Quantitative and empirical demonstration of the Matthew effect in a study of career longevity

Alexander M. Petersen[a,1], Woo-Sung Jung[b,a], Jae-Suk Yang[c], and H. Eugene Stanley[a,1]

[a]Center for Polymer Studies and Department of Physics, Boston University, Boston, MA 02215; [b]Graduate Program for Technology and Innovation Management and Department of Physics, Pohang University of Science and Technology, Pohang 790-784, Republic of Korea; and [c]Sanford C. Benstein and Co. Center for Leadership and Ethics, Columbia Business School, Columbia University, New York, NY 10027

Contributed by H. Eugene Stanley, November 10, 2010 (sent for review November 8, 2009)

The Matthew effect refers to the adage written some two-thousand years ago in the Gospel of St. Matthew: "For to all those who have, more will be given." Even two millennia later, this idiom is used by sociologists to qualitatively describe the dynamics of individual progress and the interplay between status and reward. Quantitative studies of professional careers are traditionally limited by the difficulty in measuring progress and the lack of data on individual careers. However, in some professions, there are well-defined metrics that quantify career longevity, success, and prowess, which together contribute to the overall success rating for an individual employee. Here we demonstrate testable evidence of the age-old Matthew "*rich get richer*" effect, wherein the longevity and past success of an individual lead to a cumulative advantage in further developing his or her career. We develop an exactly solvable stochastic career progress model that quantitatively incorporates the Matthew effect and validate our model predictions for several competitive professions. We test our model on the careers of 400,000 scientists using data from six high-impact journals and further confirm our findings by testing the model on the careers of more than 20,000 athletes in four sports leagues. Our model highlights the importance of early career development, showing that many careers are stunted by the relative disadvantage associated with inexperience.

career length | hazard rate | output | Poisson process | quantitative sociology

The rate of individual progress is fundamental to career development and success. In practice, the rate of progress depends on many factors, such as an individual's talent, productivity, reputation, as well as other external random factors. Using a stochastic model, here we find that the relatively small rate of progress at the beginning of the career plays a crucial role in the evolution of the career length. Our quantitative model describes career progression using two fundamental ingredients: (*i*) random forward progress "up the career ladder" and (*ii*) random stopping times, terminating a career. This model quantifies the "Matthew effect" by incorporating into ingredient (*i*) the common cumulative advantage property (1–8) that it is easier to move forward in the career the further along one is in the career. A direct result of the increasing progress rate with career position is the large disparity between the numbers of careers that are successful long tenures and the numbers of careers that are unsuccessful short stints.

Surprisingly, despite the large differences in the numbers of long and short careers, we find a scaling law that bridges the gap between the frequent short and the infrequent long careers. We test this model for both scientific and sports careers, two careers where accomplishments are methodically recorded. We analyze publication careers within six high-impact journals: *Nature*, *Science*, the *Proceedings of the National Academy of Science* (PNAS), *Physical Review Letters* (PRL), *New England Journal of Medicine* (NEJM), and CELL. We also analyze sports careers within four distinct leagues: Major League Baseball (MLB), Korean Professional Baseball, the National Basketball Association (NBA), and the English Premier League.

Career longevity is a fundamental metric that influences the overall legacy of an employee because for most individuals the measure of success is intrinsically related, although not perfectly correlated, to his or her career length. Common experience in most professions indicates that time is required for colleagues to gain faith in a newcomer's abilities. Qualitatively, the acquisition of new opportunities mimics a standard positive feedback mechanism [known in various fields as Malthusian growth, cumulative advantage, preferential attachment, a reinforcement process, the ratchet effect, and the Matthew "*rich get richer*" effect (9)], which endows greater rewards (10) to individuals who are more accomplished than to individuals who are less accomplished.

Here we use career position as a proxy for individual accomplishment, so that the positive feedback captured by the Matthew effect is related to increasing career position. There are also other factors that result in selective bias, such as the "relative age effect," which has been used to explain the skewed birthday distributions in populations of athletes. Several studies find that being born in optimal months provides a competitive advantage to the older group members with respect to the younger group members within a cohort, resulting in a relatively higher chance of succeeding for the older group members, consistent with the Matthew effect. This relative age effect is found at several levels of competitive sports ranging from secondary school to the professional level (11, 12).

In this paper we study the everyday topic of career longevity and reveal surprising complexity arising from the generic competition within social environments. We develop an exactly solvable stochastic model, which predicts the functional form of the probability density function (pdf) $P(x)$ of career longevity x in competitive professions, where we define career longevity as the final career position x after a given time duration T corresponding to the termination time of the career. Our stochastic model depends on only two parameters, α and x_c. The first parameter, α, represents the power-law exponent that emerges from the pdf of career longevity. This parameter is intrinsically related to the progress rate early in the career during which professionals establish their reputations and secure future opportunities. The second parameter, x_c, is an effective time scale that distinguishes newcomers from veterans.

Quantitative Model

In this model, every employee begins his or her career with approximately zero credibility and must labor through a common

Author contributions: A.M.P., W.-S.J., J.-S.Y., and H.E.S. designed research, performed research, contributed new reagents/analytic tools, analyzed data, and wrote the paper.

The authors declare no conflict of interest.

[1]To whom correspondence may be addressed. E-mail: amp17@physics.bu.edu or hes@bu.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1016733108/-/DCSupplemental.

development curve. At each position x in a career, there is an opportunity for progress as well as the possibility for no progress. A new opportunity, corresponding to the advancement to career position $x+1$ from career position x, can refer to a day at work or, more generally, to any assignment given by an employing body. For each particular career, the change in career position Δx has an associated time frame Δt. Optimally, an individual makes progress by advancing in career position at an equal rate as the advancing of time t so that $\Delta x \equiv \Delta t$. However, in practice, an individual makes progress Δx in a subordinate time frame, given here as the career position x. In this framework, career progress is made at a rate that is slower than the passing of work time, representing the possibility of career stagnancy.

