the empirical pdf $P(x)$ using the functional form of Eq. [5], and the significance level value f calculated from historical induction frequencies in the American Baseball Hall of Fame (HOF) in Cooperstown, NY USA. The baseball HOF has inducted 276 players out of the 14,644 players that exist in Sean Lahman's baseball database between the years 1879-2002, which corresponds to a fraction $f \equiv 0.019$. It is interesting to note that the last column, $\frac{x^*}{\sigma} \equiv \beta \approx 3.9$ for all the gamma distributions analyzed. This approximation is a consequence of the universal scaling form of the gamma function $Gamma(x) \equiv U(x/x_c)$, where the standard deviation σ of the Gamma pdf has the simple relation $\sigma = x_c\sqrt{1-\alpha}$. Hence, for a given f and α, the ratio

$$x^*/\sigma = \frac{Q^{-1}[1-\alpha,f]}{\sqrt{1-\alpha}} \tag{S4}$$

is independent of x_c. Furthermore, this approximation is valid for all statistics in MLB since α is approximately the same for all pdfs analyzed. Thus, the value $x^* \approx 4\sigma$ is a robust approximation for determining if a player's career is stellar at the $f \approx 0.02$ significance level. The highly celebrated milestone of 3,000 hits in baseball corresponds to the value $x^* = 1.26\ \beta\sigma_{hits}$. Only 27 players have exceeded this benchmark in their professional careers, while only 86 have exceeded the arbitrary 2,500 benchmark. Hence, it makes sense to set the benchmark for all milestones at a value of $x^* = \beta\sigma$ corresponding to each distribution of career metrics.

We check for consistency by comparing the extreme threshold value x^* calculated using the gamma distribution with the value x_d^* derived from the database of career statistics. Referring to the actual set of all baseball players from 1871-2006, to achieve a fame value $f_d \approx 0.019$ with respect to hits, one should set the statistical benchmark at $x_d^* \approx 2250$, which account for 146 players (this assumes that approximately half of all baseball players are not pitchers, who we exclude from this calculation of f_d). The value of $x_d^* \approx 2250$ agrees well with the value calculated from the gamma distribution, $x^* \approx 2366$. Of these 146 players with career hit tallies greater than 2250, there are 126 players who have been eligible for at least one induction round, and 82 of these players have been successfully inducted into the American baseball hall of fame. Thus, a player with a career hit tally above $x^* \approx x_d^*$ has a 65% chance of being accepted, based on just those merits alone. Repeating the same procedure for career strikeouts obtained by pitchers in baseball we obtain the milestone value $x_d^* \approx 1525$ strikeouts, and for career points in basketball we obtain the value $x_d^* \approx 16,300$ points. Nevertheless, the overall career must be taken into account, which raises the bar, and accounts for the less than perfect success rate of being voted into a hall of fame, given that a player has had a statistically stellar career in one statistical category.

III. CAREER METRICS

In Fig. 4 we plot common career metrics for success in American baseball and American basketball. Note that the exponent α for the pdf $P(z)$ of total career successes z is approximately equal to the exponent α for the pdf $P(x)$ of career longevity x (see Table S2). In this section, we provide a simple explanation for the similarity between the power law exponent for career longevity (Fig. 2) and the power law exponent for career success (Fig. 4).

Consider a distribution of longevity that is power law distributed, $P(x) \sim x^{-\alpha}$ for the entire range $1 \leq x \leq x_c < \infty$. The cutoff x_c represents the finiteness of human longevity, accounted for by the exponential decay in Eq. [7]. Also, assume that the prowess y has a pdf $P(y)$ which is characterized by a mean and standard deviation, which represent the talent level among professionals (see Ref. [S3] for the corresponding prowess distributions in major league baseball). In the first possible case, the distribution is right-skewed and approximately exponential (as in the case of home-runs). In other cases, the distributions are essentially Gaussian. Regardless of the distribution type, the prowess pdfs $P(y)$ are confined to the domain $\delta \leq y \leq 1$, where $\delta > 0$.

