Hot Hand Phenomena in Artistic, Cultural, and Scientific Careers

Lu Liu,[1,2,3] Yang Wang,[1,2] Roberta Sinatra,[4] C. Lee Giles,[3,5] Chaoming Song,[6] Dashun Wang[1,2]

[1]*Northwestern Institute on Complex Systems, Northwestern University, Evanston, IL, USA*

[2]*Kellogg School of Management, Northwestern University, Evanston, IL, USA*

[3]*College of Information Sciences and Technology, Pennsylvania State University, State College, PA, USA*

[4]*Center for Network Science and Mathematics Department, Central European University, Budapest, Hungary*

[5]*Department of Computer Science and Engineering, Pennsylvania State University, State College, PA, USA*

[6]*Department of Physics, University of Miami, Coral Gables, FL, USA*

The hot hand phenomenon, loosely defined as winning begets more winnings, highlights a specific period during which an individual's performance is substantially higher than her typical performance. While widely debated in sports[1–3], gambling[4–7], and financial markets[8–10] over the past several decades, little is known if the hot hand phenomenon applies to individual careers. Here, building on rich literature on lifecycle of creativity[11–23], we collected large-scale career histories of individual artists, movie directors and scientists, tracing the artworks, movies, and scientific publications they produced. We find that, across all three domains, hit works within a career show a high degree of temporal regularity, each career being

characterized by bursts of high-impact works occurring in sequence. We demonstrate that these observations can be explained by a simple hot-hand model we developed, allowing us to probe quantitatively the hot hand phenomenon governing individual careers, which we find to be remarkably universal across diverse domains we analyzed: The hot-hand phenomenon is ubiquitous yet unique across different careers. While the vast majority of individuals have at least one hot-hand period, hot hands are most likely to occur only once. The hot-hand period emerges randomly within an individual's sequence of works, is temporally localized, and is unassociated with any detectable change in productivity. We show that, since works produced during hot hand garner significantly more impact, the uncovered hot-hand phenomenon fundamentally drives the collective impact of an individual, ignoring which leads us to systematically over- or under-estimate the future impact of a career. These results not only deepen our quantitative understanding of patterns governing individual ingenuity and success, they may also have implications for decisions and policies involving predicting and nurturing individuals with lasting impact.

A creative career is often defined by the sequence of works an individual produces at various stages[14,15,20,22–25]. According to the Matthew effect[12,26–28], victories bring reputation and recognition that can translate into tangible assets which in turn help bring future victories. This school of thought supports the existence of hot hand in a career, which is also consistent with the innovation literature showing that peak performance clusters in time, typically occurring around mid career[11,14,24]. Yet, on the other hand, the random impact rule uncovered in arts[13,24] and sciences[13,23] predicts the opposite: The best works occur randomly within a career, primarily driven by produc-

tivity. Following this school of thought, works after a major breakthrough are not affected by what preceded them, supporting the viewpoint of regression toward the mean. The two divergent schools of thought, together with the consequential nature of this question to individual ingenuity and success, raise a fundamental question: Does the hot hand phenomenon exist in creative careers?

To answer this question, we collected systematically large-scale datasets recording career histories of individual artists, movie directors and scientists (Supplementary Information S1), allowing us to explore across three major domains involving human creativity. The first dataset (D_1) consists of auction records curated from online auction databases, allowing us to reconstruct career histories of 3,480 artists through the sequence of works they each produced, together with impacts of the artworks, approximated by hammer prices in auctions[20]. D_2 contains profiles of 6,233 movie directors recorded in the IMDB database, each career being represented by the sequence of movies he/she directed. Since metrics that quantify impacts of a movie correlate closely with each other[29], here we use the IMDB ratings to measure the goodness of a movie. Finally, our third dataset (D_3) includes publication records of 20,040 individual scientists through a large-scale name disambiguation effort that combined the Web of Science and Google Scholar datasets (Supplementary Information S1.3). The impact of each paper is measured by citations garnered after 10 years of its publication[17,21,23,30], C_{10}. Since hammer price (D_1) and C_{10} (D_3) both follow fat-tailed distributions (Fig. S3), here we take the logarithmic of these measures, i.e., $\log$(price) and $\log(C_{10})$ to approximate the goodness of an artwork and scientific publication.

Motivated by Merton's theory of multiples[12,31], we start by investigating the timing of the three most impactful works produced in each career. In a sequence of N works by an individual, we denote with N^* the position of the highest impact work within a career, N^{**} the second highest and N^{***} the third. We find each of the three highest impact works occurs randomly within a career (Fig. S4). That is, when it comes to any of the three most expensive artworks by an artist, three highest rated movies by a director, and three highest impact publications by a scientist, each of them is randomly distributed among all the works one produces. These results offer strong endorsement for the random impact rule[13,23,24], hence supporting an unpredictable view of individual creativity.

Yet, as we show next, the random impact rule observed across creative careers is only apparent, because the timing between creative works follows highly predictable patterns. Indeed, we measure the correlation between the timing of the two biggest hits within a career (e.g., N^* and N^{**}) by calculating the joint probability $P(N^*, N^{**})$, and compare it with a null hypothesis in which N^* and N^{**} each occurs at random. We find that, the normalized joint probability, $\phi(N^*, N^{**}) = P(N^*, N^{**})/(P(N^*)P(N^{**}))$, is significantly overrepresented along the diagonal elements of matrices (Figs. 1a–c), demonstrating that N^* and N^{**} are much more likely to colocate with each other than what we would expect from the random impact model. Moreover, the colocation pattern is universal across a wide range of careers we studied, including artists (Fig. 1a), movie directors (Fig. 1b) and scientists (Fig. 1c). The diagonal pattern disappears if we shuffle the order of works within each career, thereby breaking the temporal correlations between highest impact works while preserving the random impact rule (Figs. 1d–f, Fig. S6).

To quantify the temporal colocation of hits observed in Figs. 1a–c, we calculate the distance between two highest impact works for every individual, measured by the number of works produced in between, $\Delta N = N^* - N^{**}$. We compare $P(\frac{\Delta N}{N})$ of real careers with $P_S(\frac{\Delta N}{N})$ of shuffled careers by defining $R(\frac{\Delta N}{N}) = P(\frac{\Delta N}{N})/P_S(\frac{\Delta N}{N})$. For artists, movie directors, and scientists, $R(\frac{\Delta N}{N})$ all exhibits a clear peak centering around zero and decays quickly as ΔN deviates from zero (Figs. 1j–l). Indeed, the two most important works of an artist is 1.48 times more likely to occur back-to-back than expected by chance (Fig. 1j). The same is true for movie directors (Fig. 1k) and scientists (Fig. 1l), where such colocation is 1.42 and 1.57 times more likely than their baseline occurrence rate, respectively. Also important to note is the interesting fact that $R(\frac{\Delta N}{N})$ is mostly symmetric around zero (Figs. 1j–l), indicating a comparable likelihood for the biggest hit to arrive before or after the second biggest for all three types of careers. This symmetry was also captured by Figs. 1a–c, where ϕ features a roughly even split across the diagonal. The colocation patterns documented in Figs. 1a–c and 1j–l are not limited to the two highest impact works within a career. Indeed, we repeated our analyses for other pairs of hit works, such as N^* vs. N^{***} and N^{**} vs. N^{***}, uncovering the same colocation patterns (Figs. 1j–l and Fig. S5).

Do high impact works come in streaks within a career? To answer this question, we count the number of consecutive works whose goodness exceeds a certain threshold across various careers (Figs. 1m–o). Here we choose the median goodness of all works within a career as the threshold. We calculate the length of the longest streak L for each career, and measure the distribution of L across our user base in each of the three domains. We then shuffle the order of works within each career, and measure again their longest streaks L_s. We find $P(L)$ is characterized by a much

longer tail, compared with $P(L_s)$ (Figs. 1p–r), indicating real careers are characterized by long streaks of excellent works clustered together in sequence. Note that for the three types of careers, the tail part of both $P(L)$ and $P(L_s)$ follows approximately an exponential function, meaning that the likelihood to observe a longer streak diminishes rather rapidly. Hence, the deviations observed between $P(L)$ and $P(L_s)$ are rather significant (Fig. S11). To test the robustness of these results, we repeated our analyses by controlling for individual career length, and also varying our threshold used to calculate L, finding our conclusions remain the same (Supplementary Information S2.4 and S2.5).

Taken together, results presented in Fig. 1 paint a rather unexpected portrait of individual careers. Indeed, while the timing of high impact works each appears at random by itself, their relative timing, however, follows highly reproducible yet previously unknown patterns. As such, individual careers are far from being random, but characterized by bursts of high-impact works occurring in sequence. These fascinating empirical findings raise an important question: What are the mechanisms responsible for the temporal regularities observed across diverse career histories?

To unearth the fundamental mechanisms governing the patterns documented in Fig. 1, let us first consider a null model in which the goodness of works produced in a career (i.e.. $\log$(price) for artists, ratings for directors, and $\log(C_{10})$ for scientists) is drawn from a normal distribution $\mathcal{N}(\Gamma_i,\, \sigma_i^2)$, fixed for an individual. The average Γ_i characterizes typical impact of works produced by the individual, and σ_i captures the variance. This null model reproduces the fact that each hits occur randomly within a career[13,23] and the differences in typical impact between careers

(Supplementary Information S3.2, Figs. S25–S27). Yet it fails to capture any of the temporal correlations observed in Fig. 1. The main reason is illustrated in Figs. 2a–c, where we selected for illustration purposes one individual from each of the three datasets and measure the dynamics of Γ_i during his/her career. We find Γ_i is not constant throughout a career. Rather, it features deviations from a baseline performance (Γ_0) at a certain point of a career ($t_\uparrow$), elevating to a higher value Γ_H ($\Gamma_H > \Gamma_0$), which is then sustained for some time before falling back to a similar level as Γ_0 (Figs. 2a–c):

$$\Gamma(t) = \begin{cases} \Gamma_H & t_\uparrow \leq t \leq t_\downarrow \\ \\ \Gamma_0 & otherwise \end{cases}, \tag{1}$$

This observation, combined with the shortcomings of the null model, raises an intriguing hypothesis: Could a simple model based on (1) explain the temporal anomalies documented in Fig. 1?

To test this hypothesis, we apply (1) to real productivity patterns of an individual, allowing us to generatively simulate impacts of the works produced by an individual (Supplementary Information S3.3). As an individual's baseline performance is captured by Γ_0, during the period in which Γ_H operates ($t_\uparrow \leq t \leq t_\downarrow$), the individual seemingly performs at a higher level than her typical performance (Γ_0), prompting us to call this model, the hot-hand model, and correspondingly, the Γ_H period as the hot-hand period. Hence, in the hot-hand model, the goodness of works produced by an individual is drawn from two distributions $\mathcal{N}(\Gamma_0,\, \sigma^2)$ and $\mathcal{N}(\Gamma_H,\, \sigma^2)$, depending on whether the individual is within the hot-hand period. We introduce to each career one hot-hand period that occurs at random with a fixed duration and magnitude, and repeat our measurements in Fig. 1 on careers generated by the model. We find, while equation (1) only introduces a sim-

ple temporal variation, surprisingly the hot-hand model is sufficient in reproducing all empirical observations that existing modeling frameworks fail to account for from the temporal colocations among top hit works within a career (Figs. 1g–i), to their temporal distances (Figs. 1j–l), to the occurrences of long streaks of excellent works (Figs. 1p–r). For full comparisons with existing models, see Supplementary Information S4. Equation (1) assumes implicitly each individual has one hot-hand period, a hypothesis that we test later. Given the myriad factors that can affect career impacts[12–16,23,25,32,33], and the obvious diversity of careers we studied, the level of universality and accuracy demonstrated by the simple hot-hand model is rather unexpected.

The real value of the model arises, however, when we fit the model to real careers to obtain the individual specific Γ_0, Γ_H, $t_\uparrow$ and $t_\downarrow$ parameters (Supplementary Information S3.4), allowing us to probe quantitatively the hot hand phenomenon underlying artistic, cultural, and scientific careers, and helping us reveal several fundamental patterns governing individual careers:

1. The hot-hand period is ubiquitous *across* careers, yet at the same time rather unique *within* a career. We find, the vast majority of artists (91%, Fig. 2d), movie directors (82%, Fig. 2e) and scientists (90%, Fig. 2f) have at least one hot-hand period throughout their careers, documenting the practical relevance of the uncovered hot hand phenomenon. Yet, despite its ubiquity, the hot-hand period is most likely to be unique within a career. Indeed, we relax our fitting algorithm to allow for multiple hot hands (up to three) with different values of Γ_H, finding that, among those who have a hot hand, 64% of artists, 80% of directors, and 68% of scientists are best captured by one hot-hand period only (Figs. 2d–f), documenting

the precious nature of hot hand. Second acts may occur but less likely, particularly for movie directors. About 30% of artists and scientists have two hot-hand periods, but only 11% for directors. Occurrences of more than two hot hands are rare across all careers. We also find that, between those who have one or two hot hands, there is no detectable difference in terms of typical performance metrics, including impact, productivity and career length (Fig. S22), suggesting that hot hand captures an orthogonal dimension to current metrics characterizing individual careers.

2. The hot-hand period occurs randomly within a career. We estimate the beginning of hot-hand periods for artistic, cultural, and scientific careers. Denoting with $N_{\uparrow}$, the position of work produced when hot hand starts ($t_{\uparrow}$), we find that $N_{\uparrow}$ is randomly distributed in the sequence of N works within a career (Figs. 2g–i). This finding reconciles two seemingly divergent schools of thought[12,13,23], providing a further explanation for the random impact rule: If the hot-hand period occurs randomly within a career, and the highest impact works are statistically more likely to appear within the hot-hand period, then the timing of the highest impact works would also appear random.

3. Across different domains, hot hand lasts for a considerably shorter period comparing with the typical career length recorded in our database. We measure the duration distribution of hot hands ($\tau_H = t_{\uparrow} - t_{\downarrow}$), finding $P(\tau_H)$ peaks around 5.7 years for artists, 5.2 years for directors, and 3.7 years for scientists (Figs. 2j–l). Interestingly, the duration of hot hand is independent of when it occurs within a career (early, mid or late career, Figs. 2j–l). The temporally localized nature of hot hand is also captured by its proportion over career length

τ_H/T (Figs. 2j–l, insets), whose median hovers around 20% (0.17 for artists, 0.23 for directors, and 0.20 for scientists).

4. How much does an individual deviate from her typical performance during hot hand? Do people with higher Γ_0 also experience more performance gain from hot hand? To answer these questions, we explore correlations between Γ_0 and Γ_H, finding them to be well approximated by a linear relationship across three kinds of careers (Figs. 2m–o). Hence the better typical performance, the better individuals perform during their hot-hand periods. It is interesting to note that the coefficients are slightly less than 1 (0.99 for artists, 0.85 for directors, and 0.9 for scientists, Figs. 2m–o). Hence $\Delta\Gamma \equiv \Gamma_H - \Gamma_0$ decreases with Γ_0 (Figs. 2m–o insets), suggesting individuals with smaller Γ_0 benefit more from hot hand. These results are again independent of when hot hand occurs along a career (Figs. 2m–o).

5. Are individuals more productive during hot hands? Surprisingly, the answer is no. We measure the distribution of the total number of works produced during hot hand $P(N_H)$. We then construct a null distribution, by randomly picking one work out of a career and designating its production year to be the start of the hot-hand period. We find $P(N_H)$ measured in real careers well aligns with the null model's predictions for all three kinds of careers (Figs. 2p–r). Therefore, individuals show no detectable change in productivity during hot hand, despite the fact that their outputs during the period are significantly better than typical, suggesting an endogenous shift in individual creativity when hot hand occurs.

What is the impact of hot hand on individual careers? To answer this question, we focus

on scientific careers (D_3), and measure the collective impact of a scientist, $g(t)$, defined as the total number of citations over time collected by all the papers one published. Brought to spotlight by popular websites such as Google Scholar (Fig. 3a), $g(t)$ is playing an increasingly important role in driving many critical decisions, from hiring, promotion and tenure to awarding of grants and rewards. Next we show the collective impact of an individual is fundamentally governed by the uncovered hot-hand phenomenon, ignoring which would lead us to systematically under- or over-estimate the future impact of a scientist.

Many factors are known to influence the collective impact of a career, ranging from productivity[15,22,34] to citation disparity and dynamics[17,18,21,25,27,35] to temporal inhomogeneities along a career[14,22–25,36]. Since our goal is to understand impact, here we bypass the need to evaluate the inhomogeneous nature of productivity[22,23] by rearranging publication time of each paper, such that an individual produces a constant number of papers each year, denoted by n (Figs. 3b–c). To calculate $g(t)$ analytically, we need to incorporate papers' citation patterns into our hot-hand model (1). A recent study[21] suggests that citation dynamics of a paper published at time t_0 can be approximated by

$$C(t,t_0) = m\left(e^{\lambda\Phi\left(\frac{\ln(t-t_0)-\mu}{\sigma}\right)}-1\right) \equiv m\left(e^{\Gamma(t_0)\Phi\left(\frac{\ln(t-t_0)-\mu}{\sigma}\right)}-1\right), \tag{2}$$

where m is a global parameter describing the typical number of references a paper contains, and $\Phi(\cdot)$ is the cumulative normal function, characterized by μ and σ, which capture the typical citation life cycle of a paper. The paper's impact is ultimately determined by its fitness[21], λ. To adapt this model into our framework, we replace λ with $\Gamma(t_0)$, and for simplicity assume μ and σ are fixed for different papers published by an individual. The resulting model, combining (1) and (2), can be solved analytically (Supplementary Information S5.1–S5.4), allowing us to express $g(t)$ in terms

of hot-hand parameters (Γ_0, Γ_H, $t_\uparrow$, $t_\downarrow$):

$$g(t) = \underbrace{nm(e^{\Gamma_0 \Phi\left(\frac{\ln(t)-\mu}{\sigma}\right)} - 1)}_{g_0(t)} + \underbrace{\begin{cases} 0 & t < t_\uparrow \\ nm(\Gamma_H - \Gamma_0)\Phi\left(\frac{\ln(t-t_\uparrow)-\mu}{\sigma}\right) C(t, t_\uparrow) & t_\uparrow \le t < t_\downarrow \\ nm(\Gamma_H - \Gamma_0)[\Phi\left(\frac{\ln(t-t_\uparrow)-\mu}{\sigma}\right) C(t, t_\uparrow) - \\ \qquad \Phi\left(\frac{\ln(t-t_\downarrow)-\mu}{\sigma}\right) C(t, t_\downarrow)] & t \ge t_\downarrow \end{cases}}_{\Delta g(t)} . \quad (3)$$

Equation (3) consists of two terms. $g_0(t)$ captures a career's collective impact in the absence of hot hand (i.e. $\Gamma(t) = \Gamma_0$). Contributions from hot hand are encoded in $\Delta g(t)$, driven by both the timing and magnitude of hot-hand periods ($t_\uparrow$, $t_\downarrow$, Γ_H, and $\Gamma_H - \Gamma_0$). Varying hot hand parameters significantly alters the collective impact of a career (Fig. 3d).

We adopt two measures to quantify the accuracy of our model (3). To account for the inherently noisy career trajectories, we first assign an impact envelope to each individual, explicitly quantifying the uncertainty of model predictions (Fig. 3e, Supplementary Information S5.6). We measure the fraction of $g(t)$ that fall within the envelope, finding the distribution across individuals peaks close to 1 (Fig. 3f), indicating most career trajectories are well encapsulated within the predicted envelopes. The superior accuracy of our model is also captured by the Mean Absolute Percentage Error (*MAPE*) (Fig. 3g), with improvement being most pronounced for an early onset of hot hand (Fig. 3g), which is also correctly predicted by our model. Hence the hot-hand model captures a wide range of trajectories that collective impacts of scientific careers follow (Fig. 3h).

The observed accuracy prompts us to ask whether the hot-hand model is unique in its ability

to capture the impact of individual careers across diverse domains. There are several alternative hypotheses capturing different hot hand dynamics (Supplementary Information S6), each associated with possible origins of the uncovered hot hand phenomena: (*A*) A right trapezoid (Fig. S32b) captures a sudden onset of hot hand with a more gradual decline, as innovators may stumble upon a groundbreaking idea, which manifests itself in the forms of multiple artworks, movies, and publications. Hence from an evolutionary perspective, the duration of hot hand may characterize time it takes for the temporary competitive advantage to dissipate. (*B*) An isosceles trapezoid model (Fig. S32c) captures hot hand that evolves and dissolves gradually over time, which may approximate social tie dynamics, as one individual's hot hand could be the result of a fruitful, repeated collaboration[32,37]. (*C*) Furthermore, individual performance may peak at a certain point of a career, prompting us to test inverted-U shape (Fig. S32d) and tent functions (Fig. S32e). Lastly (*D*) a left trapezoid function (Fig. S32f) captures a gradual startup period with a sharp cutoff, corresponding to career opportunities that can augment impact but last for a fixed duration.

We tested hypotheses *A*–*D* systematically to describe real careers (Supplementary Information S6). Of all hypotheses considered, the proposed hot-hand model is the simplest and least flexible. Yet, surprisingly, it is the only model whose predictions are consistent with real careers (Fig. S32). The fact that none of the alternative hypotheses alone can fully account for empirical observations demonstrates the hot hand phenomena uncovered in creative careers may not be driven by one particular factor but a combination of multiple factors. Identifying its true origin requires additional experimentation and goes beyond the scope of this work. As such, hot hand uncovered in this paper should be treated in a metaphorical sense, highlighting an intriguing period

of outstanding performance tracing individual careers without implying any associated drivers for the phenomena. Yet, crucially, the findings presented in this paper hold the same, regardless of the underlying drivers.

The analytical framework presented here not only offers a new theoretical basis for our quantitative understanding of dynamical patterns governing individual career impact, it also has policy implications for comparing and evaluating scientists (Fig. S30). Indeed, for individuals whose hot hands are yet to come, ignoring hot hand may lead to underestimating their impacts (Figs. S30a–b), especially given the ubiquitous nature of hot hand (Fig. 2f). On the other hand, an early onset of hot hand leads to a high impact that peaks early but may not sustain unless a second act occurs (Fig. S30c). Given that individuals improve substantially during hot hand, the uncovered hot hand phenomenon can be particularly crucial for decisions and policies concerning long-term impact of a career.

1. Gilovich, T., Vallone, R. & Tversky, A. The hot hand in basketball: On the misperception of random sequences. *Cognitive Psychology* **17**, 295–314 (1985).

2. Bar-Eli, M., Avugos, S. & Raab, M. Twenty years of hot hand research: Review and critique. *Psychology of Sport and Exercise* **7**, 525–553 (2006).

3. Miller, J. B. & Sanjurjo, A. Surprised by the gambler's and hot hand fallacies? A truth in the law of small numbers. *IGIER Working Paper No. 552* (2016).

4. Ayton, P. & Fischer, I. The hot hand fallacy and the gambler's fallacy: Two faces of subjective

randomness? *Memory & Cognition* **32**, 1369–1378 (2004).

5. Croson, R. & Sundali, J. The gambler's fallacy and the hot hand: Empirical data from casinos. *Journal of Risk and Uncertainty* **30**, 195–209 (2005).

6. Rabin, M. & Vayanos, D. The gambler's and hot-hand fallacies: Theory and applications. *The Review of Economic Studies* **77**, 730–778 (2010).

7. Xu, J. & Harvey, N. Carry on winning: The gamblers' fallacy creates hot hand effects in online gambling. *Cognition* **131**, 173–180 (2014).

8. Hendricks, D., Patel, J. & Zeckhauser, R. Hot hands in mutual funds: Short-run persistence of relative performance, 1974–1988. *The Journal of Finance* **48**, 93–130 (1993).

9. Jagannathan, R., Malakhov, A. & Novikov, D. Do hot hands exist among hedge fund managers? An empirical evaluation. *The Journal of Finance* **65**, 217–255 (2010).

10. Kahneman, D. & Riepe, M. W. Aspects of investor psychology. *The Journal of Portfolio Management* **24**, 52–65 (1998).

11. Lehman, H. C. *Age and achievement* (Princeton, NJ, 1953).

12. Merton, R. K. The matthew effect in science. *Science* **159**, 56–63 (1968).

13. Simonton, D. K. Age and outstanding achievement: What do we know after a century of research? *Psychological Bulletin* **104**, 251 (1988).

14. Jones, B. F. & Weinberg, B. A. Age dynamics in scientific creativity. *Proceedings of the National Academy of Sciences* **108**, 18910–18914 (2011).

15. Petersen, A. M., Jung, W.-S., Yang, J.-S. & Stanley, H. E. Quantitative and empirical demonstration of the matthew effect in a study of career longevity. *Proceedings of the National Academy of Sciences* **108**, 18–23 (2011).

16. Petersen, A. M. *et al.* Reputation and impact in academic careers. *Proceedings of the National Academy of Sciences* **111**, 15316–15321 (2014).

17. Stringer, M. J., Sales-Pardo, M. & Amaral, L. A. N. Effectiveness of journal ranking schemes as a tool for locating information. *PLoS One* **3**, e1683 (2008).

18. Radicchi, F., Fortunato, S. & Castellano, C. Universality of citation distributions: Toward an objective measure of scientific impact. *Proceedings of the National Academy of Sciences* **105**, 17268–17272 (2008).

19. Ke, Q., Ferrara, E., Radicchi, F. & Flammini, A. Defining and identifying sleeping beauties in science. *Proceedings of the National Academy of Sciences* **112**, 7426–7431 (2015).