As a first step, we postulate that the stochastic process governing career progress is similar to a Poisson process, where progress is made at any given step with some approximate probability or rate. Each step forward in career position contributes to the employee's resume and reputation. Hence, we refine the process to a spatial Poisson process, where the probability of progress $g(x)$ depends explicitly on the employee's position x within the career. In our model, the progress rate $g(x) = \text{talent}(x) + \text{reputation}(x) + \text{productivity}(x) + \ldots$ represents a combination of several factors, such as the talent, reputation, and productivity at a given career position x. The criteria for the Matthew effect to apply is that the progress rate be monotonically increasing with career position, so that $g(x+1) > g(x)$. In this paper, we do not distinguish between the Matthew effect, relating mainly to the positive feedback from recognition, and cumulative advantage, which relates to the positive feedback from both productivity and recognition (2). It would require more detailed data to determine the role of the individual factors on the evolution of a career.

Employees begin their career at the starting career position $x_0 \equiv 1$ and make random forward progress through time to career position $x \geq 1$, as illustrated in Fig. 1. Career longevity is then defined as the final location $x \equiv x_T$ along the career ladder at the time of retirement T. Let $P(x|T)$ be the conditional probability that at stopping time T an individual is at the final career position x_T. For simplicity, we assume that the progress rate $g(x)$ depends only on x. As a result, $P(x|T)$ assumes the familiar Poisson form, but with the insertion of $g(x)$ as the rate parameter,

$$P(x|T) = \frac{e^{-g(x)T}[g(x)T]^{x-1}}{(x-1)!}. \quad \textbf{[1]}$$

We derive the spatial Poisson pdf $P(x|T)$ in *Appendix*. In *SI Appendix*, we further develop an alternative model where the progress rate $g(t)$ represents a career trajectory that depends on time (13).

According to the Matthew effect, it becomes easier for an individual to excel with increasing success and reputation. Hence, the choice of $g(x)$ should reflect the fact that newcomers, lacking the familiarity of their peers, have a more difficult time moving forward, whereas seasoned veterans, following from their experience and reputation, often have an easier time moving forward. For this reason we choose the progress rate $g(x)$ to have the functional form,

$$g(x) \equiv 1 - \exp[-(x/x_c)^{\alpha}]. \quad \textbf{[2]}$$

This function exhibits the fundamental feature of increasing from approximately zero and asymptotically approaching unity over some time interval x_c. Furthermore, $g(x) \sim x^{\alpha}$ for small $x \ll x_c$. In Fig. 2, we plot $g(x)$ for several values of α, with fixed $x_c = 10^3$ in arbitrary units. We will show that the parameter α is the same as the power-law exponent α in the pdf of career longevity $P(x)$, which we plot in Fig. 2, *Inset*. The random process for forward progress can also be recast into the form of random waiting times, where the average waiting time $\langle \tau(x) \rangle$ between successive steps is the inverse of the forward progress probability, $\langle \tau(x) \rangle = 1/g(x)$.

We now address the fact that not every career is of the same length. Nearly every individual is faced with the constant risk of losing his or her job, possibly as the result of poor performance, bad health, economic downturn, or even a change in the business strategy of his or her employer. Survival in the workplace requires that the individual maintain his or her performance level with respect to all possible replacements. In general, career longevity is

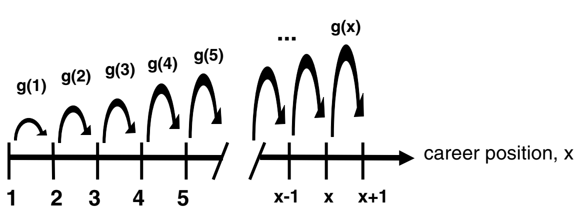

Fig. 1. Graphical illustration of the stochastic Poisson process quantifying career progress with position-dependent progress rate $g(x)$ and stagnancy rate $1 - g(x)$. A new opportunity, corresponding to the advancement to career position $x+1$ from career position x, can refer to a day at work or, even more generally, to any assignment given by an employing body. In this framework, career progress is made at a rate $g(x)$ that is slower than the passing of work time, representing the possibility of career stagnancy. The traditional Poisson process corresponds to a constant progress rate $g(x) \equiv \lambda$. Here, we use a functional form for $g(x) \equiv 1 - \exp[-(x/x_c)^{\alpha}]$ that is increasing with career position x, which captures the salient feature of the Matthew effect, that it becomes easier to make progress the further along the career. In *SI Appendix*, we further develop an alternative model where the progress rate $g(t)$ depends on time.