Assume that in any given appearance, a person can apply his/her natural prowess towards achieving a success, independent of past success. Although prowess is refined over time, this should not substantially alter our demonstration. Since not all professionals have the same career length, the career totals are in fact a combination of these two distributions as in their product. Then the career success total $z = xy$ has the distribution,

$$\begin{aligned} P(z = xy) &= \int\int dy\, dx\; P(y)P(x)\delta(xy - z) \\ &= \int\int dy\, dx P(y)P(x)\delta(x(y - z/x)) \\ &= \int dx\; P(\frac{z}{x})P(x)\frac{1}{x}\ . \end{aligned} \tag{S5}$$

This integral has three domains (Ref. [S4]),

$$P(z) \propto \begin{cases} \int_1^{z/\delta} dx\; P(\frac{z}{x})x^{-(\alpha+1)}\ , & \delta < z < 1 \\ \int_z^{z/\delta} dx\; P(\frac{z}{x})x^{-(\alpha+1)}\ , & 1 < z < x_c\delta \\ \int_z^{x_c} dx\; P(\frac{z}{x})x^{-(\alpha+1)}\ , & x_c\delta < z < x_c\ . \end{cases}$$

The first regime $\delta < z < 1$ is irrelevant, and is not observed since z is discrete in the cases analyzed here. For the first case of an exponentially distributed prowess,

$$P(z) \propto \begin{cases} z^{-\alpha}\ , & 1 < z < x_c \quad \delta \\ z^{-\alpha}\exp(-z/\lambda x_c)\ , & x_c\delta < z < x_c\ . \end{cases} \tag{S6}$$

In Ref. [S3] we mainly observe the exponential tail in the home-run distribution, as the above form suggests in the regime $x_c\delta < z < x_c$, resulting from $\delta \approx 0$ for the right-skewed home-run prowess distribution. However, in the case for a normally distributed prowess, the power law behavior of the longevity distribution is maintained for large values into the career success distribution $P(z)$, as $x_c\delta > 10^3$.

$$P(z) \propto \begin{cases} z^{-\alpha}\ , & 1 < z < x_c\delta \\ z^{-\alpha}e^{-(\frac{z}{\sigma x_c})^2/2}\ , & x_c\delta < z < x_c\ . \end{cases} \tag{S7}$$

Thus, the main result of this demonstration is that the distribution $P(z)$ maintains the power law exponent α of the career-longevity distribution, $P(x)$, when the prowess is distributed with a characteristic mean and standard deviation. This result is also demonstrated with the simplification of representing the prowess distribution $P(y)$ as an essentially uniform distribution over a reasonable domain of y, which simplifies the integral in Eq. (S5) while maintaining the inherent power law structure.

In Fig. S1 we plot the prowess distributions that correspond to the career success distributions plotted in Fig. 4. It is interesting that the competition level based on the distributions of prowess indicates that Korean and American baseball are nearly equivalent. Also, note that the prowess distributions for rebounds per minute are bimodal, as the positions of players in basketball are more specialized.

IV. A NULL MODEL WITHOUT THE MATTHEW EFFECT

In this section, we compare the predictions of our theoretical model with the predictions of a theoretical model which does not incorporate the Matthew effect. Since the Matthew effect implies that the progress rate $g(x)$ increase with career position x, we analyze the more simple model where for each individual i the progress rate $g_i(x)$ is constant,

$$g_i(x) \equiv \lambda_i\ . \tag{S8}$$

The solution to the conditional longevity pdf $P(x|\lambda_i)$ is still given by Eq. [5], taking the form

$$P(x|\lambda_i) = \frac{\lambda_i^{x-1}}{x_c(\frac{1}{x_c} + \lambda_i)^x} \approx \frac{1}{\lambda_i x_c}\, e^{-\frac{x}{\lambda_i x_c}}\ , \tag{S9}$$