20. Galenson, D. W. *Old masters and young geniuses: The two life cycles of artistic creativity* (Princeton University Press, 2011).

21. Wang, D., Song, C. & Barabási, A.-L. Quantifying long-term scientific impact. *Science* **342**, 127–132 (2013).

22. Way, S. F., Morgan, A. C., Clauset, A. & Larremore, D. B. The misleading narrative of the canonical faculty productivity trajectory. *Proceedings of the National Academy of Sciences* (2017).

23. Sinatra, R., Wang, D., Deville, P., Song, C. & Barabási, A.-L. Quantifying the evolution of individual scientific impact. *Science* **354**, aaf5239 (2016).

24. Simonton, D. K. *Scientific genius: A psychology of science* (Cambridge University Press, 1988).

25. Duch, J. *et al.* The possible role of resource requirements and academic career-choice risk on gender differences in publication rate and impact. *PloS One* **7**, e51332 (2012).

26. Merton, R. K. The matthew effect in science, ii: Cumulative advantage and the symbolism of intellectual property. *Isis* **79**, 606–623 (1988).

27. Price, D. d. S. A general theory of bibliometric and other cumulative advantage processes. *Journal of the Association for Information Science and Technology* **27**, 292–306 (1976).

28. Barabási, A.-L. & Albert, R. Emergence of scaling in random networks. *Science* **286**, 509–512 (1999).

29. Wasserman, M., Zeng, X. H. T. & Amaral, L. A. N. Cross-evaluation of metrics to estimate the significance of creative works. *Proceedings of the National Academy of Sciences* **112**, 1281–1286 (2015).

30. Uzzi, B., Mukherjee, S., Stringer, M. & Jones, B. Atypical combinations and scientific impact. *Science* **342**, 468–472 (2013).

31. Merton, R. K. Singletons and multiples in scientific discovery: A chapter in the sociology of science. *Proceedings of the American Philosophical Society* **105**, 470–486 (1961).

32. Wuchty, S., Jones, B. F. & Uzzi, B. The increasing dominance of teams in production of knowledge. *Science* **316**, 1036–1039 (2007).

33. Clauset, A., Arbesman, S. & Larremore, D. B. Systematic inequality and hierarchy in faculty hiring networks. *Science Advances* **1**, e1400005 (2015).

34. Shockley, W. On the statistics of individual variations of productivity in research laboratories. *Proceedings of the IRE* **45**, 279–290 (1957).

35. Redner, S. Citation statistics from 110 years of physical review. *Physics Today* **58**, 49 (2005).

36. Moreira, J. A., Zeng, X. H. T. & Amaral, L. A. N. The distribution of the asymptotic number of citations to sets of publications by a researcher or from an academic department are consistent with a discrete lognormal model. *PloS One* **10**, e0143108 (2015).

37. Palla, G., Barabási, A.-L. & Vicsek, T. Quantifying social group evolution. *Nature* **446**, 664–667 (2007).

Acknowledgements The authors thank A.-L. Barabasi, B. Uzzi, W. Ocasio, J. Chown, C. Jin, Y. Yin, and all members of Northwestern Institute on Complex Systems (NICO) for invaluable comments. This work is supported by the Air Force Office of Scientific Research under award number FA9550-15-1-0162 and FA9550-17-1-0089, and Northwestern University's Data Science Initiative.

Correspondence Correspondence and requests for materials should be addressed to D.W. (email: dashun.wang@northwestern.edu).

Figure captions

Figure 1: The hot-hand phenomenon in artistic, cultural and scientific careers. **a–c,** $\phi(N^*, N^{**})$, color coded, measures the joint probability of the top two highest impact works within a career for **a** artists, **b** directors, and **c** scientists. $\phi(N^*, N^{**}) > 1$ indicates two hits are more likely to colocate than random. **d–f,** we shuffle the order of each work in a career while keeping their impact intact, allowing us to measure the null hypothesis of $\phi(N^*, N^{**})$ across three domains, where N^* and N^{**} each occurs at random. The diagonal patterns in **a–c** disappear for shuffled careers. **g–i,** $\phi(N^*, N^{**})$ predicted by the hot-hand model successfully recovers the diagonal patterns observed in **a–c**. **j–l,** $R(\frac{\Delta N}{N})$ measures the temporal distance between highest impact works relative to null model's prediction. Red dots denote measurements from data, showing a clear peak around 0. Solid lines in red are predictions by the hot-hand model. Different shades of red correspond to different pairs of hit works. Blue dots denote the same measurement but on shuffled careers, and blue lines are predictions from shuffled careers generated by our model. **m–o,** Definitions of the longest streak L within a career for **m** artists, **n** directors and **o** scientists. L measures as the maximum number of consecutive works whose impacts are above the median impact of a career (horizontal dashed line denoting the threshold). Above the threshold, dots are colored in orange, and blue for below the threshold. L in the lower panel highlights the longest streak in a career. **p–r,** The distribution of the length of streaks $P(L)$ for real careers and $P(L_s)$ for shuffled careers, for **p** artists, **q** directors and **r** scientists. Red dots capture empirical observations, whereas blue dots correspond to shuffled careers. Our hot-hand model (red lines) closely reproduces $P(L)$ observed in data, demonstrating the model's validity to capture the impact beyond top three highest impact

works across different domains. The shuffled version of our model (blue lines) also well captures shuffled careers.

Figure 2: The Hot-hand Model. a–c, $\Gamma(N)$ for one **a** artist, **b** movie director and **c** scientist, selected from our data for illustration purposes. $\Gamma(N)$ is calculated by the moving average of impact with a window length $= 0.1 \times N$. **d–f,** The histogram of the number of hot-hand periods for **d** artists, **e** directors, and **f** scientists. **g–i,** $N_{\uparrow}/N$ measures the position of the work when hot hand occurs, among N works in a career. Their cumulative distributions $P(\geq \frac{N_{\uparrow}}{N})$ for **g** artists, **h** directors and **i** scientists are shown in blue dots. The red line captures the cumulative distribution when the start of hot hand $N_{\uparrow}$ is distributed randomly among N works. **j–l,** The duration distribution of the hot-hand period $P(\tau_H)$ for **j** artists, **k** directors and **l** scientists. Dots are measurements from data. Red lines are log-normal fits as guide to the eye. The median τ_H are 7.3 years for artists, 7.0 years for directors, and 4.8 years for scientists, respectively. Inset, relative duration distribution $P(\tau_H/T)$ for individuals in three domains, where T is the career length of each individual. Solid lines are lognormal fits as guide to the eye. **m–o,** The relationship between Γ_H and Γ_0 for **m** artists, **n** directors and **o** scientists, where the blue background denotes kernel density of data, dots represent binning results of data, and the red line depicts the linear fit. Within each domain, Γ_H and Γ_0 for individuals with early, middle, and late hot-hand period can be well approximated by a linear relationship. Inset, the relationship between $\Delta\Gamma$ $(= \Gamma_H - \Gamma_0)$ and Γ_0 for each domain. **p–r,** The distribution of the number of works produced during hot hand $P(N_H)$, compared with a null distribution, where we randomly pick one work as the start of the hot hand period (**p** artists,

q directors, and **r** scientists.) We use the Kolmogorov-Smirnov (KS) measure to compare $P(N_H)$ of data with the null distribution, finding that we cannot reject the hypothesis that the two distributions are drawn from the same distribution ($p > 0.05$).

Figure 3: Hot hand governs the collective impact of scientific careers. **a,** Screenshot of Albert Einstein's Google Scholar profile. **b,** Collective impact of a randomly selected scientist in our dataset D_3. The publication dates are rearranged such as one produces a constant number of papers each year (lower panel). Vertical lines in the lower panel depict when each paper is published after the rearrangement. The color indicates the order of publications, showing that the sequence of papers published in each career remains intact. The solid line indicates that the paper has been published for at least 10 years (dashed line, otherwise) **c,** For the same scientist as (b), citation patterns of each papers are shown with corresponding colors denoted in the lower panel. The collective impact of a career represents the sum of citation dynamics of all papers published by the individual. **d,** $g(t)$ modelled by (3) given different hot hand parameters (red lines). Here we use $\mu = 7.0$, $\sigma = 1.0$, $\Gamma_0 = 1.0$, and $\tau_H = 3$ years but vary $t_\uparrow$ and $t_\downarrow$. Varying hot-hand parameters of $g(t)$ allows us to reproduce a wide variety of career dynamics that cannot be captured by the null model (blue line). Inset decomposes contributions to $g(t)$ in terms of $g_0(t)$ and $\Delta g(t)$. **e,** The uncertainty envelope of $g(t)$ for an individual in our dataset, where blue dots denote data, the red line is the fitting result of equation (3), and the shaded area illustrates the predicted uncertainty measured in one standard deviation. **f** The fraction of $g(t)$ falling within the envelop P(fraction) for the null model (blue area) and our hot-hand model (red area). Fraction = 1.0 indicates the whole

$g(t)$ trajectory falls within the envelope. Our model outperforms the null model in capturing individual collective impact as P(fraction) peaks close to 1.0. **g,** The average $\langle MAPE \rangle$ of our hot-hand model and the null model for individuals with early, mid and late onset of hot hand. The difference between the two models is the largest for individuals with early hot hands and smallest for late ones. **h,** $g(t)$ of 50 randomly selected individuals in our dataset whose careers started between 1960 and 1995. Color corresponds to the year when a career started, dots denote collective impact of real careers, and solid lines capture the predictions from the hot-hand model.

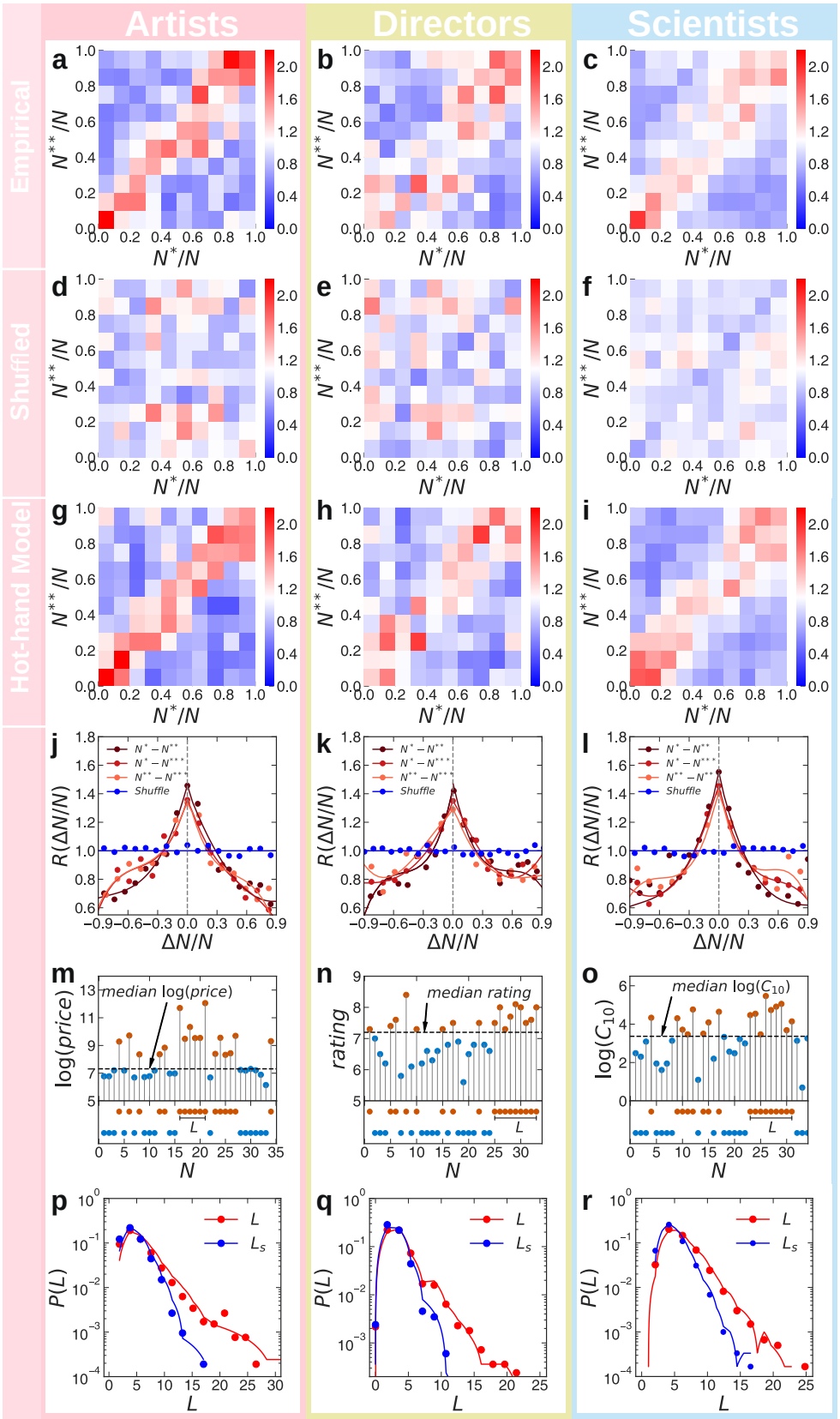
Artists
Directors
Scientists
Empirical
Shuffled
Hot-hand Model
a
b
c
d
e
f
g
h
i
j
k
l
m
n
o
p
q
r
N^*/N
N^{**}/N
$\Delta N/N$
$R(\Delta N/N)$
$N^+ - N^{++}$
$N^+ - N^{+++}$
$N^{++} - N^{+++}$
Shuffle
median log(price)
log(price)
median rating
rating
median log(C_{10})
log(C_{10})
N
L
$P(L)$
L_s

Figure 1:

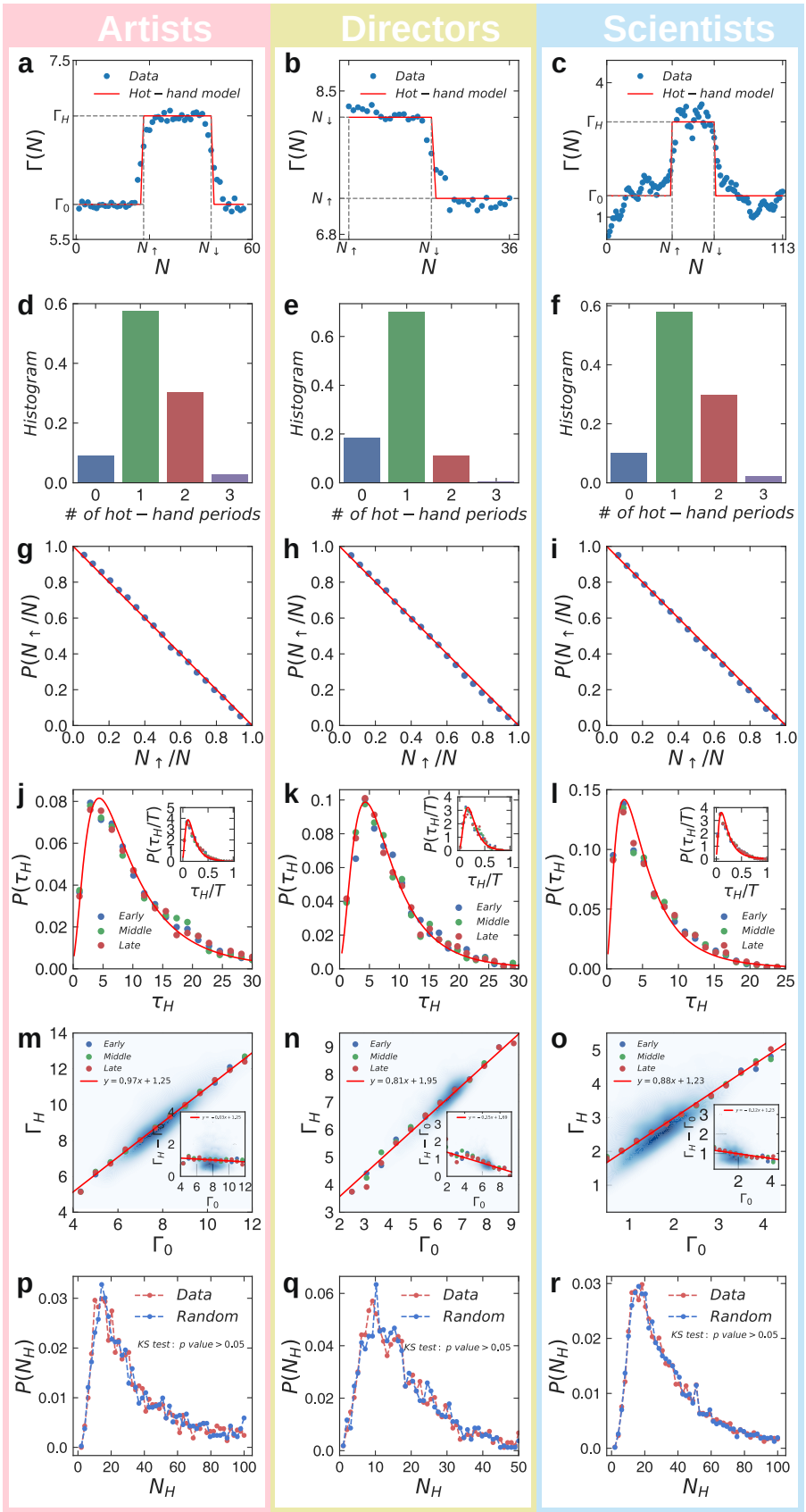
Artists
Directors
Scientists
a
b
c
Data
Hot – hand model
Γ(N)
N
d
e
f
Histogram
# of hot – hand periods
g
h
i
P(N↑/N)
N↑/N
j
k
l
P(τH)
τH
τH/T
Early
Middle
Late
m
n
o
ΓH
Γ0
y = 0,97x + 1,25
y = 0,81x + 1,95
y = 0,88x + 1,23
p
q
r
P(NH)
NH
Data
Random
KS test : p value > 0.05

Figure 2:

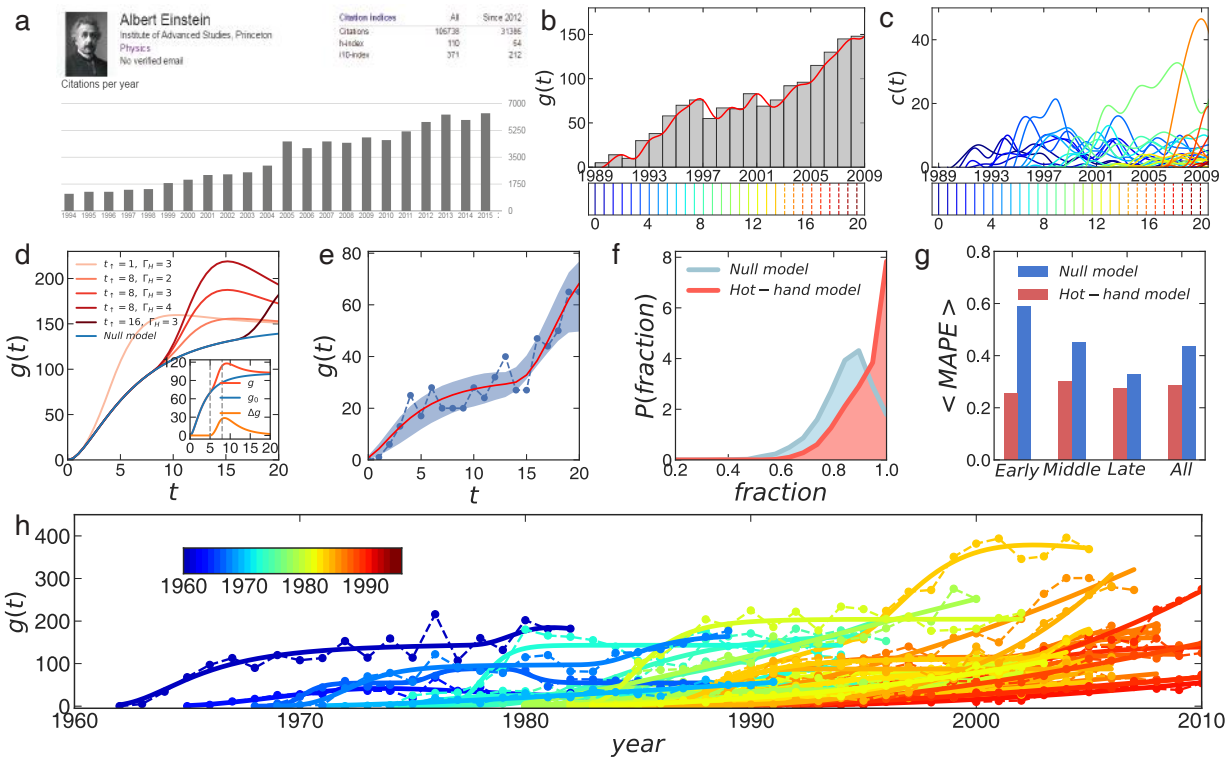
a
Albert Einstein
Institute of Advanced Studies, Princeton
Physics
No verified email
Citations per year
b
g(t)
c
c(t)
d
g(t)
t
e
g(t)
t
f
Null model
Hot – hand model
P(fraction)
fraction
g
Null model
Hot – hand model
< MAPE >
Early Middle Late All
h
g(t)
year

Figure 3:

Supplementary Information for Hot Hand Phenomena in Artistic, Cultural, and Scientific Careers

Lu Liu,[1,2,3] Yang Wang,[1,2] Roberta Sinatra,[4] C. Lee Giles,[3,5] Chaoming Song,[6] Dashun Wang[1,2*]

[1]*Northwestern Institute on Complex Systems, Northwestern University, Evanston, IL, USA*

[2]*Kellogg School of Management, Northwestern University, Evanston, IL, USA*

[3]*College of Information Sciences and Technology, Pennsylvania State University, State College, PA, USA*

[4]*Center for Network Science and Mathematics Department, Central European University, Budapest, Hungary*

[5]*Department of Computer Science and Engineering, Pennsylvania State University, State College, PA, USA*

[6]*Department of Physics, University of Miami, Coral Gables, FL, USA*

**Correspondence should be addressed to D.W. (dashun.wang@northwestern.edu)*

Contents

S1 Data Description

In this project, we compiled a comprehensive database consisting of three large-scale datasets of individual careers across three different domains: Dataset D_1 contains profiles of artists obtained from online auction databases. Dataset D_2 contains profiles of movie directors recorded in the IMDB database. Dataset D_3 contains the publication and citation records of individual scientists, obtained by combining Google Scholar and Web of Science. In this section we describe in detail how we collected and reconstructed individual career histories for the three datasets.

S1.1 Artists D_1 Among the three domains we analyzed, the success of artists is probably the most difficult to quantify, hence unsurprisingly the least studied. Indeed, apriori, it may seem that the success of an artist is inherently unquantifiable. Yet more recent developments start to suggest that several measures have the potential to be systematically collected and can be used to measure and compare the impact of artworks[20]. The various measures include auction hammer prices, being selected as illustrations in art textbooks, retrospective exhibitions, museum collections and exhibitions. Just like citations—which have by now been commonly adopted to quantify success of a paper and scientist, despite the fact that they offer at best an incomplete measure of impact that is inherently multi-faceted—these measures of artistic success each capture at best a singular dimension that is to certain degree correlated with the overall "goodness" of an artwork, with their associated limitations (Sec. S1.4). Among all measures, hammer prices are the most commonly used to quantify artistic success[20], perhaps because they reflect the values of artworks judged by art professionals and art markets, serving as a proxy for the impact of artworks. While there are

studies showing the career trajectories of artists are rather robust against different measurement choices[20,38], here to ensure our findings are consistent with the state of art, we seek to collect systematically hammer prices of artworks from different sources to reconstruct artistic careers.

We collected information on artistic careers from online art market databases, Artprice[1] and Findartinfo[2]. Both websites offer a comprehensive list of auction records for each artist, with complementary information on various kinds of fine arts ranging from old masters to contemporary art. Indeed, auction records on Artprice offer useful information that can help quantify artistic careers, including, for each artwork, its auction date, title, year of production, medium, and the price rank among all hammer prices of artworks produced by an artist. Findartinfo contains information on auction date, title, medium, and the actual hammer price for each auction. In this study, we combined the two data sources, allowing us to extract the most comprehensive information tracing individual artistic careers.

Artworks can go through multiple auctions. Artprice website contains auction records from 1983 up to now, helping us ensure that we analyze the latest sales of each artwork. Findartinfo also offers auction records from 2001 till 2015, including the actual hammer price for each record, allowing for cross check and references. Both databases are excellent in their longitudinality, containing artworks produced dating back to the Middle Ages.

In total, we collected 31,101 individual profiles from Artprice and 283,839 profiles from

[1] www.artprice.com

[2] www.findartinfo.com

Findartinfo. Artprice indexed a large number of artists, but here we focused on the top ones with more than 50 records. We conducted a comprehensive entity linking process between the two databases, aiming to match each artist in Artprice with the corresponding profile in Findartinfo. We ignore profiles from the two databases noted as "attributed to" or "attrib.", since these notions suggest sellers are not clear about the authorship of these artworks. We then cluster remaining profiles with the same last name together in each database, allowing us to match artists within a small subset for computational efficiency. We compare each artist's profile in Artprice with each profile in Findartinfo with the same last name. Two artists are considered to be the same if they satisfy the following criteria: 1) Initials of the first names are the same. If full names are available for both artists, they have to be identical; 2) They have at least one artwork with an identical title; 3) If more than one artists meet criteria 1) and 2), we pick the one with most matched titles. By applying this entity resolution procedure, we end up with a total of 5,352 matched artists. To evaluate the accuracy of our algorithm in linking the two databases, we compare the number of works for all matched artists in Findartinfo and Artprice in the same period (2001–2015), finding the number of works in the two databases to be similar (Fig. S1).