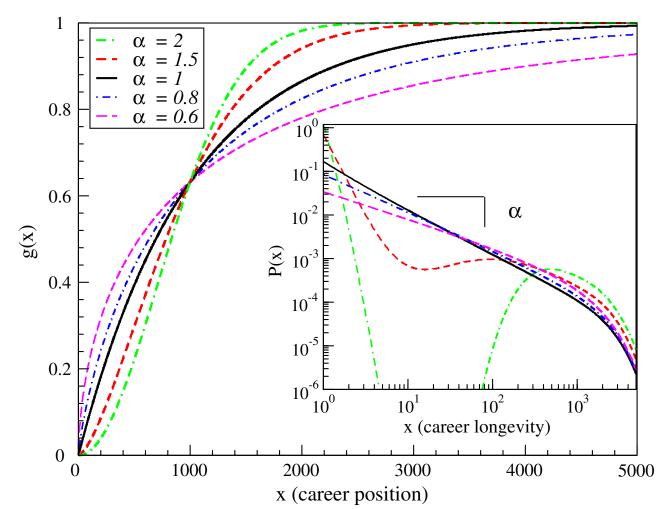

Fig. 2. Demonstration of the fundamental relationship between the progress rate $g(x)$ and the career longevity pdf $P(x)$. The progress rate $g(x)$ represents the probability of moving forward in the career to position $x+1$ from position x. The small value of $g(x)$ for small x captures the difficulty in making progress at the beginning of a career. The progress rate increases with career position x, capturing the role of the Matthew effect. We plot five $g(x)$ curves with fixed $x_c = 10^3$ and different values of the parameter α. The parameter α emerges from the small-x behavior in $g(x)$ as the power-law exponent characterizing $P(x)$. (*Inset*) Probability density functions $P(x)$ resulting from inserting $g(x)$ with varying α into Eq. **5**. The value $\alpha_c \equiv 1$ separates two distinct types of longevity distributions. The distributions resulting from concave career development $\alpha < 1$ exhibit monotonic statistical regularity over the entire range, with an analytic form approximated by the Gamma distribution Gamma($x;\alpha,x_c$). The distributions resulting from convex career development $\alpha > 1$ exhibit bimodal behavior. In the bimodal case, one class of careers is stunted by the difficulty in making progress at the beginning of the career, analogous to a "potential" barrier. The second class of careers forges beyond the barrier and is approximately centered around the crossover x_c on a log–log scale.

influenced by many competing random processes that contribute to the random termination time T of a career (14). Our model accounts for external termination factors that are not correlated to the contemporaneous productivity of a given individual. A more sophisticated model, which incorporates endogenous termination factors, e.g., termination due to sudden decrease in productivity below a given employment threshold, is more difficult to analytically model, which we leave as an open problem. The pdf $P(x|T)$ calculated in Eq. **1** is the conditional probability that an individual has achieved a career position x by his or her given termination time T. Hence, to obtain an ensemble pdf of career longevity $P(x)$, we must average over the pdf $r(T)$ of random termination times T,

$$P(x) = \int_0^\infty P(x|T)r(T)dT. \quad \textbf{[3]}$$

We next make a suitable choice for $r(T)$. To this end, we introduce the hazard rate, $H(T)$, which is the Bayesian probability that failure will occur at time $T + \delta T$, given that it has not yet occurred at time T. This is written as $H(T) = r(T)/S(T) = -\frac{\partial}{\partial T}\ln S(T)$, where $S(T) \equiv 1 - \int_0^T r(t)dt$ is the probability of a career surviving until time T. The exponential pdf of termination times,

$$r(T) = x_c^{-1}\exp[-(T/x_c)], \quad \textbf{[4]}$$

has a constant hazard rate $H(T) = \frac{1}{x_c}$ and thus assumes that external hazards are equally distributed over time. Substituting Eq. **4** into Eq. **3** and computing the integral, we obtain

$$P(x) = \frac{g(x)^{x-1}}{x_c[\frac{1}{x_c} + g(x)]^x} \approx \frac{1}{g(x)x_c} e^{-\frac{x}{g(x)x_c}}. \quad \textbf{[5]}$$

Depending on the functional form of $g(x)$, the theoretical prediction given by Eq. **5** is much different than the null model in which there is no Matthew effect, corresponding to a constant progress rate $g(x) \equiv \lambda$ for each individual.

Using the functional form given by Eq. **2**, we obtain a truncated power law for the case of concave $\alpha < 1$, resulting in a $P(x)$ that can be approximated by two regimes,

$$P(x) \propto \begin{cases} x^{-\alpha} & x \lesssim x_c \\ e^{-(x/x_c)} & x \gtrsim x_c. \end{cases} \quad \textbf{[6]}$$

Hence, our model predicts a remarkable statistical regularity that bridges the gap between very short and very long careers as a result of the concavity of $g(x)$ in early career development.

In the case of constant progress rate $g(x) \equiv \lambda$, the pdf $P(x)$ is exponential with a characteristic career longevity $l_c = \lambda x_c$. In *SI Appendix* we further consider the null model where the constant progress rate λ_i of individual i is distributed over a given range. We find again that $P(x)$ is exponential, which is quite different from the prediction given by **6**. Furthermore, we also develop a second model where the progress rate depends on a generic career trajectory $g(t)$ that peaks at a given year corresponding to the height of an individual's talent or creativity. We solve the time-dependent model in *SI Appendix* for a simple form of $g(t)$, which results in a $P(x)$ that is peaked around the maximum career length, in contrast to our empirical findings.

In order to account for aging effects, another variation of this model could include a time-dependent $H(T)$. To incorporate a nonconstant $H(T)$ one can use a more general Weibull distribution for the pdf of termination times

$$r(T) \equiv \frac{\gamma}{x_c}\left(\frac{T}{x_c}\right)^{\gamma-1}\exp\left[-\left(\frac{T}{x_c}\right)^{\gamma}\right], \quad \textbf{[7]}$$

where $\gamma = 1$ corresponds to the exponential case (15). In general, the hazard rate of the Weibull distribution is $H(T) \propto T^{\gamma-1}$, where $\gamma > 1$ corresponds to an increasing hazard rate and $\gamma < 1$ corresponds to a decreasing hazard rate. We note that the time scale x_c appears both in the definition of $g(x)$ in Eq. **2** as a crossover between early and advanced career progress rates, and also as the time scale over which the probability of survival $S(T)$ approaches 0 in the case of $\gamma \geq 1$ in Eq. **4**. It is the appearance of the quantity x_c in the definition of $S(T)$ that results in a finite exponential cutoff to the longevity distributions. Although the time scales defined in $g(x)$ and $S(T)$ could be different, we observe only one time scale in the empirical data. Hence we assume here for simplicity that the two time scales are approximately equal.

From the theoretical curves plotted in Fig. 2, *Inset*, one observes that $\alpha_c = 1$ is a special crossover value for $P(x)$, between a bimodal $P(x)$ for $\alpha > 1$, and a monotonically decreasing $P(x)$ for $\alpha < 1$. This crossover is due to the small x behavior of the progress rate $g(x) \approx x^\alpha$ for $x < x_c$, which serves as a "potential barrier" that a young career must overcome. The width x_w of the potential barrier, defined such that $g(x_w) = 1/x_c$, scales as $x_w/x_c \approx x_c^{-1/\alpha}$. Hence, the value $\alpha_c = 1$ separates convex progress ($\alpha > 1$) from concave progress ($\alpha < 1$) in early career development.