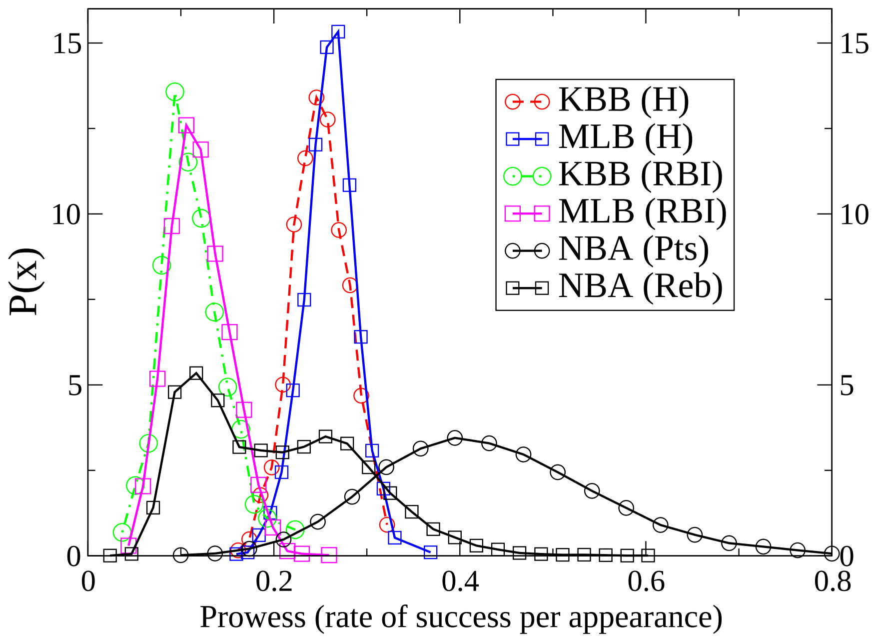

FIG. S1: Probability density functions of seasonal prowess for several career metrics. Each pdf is normally distributed, except for the bimodal curve for rebound prowess, NBA (Reb.). The bimodal distribution for Rebound prowess reflects the specialization in player positions in the sport of basketball. Furthermore, note the remarkable similarity in the distributions between American (MLB) and Korean (KBB) baseball players.

which is an exponential pdf, with a characteristic career length $l_c \equiv \lambda_i x_c$. Hence, this null model corresponds to a career progress mechanism wherein intrinsic ability, which is incorporated into the relative value of λ_i, is the dominant factor. In order to calculate the longevity pdf $P(x)$ which incorporates a distribution of intrinsic abilities across the population of individuals, we average over the conditional pdfs using a pdf $P(\lambda)$ that we assume is well-defined by a mean $\overline{\lambda}$ and standard deviation σ, consistent with what we observe for the seasonal prowess pdfs shown in Fig. S1. In the case of $P(\lambda) = Normal(\overline{\lambda}, \sigma)$, then

$$P(x) = \int_0^1 P(\lambda) P(x|\lambda) d\lambda \; \equiv \int_0^1 \frac{e^{-(\lambda-\overline{\lambda})^2/2\sigma^2}}{\sqrt{2\pi\sigma^2}} P(x|\lambda) d\lambda \; . \tag{S10}$$

For the sake of providing an analytic result, we replace $P(\lambda)$ by a uniform distribution,

$$P(\lambda) \approx \begin{cases} 0 \, , & |\lambda - \overline{\lambda}| > 2\sigma \\ \frac{1}{4\sigma} \, , & |\lambda - \overline{\lambda}| \leq 2\sigma \, , \end{cases} \tag{S11}$$

which does not change the overall result. The integral in Eq. (S10) then becomes,

$$P(x) \approx \frac{1}{4\sigma} \int_{\lambda-2\sigma}^{\lambda+2\sigma} \frac{d\lambda}{\lambda x_c} \, e^{-\frac{x}{\lambda x_c}} = \frac{1}{4\sigma x_c} [\Gamma(0, \frac{x/x_c}{\overline{\lambda}+2\sigma}) - \Gamma(0, \frac{x/x_c}{\overline{\lambda}-2\sigma})] \approx e^{-x/\overline{\lambda} x_c} \; , \tag{S12}$$

for $1 > \overline{\lambda} > 2\sigma$, where the last approximation corresponds to a relatively small σ. Thus, we find that even with a reasonable dispersion in the constant progress rates λ in a population of individuals, the pdf $P(x)$ is still exponential. Hence, our theoretical model cannot explain the empirical non-exponential form of $P(x)$ unless we incorporate the Matthew effect using $g(x)$ that increase with x.