After linking the two databases, we reconstruct the career histories of artists based on the production year of each work and its hammer price during most recent auction. If an artwork is included in the original Findartinfo dataset, the actual hammer price is used to measure the impact of the artwork. If an artwork is only included in the Artprice dataset, we know the price rank of the artwork among all artworks by the artist. Hence if we need to compute the actual hammer price, we can convert sales rank using rank-to-score conversion given the price distribution

measured in Findartinfo[39]. Note that, the uncovered hot-hand phenomena documented in the main text (Fig. 1) are independent of whether we use price or rank because the choice does not affect the measurements of relative positions between hit works within each individual. When an artwork was produced during several years, we use the last year as its year of production, corresponding to the year in which the work was finalized. For our final dataset, we selected artists with at least 15 works and 10 years of career length, resulting in 3,480 artists, with careers dating back to as far as 1460.

S1.2 Movie directors D_2 The Internet Movie Database (IMDB)[3] is the largest movie database around the world, containing information about over one million movies, spanning over 20 genres from 1874 to present. Each movie in the database includes detailed information such as title, release time, casts, crews, an average user rating and the number of votes. IMDB also contains over eight million personal profiles with unique identifiers, each containing personal information, a list of works she was involved in, and specific roles in these works, such as being a director, editor, writer, actress, etc. Although a movie is inherently a product of complex collaborative efforts often involving a large number of individuals, the director of the movie is commonly considered to play a prominent role in cinematic creativity[40,41].

To this end, we gathered 513,306 movie records and profiles of 20,592 directors from IMDB, including those who serve as an assistant director and art director. IMDB provides a rating system for each movie that ranges from one to ten, reflecting the "crowd wisdom" of users, adjusted by

[3] http://www.imdb.com/interfaces/

the weighting algorithms developed by IMDB to avoid vote stuffing. Previous work has found that metrics quantifying the impact of a movie largely correlate with each other[29]. Here, we use the IMDB rating to approximate the impact of a movie, and construct the sequence of works with their impacts for each director in our dataset. We focused on movies released before 2017 with more than 5 votes. To select directors with long enough career histories, we focused on those who have at least 15 movies and 10 years of career length, resulting in a total of 6,233 directors, whose careers dating back to as far as 1890 (D_2).

S1.3 Scientists D_3 For studies on scientific careers, automated name disambiguation in large-scale scholarly datasets remains a challenging problem[42,43]. In this study we perform a large-scale disambiguation effort by combining two large datasets, Google Scholar (GS) and Web of Science (WoS). Google offers scholar profile services for individual scientists to create, maintain, and update their own publication records, assisted by its disambiguation algorithms. Users can adjust the publication records recommended by Google, further ensuring the accuracy and reliability of each profile. Hence, GS offers a comprehensive dataset of individual scientific profiles across different domains that should be at least as accurate as the state of the art, with additional two levels of assurance. As such it has the potential to catalyze more and more research on scientific careers[23,44,45].

To this end, we crawled over 240,000 public profiles from GS in summer 2015. Each GS profile contains the publication records of a scientist, including for each paper its title, publication year, journal, author(s) and cumulative citation count within GS database calculated up to the point

when the data was collected. GS encourages users to enter other information including affiliation, research interest, homepage, and collaborators. Users with verified academic email address are noted in their GS pages. Here, we choose those with verified email address and co-authors as they are considered well-maintained. We further remove individuals with more than 1,000 publications in GS, finding most of them have Asian names which are the most difficult to disambiguate. Here again the goal is to follow closely the same procedures used by existing studies in this area[23,44,45], so that every finding in our paper is immediately comparable with previous literature to ensure our results are consistent hence new findings reliable given the existing literature.

Since GS only provides the cumulative citation count of each paper by the time of data collection, to study the dynamical impact of each scientist we linked GS with the WoS dataset, which provides comprehensive citation records of around 46 million journal papers published after 1900. For each scientist, we match each publication in her GS profile with the corresponding paper in WoS. We conducted a comprehensive linkage process that takes into account not only the author name and paper title, but also the metadata available for each publication. For each scientist in GS we first consider a pool of WoS papers having an author with the same last name. Then for each publication in her GS profile, we calculated the cosine similarity between the title of each GS record and each WoS paper in the subset. We then consider a record in GS profile matched with a WoS publication if the following criteria are met: 1) The WoS paper was published within ± 2 years of the GS publication year; 2) There is at least one co-author sharing the same last name in GS and WoS if the paper has more than one authors; 3) The cosine similarity between titles is higher than 0.5 after removing stop words. If there are multiple papers, we choose the one

with the most similar title (highest cosine similarity). If, using these criteria, we could not find a corresponding paper in WoS, we consider the GS record is not matched to ensure the quality of the resulting dataset. We choose scientists with at least 15 papers and 20 years of career length, resulting in 20,040 profiles for our analyses (D_3).

To approximate the impact of each paper, we follow recent studies[17,21,23] to calculate the number of citations the paper received after 10 years of publication C_{10}. Previous studies have shown that for WoS dataset the average citation counts of a paper increase over time[18,23,46], which we verified to be the case for our dataset (Fig. S2). To make sure our results are not affected by this temporal effect, we follow previous studies[18,23] and use a rescaled C_{10}, defined as $C_{10}/\langle C_{10}\rangle_y$, to gauge the impact of a paper, where $\langle C_{10}\rangle_y$ is the average impact of all papers in WoS published in the same year y. We report in S2 and Fig. 1 the results based on the rescaled C_{10}. We find the use of raw or rescaled C_{10} does not alter our conclusions.

S1.4 Data limitations Although the datasets curated in our study are among the largest collections of individual careers, our data are not without limitations. While auction data are most commonly used to quantify impact of an artwork, it is important to understand a potential bias involved in using auction data to measure and compare the relative value of an artwork[20]. This issue is of particular relevance for famous artists, whose work is eagerly sought after by museums. While museums sometimes take part in auctions, they are in general less likely to sell a collection. The "museum bias" may also affect works differently. For example, if we consider the probability that an oil painting may be owned by a museum, late paintings tend to have a higher probability

than early paintings. Also museums may not take paintings randomly from a career, but are rather biased towards the best works. Despite these limitations, one possible reason why auction prices remain the best measures to quantify the importance of artworks is rooted in the fact that they are significantly better than alternatives. For example, other measures like museum exhibitions mainly focus on masterpieces of famous artists, limiting the scope and data availability for quantitative analysis.

Datasets about movies and scientific publications are more frequently studied quantitatively than artworks, hence metrics used in those cases are better defined. But it is still important to remember that potential biases also exist. For example, the IMDB rating could be biased towards the judgement of general audience, whereas professional film critics may have different opinions. Papers could be highly cited for many reasons. There are many scientific discoveries that are ground-breaking but have few citations.

Lastly, for any study regarding individual careers, name disambiguation is always an issue that should not be overlooked. For the three domains analyzed here, name disambiguation issues may potentially affect results regarding artists and scientists, but less so for movie directors, as they are associated with unique identifiers. Another issue when studying individual careers is that, as we need enough records to study each individual, the insights obtained are inherently biased towards productive individuals who have produced much with long enough careers.

Amid all the limitations described in this section, it should be clear, however, that these limitations are by no means unique to our study; rather, they apply to most, if not all studies in

this domain. In many ways, research progress on individual careers is hindered by the inherent difficulties to collect and curate high-resolution individual career histories from diverse domains. Hence another contribution of our paper is to make available to broad research communities the comprehensive datasets we collected in this paper. By doing so, the hope is that these datasets could be an important asset for researchers from many different disciplines as they significantly improve the data-scarce situation, allowing researchers to build and develop new findings. To this end, we created a dedicated website to host datasets and descriptions which can be accessed here: http://personal.psu.edu/lpl5107/data/.

S2 Empirical Measurements

S2.1 Impact distribution We use auction hammer price, IMDB rating, and paper citation in 10 years C_{10}, to approximate the impact of works for artists, movie directors, and scientists, respectively. To study the impact across three domains, we measure the distribution of hammer prices, IMDB ratings and paper citation count in our datasets. The hammer price for all artworks in our dataset can be approximated by a log-normal distribution (Fig. S3a)

$$P(x) = \frac{1}{x\sqrt{2\pi}\sigma} e^{-\frac{(\log x - \mu)^2}{2\sigma^2}} \tag{S1}$$

We fit Eq. S1 to $P(\text{price})$ measured from data, and obtain the estimated parameters $\mu = 7.91$ and $\sigma = 1.55$. Similarly, we find both raw and rescaled C_{10} in our dataset also follow log-normal distributions (Figs. S3c–d) consistent with previous studies[17,18,23]. In contrast, the IMDB rating follows a normal distribution ranging between 1 and 10 (Fig. S3b). Hence, to study the impact across three domains, we define a goodness parameter Γ for artists, directors and scientists as

$\log$(price), IMDB ratings and $\log(C_{10})$, respectively.

S2.2 Random impact rule One school of thought on the lifecycle of creative careers suggests a hit work within individual career is largely driven by productivity, having a constant probability to appear in unit of works within each career[13,23,47,48]. To verify this hypothesis in our dataset, we study the position N^i of each top i highest impact work in the sequence of N works within a career, and measure the complementary cumulative distribution $P(\geq N^i/N)$ within the sequence of works produced by individuals in different domains. We find for artists, $P(\geq N^i/N)$ for each of the top three hit works decreases linearly as $(N^i/N)^{-1}$, corresponding to a uniform $P(N^i/N)$ (Fig. S4a). Similarly, we find the same pattern of $P(\geq N^i/N)$ for directors and scientists (Figs. S4b–c). Hence, when it comes to any of the three most expensive artworks by an artist, three highest rated movies by a director, and three highest impact publications by a scientist, each of them appears randomly distributed within a career, suggesting the random impact rule[13,23,47,48] applies across three different domains.

S2.3 Φ for other pairs of hit works In the main text, we observe that the normalized joint probability $\Phi(N^*, N^{**})$ is overrepresented along the diagonal for artists, directors and scientists, suggesting the colocation of the top two hit works within a career. To study if the colocation applies to other pairs of hit works, we calculate the normalized joint probability $\Phi(N^*, N^{***})$ for the highest and third highest impact works, and $\Phi(N^{**}, N^{***})$ for the second and third highest impact works within each career across three domains. We first measure $\Phi(N^*, N^{***})$ and $\Phi(N^{**}, N^{***})$ in real careers (Fig. S5), finding a similar diagonal pattern for $\Phi(N^*, N^{***})$ and $\Phi(N^{**}, N^{***})$, indicating

the temporal collocation applies to other pairs of hit works. $\Phi(N^*, N^{***})$ and $\Phi(N^{**}, N^{***})$ both feature roughly even split across the diagonal, corresponding to an equal probability for hits to appear before or after another hit. However, if we shuffle the order of works within each career while keeping their impact intact, the diagonal pattern of $\Phi(N^*, N^{***})$ and $\Phi(N^{**}, N^{***})$ disappears across three domains (Fig. S6), suggesting the colocation of any pair of hit works observed in data cannot be explained by the random impact rule.

S2.4 Measurement under different career length The empirical observations in Fig. 1 are based on artists and directors with at least 10 years of career length, and scientists with at least 20 years of career length. To study if the random impact rule and the temporal correlation of hit works are influenced by the career length we measured, we selected two groups of individuals with longer career length in each domain: artists with at least 20 and 30 years of career length, directors with at least 20 and 30 years of career length, and scientists with at least 30 and 40 years of career length, and repeated the same analyses. First, we find the $P(\geq N^i/N)$ still follows a uniform distribution given different career length for artists, directors and scientists (Fig. S7), indicating the results are robust against the various career length we measured. Second, we calculated the temporal distance $R(\Delta N^*/N)$ for the hit and second hit within a career, where ΔN^* is defined as $N^* - N^{**}$. We find $R(\Delta N^*/N)$ peaks around zero, suggesting the colocation of hit works is not influenced by career length. Third, we find the tail of both $P(L)$ and $P(L_S)$ follows an exponential distribution. $P(L)$ features a much wider than $P(L_S)$ (Fig. S9), demonstrating the results on long streaks remain the same in careers of different length.

S2.5 $P(L)$ under different threshold We calculated the streak of works above the median impact within a career in Fig. 1, finding $P(L)$ observed in real careers has a longer tail than that of shuffled careers. To test if the conclusion is robust to different choices of impact threshold, we changed the threshold to the mean or the geometric mean of impact within each career for artists, directors, and scientists. We recovered similar results for $P(L)$ and $P(L_S)$ given different impact threshold (Fig. S10).

S2.6 Difference between $P(L)$ and $P(L_S)$ To quantify the different probabilities to observe streaks for real and shuffled careers (Figs. 1p–r), we measured $P(L)/P(L_S)$, capturing how much more likely it is to observe streaks than random at different streak length. Since we care about the probability of long streaks, we focus on the tail of $P(L)$ and $P(L_S)$, and fit their tails to an exponential distribution. We compared the tail difference of $P(L)$ and $P(L_S)$ based on the fitting results across three domains (Fig. S11). We find for artists and scientists, the probability for an individual to produce 20 consecutive high impact works above the median impact within a career, is around 20 times more likely than shuffled careers. The probability is over 100 times for directors.

S3 Hot-hand Model

S3.1 Review of hot hand studies in the literature The debate on hot hand dates back to 1985, when Gilovich *et al.* presented that the hot-hand belief on basketball players is merely a cognitive bias of random process[1]. Since then, hot hand has been measured and reported by independent research groups, each using different datasets and statistical models, mainly focused on sports,

financial markets and gambling. These studies tend to answer two fundamental questions: 1) uncovering statistical evidence on whether the hot hand exists[1–3], and 2) psychological origins of hot hand[1,49–52]. Here we conduct a comprehensive review of the existing literature related to the first question. Table S1 reviews selected empirical studies of whether hot hand exists. Table S2 reviews different mathematical models to detect hot hand.

Our study is complementary to existing studies on hot hand summarized in Table S1 and S2. We treat each creative career as a sequence of works, and by quantifying the importance of theses works, we ask the question of whether hot hand exists in an individual careers. In this respect, our paper serves as a bridge that connects two distinct areas that have been pursued largely in parallel—on one hand, the extant literature on the lifecycle of creativity[11–23], including the recent surge of interest in science of science[14–16,22,23,33,44], and on the other hand, studies of hot hand which have primarily been situated within the psychology literature[1,2,4,7,52]. As such, our study adds value to both areas. On the lifecycle of creativity, our paper documents an important yet previous unknown pattern underlying individual careers, with critical implications for both innovators and institutions that support them. On the other hand, by documenting the existence of hot hand in individual careers, our study helps extend this important line of enquiry to broader domains.

S3.2 Null model To uncover the regularities behind the empirical observations, we first introduce a null model motivated by the random impact rule[13,23,47,48]. That is, we assume each individual i produces a sequence of N works whose impact (i.e.. $\log(\text{price})$ for artists, ratings for directors, and $\log(C_{10})$ for scientists) is randomly drawn from an impact distribution $P(\Gamma) = \mathcal{N}(\Gamma_i, \sigma_i^2)$,

where Γ_i is a constant goodness parameter specific to each individual, and σ_i reflects the impact fluctuation within each career. For simplicity, we assume σ_i to be the same for each individual in a domain. The null model allows us to simulate impacts of the works within a career. For each individual, we use real productivity N as input, and assume Γ_i as the average impact measured from each individual, and $\sigma_i = 1.0$ to be a constant for all the individuals. We repeated the same measurements of $P(\geq N^i/N)$ and Φ for the null model. We find the null model can reproduce the random impact rule of the top three hit works across three domains (Fig. S12), while it fails to capture any temporal clustering among hits (Fig. S13), demonstrating that there are other factors affecting individual careers.

S3.3 The hot-hand model as a generative framework The failure of the null model prompts us to abandon our hypothesis that Γ_i is constant with each career. Indeed, each hit work is random, while their relative position is not, indicating the presence of a period of outstanding performance (Γ_H) that appears randomly with a career. Using real productivity N as input, the hot-hand model allows us to generatively simulate the impacts of works produced by an individual. For each individual i, the impact of a work is randomly drawn from a normal distribution $\mathcal{N}(\Gamma_H,\, \sigma_i^2)$ if it is produced during hot hand, or $\mathcal{N}(\Gamma_0,\, \sigma_i^2)$ otherwise. To define a random hot-hand period in each career, we randomly pick one work out of the sequence of N works she produced, and denote its year of production as $t_\uparrow$, marking the start of the hot-hand period. For simplicity, we assume Γ_0, Γ_H, σ_i, and $\tau_H = t_\uparrow - t_\downarrow$ to be the same for each individual in a domain. The result reported in Fig. 1 is based on the following parameters: $\Gamma_0 = 6.9$, $\sigma = 1.1$ and $\tau_H = 6$ years for artists; $\Gamma_0 = 6.5$, $\sigma = 1.1$ and $\tau_H = 6$ years for directors; and $\Gamma_0 = 3.0$, $\sigma = 1.3$ and $\tau_H = 4$ years for scientists,

with $\Gamma_H = \Gamma_0 + 1.0$ for individuals in all three domains. Although it is a simple generative model with four parameters, our hot-hand model can reproduce all the empirical findings measured from a variety of individual careers.

S3.4 Inferring model parameters In order to test how well our hot-hand model matches empirical data, we need to estimate the model parameters for each individual. We show in this section that the impact of works within a career can be captured by a time-dependent variable $\Gamma(N)$, defined as the average of impact calculated by the moving average with a window of size $\Delta N = \max(5, 0.1N_T)$ over the sequence of works by each individual, where N_T is the total number of works. In order to capture the average performance during a period, we assume the window size to account for 10% of all the works one produced, and set the lower bound of ΔN as 5 to calculate $\Gamma(N)$ with enough statistics. To remove any potential boundary effect, we calculate $\Gamma(N)$ from $\Delta N/2$, whose value is defined by the average Γ between 0 and ΔN, and then move the sliding window one work per step until $N_T - \Delta N/2$. $\Gamma(N)$ reflects the average impact of works between $N - \Delta N/2$ and $N + \Delta N/2$. We find for scientists, calculating the raw or the rescaled C_{10} does not affect the trend of $\Gamma(N)$ (Fig. S14a). Indeed, we measured the Pearson correlation between the $\Gamma(N)$ sequence calculated from raw $\log(C_{10})$ and rescaled $\log(C_{10})$ for each individual, finding the distribution of the correlation coefficient $P(\rho)$ peaks around 1 (Fig. S14b), suggesting the two sequences are highly correlated and the rescale does not change the trend of $\Gamma(N)$ much. Since the raw $\log(C_{10})$ is easier to interpret, here we report results related to $\Gamma(N)$ based on the raw $\log(C_{10})$.

For each individual in our dataset, we used a piecewise function to fit the sequence of $\Gamma(N)$ measured from real careers. Specifically, we relax the number of hot hands to at most three, and allow the Γ_H of each hot-hand period to be different. We used ordinary least square (OLS) to fit the piecewise function to data. To overcome the over-fitting problem, we added the $L1$ regularization term to the cost function that penalizes the number of hot hands. For each individual, we repeated the fitting procedures for 20 realizations, and selected the results with the smallest cost. We show more fitting results for individuals across three domains in Figs. S15–S17. To define hot hand in each career, we assume the smallest fitted $\Gamma(N)$ as Γ_0 for an individual. To make sure the fitted hot hand reflects a substantial period of improved performance, we set a threshold for both the duration and the magnitude of hot hand. Specifically, we define a hot-hand period if $\Gamma_H - \Gamma_0$ is larger than the inherent noise within a career (standard deviation of $\Gamma(N)$), and Γ_H lasts for more than 5 works.

S3.5 Model evaluation: Fitting performance To systematically evaluate the goodness of fit for the procedure described above, we measured the difference between fitted and real $\Gamma(N)$ by calculating the coefficient of determination R^2. To study if R^2 between the fitted and real $\Gamma(N)$ can be explained by the inherent noise of $\Gamma(N)$ sequence, we calculated the expected value of R^2 generated by the noise in impact. To do so, we used fitted $\Gamma(N)$ as input, and simulated the impacts of works within a career for each individual, by assuming the impact of each work is drawn randomly from a normal distribution $\mathcal{N}(\Gamma_H,\, \sigma_s^2)$ if it was produced during hot hand, or from $\mathcal{N}(\Gamma_0,\, \sigma_s^2)$, otherwise. We assume σ_s to be the same for all individuals in each domain.

To determine the value of σ_s, we calculated the difference σ between the real $\Gamma(N)$ and fitted $\Gamma(N)$, and measured the distribution $P(\sigma)$ for all individuals in each domain. We find $P(\sigma)$ follows a normal distribution that peaks around zero for individuals across three domains (Fig. S18). The standard deviation for $P(\sigma)$ is 0.186 for artists, 0.229 for directors, and 0.189 for scientists. We approximate the standard deviation for $P(\sigma)$ as the noise in each domain, respectively. Using these numbers as input, we simulated the impacts of works within a career for 1000 realizations, allowing us to calculate a distribution $P(R^2)$ for each individual. To study if R^2 of real and fitted $\Gamma(N)$ can be explained by noise, we define a baseline R^2 corresponding to the lowest 5% of all simulated R^2 (p-value $= 0.05$). If the R^2 of data and fitted $\Gamma(N)$ is larger than the baseline R^2, we assume the error is mainly generated by noise, and assume $\Gamma(N)$ is well captured by the hot-hand model (Fig. S19). We find for individuals in our dataset, over 69% artists, 80% directors and 75% scientists have R^2 larger than the baseline R^2, suggesting the hot-hand model captures the majority of individuals in our dataset. We also compared the fitting performance of the hot-hand model with the null model by calculating the adjusted R^2 and Bayesian information criterion (*BIC*), both penalizing the number of parameters in the model. Compared with the null model, we find the hot-hand model systematically has larger adjusted R^2 and smaller *BIC* for individuals across three domains (Fig. S20), suggesting the hot-hand model better captures the dynamics of $\Gamma(N)$ than the null model.

S3.6 Model validation: Impact before and after hot hand The hot-hand model assumes that after hot hand, the individual performance returns back to one's normal performance. To test this assumption, we measured for each individual the average impact of all works produced during the

normal period before and after hot hand. We calculated the distribution of the difference between average impact $P(\Delta\langle\Gamma\rangle)$, where $\Delta\langle\Gamma\rangle = \langle\Gamma\rangle_{after} - \langle\Gamma\rangle_{before}$. We find $P(\Delta\langle\Gamma\rangle)$ follows a normal distribution peaks around zero (Fig. S21) for individuals across three domains, demonstrating that there is no systematic difference between the average impact before and after hot hand.

S3.7 Comparing individuals with different numbers of hot hands In this section, we discuss if there is any difference among individuals with zero, one and more than one hot-hand periods across three domains. We first compared the distribution of average impact $P(\Gamma)$, the number of works $P(N)$, and the career length $P(T)$ for individuals with and without hot hand, finding that there is virtually no difference in terms of impact, productivity and career length (Fig. S22). We further compared the distributions of Γ_0 and Γ_H for individuals with different number of hot-hand periods (Fig. S23). We find again they are characterized by very similar distributions across all three domains.

S3.8 Comparing across different scientific disciplines To compare the hot-hand phenomenon for scientists from different disciplines, we identify each scientist her primary discipline by using the journal categories provided by WoS[18]. D_3 contains publications from 145 different categories in WoS ranging from Acoustic to Zoology. For each scientist, we counted the number of papers published in each of the 145 categories. When a journal belongs to multiple categories, we counted each paper multiple times in all related categories. We consider the category with the most publications within a scientific career as her primary research discipline. Hence, each scientist in our dataset is assigned to one discipline. Here we study the top 30 disciplines with the most scientists

in our datasets. We find the ratio of scientists with one hot hand to be stable across disciplines, accounting for roughly 60% within each discipline (Fig. S24a), where the ratio is highest (up to 70%) for scientists from oncology, and the lowest (around 50%) for scientists from ophthalmology. We also calculated the average duration of hot hand $\langle \tau_H \rangle$ for scientists from 30 disciplines (Fig. S24b), finding scientists from geophysics have on average the longest hot hand (around 7.5 years), whereas scientists from chemistry have the shortest hot hand on average (around 3 years). These results demonstrate that the uncovered hot hand phenomenon is robust and general across scientific disciplines.

S4 Relationship with Existing Models

In this section, we discuss the relationship between the proposed hot-hand model with previous studies on individual impact. Here we focus on the most recent model of individual careers, namely the Q-model[23]. Different from the assumption that scientists with similar productivity have indistinguishable impact, the Q-model suggests the existence of a hidden parameter Q unique to each individual. Although building upon a similar conceptual basis as the hot-hand model, the Q-model is not sufficient to capture the impact dynamics within each career. This is because the two models were built for different purposes: the Q-model is mainly designed to capture the impact differences *across* individuals, focusing on the overall performance of an individual rather than the how impact changes *within* a career, which is the focus of our model. Next we show in this section the mathematical consistency between the hot-hand model and the Q-model. As such, the hot-hand model not only helps us explain the temporal regularities documented in the main text which are

not anticipated by the Q-model; it is also able to reproduce independently all the predictions the Q-model makes on individual careers.