In the case $\alpha > 1$, one class of careers is stunted by the barrier, whereas the other class of careers excels, resulting in a bimodal $P(x)$. In the case $\alpha < 1$, it is relatively easier to make progress in the beginning of the career. It has been shown in ref. 16 that random stopping times can explain power-law pdfs in many stochastic systems that arise in the natural and social sciences, with predicted exponent values $\alpha \geq 1$. Our model provides a mechanism that predicts truncated power-law pdfs with scaling exponents $\alpha \leq 1$, where the truncation is a requirement of normalization. Moreover, our model provides a quantitative meaning for the power-law exponent α characterizing the probability density function.

Empirical Evidence

The two essential ingredients of our stochastic model, namely random forward progress and random termination times corresponding to a stochastic hazard rate, are general and should apply in principle to many competitive professions. The individuals, some who are championed as legends and stars, are judged by their performances, usually on the basis of measurable metrics for longevity, success, and prowess, which vary between professions.

In scientific arenas, and in general, the metric for career position is difficult to define, even though there are many conceivable metrics for career longevity and success (17–19). We compare author longevity within individual journals, which mimic an arena for competition, each with established review standards that are related to the journal prestige. As a first approximation, the career longevity of a given author within a particular high-impact journal can be roughly measured as the duration between an author's first and last paper in that journal, reflecting his or her ability to produce at the top tiers of science. This metric for longevity should *not* be confused with the career length of the scientist, which is probably longer than the career longevity within any particular journal. Following standard lifetime data analysis methods (20), we collect "completed" careers from our dataset. The publication data we collect for each journal begins at year $Y_0 = 1958$ for all journals except for CELL (for which $Y_0 = 1974$), and ends at year $Y_f = 2008$.

For each scientific career i, we calculate $\langle\Delta\tau_i\rangle$, the average time between publications in a particular journal. A journal career that begins with a publication in year $y_{i,0}$ and ends with a publication in year $y_{i,f}$ is considered "complete" if the following two criteria are met: (*I*) $y_{i,f} \leq Y_f - \langle\Delta\tau_i\rangle$ and (*ii*) $y_{i,0} \geq Y_0 + \langle\Delta\tau_i\rangle$. These criteria help eliminate from our analysis incomplete careers that possibly began before Y_0 or ended after Y_f. We then estimate the career length within journal j as $L_{i,j} = y_{i,f} - y_{i,0} + 1$, with a year allotted for publication time, and do not consider careers with $y_{i,f} = y_{i,0}$. This reduces the size of each journal dataset by approximately 25% (for a description of data and methods, see *SI Appendix* (Sec. I and Table S1)).

In ref. 21 we further analyze the scientific careers of the authors in these six journal databases. In order to account for time-dependent and discipline-dependent factors that affect both success and productivity measures, we develop normalized metrics for career success ("citation shares") and productivity ("papers shares"). We also find further evidence of the Matthew effect by analyzing the interpublication time $\tau(x)$ that decreases with increasing publication x for individual authors within a given journal. Thus, we conclude that publication in a particular journal is facilitated by previous publications in the journal, corresponding to an increasing reputation within the given journal (22). Several other metrics for quantifying career success (18, 23), such as the h index (17) and generalizations (24, 25), along with methods for removing time- and discipline-dependent citation factors (26) have been analyzed in the spirit of developing unbiased rating systems for scientific achievement.

In athletic arenas, the metrics for career position, success, and success rate are easier to define (27). In general, a career position in sports can be measured by the cumulative number of in-game opportunities a player has obtained. In baseball, we define an opportunity as an "at bat" (AB) for batters and an "inning pitched in outs" (IPO) for pitchers, whereas in basketball and soccer, we define the metrics for opportunity as "minutes played" and "games played," respectively.

In Fig. 3 we plot the distributions of career longevity for 20,000 professional athletes in four distinct leagues and roughly 400,000 scientific careers in six distinct journals (data are publicly available at refs. 28 and 29). We observe universal statistical regularity corresponding to $\alpha < 1$ in the career longevity distributions for three distinct sports and several high-impact journals (see Table S2 for a summary of least squares parameters). The disparity in career lengths indicates that it is very difficult to sustain a competitive professional career, with most individuals making their debut and finale over a relatively short time interval. For instance, we find that roughly 3% of baseball pitchers have a career length in the MLB of one inning pitched or less, whereas we also find that roughly 3% of basketball players have a career length in the NBA of less than 12 in-game minutes. Yet, despite the relatively high frequency of short careers, there are also instances of careers that are extremely long, corresponding to roughly the entire productive lifetime of the individual. The statistical regularity that bridges the gap between the "one-hit wonders" and the "iron horses" indicates that there are careers of every length between the minimum and the maximum career length, with a smooth and monotonic relation quantifying the relative frequencies of the careers in between. Furthermore, we find that stellar careers are not anomalies, but rather, as predicted by our model, the outcome of the cumulative advantage in competitive professions. The properties of the cumulative advantage process are also compounded by an individual's "sacred spark" factor (2) that accounts for his or her relative level of talent and/or professional drive, which also factors into career longevity.