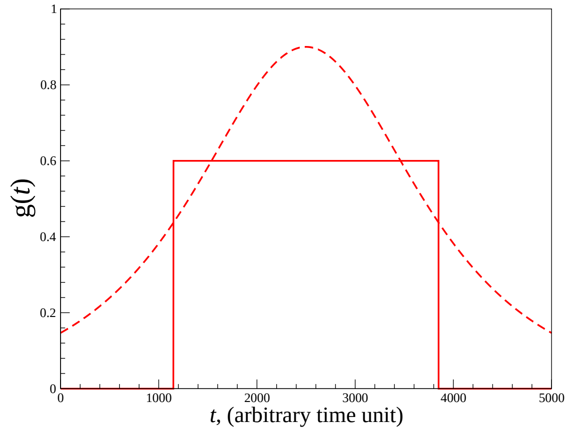

FIG. S2: A graphical illustration of a hypothetical career progress trajectory $g(t) = a\ \mathrm{sech}[(t-t^*)/w]$ (dashed red line), with amplitude $a = 0.9$, peak time $t^* = 2500$, and width $w = 1000$, in arbitrary time units. As an approximation, in order to provide an analytic solution to the model, we approximate $g(t)$ by a uniform plateau function $g(t) \approx \gamma[H(t-t_1) - H(t-t_2)]$ (solid red line), as in Eq. (S18), where $H(t)$ is the standard Heavyside step function.

V. A NULL MODEL WITH TIME-DEPENDENT CAREER TRAJECTORY

In this section, we develop a career progress model where the progress rate $g(t)$ is time-dependent instead of being position-dependent $g(x)$, as in the previous sections. We use a time dependent career trajectory to capture the non-monotonic peaks in key productivity factors, e.g. creativity and talent, that are observed for various creative careers [S6]. In Fig. S2 we show a generic $g(t)$ which peaks at a variable time t^*, and has an amplitude a related to the individual's underlying talent. The regime in which $g(t)$ is increasing reflects the learning curve associated with a difficult endeavor, whereas the regime in which $g(t)$ is decreasing reflects e.g. aging factors and the upper limit to the finite resources which facilitate improvement.

In analogy to Eq. [10], the master equation for the evolution of career progress is

$$\frac{\partial P(x+1,t)}{\partial t} = g(t)P(x,t) - g(t)P(x+1,t)\ , \tag{S13}$$

where $g(t)$ is an arbitrary function which quantifies the forward progress rate at time t. To solve for $P(x,t)$, we define the "integrated time" τ given by,

$$\tau \equiv \int_0^t dt' g(t')\ . \tag{S14}$$

Hence, we write Eq. (S13) as,

$$\frac{\partial P(x+1,\tau)}{\partial \tau} = P(x,\tau) - P(x+1,\tau)\ , \tag{S15}$$

which along with the initial condition $P(x+1,\tau) = P(x+1,t) = \delta_{x,0}$, has the solution,

$$P(x,\tau) = \frac{e^{-\tau}\tau^{x-1}}{(x-1)!}\ . \tag{S16}$$

As previously described in the main text, we obtain the unconditional probability density function $P(x)$ of career longevity x from the conditional pdf $P(x|T) = P(x, t \equiv T)$ using a pdf of random termination times $r(T)$,

$$P(x) = \int_0^\infty P(x|T)r(T)dT\ , \tag{S17}$$

where we use the exponential pdf $r(T) = x_c^{-1}\exp[-(T/x_c)]$ for the demonstration of a career termination model with constant hazard rate, corresponding to the Laplace transform of $P(x|T)$ in the variable $s = 1/x_c$. The integral in Eq. (S17) is typically difficult to calculate given the time-dependence of the progress rate.

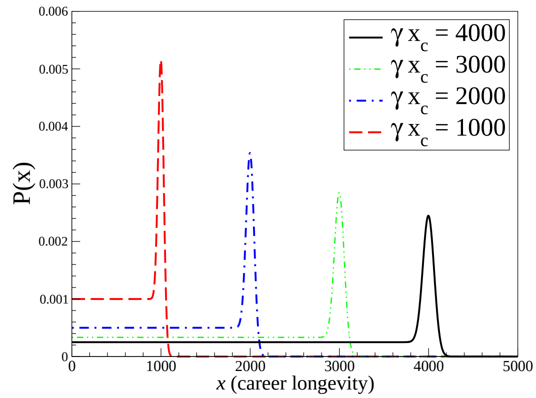

FIG. S3: Exact solutions for $P(x)$ with time-dependent career trajectory $g(t)$ defined in Eq. (S21), for the case of $t_1 = 0$, $x_c = t_2$, and $\gamma x_c = \{1000, 2000, 3000, 4000\}$.