S4.1 Connecting Γ and Q The Q-model assumes C_{10} of a paper α for a scientist i, is determined by the multiplication of two factors[23]:

$$C_{10,i\alpha} = Q_i p_\alpha, \tag{S2}$$

where Q_i is a individual-specific parameter, and p_α is the luck component that forms the same distribution for every individual. The Q-model assumes Q_i to be a constant over time for each individual, and p_α follows a log-normal distribution that $P(\log p) = \mathcal{N}(\mu_p, \sigma_p)$. Hence, we can take the logarithms of Eq. S2 that $\log(C_{10,i\alpha}) = \log(Q_i) + \log(p_\alpha)$. Noting that C_{10}, p_α and Q_i all follow log-normal distributions[23], the impact of a paper produced by a scientist i is randomly drawn from a normal distribution

$$P(\log C_{10,i\alpha}) = \mathcal{N}(\mu_p + \log Q_i, \sigma_p), \tag{S3}$$

whose mean value is modulated by the individual-specific Q_i parameter. Comparing Eq. S3 with the definition of Γ parameter in Sec. S2.1, we find the Γ parameter can be expressed as

$$\Gamma_{i\alpha} = \log(Q_i) + \mu_\alpha. \tag{S4}$$

Hence, the Γ parameter combines the two parts in the Q-model, reflecting both the individual differences and the luck component. It is also interesting to note that since the duration of hot hand is relatively short compared with the typical career length, the hot-hand model can be approximated as the Q-model in the long run. As such, the hot-hand model is also consistent with other predictions the Q-model makes regarding individual careers, as we show in the next sections.

S4.2 Q-model validation for scientists In this section, we validate the Q-model for scientists in our dataset, demonstrating the hot-hand model successfully reproduces additional predictions by the Q-model on impacts across scientists. To validate the Q-model in our dataset, we first calculated the distribution of impact and productivity, finding both $P(C_{10})$ and $P(N)$ follow log-normal distributions, which verifies the assumptions of the Q-model (Fig. S3c, Fig. S25a). We then used the maximum likelihood method to estimate the parameters μ and Σ of the trivariate normal distribution $P(\log p, \log Q, \log N) \sim \mathcal{N}(\mu, \Sigma)$ in the Q-model. The estimated parameters for the joint probability $P(\log p, \log Q, \log N)$ is

$$\begin{aligned} \mu &= (\mu_p, \mu_Q, \mu_N) = (1.39, 0.72, 3.43) \\ \Sigma &= \begin{pmatrix} \sigma_p^2 & \sigma_{p,Q} & \sigma_{p,N} \\ \sigma_{p,Q} & \sigma_Q^2 & \sigma_{Q,N} \\ \sigma_{p,N} & \sigma_{Q,N} & \sigma_N^2 \end{pmatrix} = \begin{pmatrix} 1.67 & 0.009 & 0.002 \\ 0.009 & 0.56 & 0.11 \\ 0.002 & 0.11 & 0.77 \end{pmatrix} \end{aligned} \tag{S5}$$

The matrix Eq. S5 is consistent with the Q-model that $\sigma_{p,Q} = \sigma_{p,N} \approx 0$. The estimated Q parameter also follows a log-normal distribution (Fig. S25b).

The Q-model successfully explains the impact differences across individual scientists by capturing the scaling of $\langle C_{10}^* \rangle$ with productivity N and with the logarithm of the average impact without the hit $\langle C_{10}^{-*} \rangle$[23]. To validate the Q-model in our dataset, we calculated the predictions of the Q-model by using Eq. S5, and compared them with the empirical observations, finding the results of the Q-model are aligned with data for the relationship between $\langle C_{10}^* \rangle$ and N, $\langle C_{10}^* \rangle$ and $\langle C_{10}^{-*} \rangle$ (Figs. S25c–d). To test if the hot-hand model can reproduce these predictions made by the Q-model, we simulated the impacts of each scientist using the generative hot-hand model described

in Sec. S3.3, and repeated the analyses for the correlation between $\langle C_{10}^{*} \rangle$ and N, and $\langle C_{10}^{*} \rangle$ and $\langle C_{10}^{-*} \rangle$. We find the hot-hand model can also capture these empirical observations on the impact differences across scientists (Figs. S25c–d), suggesting the hot-hand model is consistent with the Q-model when comparing impacts across different scientists.

S4.3 Q-model validation for artists and directors To study if the Q-model can be extended to model the impacts of works for artists and movie directors, we repeated the procedures used to estimate the Q-model parameters for scientists. We assume the auction price of each work is generated from a joint probability of $p(p, N, Q)$, where p captures the global distribution of price. To obtain the parameters for $p(p, N, Q)$, we first measured the distribution of auction price $P(\text{price})$ for all artworks, and productivity $P(N)$ for all artists in our dataset, finding they both follow log-normal distributions (Fig. S3a, Fig. S26a). The maximum-likelihood estimation allows us to calculate the parameters for the joint probability $p(\hat{p}, \hat{N}, \hat{Q})$, where $\hat{p} = \log p, \hat{N} = \log N$, and $\hat{Q} = \log Q$. Similarly, we find $p(\hat{p}, \hat{N}, \hat{Q})$ follows a trivariate normal distribution

$$\begin{aligned} \mu &= (\mu_p, \mu_Q, \mu_N) = (6.5, 1.8, 4.03), \\ \Sigma &= \begin{pmatrix} \sigma_p^2 & \sigma_{p,Q} & \sigma_{p,N} \\ \sigma_{p,Q} & \sigma_Q^2 & \sigma_{Q,N} \\ \sigma_{p,N} & \sigma_{Q,N} & \sigma_N^2 \end{pmatrix} = \begin{pmatrix} 1.43 & 0.28 & 0.52 \\ 0.28 & 1.26 & 0.25 \\ 0.52 & 0.25 & 1.41 \end{pmatrix} \end{aligned} \tag{S6}$$

Noting that different from scientists, Q and N, N and p have positive correlations. The estimated Q parameter follows a log-normal distribution as well (Fig. S26b). We find the Q-model's predictions for artists are in agreement with data in terms of the correlation between N and $\log \text{price}^{*}$, and the correlation between $\langle \log \text{price}^{-*} \rangle$ and $\log \text{price}^{*}$ (Figs. S26c–d). Hence for artists, the Q parameter

can be calculated as $Q = e^{<\log price> - \mu_{price}}$, where $u_{price} = 6.5$.

Similarly, we repeated the same analyses for movie directors. We find $P(\text{rating})$ follows a normal distribution, and $P(N)$ follows a log-normal distribution (Fig. S3b, Fig. S27a), allowing us to calculate the parameters for the joint probability $p(p, \hat{N}, \hat{Q})$, where $\hat{N} = \log N$, and $\hat{Q} = \log Q$, where p is the global distribution for ratings. Similarly, $p(p, \hat{N}, \hat{Q})$ follows a trivariate normal distribution whose

$$\begin{aligned} \mu &= (\mu_p, \mu_Q, \mu_N) = (5.9, 0.9, 2.8) \\ \Sigma &= \begin{pmatrix} \sigma_p^2 & \sigma_{p,Q} & \sigma_{p,N} \\ \sigma_{p,Q} & \sigma_Q^2 & \sigma_{Q,N} \\ \sigma_{p,N} & \sigma_{Q,N} & \sigma_N^2 \end{pmatrix} = \begin{pmatrix} 1.3 & 0.25 & 0.38 \\ 0.25 & 0.8 & 0.22 \\ 0.38 & 0.22 & 1.0 \end{pmatrix} \end{aligned} \tag{S7}$$

The estimated Q parameter follows a log-normal distribution as well (Fig. S27b). The Q-model predictions for directors are also consistent with data (Figs. S27c–d).

S5 Modeling the Collective Impact of Individual Scientists

S5.1 Definition of $g(t)$ Individual collective impact, $g(t)$, is defined as the sum of citations at each year to all publications one has produced:

$$g(t) = \sum_{i=1}^{N(t)} c_i(t - t_i), \tag{S8}$$

where t measures the career age, defined as the number of years after her first publication. $N(t)$ is the total number of papers published up to time t. t_i is the publication time of her i^{th} paper, whose yearly citations are denoted as c_i. $g(t)$ is driven by the interplay of three factors: productivity

$N(t)$, the citation dynamics of each publication c_i, and how publications and their impacts are distributed within a career. Due to the non-homogeneous nature of productivity, we bypass the need to evaluate $N(t)$ by measuring the career stage in unit of publications. That is, we keep the total number of publications in 20 years N_T ($T = 20$) of each scientist and their order, but rearrange the publication time, such that a scientist produces the same number of papers every year. After this procedure, a scientist publishes constant n papers per year, and $g(t)$ is defined as

$$g(t) = \sum_{i=1}^{N_T} c_i(t - \lfloor \frac{i}{N_T} T \rfloor), \tag{S9}$$

where $\lfloor \frac{i}{N_T} T \rfloor$ measures the publication year of the i^{th} paper. Note that, in some cases, papers published in late careers may be shifted to earlier career stage, whose citation records are not long enough to cover the rest of the career. In such case, we consider $g(t)$ up to the career stage that is not influenced by the boundary effect.

S5.2 Incorporating $\Gamma(t)$ into citations Next, we build a mathematical model for $g(t)$ by incorporating into the hot-hand model. We combine $\Gamma(t)$ with a previous model describing the citation dynamics, namely the WSB model[21], allowing us to model how publications and their impacts are distributed within a career. The WSB model captures the cumulative citation of a paper $C_i(t)$ by three parameters: fitness λ_i, immediacy σ_i, and longevity μ_i

$$C(t, t_i) = m \left(e^{\lambda_i \Phi\left(\frac{\ln(t-t_i)-\mu_i}{\sigma_i}\right)} - 1 \right), \tag{S10}$$

where t and t_i are defined by Eq. S8. $\Phi(\xi)$ is the cumulative normal distribution that follows

$$\Phi(\xi) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\xi} \phi(x) \mathrm{dx}, \tag{S11}$$

The constant m denotes the average number of references a paper contains. Here we assume $m = 30$ to be consistent with previous studies[21,53]. By replacing fitness λ_i with $\Gamma(t_i)$, Eq S10 can be expressed as

$$C(t, t_i) = m \left(e^{\Gamma(t_i)\Phi\left(\frac{\ln(t-t_i)-\mu_i}{\sigma_i}\right)} - 1 \right), \tag{S12}$$

where $\Gamma(t_i)$ represents the goodness parameter when a paper was published that can be measured from $\langle \log(C_{10}) \rangle_{t_i}$. $\langle \cdot \rangle_{t_i}$ represents the average around t_i. Eq.S12 is consistent with the WSB model by using $\Gamma(t_i)$ to reflect the fitness parameter λ_i in the WSB model[21].

S5.3 $g(t)$ **under the null model** We first discuss $g(t)$ under the null model, denoted as $g_0(t)$. The null model assumes Γ remains constant across a career ($\Gamma(t) \equiv \Gamma_0$). Integrating Eq. S12 over t allows us to obtain the cumulative individual citation $G_0(t)$ under the null model:

$$\begin{aligned} G_0(t) &= \sum_{i=1}^{N_T} C_i(t, t_i) \\ &\approx N_T \int_0^t \int_{\mu',\sigma'} C(t, t') P(\mu', \sigma') \mathrm{d}t' \mathrm{d}\mu' \mathrm{d}\sigma' \\ &= \int_0^t \int_{\mu',\sigma'} nC(t, t') \delta(\mu') \delta(\sigma') \mathrm{d}t' \mathrm{d}\mu' \mathrm{d}\sigma', \\ &= \int_0^t m \left(e^{\Gamma_0 \Phi\left(\frac{\ln(t-t')-\mu}{\sigma}\right)} - 1 \right) \mathrm{d}t' \end{aligned} \tag{S13}$$

where $N_T = \int_0^t n \mathrm{d}t'$ and n is the publication rate at each time stamp. t denotes the career stage of a scientist. t' is the time when each paper was published. $C(t, t')$ is the cumulative paper citation which follows Eq. S12. For simplicity, we assume in Eq. S13 that each scientist has constant μ and

σ for all the publications. Hence, $g_0(t)$ can be expressed as

$$\begin{aligned} g_0(t) &= \frac{\mathrm{d}}{\mathrm{d}t} G_0(t) \\ &= n \frac{\mathrm{d}}{\mathrm{d}t} \int_0^t m \left(e^{\Gamma_0 \Phi\left(\frac{\ln(t-t')-\mu}{\sigma}\right)} - 1 \right) \mathrm{dt}' \\ &= nm \left(e^{\Gamma_0 \Phi\left(\frac{\ln(t)-\mu}{\sigma}\right)} - 1 \right). \end{aligned} \tag{S14}$$

$g_0(t)$ follows a monotonically increasing curve over time (Fig. 3d, Fig. S28) which cannot account for the variety of $g(t)$ observed in real careers (Fig. 3d). Eq. S14 suggests $g_0(t)$ is captured by individual goodness parameter Γ_0 and μ and σ capturing paper citation dynamics. Comparing Eq. S12 with Eq. S14 we find $g_0(t)$ is the cumulative citation of papers published at $t = 0$.

S5.4 $g(t)$ **under the hot-hand model** Next we calculate $g(t)$ under the hot-hand model by incorporating a time dependent $\Gamma(t)$ following Eq. 1:

$$\begin{aligned} g(t) &= n \frac{\mathrm{d}}{\mathrm{d}t} G(t) \\ &= n \frac{\mathrm{d}}{\mathrm{d}t} \int_0^t m \left(e^{\Gamma(t') \Phi\left(\frac{\ln(t-t')-\mu}{\sigma}\right)} - 1 \right) \mathrm{dt}' \\ &= n \int_0^t \frac{\partial}{\partial t} m e^{\Gamma(t') \Phi\left(\frac{\ln(t-t')-\mu}{\sigma}\right)} \mathrm{dt}' \\ &= nm \left[\int_0^t e^{\Gamma(t') \Phi\left(\frac{\ln(t-t')-\mu}{\sigma}\right)} \Phi\left(\frac{\ln(t-t')-\mu}{\sigma} \right) \frac{\mathrm{d}}{\mathrm{d}t'} \Gamma(t') \mathrm{d}t' - \int_0^t \frac{\partial}{\partial t'} e^{\Gamma(t') \Phi\left(\frac{\ln(t-t')-\mu}{\sigma}\right)} \mathrm{d}t' \right] \\ &= nm \int_0^t e^{\Gamma(t') \Phi\left(\frac{\ln(t-t')-\mu}{\sigma}\right)} \Phi\left(\frac{\ln(t-t')-\mu}{\sigma} \right) \frac{\mathrm{d}}{\mathrm{d}t'} \Gamma(t') \mathrm{d}t' + nm \left(e^{\Gamma_0 \Phi\left(\frac{\ln(t)-\mu}{\sigma}\right)} - 1 \right). \end{aligned} \tag{S15}$$

Similarly, we assume constant μ and σ for each scientist. For simplicity, Eq. S15 assumes each individual has one hot-hand period. The second term in Eq. S15 represents $g_0(t)$ under the null

model (Eq. S14). The derivatives of Eq. 1 can be expressed as

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t'}\Gamma(t') &= \frac{\mathrm{d}}{\mathrm{d}t'}[\Gamma_0 + (\Gamma_H - \Gamma_0)[\mathcal{H}(t'-t_\uparrow) - \mathcal{H}(t'-t_\downarrow)] \\ &= (\Gamma_H - \Gamma_0)[\delta(t'-t_\uparrow) - \delta(t'-t_\downarrow))].\end{aligned} \tag{S16}$$

By replacing $\Gamma(t')$ in Eq. S15 with Eq. 1 and incorporating Eq. S16, we obtain the expression of the first term in Eq. S15, denoted as $\Delta g(t)$

$$\Delta g(t) = \begin{cases} 0 & t < t_\uparrow \\ nm(\Gamma_H - \Gamma_0)\Phi\left(\frac{\ln(t-t_\uparrow)-\mu}{\sigma}\right) e^{\Gamma_H \Phi\left(\frac{\ln(t-t_\uparrow)-\mu}{\sigma}\right)} & t_\uparrow \le t < t_\downarrow \\ nm(\Gamma_H - \Gamma_0)\left(\Phi\left(\frac{\ln(t-t_\uparrow)-\mu}{\sigma}\right) e^{\Gamma_H \Phi\left(\frac{\ln(t-t_\uparrow)-\mu}{\sigma}\right)} - \Phi\left(\frac{\ln(t-t_\downarrow)-\mu}{\sigma}\right) e^{\Gamma_H \Phi\left(\frac{\ln(t-t_\downarrow)-\mu}{\sigma}\right)}\right) & t \ge t_\downarrow \end{cases} \tag{S17}$$

Hence, $g(t)$ under the hot-hand model can be expressed as

$$g(t) = \underbrace{nm(e^{\Gamma_0 \Phi\left(\frac{\ln(t)-\mu}{\sigma}\right)} - 1)}_{g_0(t)} + \underbrace{\begin{cases} 0 & t < t_\uparrow \\ nm(\Gamma_H - \Gamma_0)\Phi\left(\frac{\ln(t-t_\uparrow)-\mu}{\sigma}\right) e^{\Gamma_H \Phi\left(\frac{\ln(t-t_\uparrow)-\mu}{\sigma}\right)} & t_\uparrow \le t < t_\downarrow \\ nm(\Gamma_H - \Gamma_0)\left(\Phi\left(\frac{\ln(t-t_\uparrow)-\mu}{\sigma}\right) e^{\Gamma_H \Phi\left(\frac{\ln(t-t_\uparrow)-\mu}{\sigma}\right)}\right. & \\ \qquad \left. -\Phi\left(\frac{\ln(t-t_\downarrow)-\mu}{\sigma}\right) e^{\Gamma_H \Phi\left(\frac{\ln(t-t_\downarrow)-\mu}{\sigma}\right)}\right) & t \ge t_\downarrow \end{cases}}_{\Delta g(t)}, \tag{S18}$$

Eq. S18 consists of two parts: $g_0(t)$ captures the cumulative citation of papers published when $\Gamma(t_i) = \Gamma_0$, and $\Delta g(t)$ captures the citations contributed by papers published during hot hand when $\Gamma(t_i) = \Gamma_H$. Fig. 4d illustrates the influence of hot hand on individual collective impact: Before hot hand, $\Delta g(t)$ has no impact on individual citation and $g(t) = g_0(t)$. During hot hand, $g(t)$ experiences rapid growth due to Γ_H, and the trend continues for a while to reach its peak even

after the end of the hot-hand period. Eq. S18 allows us to model a wide range of $g(t)$ that cannot be expected by the null model (Fig. 3d, Fig. 3h, Figs. S29–S30). $g(t)$ is determined by the timing of the hot-hand period ($t_\uparrow$ and $t_\downarrow$), the level of performance (Γ_0 and Γ_H), and one's typical paper citation dynamics (μ and σ).

S5.5 Parameter estimation In order to test how well Eq. S18 matches empirical data, we need to estimate the parameters Γ_0, Γ_H, $t_\uparrow$, $t_\downarrow$, μ and σ for each scientist given $g(t)$. For each individual in our dataset, we used ordinary least square (OLS) and fitted to Eq. S18 the $g(t)$ measured from real careers by minimizing the mean squared error (*MSE*) between data and the model predictions. We set the boundary conditions for the fitted parameters as $t_\uparrow > 0$, $t_\downarrow > 0$, $0 < \Gamma_0 \leq \Gamma_H \leq 10$, $0 \leq \mu \leq 8.5$ and $\sigma > 0$. We assume Γ_H cannot exceed 10 to include individuals whose average citation per paper during hot hand $\langle C_{10} \rangle \leq 22026$.

S5.6 Model validation To systematically evaluate the accuracy of our model, we generate an uncertainty envelop of $g(t)$ for each individual for the model's predictions. To generate the uncertainty envelop, we simulated $g(t)$ for each individual by assigning a Gaussian noise $\mathcal{N}(0,\, \sigma_s^2)$ to the fitted Γ_0 and Γ_H. That is, for each paper i we randomly draw its Γ_i from a normal distribution, with Γ_0 as the mean if the paper is published during the normal period or Γ_H during hot hand. The standard deviation σ_s represents the inherent noise of the goodness parameter defined by Sec. S3.5. We assume other parameters to be the same as fitting results. We simulated $g(t)$ of each individual for 1000 realizations, allowing us to obtain a distribution of $g(t)$ at each time step. The envelop at t is defined as one standard deviation of all the values of $g(t)$. Similarly, we repeated the same

procedures for the null model to calculate its uncertainty envelop. We then measured each model's performance by calculating the fraction of $g(t)$ falling within the envelope. Our hot-hand model outperforms the null model as more data fall within the envelop (Fig. 3f). We also compared the distribution of the Mean Absolute Percentage Error (*MAPE*) between data and model's predictions, finding the hot-hand model systematically outperforms the null model under different evaluation metrics (Fig. S31).

S6 Testing Alternative Hypotheses

The observed accuracy of the hot-hand model prompts us to ask whether it is unique in its ability to capture the impacts of individual careers across different domains. To answer this question, we propose in the main text four alternative hypotheses corresponding to different hot hand dynamics: (*A*) A right trapezoid model (Fig. S32b) captures a sudden onset of hot hand with a more gradual decline. (*B*) An isosceles trapezoid model (Fig. S32c) captures hot hand that evolves and dissolves gradually over time. (*C*) Inverted-U shape (Fig. S32d) and tent functions (Fig. S32e) captures the peak performance at a certain point of a career. (*D*) A left trapezoid function (Fig. S32f) captures a gradual startup period with a sharp cutoff.

We test the validity of the four alternative hypotheses (A–D) by comparing each model's prediction with empirical observations on the relative order of the top six hits within a career. The symmetric patterns of Φ and $R(\Delta N/N)$ observed in real careers suggest a hit is equal likely to appear before or after another hit. We find such randomness on the relative order among hits is

not just limited to the top three. Indeed, we measured the position of the top three hits $\tilde{N}$ relative to the top six hits of the career, and compute $P(\tilde{N})$ for each of the three hits for artists, directors and scientists. We find a lack of predictive patterns for $P(\tilde{N})$ across three domains, suggesting the relative order among the top six hits in real careers are random (Fig. S32g, Fig. S32n, Fig. S32u).

To test if hypotheses A–D can reproduce the randomness among the top six hits within a career, we compared $P(\tilde{N})$ predicted by each hypothesis with the corresponding distributions measured from real careers. Using real productivity N and average impact of each career as input, we first generated ensembles of career histories under different models (Figs. 32a–f): For each individual i, the impact of a work is randomly drawn from a normal distribution $\mathcal{N}(\Gamma(N),\, \sigma_s^2)$, where $\Gamma(N)$ is defined by each model (Figs. 32a–f), and the standard deviation σ_s is defined by Sec. S3.5 (Fig. S18). For simplicity, we assume Γ_0 for each individual as the mean impact measured from data, and $\Gamma_H = \Gamma_0 + 1.0$ for all tested models (Figs. 32a–f). For the quadratic and tent functions, we assume the peak value of $\Gamma(N)$ as Γ_H for each individual. To define a random hot-hand period in each career, we randomly pick one work out of the sequence of N works one produced, marking it the start of the hot-hand period. Since the typical duration of hot hand accounts for 20% of total N works, here we assume a fixed length of Γ_H period $(0.2N)$ for the right trapezoid, left trapezoid, and isosceles trapezoid model. We assume the hypotenuses of each trapezoid model to be $0.2N$, ensuring the impact can gradually change between Γ_0 and Γ_H under different hypotheses. To keep the duration of hot hand to be the same for different hypotheses, we assume a $0.4N$ length of hot hand for hypothesis C and the hot-hand model, respectively.

We then calculated $P(\tilde{N})$ predicted by each hypothesis, finding $P(\tilde{N})$ for hypotheses A–D shows different trends on the relative position of top six hits (Figs. S32h–m). For example, the top two hits of the isosceles trapezoid, quadratic, and tent model always appear in the middle of the top six hits, and their probability decreases when deviating from the middle (Figs. S32j–l). For the right trapezoid model, the top three hits always appear before the other hits (Fig. S32i); whereas the case reverses for the left trapezoid model (Fig. S32m). Among these hypotheses, only the hot-hand model can reproduce the randomness of the top six hits (Fig. S32h). To quantify the difference between data and each model's prediction, we performed the Kolmogorov-Smirnov test to measure whether we can reject the hypothesis that predictions and data are drawn from the same distribution. We find $P(\tilde{N})$ predicted by the hot-hand model is the only one that captures the randomness of the top six hits within a career (Fig. S32h), while $P(\tilde{N})$ for hypotheses A–D is statistically different from empirical observations (Figs. S32i–m). The results reported in Figs. S32h–m are based on the simulations using artists' profiles as input. Similarly, we repeated the same procedures using profiles of directors (Figs. S32o–t) and scientists (Figs. S32v–aa) as well, finding similar patterns of $P(\tilde{N})$ under each hypothesis. Hence, our conclusions are robust to three different domains.