The exponential cutoff in $P(x)$ that follows after the crossover value x_c arises from the finite human lifetime and is reminiscent of any real system where there are finite-size effects that dominate the asymptotic behavior. The scaling regime is less pronounced in the curves for journal longevity. This results from the granularity of our dataset, which records publications by year

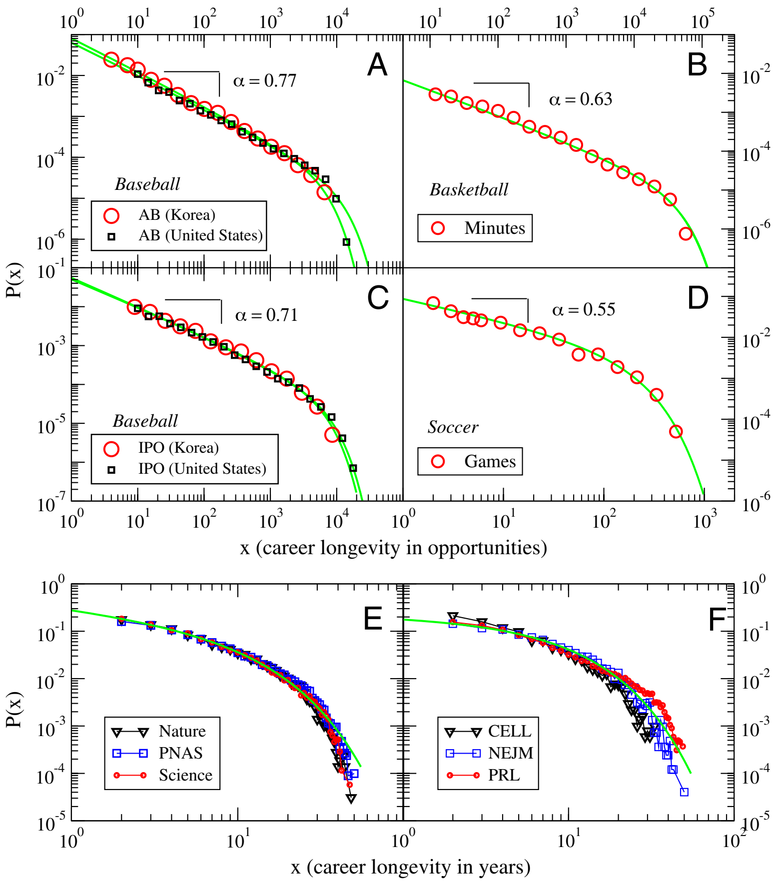

Fig. 3. Extremely right-skewed pdfs $P(x)$ of career longevity in several high-impact scientific journals and several major sports leagues. We analyze data from American baseball (Major League Baseball) over the 84-year period 1920–2004, Korean Baseball (Korean Professional Baseball League) over the 25-year period 1982–2007, American basketball (National Basketball Association and American Basketball Association) over the 56-year period 1946–2004, and English soccer (Premier League) over the 15-year period 1992–2007, and several scientific journals over the 42-year period 1958–2000. Solid curves represent least-squares best-fit functions corresponding to the functional form in Eq. **5**. (*A*) Baseball fielder longevity measured in at-bats (pitchers excluded): we find $\alpha \approx 0.77$, $x_c \approx 2{,}500$ (Korea) and $x_c \approx 5{,}000$ (United States). (*B*) Basketball longevity measured in minutes played: we find $\alpha \approx 0.63$, $x_c \approx 21{,}000$ minutes. (*C*) Baseball pitcher longevity measured in IPO: we find $\alpha \approx 0.71$, $x_c \approx 2{,}800$ (Korea), and $x_c \approx 3{,}400$ (United States). (*D*) Soccer longevity measured in games played: we find $\alpha \approx 0.55$, $x_c \approx 140$ games. (*E* and *F*) High-impact journals exhibit similar longevity distributions for the "journal career length," which we define as the duration between an author's first and last paper in a particular journal. Deviations occur for long careers due to dataset limitations (for comparison, least-square fits are plotted in *E* with parameters $\alpha \approx 0.40$, $x_c = 9$ years and in *F* with parameters $\alpha \approx 0.10$, $x_c = 11$ years). These statistics are summarized in *SI Appendix* (Table S2).

APPLIED PHYSICAL SCIENCES

SOCIAL SCIENCES

only. A finer time resolution (e.g., months between first and last publication) would likely reveal a larger scaling regime. However, regardless of the scale, one observes the salient feature of there being a large disparity between the frequency of long and short careers.

In science, an author's success metric can be quantified by the total number of papers or citations in a particular journal. Publication careers have the important property that the impact of scientific work is time dependent. Where many papers become outdated as the scientific body of knowledge grows, there are instances where "late-blooming" papers make significant impact a considerable time after publication (30). Accounting for the time-dependent properties of citation counts, in ref. 21 we find that the pdf of total number of normalized citation shares for a particular author in a single journal over his or her entire career follows the asymptotic power law $P(z)dz \sim z^{-2.5}dz$ for the six journals analyzed here.

In sports, however, career accomplishments do not wax or wane with time. In Fig. 4 we plot the pdf $P(z)$ of career success z for common metrics in baseball and basketball. Remarkably, the power-law regime for $P(z)$ is governed by a scaling exponent that is approximately equal to the scaling exponent of the longevity pdf $P(x)$. In *SI Appendix*, we show that the pdf $P(z)$ of career success z follows directly from a simple Mellin convolution of the pdf $P(x)$ for longevity x and the pdf $P(y)$ of prowess y.

The Gamma pdf $P(x) = \mathrm{Gamma}(x;\alpha,x_c) \propto x^{-\alpha}e^{-x/x_c}$ is commonly employed in statistical modeling and can be used as an approximate form of **6**. One advantage to the Gamma pdf is that it can be inverted in order to study extreme statistics corresponding to rare stellar careers. In *SI Appendix* and in ref. 31, we further analyze the relationship between the extreme statistics of the Gamma pdf and the selection processes for *Hall of Fame* museums. In general, the statistical regularity of these distributions allows one to establish robust milestones, which could be used for setting the corresponding financial rewards and pay scales, within a particular profession. Interestingly, we also find in ref. 31 that the pdfs for career success in MLB are stationary even if we quantitatively remove the time-dependent factors that can relatively inflate or deflate measures for success. This stationarity implies that the right-skewed statistical regularity we observe in $P(z)$ arises from both the intrinsic talent and the longevity of professional athletes and does not result from changes in technology, economic factors, training improvements, etc. In the case of MLB, this detrending method allows one to compare the accomplishments of baseball players across historical eras and, in particular, can help to interpret and quantify the relative achievements of players from the recent "steroids era."