Simonton [S6] finds that the annual productivity of creative products or ideas has a trajectory that is peaked around a given characteristic time t^* into a given profession. This peak is determined by two model parameters quantifying "ideation" and "elaboration" rates, and two additional parameters quantifying initial creative potential and the age at career onset. To demonstrate the solution to our null model, we use an simplified functional form for $g(t)$ corresponding to a uniform distribution over the interval $t \in [t_1, t_2]$,

$$g(t) \approx \begin{cases} 0\ , & t < t_1 \\ \gamma\ , & t \in [t_1, t_2] \\ 0\ , & t > t_2\ , \end{cases} \tag{S18}$$

where t_1 is the "breakout" year of the career, t_2 corresponds to the year in which the individual's productivity declines rapidly, and $0 \le \gamma \le 1$ is the intrinsic potential or talent of the given individual, and the time duration $t_2 - t_1$ is the precocity of the given individual. Hence, the corresponding integrated time τ is given by

$$\tau \equiv \int_0^t dt' g(t') = \begin{cases} 0\ , & t < t_1 \\ \gamma(t - t_1)\ , & t \in [t_1, t_2] \\ \gamma(t_2 - t_1)\ , & t > t_2\ . \end{cases} \tag{S19}$$

Then Eq. (S17) becomes,

$$\begin{aligned} P(x) &= \int_{t_1}^{t_2} dT e^{-\gamma(T-t_1)} \frac{[\gamma(T-t_1)]^{x-1}}{(x-1)!} x_c^{-1} e^{-T/x_c} + \int_{t_2}^{\infty} dT e^{-\gamma(t_2-t_1)} \frac{[\gamma(t_2-t_1)]^{x-1}}{(x-1)!} x_c^{-1} e^{-T/x_c} \\ &= \frac{e^{-t_1/x_c}}{\gamma x_c} \Big(\frac{1}{1+1/\gamma x_c}\Big)^x \Big[1 - \frac{\Gamma(x, \gamma(t_2-t_1))}{\Gamma(x)}\Big] + e^{-\gamma(t_2-t_1)} \frac{[\gamma(t_2-t_1)]^{x-1}}{\Gamma(x)} e^{-t_2/x_c}\ . \end{aligned} \tag{S20}$$

In the limit $t_1 \to 0$ and with $t_2 \equiv x_c$, the functional form of $P(x)$ has only one parameter, the product $\gamma x_c \gg 1$, so that

$$P(x) = \frac{1}{\gamma x_c} \Big[1 - \frac{\Gamma(x, \gamma x_c)}{\Gamma(x)}\Big] + e^{-(\gamma x_c + 1)} \frac{[\gamma x_c]^{x-1}}{\Gamma(x)} \tag{S21}$$

In Fig. S3 we plot $P(x)$ for several values of the parameter γx_c, where each curve demonstrates two common features, (i) a uniform distribution of career longevity x for $1 \le x \lesssim \gamma x_c$, and (ii) a sharp peak that is centered around $x = \gamma x_c$ which corresponds to approximately 10% of careers which are stellar. Averaging the $P(x)$ over a distribution $P(\gamma)$ of talent values γ that is approximately normal, as in the case of the prowess pdfs in Fig. S1, would result in a qualitatively similar $P(x)$ which is peaked around the value $x \approx \overline{\gamma} x_c$. The resulting distribution would be essentially "bimodal", with one mode corresponding to "stellar" careers distributed for $x \approx \overline{\gamma} x_c$, and a mode corresponding to less-substantial careers for $x \lesssim \overline{\gamma} x_c$, just as in the case of the convex progress rate for $\alpha > 1$, both of which do not agree with the statistical regularity in the empirical data (Fig. 3) which occurs over several orders of magnitude.

In our model, we assume that termination is due to external factors. A more complex model might include the possibility that termination is due to endogenous factors, e.g. a reduced level of productivity below a predetermined employment threshold at any given time. This type of endogenous termination is more difficult to model, since it correlates the progress $\delta x/\delta t$ with the termination probability $r(T)$, whereas above they are assumed to evolve independently. We leave this more complex model as an open avenue of research.