Table S1: Empirical evidence of whether hot hand exists

Year	Paper	Field	Hot hand	Data
1985	Gilovich et al. (1985)[1]	Basketball	No	Field goal data for 9 NBA player during the 1980 –1981 season, free-throw data for 9 NBA players during 1980–1982 seasons, a controlled shooting experiment with 26 Cornell students
1989	Larkey et al. (1989)[54]	Basketball	Yes	18 players in 39 NBA games in 1987–1988 season
1993	Albright (1993)[55]	Baseball	No	40 MLB players in 1987–1990 seasons
1993	Albert (1993)[56]	Baseball	Yes	Same as Albright (1993)[55]
1994	Frohlich (1994)[57]	Baseball	No	MLB No-hit games in 1989–1993
1995	Wardrop (1995)[58]	Basketball	Inconclusive	Free throw data for 9 members of the Boston Celtics during 1980 - 1982 seasons
1999	Wardrop (1999)[59]	Basketball	Yes	A shooting experiment for a member on the UW-Madison women's varsity basketball team
2000	Vergin (2000)[60]	Basketball, Baseball	No	29 NBA team wins and losses in 1996–1998 seasons, 28 MLB team wins and losses in 1996 season
2001	Albert and Williamson (2001)[61]	Basketball, Baseball	Yes	Same as Wardrop (1999)[59], hitting data for Javy Lopez for the 1998 MLB season
2001	Klaassen and Magnus (2001)[62]	Tennis	Yes	Point-to-point data from Wimbledon singles in 1992–1995
2003	Koehler and Conley (2003)[63]	Basketball	No	NBA long distance shootout contest in 1994–1997
2003	Smith (2003)[64]	Horse Pitching	Yes	62 pitchers in 2000–2001 World Championships
2004	Dorsey-Palmateer and Smith (2004)[65]	Bowling	Yes	43 players from PBA in 2002–2003 season
2005	Clark (2005)[66]	Golf	Yes	Hole-to-hole scores from PGA tour in 1997–1998
2008	Bocskocsky et al. (2008)[67]	Basketball	Yes	83,000 shots from the 2012–2013 NBA season
2008	Albert (2008)[68]	Baseball	Yes	All regular baseball players during the 2005 MBL season
2010	Arkes (2010)[69]	Basketball	Yes	Free throw data during the 2005–06 NBA season
2011	Yaari and Eisenmann (2011)[70]	Basketball	Yes	Free throws of five NBA regular seasons from 2005 to 2009
2012	Yaari and David (2012)[71]	Bowling	Yes	Frame by frame games for 100 top players in PBA from 2002 to 2011
2012	Aharoni and Sarig (2012)[72]	Basketball	Yes	1,218 games in 2004–2005 NBA season
2012	Raab and Gula (2012)[73]	Volleyball	Yes	26 top players' offensive performance in Germany s first-division volleyball league
2012	Gabel and Redner (2012)[74]	Basketball	No	6,087 games from the 2006/07 – 2009/10 seasons of NBA
2014	Miller and Sanjurjo (2014)[75]	Basketball	Yes	Controlled shooting experiment for players from the semi-professional basketball team Santo Domingo de Betanzos
2014	Csapo and Raab (2014)[76]	Basketball	Yes	666 NBA games from the 2011-12 to 2013-14 seasons
2015	Miller and Sanjurjo (2015)[77]	Basketball	Yes	NBA Three-Point Contest
2015	Rosenqvist and Skans (2015)[78]	Golf	Yes	Male European PGA tournaments in 2000–2012
2015	Parsons and Rohde (2015)[79]	Football/Soccer	Yes	20 teams from EPL in 2010–2013 seasons

2015	Miller and Sanjurjo (2015)[77]	Baseball	Yes	NBA Three point shooting contest in 1986–2005
2016	Miller and Sanjurjo (2016)[3]	Basketball	Yes	Same as Miller and Sanjurjo (2014)[75]
2017	Green and Zwiebel (2017)[80]	Baseball	Yes	MLB players' batter
1993	Hendricks (1993)[8]	Mutual funds	Yes	Risk-adjusted mutual-fund returns in 1975-1988
1995	Malkiel (1995)[81]	Equity mutual funds	Inconclusive	All risk-adjusted mutual-fund returns in 1971–1991
1997	Carhart (1997)[82]	Equity mutual funds	No	All risk-adjusted equity mutual-fund returns in 1962–1993
2008	Huij et al. (2008)[83]	Bond funds	Yes	Risk-adjusted bond fund returns in 1990–2003
2010	Jagannathan et al. (2010)[9]	Hedge funds	Yes	Risk-adjusted hedge fund returns in 1996–2005

Table S2: Statistical methods to detect hot hand

Method	**Paper**
Conditional probablity	Gilovich et al. (1985)[1], Larkey et al. (1989)[54], Wardrop (1995)[58], Koehler and Conley (2003), Smith (2003)[64], Koehler and Conley (2003)[63], Dorsey-Palmateer and Smith (2004)[65], Clark (2005)[66], Albert (2008)[68], Yaari and Eisenmann (2011)[70], Aharoni and Sarig (2012)[72], Raab and Gula (2012)[73], Csapo and Raab (2014)[76], Miller and Sanjurjo (2016)[3], Malkiel (1995)[81]
Serial correlation	Gilovich et al. (1985)[1], Larkey et al. (1989)[54], Wardrop (1999)[59], Klaassen and Magnus (2001)[62], Smith (2003)[64], Raab and Gula (2012)[73], Gabel and Redner (2012)[74], Green and Zwiebel (2017)[80], Hendricks (1993)[8]
Number of runs	Gilovich et al. (1985)[1], Wardrop (1999)[59], Albright (1993)[55], Vergin (2000)[60], Koehler and Conley (2003)[63], Smith (2003)[64], Albert (2008)[68], Raab and Gula (2012)[73], Csapo and Raab (2014)[76], Miller and Sanjurjo (2014)[75], Parsons and Rohde (2015)[79], Miller and Sanjurjo (2015)[77]
Stationary of successful rate	Gilovich et al. (1985)[1], Frohlich (1994)[57], Wardrop (1999)[59], Albert (2008)[68], Yaari and Eisenmann (2011)[70], Yaari and David (2012)[71], Gabel and Redner (2012)[74]
hidden Markov model	Albright (1993)[55], Albert (1993)[56], Albert and Williamson (2001)[61], Albert (2008)[68], Wetzels *et al.*(2016)[84]
Long streaks	Larkey et al. (1989)[54], Vergin (2000)[60], Frohlich (1994)[57],Dorsey-Palmateer and Smith (2004)[65], Albert (2008)[68], Aharoni and Sarig (2012), Csapo and Raab (2014)[76], Miller and Sanjurjo (2014)[75], Miller and Sanjurjo (2015)[77]
Regression	Albright (1993)[55], Bocskocsky et al. (2008)[67], Arkes (2010)[69], Aharoni and Sarig (2012)[72], Csapo and Raab (2014)[76], Parsons and Rohde (2015)[79], Hendricks (1993)[8], Carhart (1997)[82], Huij et al. (2008)[83], Jagannathan et al. (2010)[9]
Time interval between scoring event	Gabel and Redner (2012)[74]
Regression discontinuity design	Rosenqvist and Skans (2015)[78]

References

1. Gilovich, T., Vallone, R. & Tversky, A. The hot hand in basketball: On the misperception of random sequences. *Cognitive Psychology* **17**, 295–314 (1985).

2. Bar-Eli, M., Avugos, S. & Raab, M. Twenty years of hot hand research: Review and critique. *Psychology of Sport and Exercise* **7**, 525–553 (2006).

3. Miller, J. B. & Sanjurjo, A. Surprised by the gambler's and hot hand fallacies? A truth in the law of small numbers. *IGIER Working Paper No. 552* (2016).

4. Ayton, P. & Fischer, I. The hot hand fallacy and the gambler's fallacy: Two faces of subjective randomness? *Memory & Cognition* **32**, 1369–1378 (2004).

5. Croson, R. & Sundali, J. The gambler's fallacy and the hot hand: Empirical data from casinos. *Journal of Risk and Uncertainty* **30**, 195–209 (2005).

6. Rabin, M. & Vayanos, D. The gambler's and hot-hand fallacies: Theory and applications. *The Review of Economic Studies* **77**, 730–778 (2010).

7. Xu, J. & Harvey, N. Carry on winning: The gamblers' fallacy creates hot hand effects in online gambling. *Cognition* **131**, 173–180 (2014).

8. Hendricks, D., Patel, J. & Zeckhauser, R. Hot hands in mutual funds: Short-run persistence of relative performance, 1974–1988. *The Journal of Finance* **48**, 93–130 (1993).

9. Jagannathan, R., Malakhov, A. & Novikov, D. Do hot hands exist among hedge fund managers? An empirical evaluation. *The Journal of Finance* **65**, 217–255 (2010).

10. Kahneman, D. & Riepe, M. W. Aspects of investor psychology. *The Journal of Portfolio Management* **24**, 52–65 (1998).

11. Lehman, H. C. *Age and achievement* (Princeton, NJ, 1953).

12. Merton, R. K. The matthew effect in science. *Science* **159**, 56–63 (1968).

13. Simonton, D. K. Age and outstanding achievement: What do we know after a century of research? *Psychological Bulletin* **104**, 251 (1988).

14. Jones, B. F. & Weinberg, B. A. Age dynamics in scientific creativity. *Proceedings of the National Academy of Sciences* **108**, 18910–18914 (2011).

15. Petersen, A. M., Jung, W.-S., Yang, J.-S. & Stanley, H. E. Quantitative and empirical demonstration of the matthew effect in a study of career longevity. *Proceedings of the National Academy of Sciences* **108**, 18–23 (2011).

16. Petersen, A. M. *et al.* Reputation and impact in academic careers. *Proceedings of the National Academy of Sciences* **111**, 15316–15321 (2014).

17. Stringer, M. J., Sales-Pardo, M. & Amaral, L. A. N. Effectiveness of journal ranking schemes as a tool for locating information. *PLoS One* **3**, e1683 (2008).

18. Radicchi, F., Fortunato, S. & Castellano, C. Universality of citation distributions: Toward an objective measure of scientific impact. *Proceedings of the National Academy of Sciences* **105**, 17268–17272 (2008).

19. Ke, Q., Ferrara, E., Radicchi, F. & Flammini, A. Defining and identifying sleeping beauties in science. *Proceedings of the National Academy of Sciences* **112**, 7426–7431 (2015).

20. Galenson, D. W. *Old masters and young geniuses: The two life cycles of artistic creativity* (Princeton University Press, 2011).

21. Wang, D., Song, C. & Barabási, A.-L. Quantifying long-term scientific impact. *Science* **342**, 127–132 (2013).

22. Way, S. F., Morgan, A. C., Clauset, A. & Larremore, D. B. The misleading narrative of the canonical faculty productivity trajectory. *Proceedings of the National Academy of Sciences* (2017).

23. Sinatra, R., Wang, D., Deville, P., Song, C. & Barabási, A.-L. Quantifying the evolution of individual scientific impact. *Science* **354**, aaf5239 (2016).

24. Simonton, D. K. *Scientific genius: A psychology of science* (Cambridge University Press, 1988).

25. Duch, J. *et al.* The possible role of resource requirements and academic career-choice risk on gender differences in publication rate and impact. *PloS One* **7**, e51332 (2012).

26. Merton, R. K. The matthew effect in science, ii: Cumulative advantage and the symbolism of intellectual property. *Isis* **79**, 606–623 (1988).

27. Price, D. d. S. A general theory of bibliometric and other cumulative advantage processes. *Journal of the Association for Information Science and Technology* **27**, 292–306 (1976).

28. Barabási, A.-L. & Albert, R. Emergence of scaling in random networks. *Science* **286**, 509–512 (1999).

29. Wasserman, M., Zeng, X. H. T. & Amaral, L. A. N. Cross-evaluation of metrics to estimate the significance of creative works. *Proceedings of the National Academy of Sciences* **112**, 1281–1286 (2015).

30. Uzzi, B., Mukherjee, S., Stringer, M. & Jones, B. Atypical combinations and scientific impact. *Science* **342**, 468–472 (2013).

31. Merton, R. K. Singletons and multiples in scientific discovery: A chapter in the sociology of science. *Proceedings of the American Philosophical Society* **105**, 470–486 (1961).

32. Wuchty, S., Jones, B. F. & Uzzi, B. The increasing dominance of teams in production of knowledge. *Science* **316**, 1036–1039 (2007).

33. Clauset, A., Arbesman, S. & Larremore, D. B. Systematic inequality and hierarchy in faculty hiring networks. *Science Advances* **1**, e1400005 (2015).

34. Shockley, W. On the statistics of individual variations of productivity in research laboratories. *Proceedings of the IRE* **45**, 279–290 (1957).

35. Redner, S. Citation statistics from 110 years of physical review. *Physics Today* **58**, 49 (2005).

36. Moreira, J. A., Zeng, X. H. T. & Amaral, L. A. N. The distribution of the asymptotic number of citations to sets of publications by a researcher or from an academic department are consistent with a discrete lognormal model. *PloS One* **10**, e0143108 (2015).

37. Palla, G., Barabási, A.-L. & Vicsek, T. Quantifying social group evolution. *Nature* **446**, 664–667 (2007).

38. Galenson, D. W. & Lenzu, S. Pricing genius: the market evaluation of innovation. *Journal of Applied Economics* **19**, 219–248 (2016).

39. Blumm, N. *et al.* Dynamics of ranking processes in complex systems. *Physical Review Letters* **109**, 128701 (2012).

40. Zickar, M. J. & Slaughter, J. E. Examining creative performance over time using hierarchical linear modeling: An illustration using film directors. *Human Performance* **12**, 211–230 (1999).

41. Simonton, D. K. Group artistic creativity: Creative clusters and cinematic success in feature films. *Journal of Applied Social Psychology* **34**, 1494–1520 (2004).

42. Han, H., Zha, H. & Giles, C. Name disambiguation spectral in author citations using a K-way clustering method. In *Proceedings of the 5TH ACM/IEEE Joint Conference on Digital Libraries*, 334–343 (IEEE, 2005).

43. Treeratpituk, P. & Giles, C. L. Disambiguating authors in academic publications using random forests. In *Proceedings of the 9th ACM/IEEE-CS Joint Conference on Digital Libraries*, 39–48 (ACM, 2009).

44. Radicchi, F. & Castellano, C. Analysis of bibliometric indicators for individual scholars in a large data set. *Scientometrics* **97**, 627–637 (2013).

45. Guevara, M. R., Hartmann, D., Aristarán, M., Mendoza, M. & Hidalgo, C. A. The research space: Using career paths to predict the evolution of the research output of individuals, institutions, and nations. *Scientometrics* **109**, 1695–1709 (2016).

46. Pan, R. K., Petersen, A. M., Pammolli, F. & Fortunato, S. The memory of science: Inflation, myopia, and the knowledge network. *arXiv preprint arXiv:1607.05606* (2016).

47. Dennis, W. Creative productivity between the ages of 20 and 80 years. *Journal of Gerontology* **21**, 1–8 (1966).

48. Simonton, D. K. Creative productivity: A predictive and explanatory model of career trajectories and landmarks. *Psychological Review* **104**, 66 (1997).

49. Wilke, A. & Barrett, H. C. The hot hand phenomenon as a cognitive adaptation to clumped resources. *Evolution and Human Behavior* **30**, 161–169 (2009).

50. Sun, Y. & Wang, H. Gambler's fallacy, hot hand belief, and the time of patterns. *Judgment and Decision Making* **5**, 124 (2010).

51. Oskarsson, A. T., Van Boven, L., McClelland, G. H. & Hastie, R. What's next? Judging sequences of binary events. *Psychological Bulletin* **135**, 262 (2009).

52. Alter, A. L. & Oppenheimer, D. M. From a fixation on sports to an exploration of mechanism: The past, present, and future of hot hand research. *Thinking & Reasoning* **12**, 431–444 (2006).

53. Shen, H., Wang, D., Song, C. & Barabási, A.-L. Modeling and predicting popularity dynamics via reinforced poisson processes. In *AAAI*, vol. 14, 291–297 (2014).

54. Larkey, P. D., Smith, R. A. & Kadane, J. B. It's okay to believe in the hot hand. *Chance* **2**, 22–30 (1989).

55. Albright, S. C. A statistical analysis of hitting streaks in baseball. *Journal of the American Statistical Association* **88**, 1175–1183 (1993).

56. Albert, J. A statistical analysis of hitting streaks in baseball - comment. *Journal of the American Statistical Association* **88**, 1184–1188 (1993).

57. Frohlich, C. Baseball: Pitching no-hitters. *Chance* **7**, 24–30 (1994).

58. Wardrop, R. L. Simpson's paradox and the hot hand in basketball. *The American Statistician* **49**, 24–28 (1995).

59. Wardrop, R. L. Statistical tests for the hot-hand in basketball in a controlled setting. *American Statistician* **1**, 1–20 (1999).

60. Vergin, R. Winning streaks in sports and the misperception of momentum. *Journal of Sport Behavior* **23**, 181 (2000).

61. Albert, J. & Williamson, P. Using model/data simulations to detect streakiness. *The American Statistician* **55**, 41–50 (2001).

62. Klaassen, F. J. & Magnus, J. R. Are points in tennis independent and identically distributed? Evidence from a dynamic binary panel data model. *Journal of the American Statistical Association* **96**, 500–509 (2001).

63. Koehler, J. J. & Conley, C. A. The hot hand myth in professional basketball. *Journal of Sport and Exercise Psychology* **25**, 253–259 (2003).

64. Smith, G. Horseshoe pitchers hot hands. *Psychonomic Bulletin & Review* **10**, 753–758 (2003).

65. Dorsey-Palmateer, R. & Smith, G. Bowlers' hot hands. *The American Statistician* **58**, 38–45 (2004).

66. Clark III, R. D. Examination of hole-to-hole streakiness on the PGA tour. *Perceptual and Motor Skills* **100**, 806–814 (2005).

67. Bocskocsky, A., Ezekowitz, J. & Stein, C. The hot hand: A new approach to an old fallacy. In *8th Annual Mit Sloan Sports Analytics Conference* (2014).

68. Albert, J. Streaky hitting in baseball. *Journal of Quantitative Analysis in Sports* **4** (2008).

69. Arkes, J. Revisiting the hot hand theory with free throw data in a multivariate framework. *Journal of Quantitative Analysis in Sports* **6**, 2 (2010).

70. Yaari, G. & Eisenmann, S. The hot (invisible?) hand: Can time sequence patterns of success/failure in sports be modeled as repeated random independent trials? *PloS One* **6**, e24532 (2011).

71. Yaari, G. & David, G. "Hot Hand" on strike: Bowling data indicates correlation to recent past results, not causality. *PLoS One* **7**, e30112 (2012).

72. Aharoni, G. & Sarig, O. H. Hot hands and equilibrium. *Applied Economics* **44**, 2309–2320 (2012).

73. Raab, M., Gula, B. & Gigerenzer, G. The hot hand exists in volleyball and is used for allocation decisions. *Journal of Experimental Psychology: Applied* **18**, 81 (2012).

74. Gabel, A. & Redner, S. Random walk picture of basketball scoring. *Journal of Quantitative Analysis in Sports* **8** (2012).

75. Miller, J. B. & Sanjurjo, A. A cold shower for the hot hand fallacy. *IGIER Working Paper 518* (2014).

76. Csapo, P. & Raab, M. "Hand down, man down." Analysis of defensive adjustments in response to the hot hand in basketball using novel defense metrics. *PloS one* **9**, e114184 (2014).

77. Miller, J. B. & Sanjurjo, A. Is it a fallacy to believe in the hot hand in the NBA three-point contest? *IGIER Working Paper 548* (2015).

78. Rosenqvist, O. & Skans, O. N. Confidence enhanced performance? The causal effects of success on future performance in professional golf tournaments. *Journal of Economic Behavior & Organization* **117**, 281–295 (2015).

79. Parsons, S. & Rohde, N. The hot hand fallacy re-examined: New evidence from the English Premier League. *Applied Economics* **47**, 346–357 (2015).

80. Green, B. & Zwiebel, J. The hot-hand fallacy: Cognitive mistakes or equilibrium adjustments? Evidence from Major League Baseball. *Management Science* (2017).

81. Malkiel, B. G. Returns from investing in equity mutual funds 1971 to 1991. *The Journal of Finance* **50**, 549–572 (1995).

82. Carhart, M. M. On persistence in mutual fund performance. *The Journal of finance* **52**, 57–82 (1997).

83. Huij, J. & Derwall, J. "Hot hands" in bond funds. *Journal of banking & finance* **32**, 559–572 (2008).

84. Wetzels, R. *et al.* A bayesian test for the hot hand phenomenon. *Journal of Mathematical Psychology* **72**, 200–209 (2016).

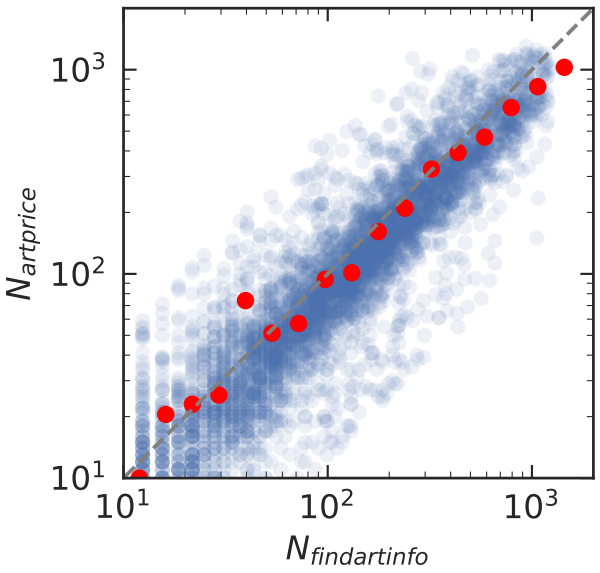

Figure S1: **Entity linking result of two auction databases.** The relationship of the number of auction records in Findartinfo and Artprice for matched artists in log-log scale, where $N_{findartinfo}$ is the number of records in Findartinfo and $N_{artprice}$ is the number of records in Artprice. Blue dots denote data, red dots denote the logarithmic binning of the scattered data, and the dashed grey line indicates $y = x$.

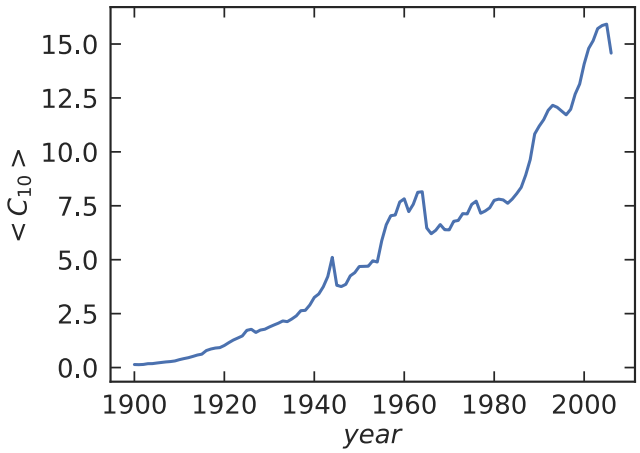

Figure S2: **Citation inflation in WoS.** The average impact of papers published in the same year, quantified with number of citations 10 years after publication, steadily grows as a function of the publication year ranging from 1900 to 2004.

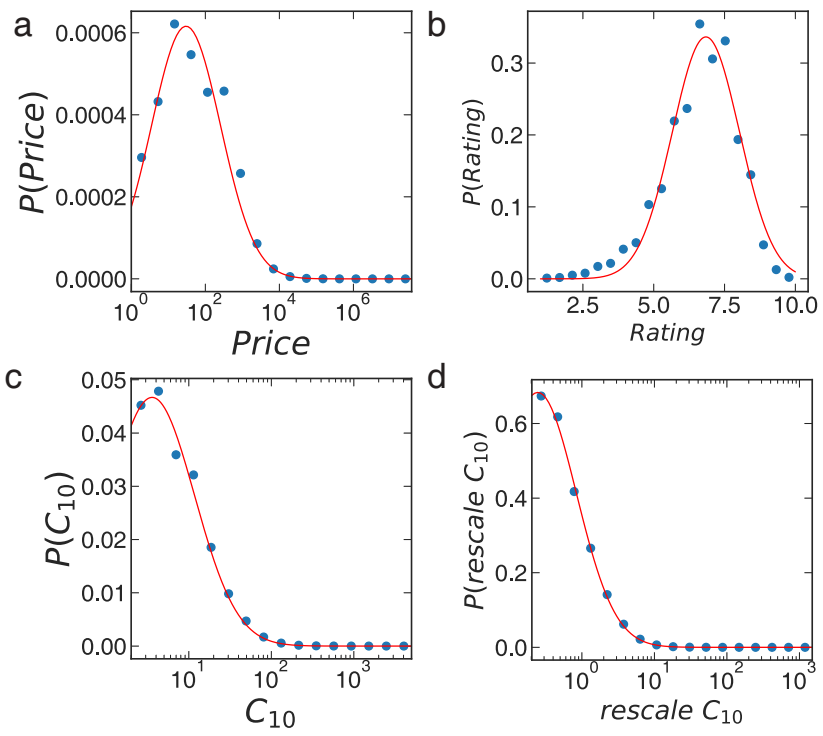

Figure S3: **Impact distribution.** (a) The distribution of auction price $P(\text{Price})$ for artists. Blue dots denote data, and the red line is a log-normal distribution with average $\mu = 8.0$ and standard deviation $\sigma = 2.15$. (b) The distribution of movie rating $P(\text{Rating})$ for directors. The red line is a normal distribution with average $\mu = 6.8$ and standard deviation $\sigma = 1.19$. (c–d), The distribution of (c) raw and (d) rescaled C_{10} for scientists. The red line is a log-normal distribution, with $\mu = 2.7$ and $\sigma = 1.18$ for (c) and $\mu = 0.1$ and $\sigma = 1.42$ for (d).