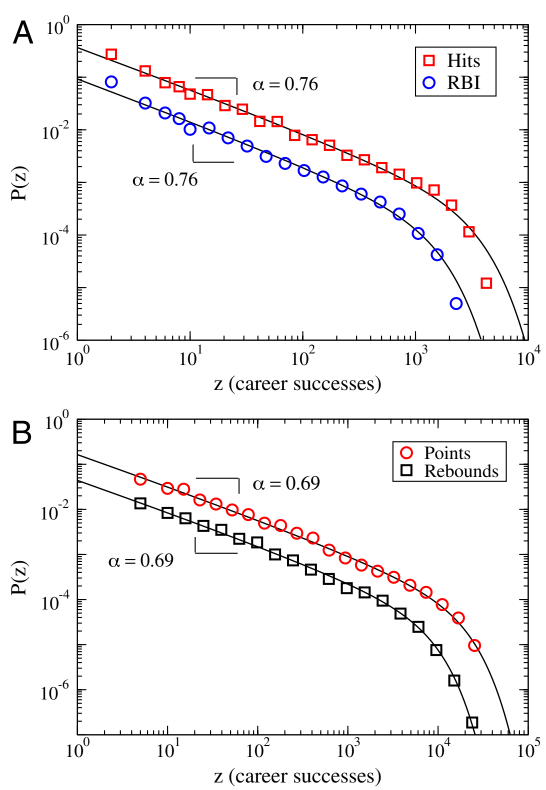

Fig. 4. Probability density function $P(z)$ of common metrics for career success, z. Solid curves represent best-fit functions corresponding to Eq. **5**. (*A*) Career batting statistics in American baseball: $x_c^{\mathrm{Hits}} \approx 1{,}200$, $x_c^{\mathrm{RBI}} \approx 600$, (RBI = runs batted in). (*B*) Career statistics in American basketball: $x_c^{\mathrm{Points}} \approx 8{,}000$, $x_c^{\mathrm{Rebounds}} \approx 3{,}500$. For clarity, the top set of data in each plot has been multiplied by a constant factor of four in order to separate overlapping data.

In summary, a wealth of data recording various facets of social phenomena has become available in recent years, allowing scientists to search for universal laws that emerge from human interactions (32). Theoretical models of social dynamics, employing methods from statistical physics, have provided significant insight into the various mechanisms that can lead to emergent phenomena (33). An important lesson from complex system theory is that oftentimes the details of the underlying mechanism do not affect the macroscopic emergent phenomena. For baseball players in Korea and the United States, we observe remarkable similarity between the pdfs of career longevity (Fig. 3) and the pdfs of prowess (Fig. S1), despite these players belonging to completely distinct leagues. This fact is consistent with the hypothesis that universal stochastic forces govern career development in science, professional sports, and presumably in a large class of competitive professions.

In this paper we demonstrate strong empirical evidence for universal statistical laws that describe career progress in competitive professions. Universal phenomena also occur in many other social complex systems where regularities arise despite the complexity of the human interactions and the spatiotemporal dynamics (34–47). Stemming from the simplicity of the assumptions, the stochastic model developed in this paper could conceivably apply elsewhere in society, such as the duration of both platonic and romantic friendships. Indeed, long relationships are harder to break than short ones, with random factors inevitably terminating them forever. Also, supporting evidence for the applicability of this model can be found in the similar truncated power-law pdfs with $\alpha < 1$, which describe the dynamics of connecting within online social networks (43).

Appendix: The Spatial Poisson Distribution

The master equation for the evolution of $P(x,N)$ is

$$P(x+1,N+1) - P(x+1,N) = f(x)P(x,N) - f(x+1)P(x+1,N), \quad [8]$$

with initial condition

$$P(x+1,0) = \delta_{x,0}. \quad [9]$$

Here $f(x)$ represents the probability that an employee obtains another future opportunity given his or her resume at career position x. We next write the discrete-time discrete-space master equation in the continuous-time discrete-space form

$$\frac{\partial P(x+1,t)}{\partial t} = g(x)P(x,t) - g(x+1)P(x+1,t), \quad [10]$$

where $g(x) = f(x)/\delta t$ and $t = N\delta t$ (for an extensive discussion of master equation formalism, see ref. 48). Taking the Laplace transform of both sides, one obtains

$$sP(x+1,s) - P(x+1,t=0) = g(x)P(x,s) - g(x+1)P(x+1,s). \quad [11]$$

From the initial condition in Eq. **9**, we see that the second term above vanishes for $x \geq 1$. Solving for $P(x+1,s)$ we obtain the recurrence equation

$$P(x+1,s) = \frac{g(x)}{s+g(x+1)} P(x,s). \quad [12]$$

If the first derivative $\frac{d}{dx}g(x)$ is relatively small, we can replace $g(x+1)$ with $g(x)$ in the equation above. Then, one can verify the ansatz

$$P(x,s) = \frac{g(x)^{x-1}}{[s+g(x)]^x}, \quad [13]$$

which is the Laplace transform of the spatial Poisson distribution $P[x,t;\lambda = g(x)]$ as in ref. 49. The Laplace transform is defined as $L\{f(t)\} = f(s) = \int_0^\infty dt f(t) e^{-st}$. Inverting the transform we obtain

$$P(x,t) = \frac{e^{-g(x)t}[g(x)t]^{x-1}}{(x-1)!}. \quad [14]$$

Hence, Eq. **14** corresponds to the pdf of final career position x observed at a particular time t. Because not all careers last the same length of time, we define the time $t \equiv T$ to be a conditional stopping time that characterizes a given subset of careers that lasted a time duration T. We average over a distribution $r(T)$ of stopping times to obtain the empirical longevity pdf $P(x)$ in Eq. **5**, which is equivalent to Eq. **13**, so that $P(x)$ is comprised of careers with varying T.