TABLE S1: Summary of data sets for each journal. Total number N of unique (but possibly degenerate) name identifications. N^* is the number of unique name identifications after pruning the data set of incomplete careers.

Journal	Years	Articles	Authors, N	N^*
Nature	1958-2008	65,709	130,596	94,221
Science	1958-2008	48,169	109,519	82,181
PNAS	1958-2008	84,520	182,761	118,757
PRL	1958-2008	85,316	112,660	72,102
CELL	1974-2008	11,078	31,918	23,060
NEJM	1958-2008	17,088	66,834	49,341

TABLE S2: Data summary for the pdfs of career statistical metrics. The values α and x_c are determined for each career longevity pdf $P(x)$ and each career success pdf $P(z)$ via least-squares method using the functional form given by Eq. [5]. We calculate the Gamma pdf average $\langle x \rangle$, the standard deviation σ, and the extreme threshold value x^* at the $f = 0.019$ significance level using the corresponding values of α and x_c. The units for each metric are indicated in parenthesis alongside the league in the first column.

For publication distributions, the career longevity metric x is measured in years.

Professional League,	Least-square values		Gamma pdf values				
(success metric)	α	x_c	$\langle x \rangle$	σ	x^*	$\frac{x^*}{\langle x \rangle}$	$\frac{x^*}{\sigma}$
MLB, (H)	0.76 ± 0.02	1240 ± 150	300	610	2400	7.8	3.9
MLB, (RBI)	0.76 ± 0.02	570 ± 80	140	280	1100	7.8	3.9
NBA, (Pts)	0.69 ± 0.02	7840 ± 760	2400	4400	17000	7.0	3.9
NBA, (Reb)	0.69 ± 0.02	3500 ± 130	1100	2000	7600	6.9	3.9

Professional League,	Least-square values		Gamma pdf values				
(opportunities)	α	x_c	$\langle x \rangle$	σ	x^*	$\frac{x^*}{\langle x \rangle}$	$\frac{x^*}{\sigma}$
KBB, (AB)	0.78 ± 0.02	2600 ± 320	580	1200	4700	8.2	3.9
MLB, (AB)	0.77 ± 0.02	5300 ± 870	1200	2500	9700	8.1	3.9
MLB, (IPO)	0.72 ± 0.02	3400 ± 240	950	1800	6900	7.3	3.9
KBB, (IPO)	0.69 ± 0.02	2800 ± 160	840	1500	5900	7.0	3.9
NBA, (Min)	0.64 ± 0.02	20600 ± 1900	7700	12600	48800	6.4	3.9
UK, (G)	0.56 ± 0.02	138 ± 14	61	92	360	5.8	3.9

Academic Journal,	Least-square values	
(career length in years)	α	x_c
Nature	0.38 ± 0.03	9.1 ± 0.2
PNAS	0.30 ± 0.02	9.8 ± 0.2
Science	0.40 ± 0.02	8.7 ± 0.2
CELL	0.36 ± 0.05	6.9 ± 0.2
NEJM	0.10 ± 0.02	10.7 ± 0.2
PRL	0.31 ± 0.04	9.8 ± 0.3

[S1] Huber JC (1998) Inventive Productivity and the Statistics of Exceedances. *Scientometrics* **45**: 33.

[S2] Radicchi F, Fortunato S, Markines B, Vespignani A (2009) Diffusion of scientific credits and the ranking of scientists. Phys. Rev. E **80**, 056103.

[S3] Petersen AM, Jung W-S & Stanley HE (2008) On the distribution of career longevity and the evolution of home-run prowess in professional baseball. *Europhysics Letters* **83**, 50010.

[S4] Glen A, Leemis L & Drew J (2004) Computing the distribution of the product of two continuous random variables. *Computational Stat. & Data Analysis* **44**, 451.

[S5] Petersen AM, Penner O, Stanley HE (2010) Detrending career statistics in professional Baseball: accounting for the Steroids Era and beyond. e-print arXiv:1003.0134.

[S6] Simonton DK (1997) Creative productivity: A predictive and explanatory model of career trajectories and landmarks. *Psychological Review* **104**: 66-89.