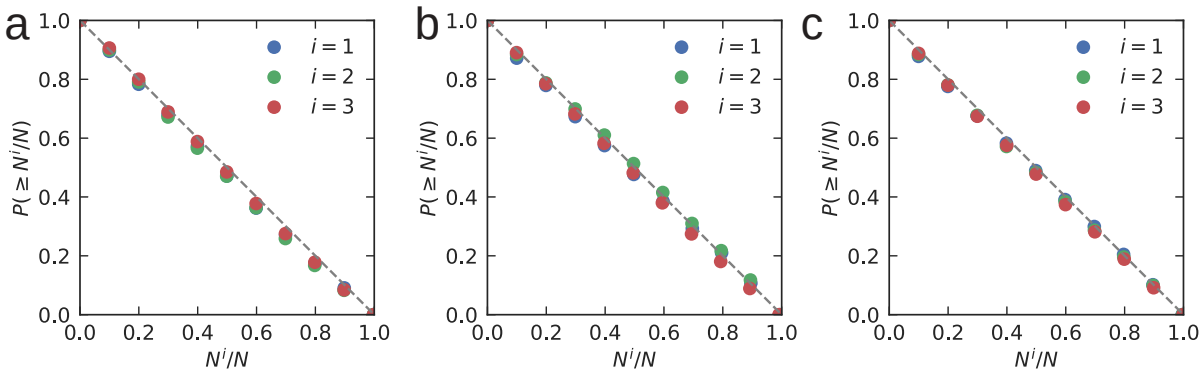

Figure S4: **Random impact rule.** $P(\geq N^i/N)$ of the top three hit works within a career for (a) artists, (b) directors, and (c) scientists. N^i denotes the order of the i^{th} hit work within a career. The color denotes different hit works, and the dashed grey line denotes $P(\geq N^i/N)$ for a uniform distribution.

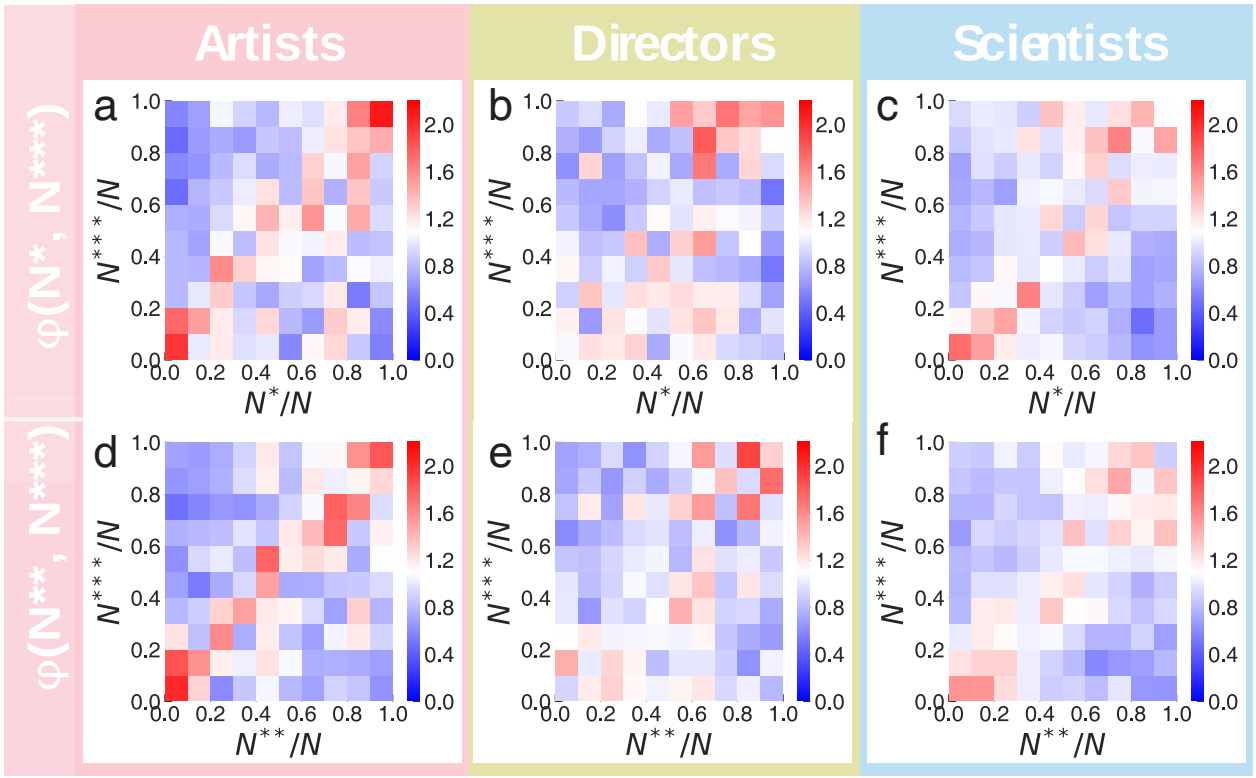

Figure S5: Φ **for other pairs of hit works.** (a–c), $\Phi(N^{*}, N^{***})$ of real careers for (a) artists, (b) directors, and (c) scientists. (d–f), $\Phi(N^{**}, N^{***})$ of real careers for (d) artists, (e) directors, and (f) scientists. We find the same patterns along the diagonal for $\Phi(N^{*}, N^{***})$ and $\Phi(N^{**}, N^{***})$.

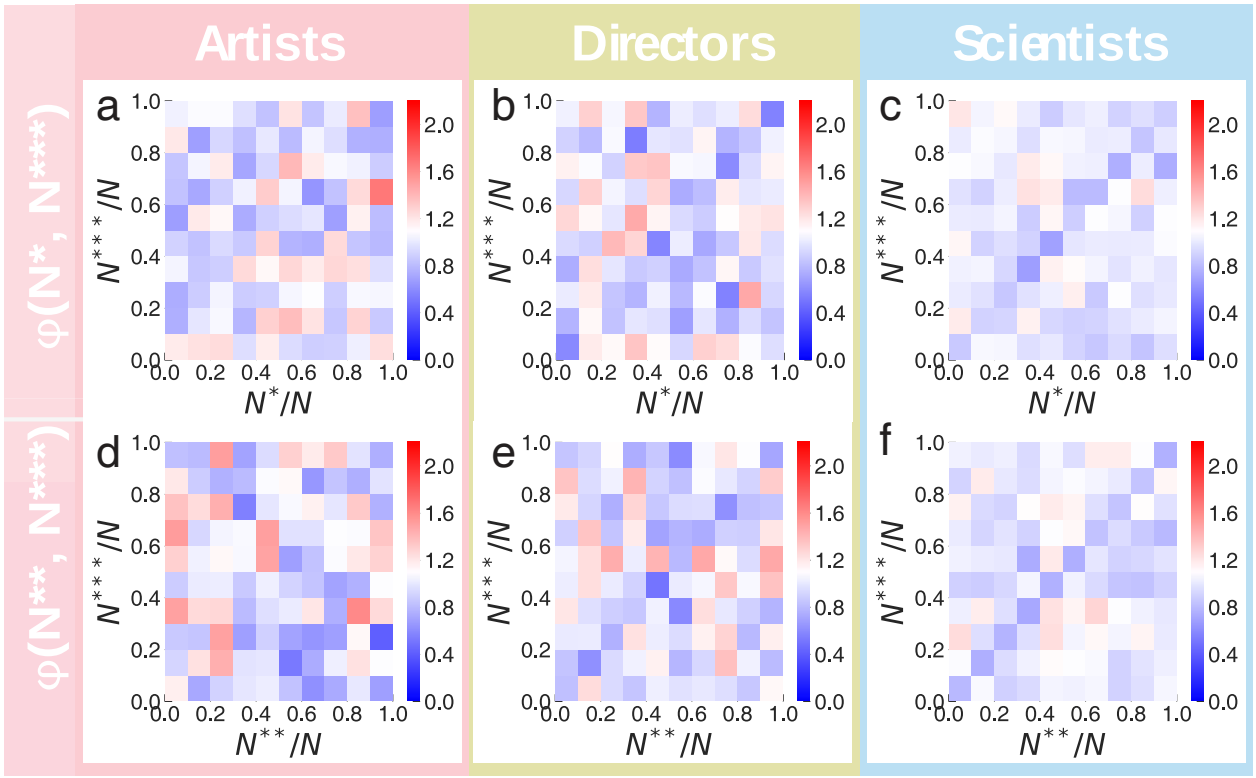

Figure S6: Φ **for shuffled careers.** (a–c), $\Phi(N^{*}, N^{***})$ of shuffled careers for (a) artists, (b) directors, and (c) scientists. (d–f), $\Phi(N^{**}, N^{***})$ of shuffled careers for (d) artists, (e) directors, and (f) scientists. We find for shuffled careers the patterns along the diagonal for $\Phi(N^{*}, N^{***})$ and $\Phi(N^{**}, N^{***})$ disappear.

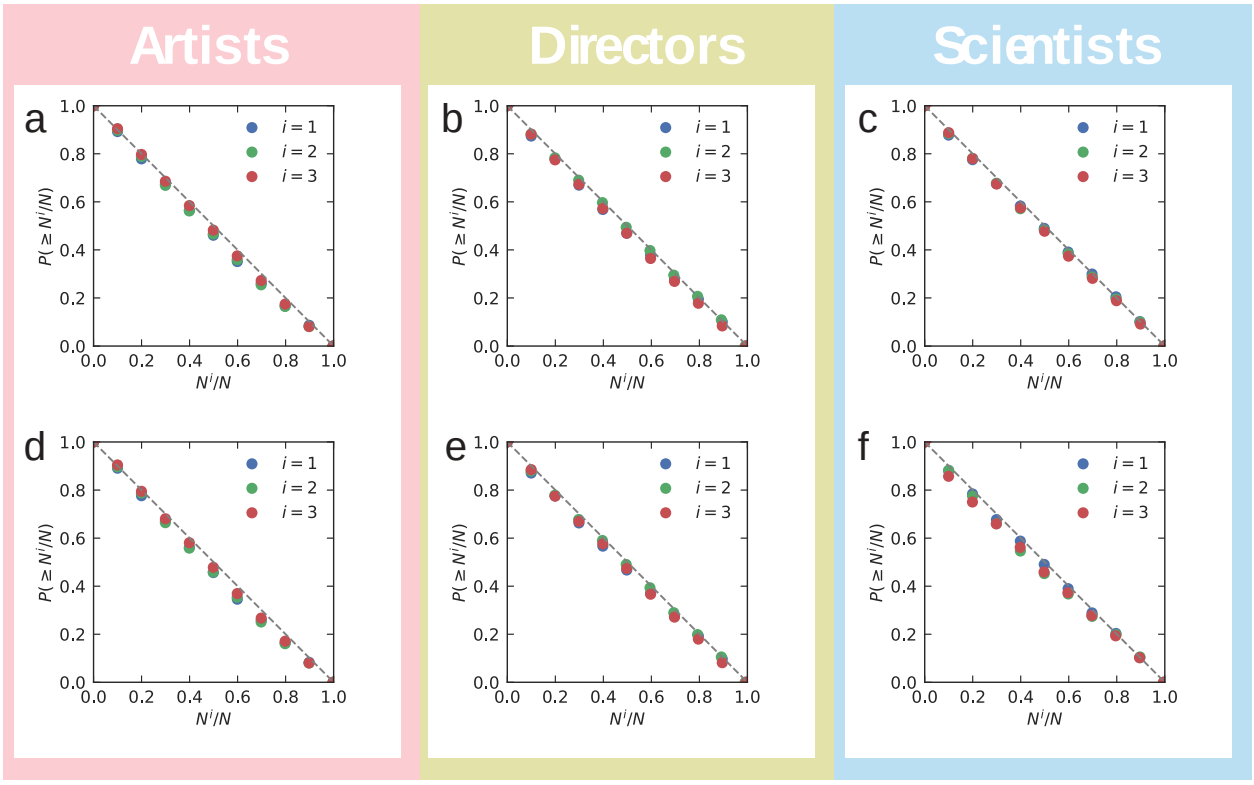

Figure S7: **Random impact rule under different career length.** $P(\geq N^i/N)$ for individuals under different career length: Artists with at least (a) 20 years and (d) 30 years of career length, directors with at least (b) 20 years and (e) 30 years of career length, and scientists with at least (c) 30 years and (f) 40 years of career length. N^i denotes the order of the i^{th} hit work within a career. The color denotes different hit works, and the dashed grey line denotes $P(\geq N^i/N)$ for a uniform distribution.

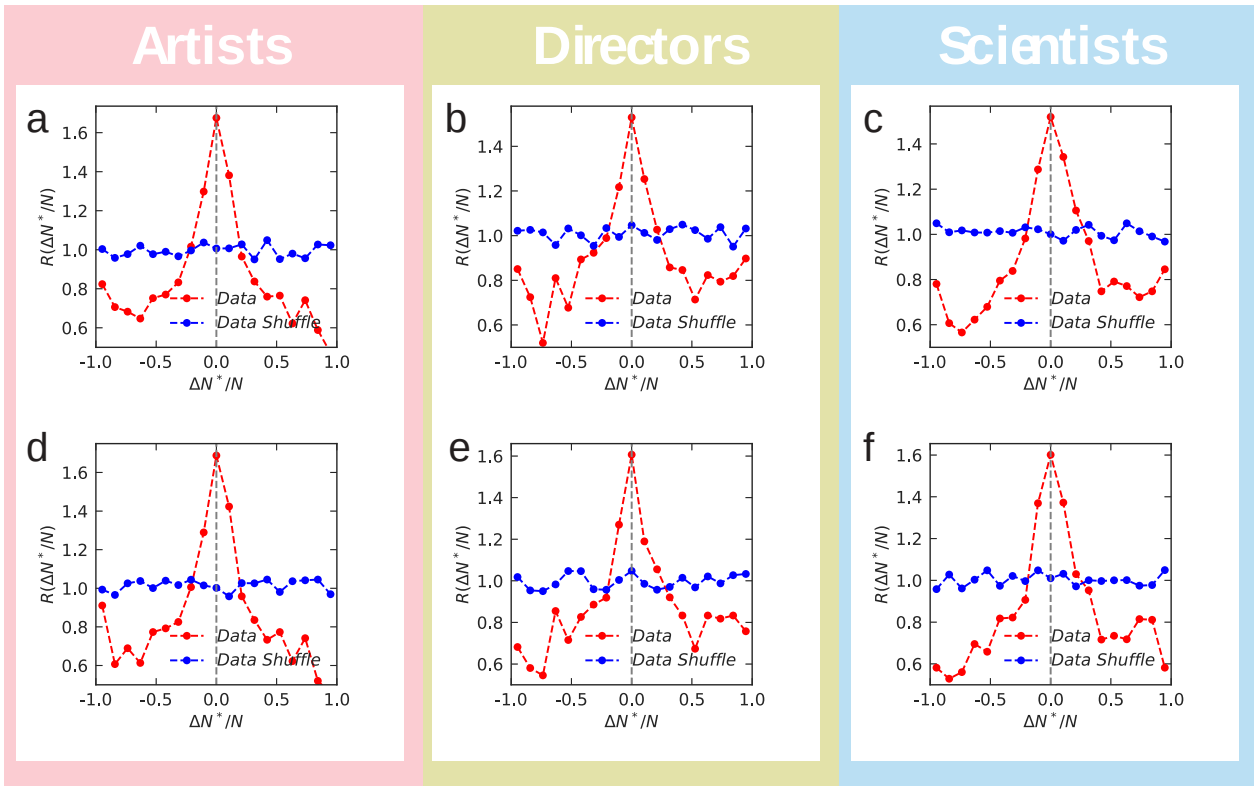

Figure S8: $R(\Delta N^*/N)$ **under different career length.** $R(\Delta N^*/N)$ for individuals under different career length: Artists with at least (a) 20 years and (d) 30 years of career length, directors with at least (b) 20 years and (e) 30 years of career length, and scientists with at least (c) 30 years and (f) 40 years of career length, where $\Delta N^* = N^* - N^{**}$. Red dots denote data and blue dots denote shuffled careers. We find the $R(\Delta N^*/N)$ still peaks around zero for individuals with different career length.

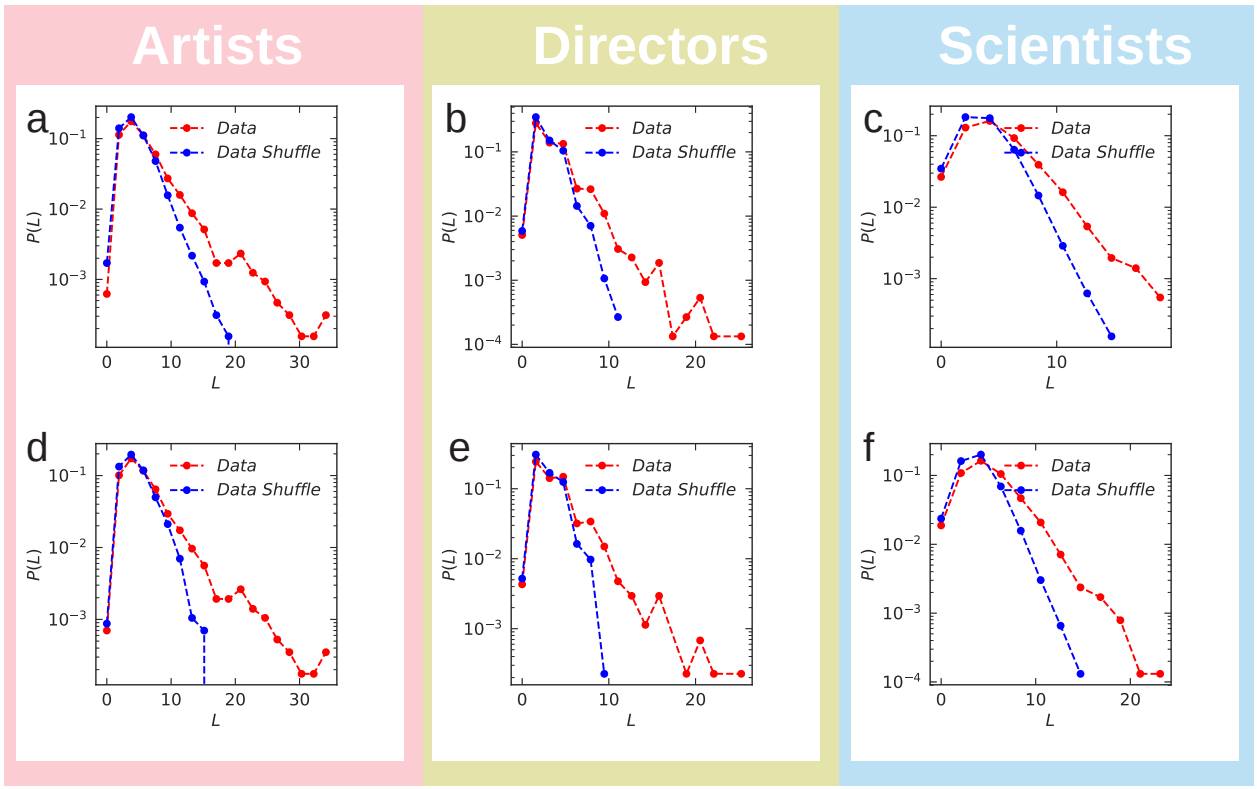

Figure S9: **Streak length under the different career length.** $P(L)$ for individuals under different career length: Artists with at least (a) 20 years and (d) 30 years of career length, directors with at least (b) 20 years and (e) 30 years of career length, and scientists with at least (c) 30 years and (f) 40 years of career length. Red dots denote data and blue dots denotes shuffled careers. We find the $P(L)$ of data has exponential tail that is wider than the shuffled careers for individuals with different career length.

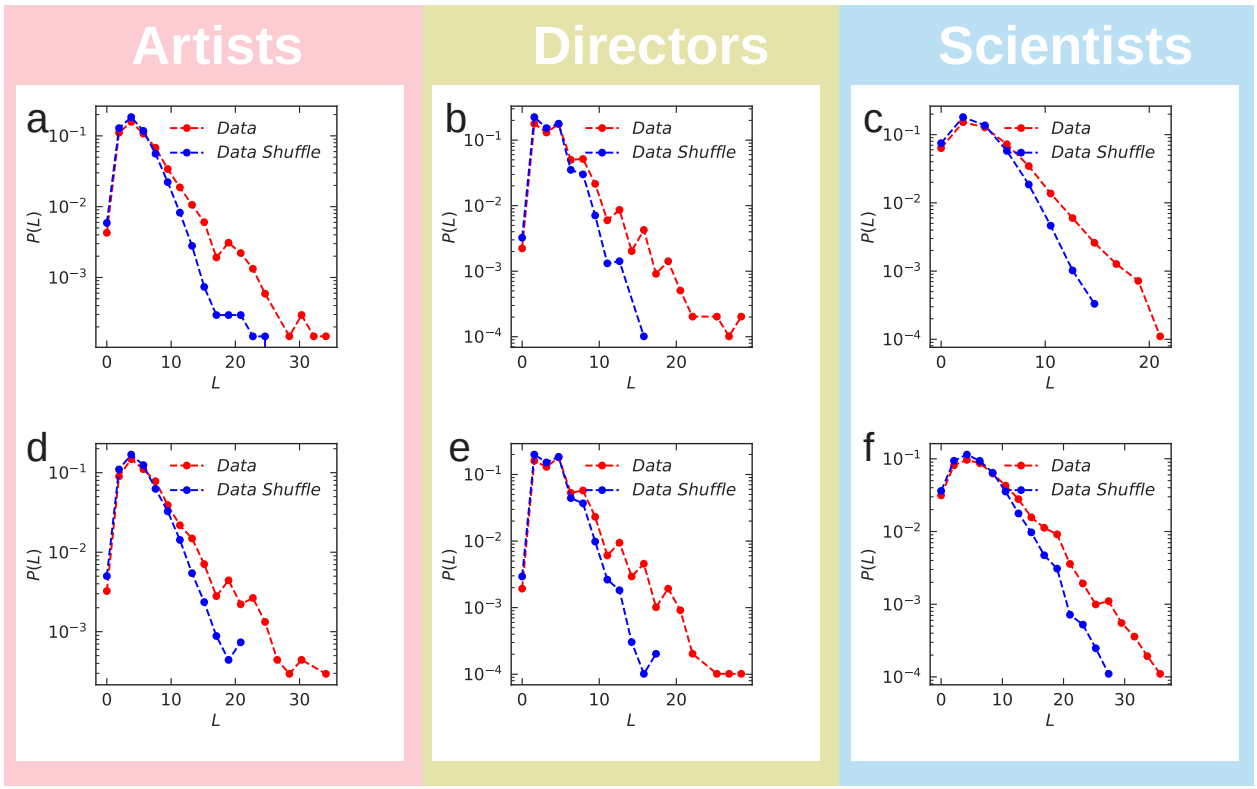

Figure S10: **Streak length under different thresholds**. (a–c), $P(L)$ and $P(L_S)$ when the threshold is the mean of impact within a career for (a) artists, (b) directors and (c) scientists. (d–f) $P(L)$ and $P(L_S)$ when the threshold is the geometric mean of impact within a career for individuals across three domains.

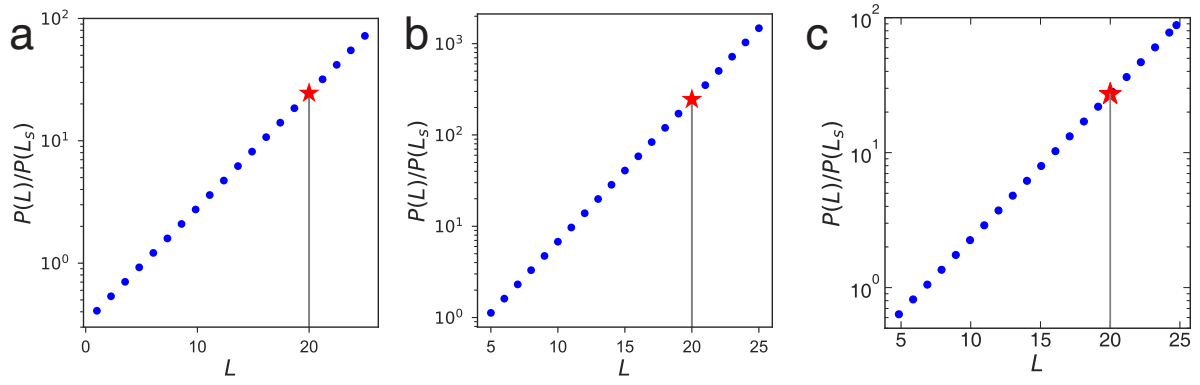

Figure S11: **The difference of $P(L)$ and $P(L_S)$.** (a–c), $P(L)/P(L_S)$ for (a) artists, (b) director, and (c) scientists in semi-log scale. The vertical line denotes $L = 20$, and the red star denotes the probability difference when $L = 20$.