ACKNOWLEDGMENTS. We thank P. Krapivsky, G. Viswanathan, G. Paul, F. Wang, and J. Tenenbaum for insight and helpful comments. A.M.P. and H.E.S. thank the Office of Navy Research and Defense Threat Reduction Agency for financial support, W.S.J. was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology Grant 2010-0021987, and J.-S.Y. thanks Grant NRF-2010-356-B00016 for financial support.

1. Merton RK (1968) The Matthew effect in science. *Science* 159:56–63.
2. Allison PD, Stewart JA (1974) Productivity differences among scientists: Evidence for accumulative advantage. *Am Sociol Rev* 39:596–606.
3. De Solla Price D (1976) A general theory of bibliometric and other cumulative advantage processes. *J Am Soc Inf Sci* 27:292–306.
4. Allison PD, Long SL, Krauze TK (1982) Cumulative advantage and inequality in science. *Am Sociol Rev* 47:615–625.
5. Merton RK (1988) The Matthew effect in science, II: Cumulative advantage and the symbolism of intellectual property. *ISIS* 79:606–623.
6. Walberg JH, Tsai S (1983) Matthew effects in education. *Am Educ Res J* 20:359–373.
7. Stanovich KE (1986) Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. *Read Res Quart* 21:360–407.
8. Bonitz M, Bruckner E, Scharnhorst A (1997) Characteristics and impact of the Matthew effect for countries. *Scientometrics* 40:407–422.
9. "For to all those who have, more will be given, and they will have an abundance; but from those who have nothing, even what they have will be taken away." Matthew 25:29, New Revised Standard Version.
10. Cole S, Cole JR (1967) Scientific output and recognition: A study in the operation of the reward system in science. *Am Sociol Rev* 32:377–390.
11. Helsen WF, Starkes JL, Van Winckel J (1998) The influence of relative age on success and dropout in male soccer players. *Am J Hum Biol* 10:791–798.
12. Musch J, Grondin S (2001) Unequal competition as an impediment to personal development: A review of the relative age effect in sport. *Dev Rev* 21:147–167.
13. Simonton DK (1997) Creative productivity: A predictive and explanatory model of career trajectories and landmarks. *Psychol Rev* 104:66–89.
14. Segalla M, Jacobs-Belschak G, Müller C (2001) Cultural influences on employee termination decisions: Firing the Good, Average or the Old? *Eur Manag J* 19:58–72.
15. Lawless JF (2003) *Statistical Models and Methods for Lifetime Data* (Wiley, New York), 2nd Ed.
16. Reed WJ, Hughes BD (2002) From gene families and genera to incomes and internet file sizes: Why power laws are so common in nature. *Phys Rev E* 66:067103.
17. Hirsch JE (2005) An index to quantify an individual's scientific research output. *Proc Natl Acad Sci USA* 102:16569–16572.
18. Radicchi F, Fortunato S, Markines B, Vespignani A (2009) Diffusion of scientific credits and the ranking of scientists. *Phys Rev E* 80:056103.
19. Davies JA (2002) The individual success of musicians, like that of physicists, follows a stretched exponential distribution. *Eur Phys J B* 27:445–447 10.1140/epjb/ e2002-00176-y.
20. Huber JC (1999) Inventive productivity and the statistics of exceedances. *Scientometrics* 45:33–53.
21. Petersen AM, Wang F, Stanley HE (2010) Methods for measuring the citations and productivity of scientists across time and discipline. *Phys Rev E* 81:036114.
22. Castillo C, Donato D, Gionis A (2007) *Estimating Number of Citations Using Author Reputation* (Springer, Berlin).
23. Sidiropoulos A, Manolopoulos Y (2006) Generalized comparison of graph-based ranking algorithms for publications and authors. *J Syst Software* 79:1679–1700.
24. Sidiropoulos A, Katsaros D, Manolopoulos Y (2007) Generalized Hirsch h-index for disclosing latent facts in citation networks. *Scientometrics* 72:253–280.
25. Batista PD, Campiteli MG, Martinez AS (2006) Is it possible to compare researchers with different scientific interests? *Scientometrics* 68:179–189.
26. Radicchi F, Fortunato S, Castellano C (2008) Universality of citation distributions: Toward an objective measure of scientific impact. *Proc Natl Acad Sci USA* 105:17268–17272.
27. Petersen AM, Jung W-S, Stanley HE (2008) On the distribution of career longevity and the evolution of home-run prowess in professional baseball. *Europhys Lett* 83:50010.
28. Sean Lahman's Baseball Archive: http://baseball1.com/index.php; Korean Professional Baseball League: http://www.inning.co.kr; Database Sports Basketball Archive: http://www.databasebasketball.com/; Barclays Premier League: http://www.premierleague.com/.
29. ISI Web of Knowledge: www.isiknowledge.com/.
30. Redner S (2005) Citation statistics from 110 years of Physical Review. *Phys Today* 58(6):49–54.
31. Petersen AM, Penner O, Stanley HE (2010) Methods for detrending success metrics to account for inflationary and deflationary factors. *Eur Phys J B* 10.1140/epjb/e2010-10647-1.
32. Lazer D, et al. (2009) Computational social science. *Science* 323:721–723.
33. Castellano C, Fortunato S, Loreto V (2009) Statistical physics of social dynamics. *Rev Mod Phys* 81:591–646.
34. Liljeros F, et al. (2001) The web of human sexual contacts. *Nature* 411:907–908.
35. de Blasio BF, Svensson A, Liljeros F (2007) Preferential attachment in sexual networks. *Proc Natl Acad Sci USA* 104:10762–10767.
36. Newman MEJ (2005) Power laws, Pareto distributions and Zipf's law. *Contemp Phys* 46:323–351.
37. Farmer JD, Shubik M, Smith E (2005) Is economics the next physical science? *Phys Today* 58(9):37–42.
38. Barabási AL (2005) The origin of bursts and heavy tails in human dynamics. *Nature* 435:207–211.
39. Watts DJ, Strogatz SH (1998) Collective dynamics of small-world networks. *Nature* 393:440–442.
40. González MC, Hidalgo CA, Barabási AL (2008) Understanding individual human mobility patterns. *Nature* 453:779–782.
41. Malmgren RD, et al. (2008) A Poissonian explanation for heavy tails in e-mail communication. *Proc Natl Acad Sci USA* 105:18153–18158.
42. Crane R, Sornette D (2008) Robust dynamic classes revealed by measuring the response function of a social system. *Proc Natl Acad Sci USA* 105:15649–15653.
43. Leskovec J, et al. (2008) Microscopic evolution of social networks. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD'08* (Association for Computing Machinery, New York), pp 462–470.
44. Petersen AM, Wang F, Havlin S, Stanley HE (2010) Quantitative law describing market dynamics before and after interest rate change. *Phys Rev E* 81:066121.
45. Jung W-S, Wang F, Stanley HE (2008) Gravity model in the Korean highway. *Europhys Lett* 81:48005.
46. Lee K, Jung W-S, Park JS, Choi MY (2008) Statistical analysis of the Metropolitan Seoul Subway System: Network structure and passenger flows. *Physica A* 387:6231–6234.
47. Song C, Koren T, Wang P, Barabási AL (2010) Modeling the scaling properties of human mobility. *Nat Phys* 6:818–823.
48. Redner S (2001) *A Guide to First-Passage Processes* (Cambridge Univ Press, Cambridge, UK).
49. Larson RC, Odoni AR (2007) *Urban Operations Research* (Dynamic Ideas, Charlestown, MA), 2nd Ed.