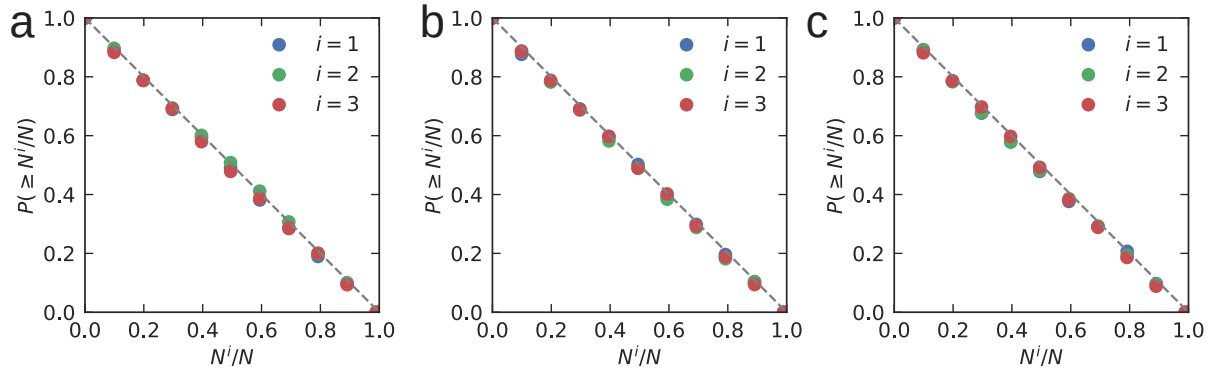

Figure S12: **Random impact rule under the null model.** $P(\geq N^i/N)$ under the null model prediction for (a) artists, (b) directors, and (c) scientists. The null model can reproduce the randomness of the top three highest impact works in a career.

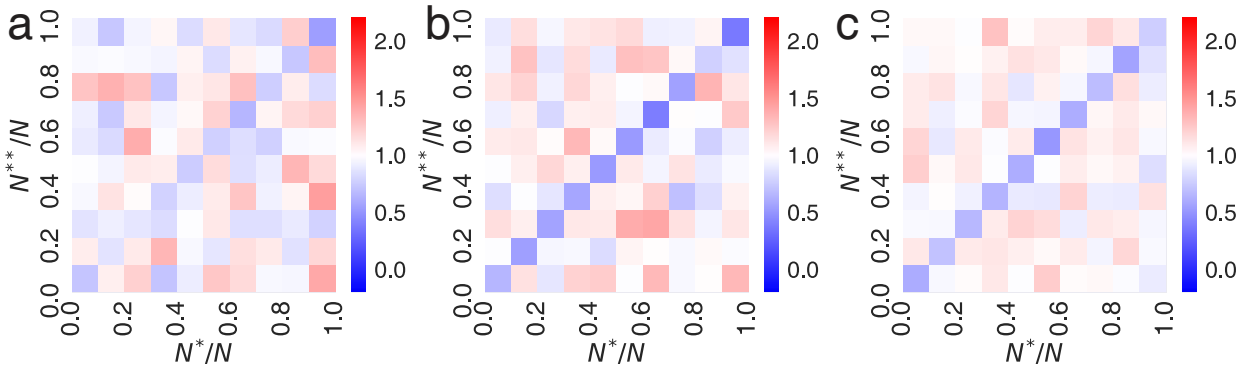

Figure S13: $\Phi(N^*, N^{**})$ **predicted by the null model.** The normalized joint probability $\Phi(N^*, N^{**})$ for the top two hit works within a career under the null model prediction for (a) artists, (b) directors, and (c) scientists. The null model cannot reproduce the temporal colocation of the two highest impact works within a career.

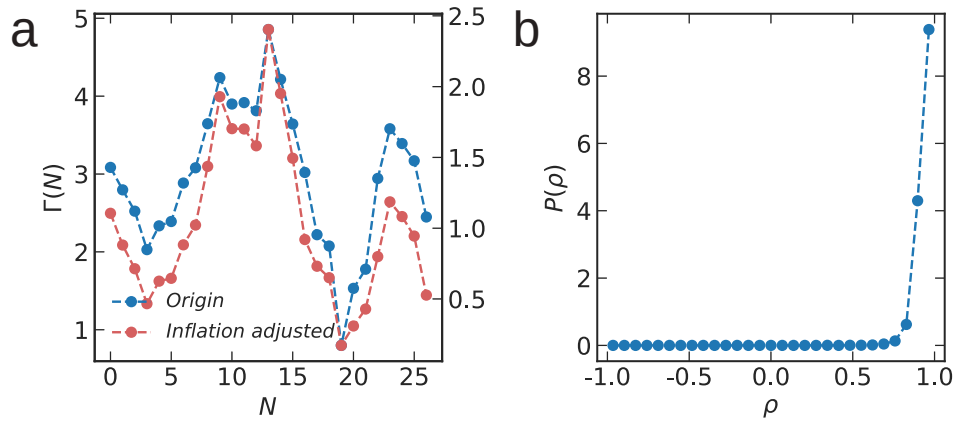

Figure S14: **Comparison of** $\Gamma(N)$ **calculated from raw and rescale** $\log C_{10}$**.** (a) The $\Gamma(N)$ sequence for a scientist in our dataset calculated from raw C_{10} (blue dots) and rescaled C_{10} (red dots). (b) The distribution of Pearson correlation $P(\rho)$ for $\Gamma(N)$ sequence calculated from the raw C_{10} and rescaled C_{10}. $P(\rho)$ peaks around 1.0, with mean value 0.93.

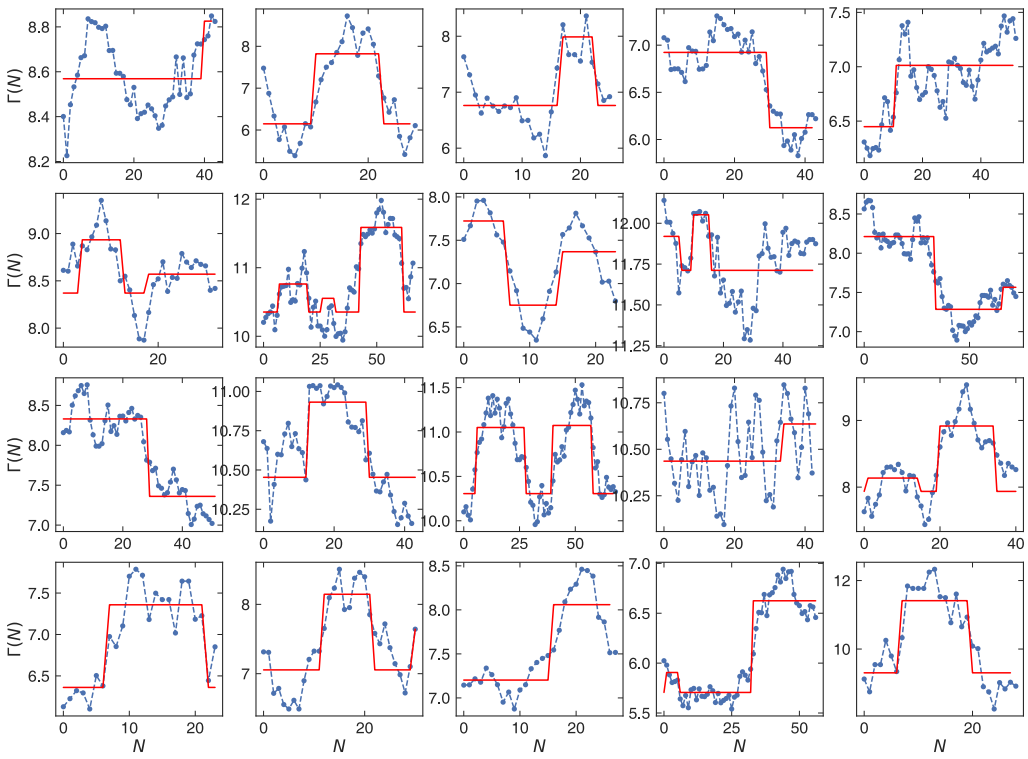

Figure S15: **The fitting results of 20 randomly selected artists.** Each subplot denotes the $\Gamma(N)$ sequence of an individual in our dataset, where blue dots denote data and red lines denote the best fitting result of each individual.

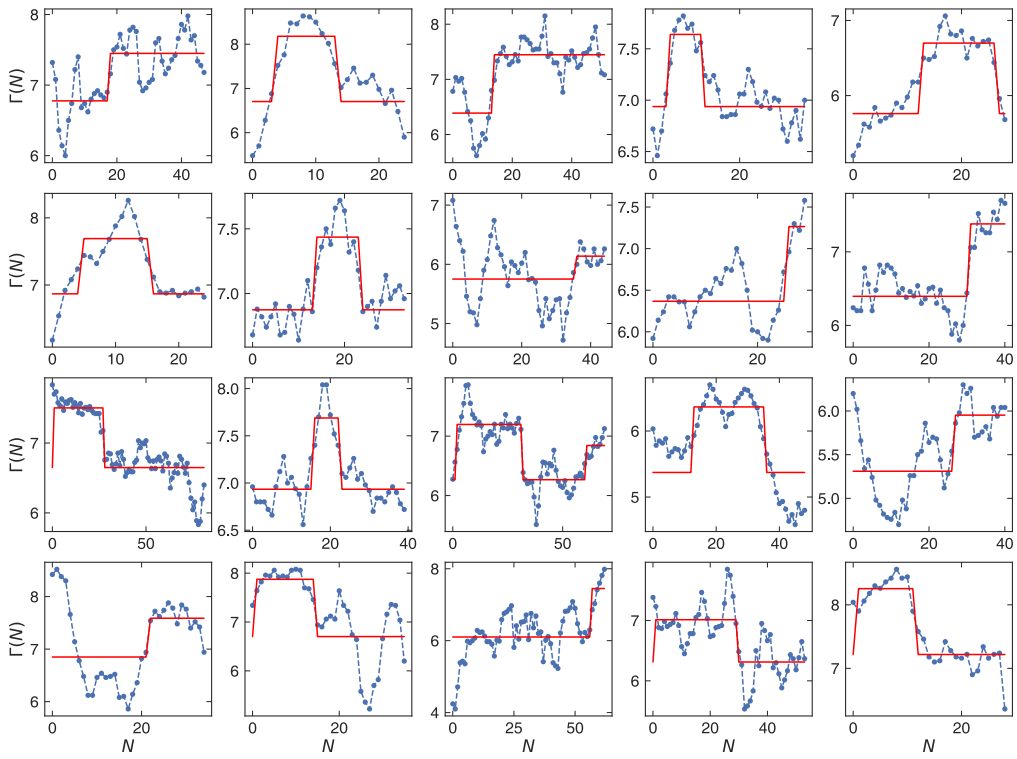

Figure S16: **The fitting results for 20 randomly selected directors.** Each subplot denotes the $\Gamma(N)$ sequence of an individual in our dataset, where blue dots denote data and red lines denote the best fitting result of each individual.

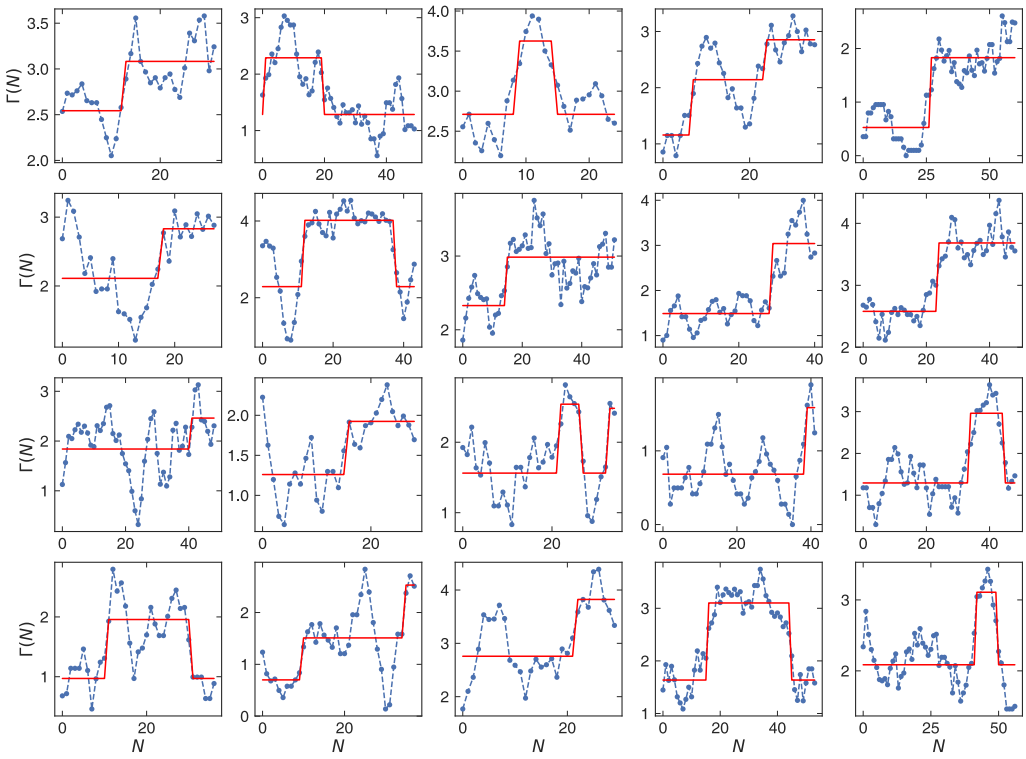

Figure S17: **The fitting results for 20 randomly selected scientists.** Each subplot denotes the $\Gamma(N)$ sequence of an individual in our dataset, where blue dots denote data and red lines denote the best fitting result of each individual.

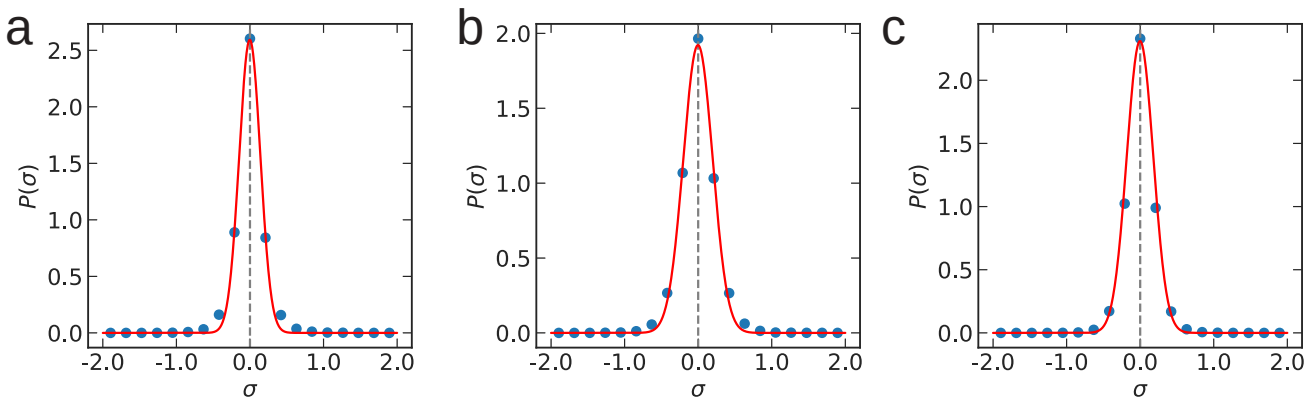

Figure S18: **Measuring the noise of $\Gamma(N)$.** The distribution $P(\sigma)$ for (a) artists, (b) directors, and (c) scientists, where σ is the difference between real and fitted $\Gamma(N)$ sequence for all points in a domain. The blue dots denote data and the red line is a normal distribution. $P(\sigma)$ peaks around zero for the three domains. The standard deviation for the normal distribution is 0.186 for artists, 0.229 for directors, and 0.189 for scientists, respectively.

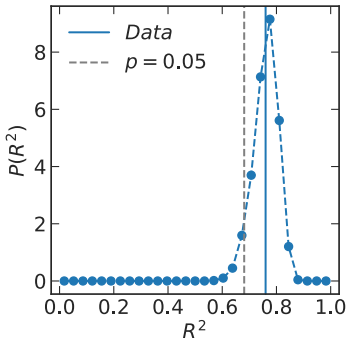

Figure S19: **An illustration of R^2 distribution.** The distribution $P(R^2)$ of simulated careers for an individual in our dataset. The blue dots denote $P(R^2)$ calculated from 1000 simulated careers, the vertical blue line denotes R^2 between data and fitted $\Gamma(N)$, and the dashed grey line is the R^2 when p-value is 0.05.

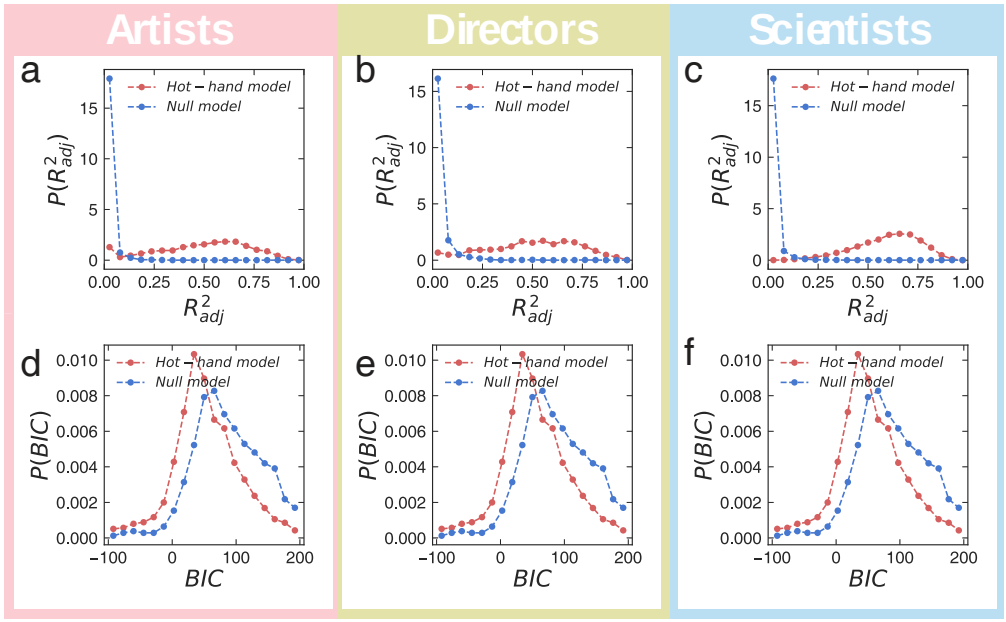

Figure S20: **Model evaluation.** (a–c), Distribution for adjusted R^2 of the hot-hand model and the null model for (a) artists, (b) directors, and (c) scientists. (d–f), BIC distribution of the hot-hand model and the null model. Red dots denote the hot-hand model and blue dots denote the null model.

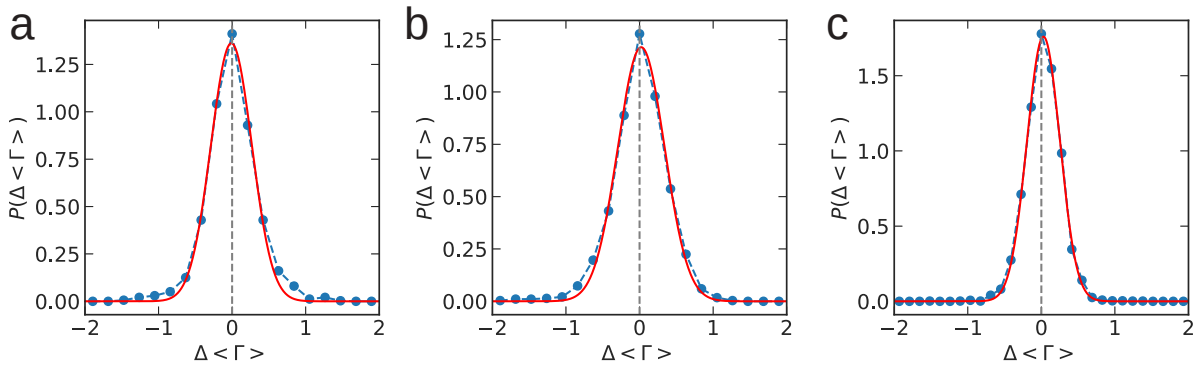

Figure S21: **Impact before and after hot hand.** The distribution of $P(\Delta\langle\Gamma\rangle)$ for (a) artists, (b) directors, and (c) scientists, where $\Delta\langle\Gamma\rangle$ measures the average impact difference before and after a hot-hand period. The blue dots denote data, and the red line is a normal distribution. The mean value of the normal distribution is 0.01 for artists, -0.02 for directors, and 0.02 for scientists, respectively.

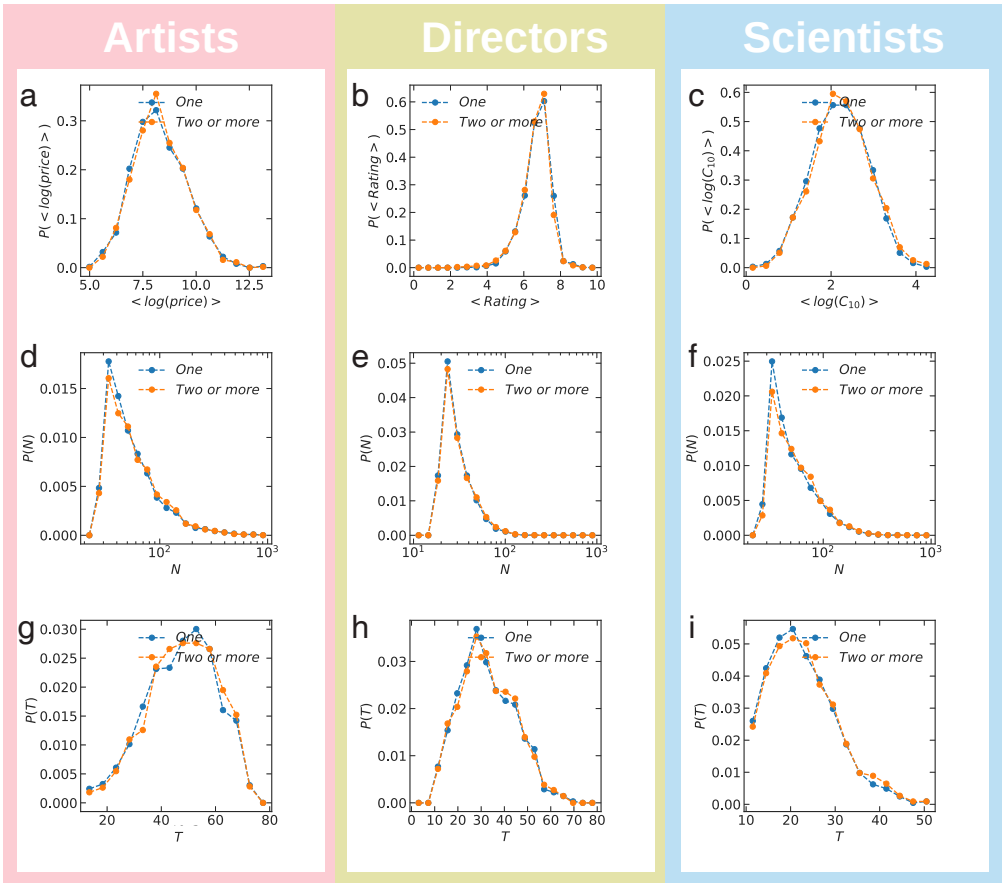

Figure S22: **Individuals with one and more than one hot hands.** (a–c), The distribution of average impact for individuals with one, and more than one hot-hand periods for (a) artists, (b) directors, and (c) scientist. (d–f), The distribution of number of works $P(N)$ within a career for individuals with one and more than one hot-hand periods across three domains. (g–i), The distribution of career length $P(\tau)$ for individuals with one and more than one hot-hand periods across three domains. The blue dot denotes individuals with one hot-hand period, and the orange dot denote the ones with at least two hot-hand periods.

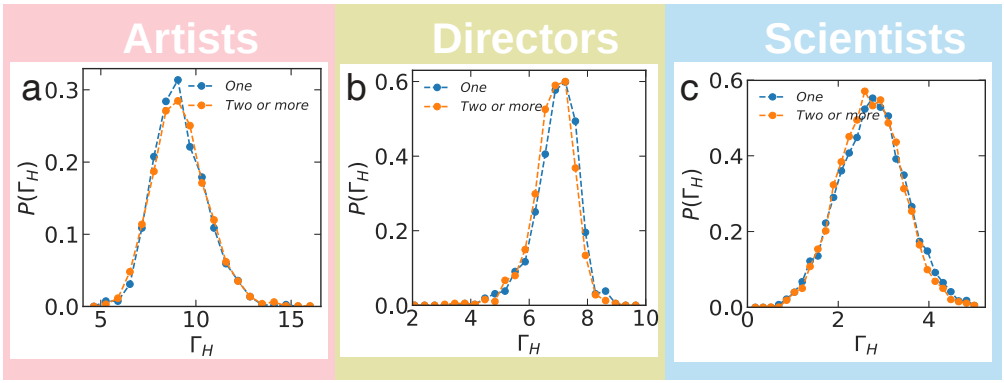

Figure S23: Γ **for individuals with different numbers of hot hands.** (a–c), $P(\Gamma_H)$ of individuals with one and more than one hot-hand periods, showing that the two groups have similar distributions across three domains.

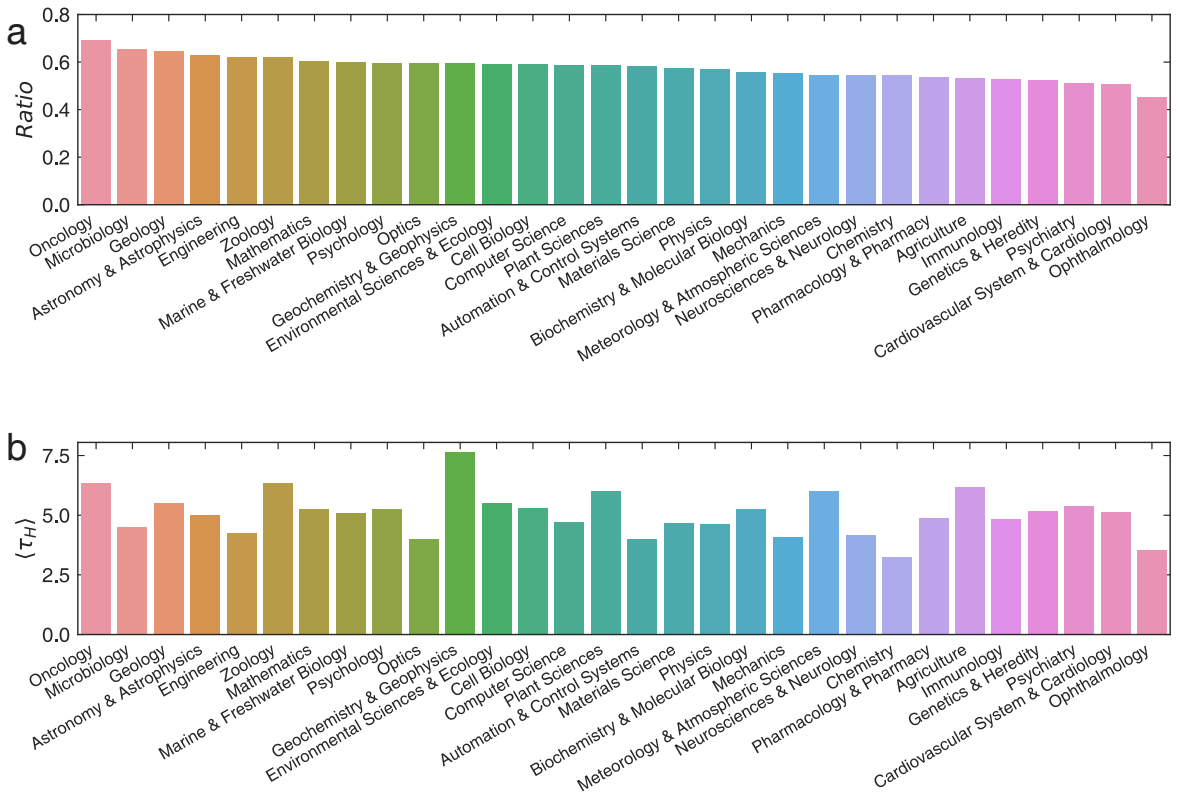

Figure S24: **Hot hand for different scientific disciplines.** (a) The proportion of scientists with one hot-hand period in each discipline. (b) The average duration $\langle \tau_H \rangle$ for scientists within each discipline.

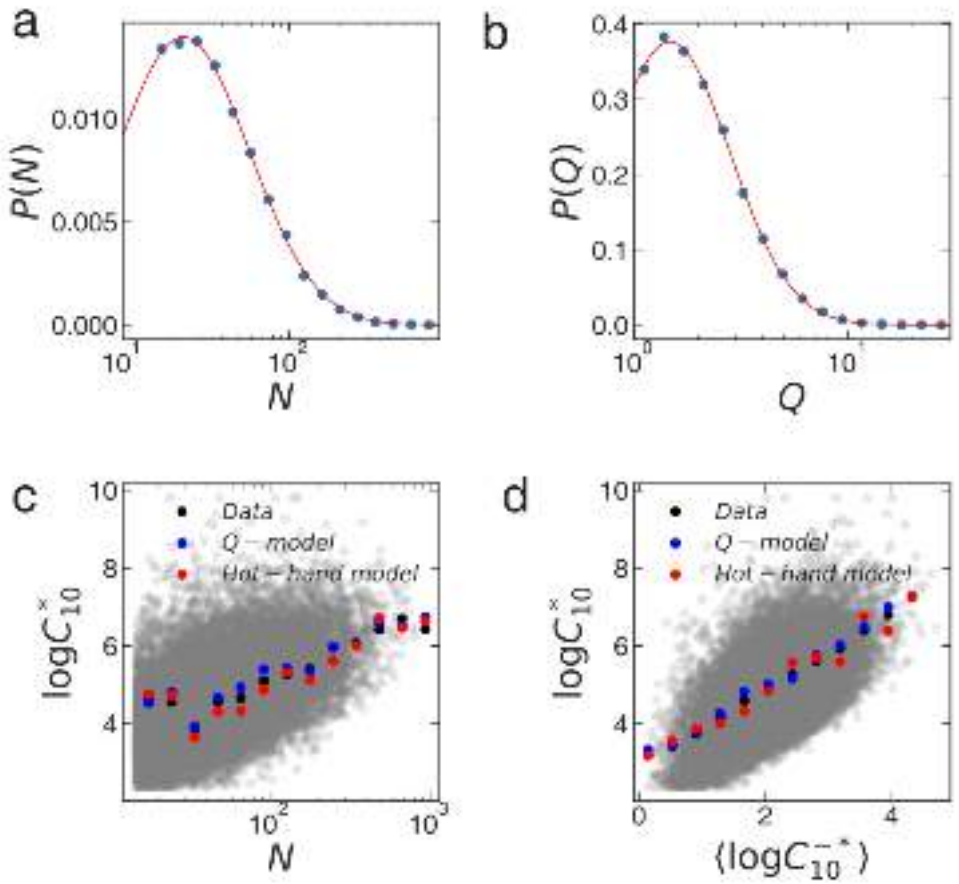

Figure S25: **Q model validation for scientists.** (a) The distribution $P(N)$ for scientists, where dots denote data, and the red line corresponds to a log-normal function with average $\mu = 3.9$ and standard deviation $\sigma = 0.84$. (b) The distribution $P(Q)$ for scientists. The red line corresponds to a log-normal function with $\mu = 0.8$ and $\sigma = 0.58$ (c) The highest impact $\log C^*_{10}$ versus the number of works N within a career. Each grey dot corresponds to an artist. The black circles are the logarithmic binning of the scattered data. The blue and red circles represent the prediction of the Q-model and the hot-hand model, respectively. (d) $\log C^*_{10}$ versus $\langle \log C^{-*}_{10} \rangle \rangle$. Each grey dot corresponds to an artist, where $\langle \log C^{-*}_{10} \rangle \rangle$ is the average impact within a career without $\log C^*_{10}$.

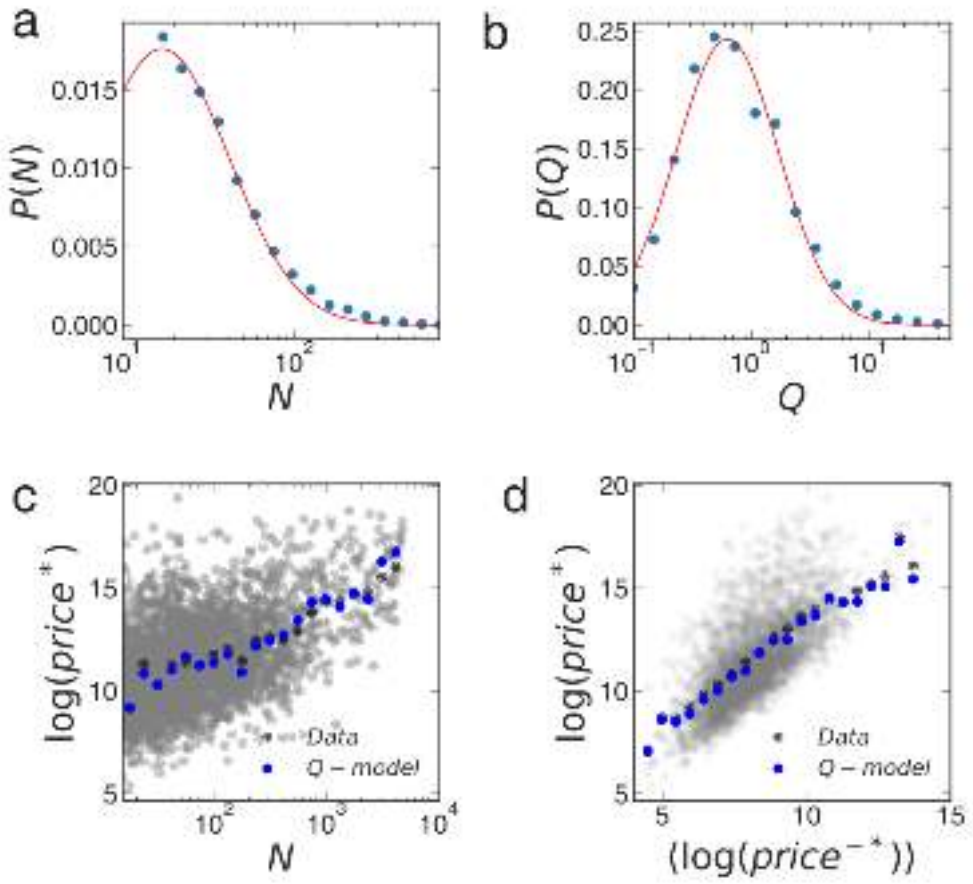

Figure S26: **Q model validation for artists.** (a) The distribution $P(N)$ for artists, where dots denote data, and the red line corresponds to a log-normal function with average $\mu = 3.6$ and standard deviation $\sigma = 0.90$. (b) The distribution $P(Q)$ for artists. The red line corresponds to a log-normal function with $\mu = 1.1$ and $\sigma = 1.21$ (c) The highest impact $\log(\text{price}^*)$ versus the number of works N within a career. Each grey dot corresponds to an artist. The black circles are the logarithmic binning of the scattered data. The blue circles represent the prediction of the Q-model. (d) $\log(\text{price}^*)$ versus $\langle \log(\text{price}^{-*}) \rangle$. Each grey dot corresponds to an artist, where $\langle \log(\text{price}^{-*}) \rangle$ is the average impact within a career without $\log(\text{price}^*)$.

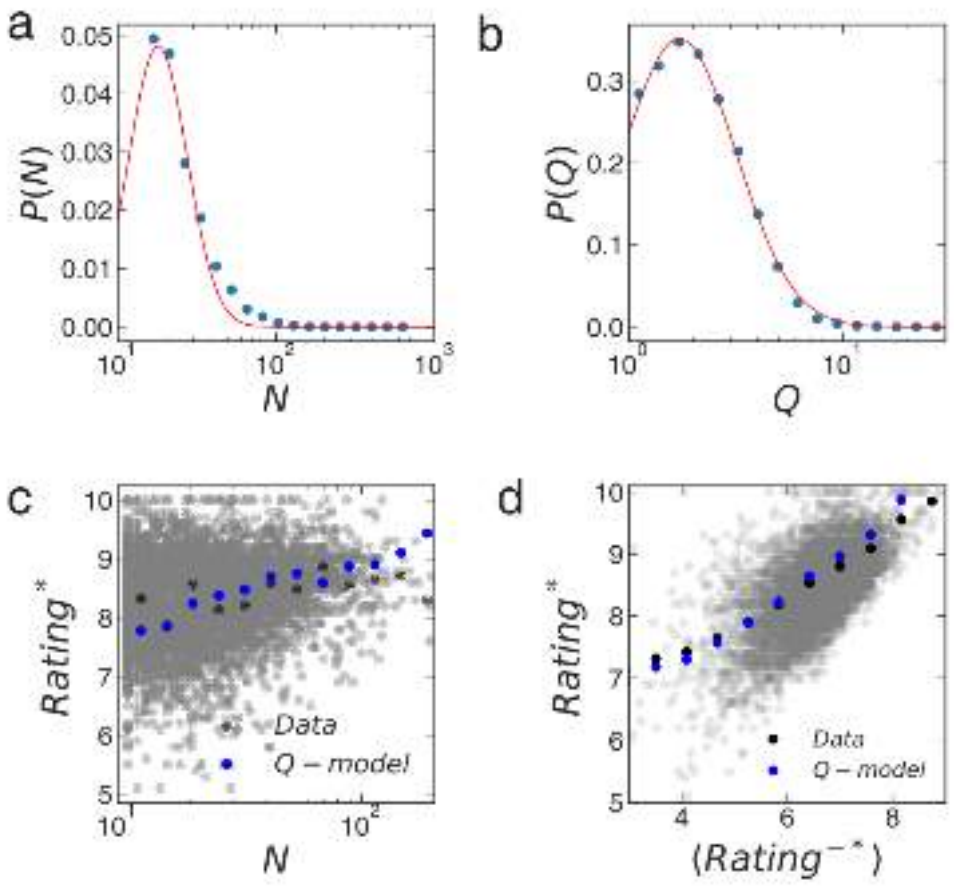

Figure S27: **Q model validation for directors.** (a) The distribution $P(N)$ for directors, where dots denote data, and the red line corresponds to a log-normal function with average $\mu = 3.1$ and standard deviation $\sigma = 0.42$. (b) The distribution $P(Q)$ for directors. The red line corresponds to a log-normal function with $\mu = 0.9$ and $\sigma = 0.56$ (c) The highest rating Rating* versus the number of works N within a career. Each grey dot corresponds to a director. The black circles are the logarithmic binning of the scattered data. The blue circles represent the prediction of the Q-model. (d) Rating* versus $\langle \text{Rating}^{-*} \rangle$. Each grey dot corresponds to a director, where $\langle \text{Rating}^{-*} \rangle$ is the average ratings within a career without Rating*.

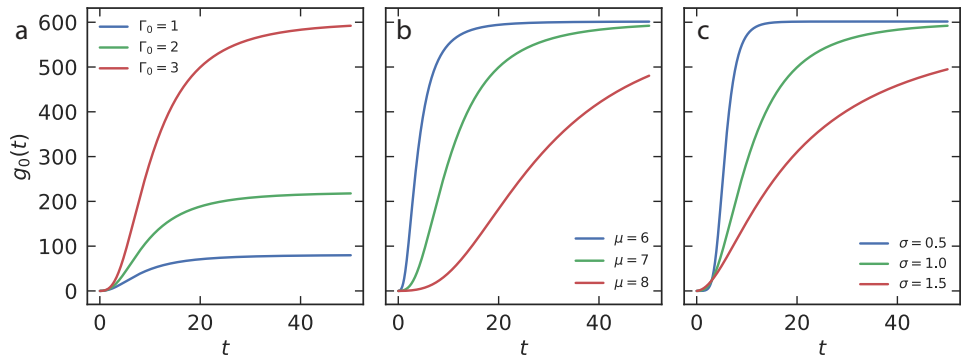

Figure S28: $g(t)$ **under the null model.** (a–c), The behaviour of $g_0(t)$ under the null model prediction with different (a) Γ_0, (b) μ, and (c) σ parameters. We use $\Gamma_0 = 3.0$, $\mu = 7.0$, and $\sigma = 1.0$ as input.

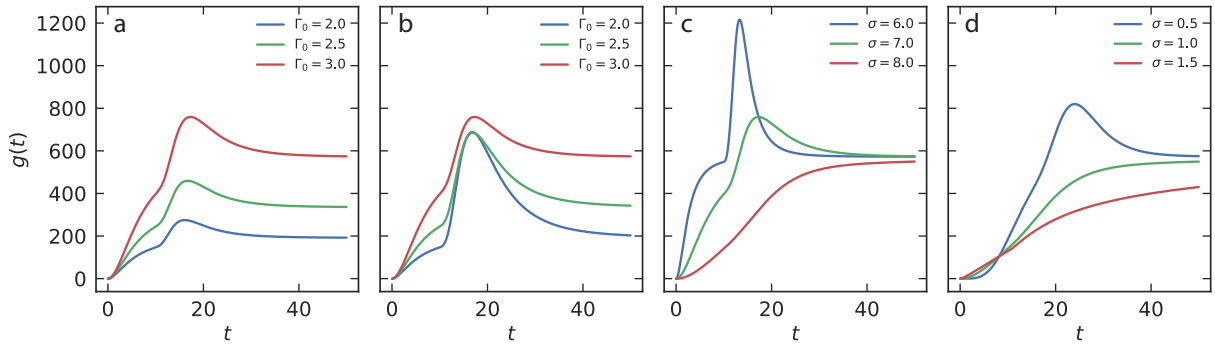

Figure S29: $g(t)$ **under the hot-hand model.** (a–d), $g(t)$ under the hot-hand model prediction with different (a) Γ_0, (b) $\Delta\Gamma$, (c) μ, and (d) σ parameters. For (a–c) we use $\Gamma_H = 4.0$, $\Gamma_0 = 3.0$, $\mu = 7.0$, $\sigma = 1.0$, $t_\uparrow = 10$ years and $t_\downarrow = 12$ years as input. We use $\Gamma_H = \Gamma_0 + 1.0$ for (a).

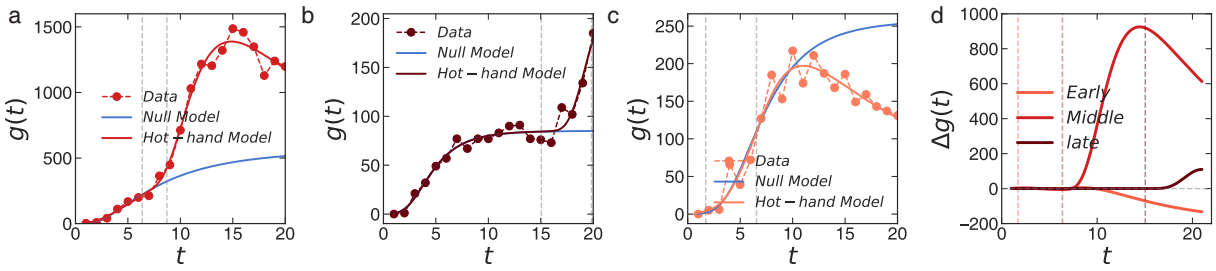

Figure S30: **Policy implications of** $g(t)$. (a–c), Individuals in our dataset with (a) mid, (b) late, and (c) early onset of the hot-hand period. Red dots denote data, the blue line is the null model's prediction based on early performance, and the red line captures the predictions from the hot-hand model, with dashed grey lines denoting the start and end of hot-hand periods. (d) The difference $\Delta g(t)$ between our hot-hand model and the null model for each individual, where dashed lines with corresponding color denotes the start of a hot-hand period. (d) illustrates the systematic discrepancies in predicting individual's future impact, if we ignore the uncovered hot hand phenomena.

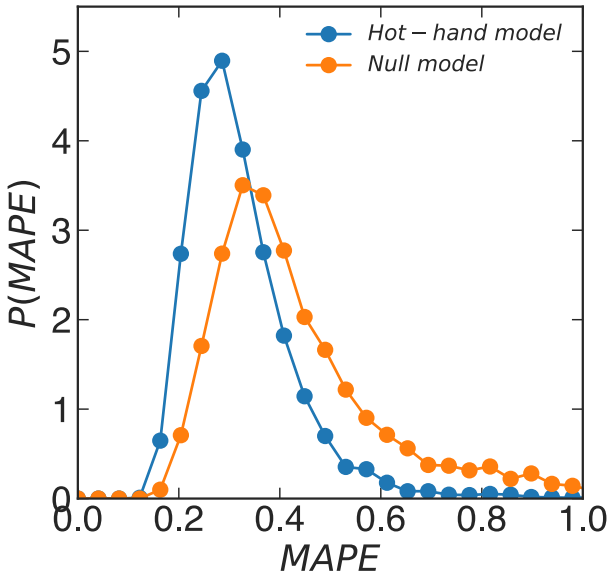

Figure S31: **The performance of different models.** The distribution $P(MAPE)$ for the null model (orange dots) and our hot-hand model (blue dots). The hot-hand model systematically outperforms the null model.

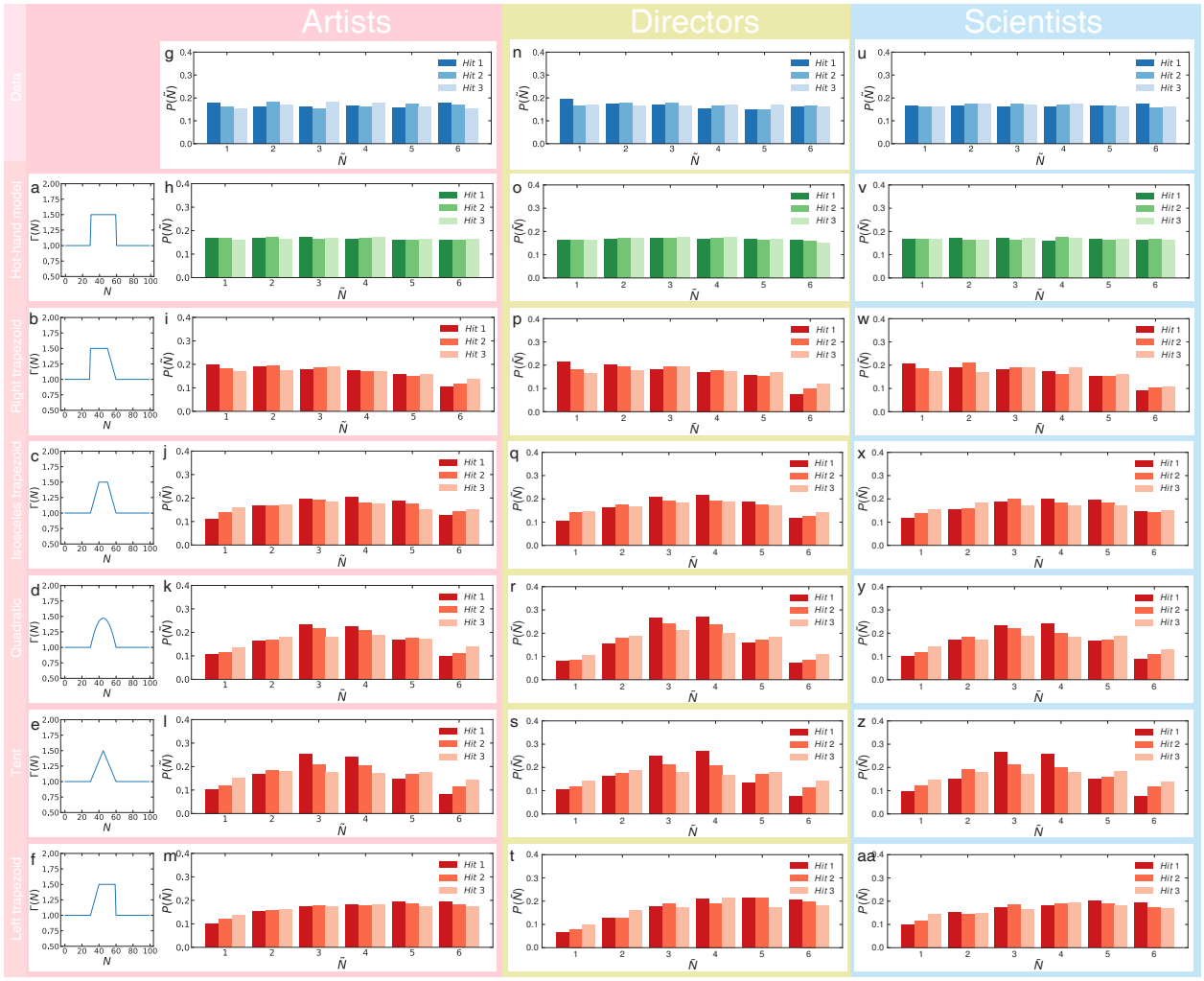
Artists
Directors
Scientists
Data
Hot-hand model
Right trapezoid
Isosceles trapezoid
Quadratic
Tent
Left trapezoid
Hit 1
Hit 2
Hit 3
Γ(N)
N
P(Ñ)
Ñ

Figure S32: **Alternative models of hot hands.** (a–f), An illustration of $\Gamma(N)$ for the (a) hot-hand model, (b) right trapezoid function, (c) isosceles trapezoid function, (d) quadratic function, (e) tent function, and (f) left trapezoid function. (g) The distribution of the relative position $P(\tilde{N})$ of the top three hit works among the top six hits within a career for artists. The shades of color correspond to different hits.The relative position $\tilde{N} = 1$ means the hit paper appears first among the top six hit, whereas $\tilde{N} = 6$ means the last one to appear. The relative order among the top six hits are random. The conclusion remains the same for (n) directors and (u) scientists. (h-m), $P(\tilde{N})$ predicted by corresponding model shown in (a–f) using artists' profiles as input. We again use shades to color different hits. We measure the statistical difference between data and the model's predictions, using the p-value of KS test for discrete distributions. We color the bars green if we cannot reject the hypothesis that data and model's prediction come from the same distributions (p-value ≥ 0.05), and color them red otherwise. Among all alternative models considered, the hot-hand model turns out to be the only model that successfully reproduces the randomness observed in the top six hit works in (a). While the alternative functions show different trend of probability, contradicting to the randomness measured from data. We repeated the analyses using (o–t) directors' profiles and (v–aa) scientists' profiles as input, finding our results are robust to different domains.