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,

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

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,

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

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 $x_c\delta > 10^3$.

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

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,

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

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

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

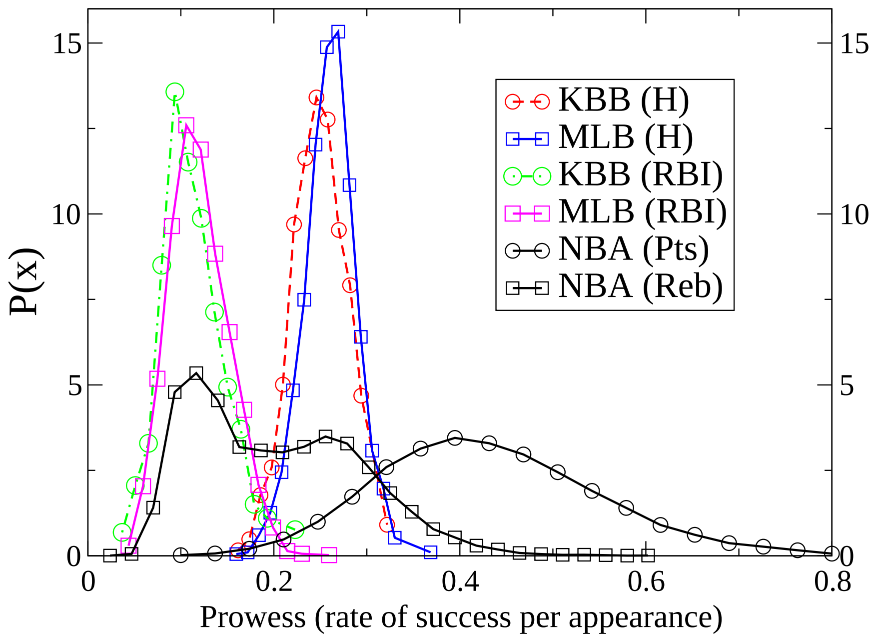

FIG. S1: Probability density functions of seasonal prowess for several career metrics. Each pdf is normally distributed, except for the bimodal curve for rebound prowess, NBA (Reb.). The bimodal distribution for Rebound prowess reflects the specialization in player positions in the sport of basketball. 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.

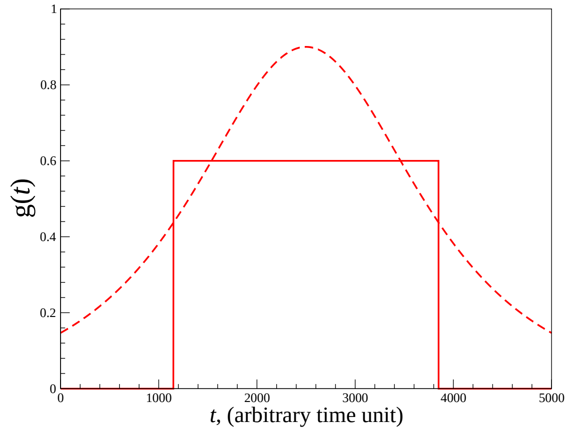

FIG. S2: A graphical illustration of a hypothetical career progress trajectory $g(t) = a\ \text{sech}[(t - t^*)/w]$ (dashed red line), with amplitude $a = 0.9$, peak time $t^* = 2500$, and width $w = 1000$, in arbitrary time units. As an approximation, in order to provide an analytic solution to the 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.

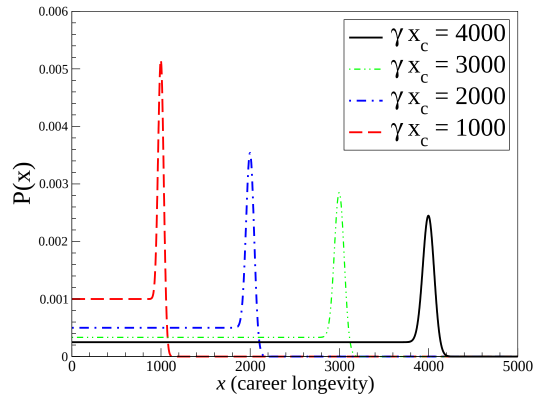

FIG. S3: Exact solutions for $P(x)$ with time-dependent career trajectory $g(t)$ defined in Eq. (S21), for the case of $t_1 = 0$, $x_c = t_2$, and $\gamma x_c = \{1000, 2000, 3000, 4000\}$.

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 $0 \leq \gamma \leq 1$ 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 $1 \leq x \lesssim \gamma x_c$, 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.