https://doi.org/10.1038/s44260-026-00078-y

On the nature of human performance in competitive endeavors

Check for updates

Vitalii Zhukov[1], Panagiotis Tsiamyrtzis[2,3] & Ioannis Pavlidis[1] ✉

Heavy-tailed success is common in human competition, but it is unclear when it signals runaway dominance versus fair opportunity for skill to accumulate. We study outcome distributions in three arenas: World War II Luftwaffe fighter aces (victories), U.S. biology and computer science faculty competing for NIH/NSF awards, and U.S. Olympic swimmers and French Olympic fencers (medal totals). For each domain we analyze system-total distributions over the full record, then partition the same data into periods aligned with institutional eras to test whether tail shape is stable or shifts over time. Using a common tail-frontier scan, we fit three discrete upper-tail models—discrete lognormal (dLN), Zipf, and shifted geometric—over varying retained fractions. Where rules are stable and participants enjoy sustained chances to compete, upper tails consistently concentrate around a dLN regime: heavy but sub-power-law, consistent with repeated multiplicative gains under what we term Relative-Fairness, where skill has a fighting chance to accumulate. Time-partitioned analyses probe falsifiability: relaxing selectivity or temporarily doubling resources shifts tails toward a thinner, geometric-like regime, while episodic dominance yields localized Zipf episodes. Stress tests that vary roster size and competition tier under fixed rules show that tail shape distinguishes chance-dominated, relatively fair, and dominance-driven regimes.

Human societies have long staged an extraordinary range of competitive systems, from funding races in science and tournaments in elite sport to mortal duels in warfare and other arenas. In all of these settings, a small minority of participants often accumulate a disproportionate share of success—grants, medals, or confirmed victories—while most achieve modest totals. Heavy-tailed performance distributions of this kind have been documented in scientific careers, elite sport, and martial success[1–5], but the underlying generative mechanisms remain contested. Do these heavy tails primarily signal runaway cumulative advantage, or can they also arise under rigorous but relatively fair competition, where entrants are already highly selected and skill has repeated chances to accumulate? And how do outcome distributions change when institutions broaden access and weaken selectivity? Three large strands of literature offer complementary but distinct answers to these questions.

First, the *Science of Science* tradition emphasizes cumulative advantage and the *Matthew effect* in scientific careers: early recognition and access to resources can compound into very heavy, sometimes power-law-like tails in citations, awards, and career length[6]. This work documents highly skewed impact distributions and shows that chance and timing can interact with ability to generate outsized careers, suggesting that scientific competition often operates in a rich-get-richer regime. Related contributions on research stars and star scientists highlight how a small number of individuals can reshape entire domains[7,8].

Second, research in *strategy and industrial dynamics* documents persistent performance heterogeneity among firms and business units. Some aggregates (such as firm sizes) approach Zipf-like behavior[9,10], whereas *flow* outcomes (such as annual growth, returns, or productivity changes) often exhibit lognormal bodies with lighter, sometimes Laplace-like, tails[11–14]. This literature links tail shape to organizational capabilities, market structure, and competitive interaction, and shows that not every setting is dominated by runaway winners; in many cases, firms differ in their ability to convert repeated opportunities into performance without collapsing into a single dominant player.

Third, *human capital and labor economics* models connect experience, specialization, and skill accumulation to payoffs in labor markets. Tournament and superstar theories show how rank-based rewards and winner-take-all pay schemes can create highly skewed income and performance distributions even when underlying ability differences are modest[15]. Other models, such as those based on learning curves, multiplicative wage growth, and career specialization, naturally produce lognormal or lognormal-plus-Pareto distributions of earnings and output[16–18].

[1]University of Houston, Department of Computer Science, Houston, TX, USA. [2]Politecnico di Milano, Department of Mechanical Engineering, Milan, Italy. [3]Department of Statistics, Athens University of Economics and Business, Athens, Greece. ✉e-mail: ipavlidis@uh.edu

Taken together, these frameworks do not point to a single universal tail form; instead they suggest how outcome distributions can vary with opportunity structure. Strong cumulative advantage and winner-take-all rewards can generate very heavy, near-power-law upper tails. Multiplicative skill and experience accumulation under demanding, repeated competition leads to heavy but sub-power-law distributions, often lognormal or lognormal-plus-Pareto. Conceptually, when access is broad and both selection and prize spreads are relatively weak, one would expect much thinner upper tails in which repeated trials contribute less to long-run inequality. For what follows, it is useful to organize these ideas into three stylized regimes:

- **Runaway cumulative advantage**. A small set of actors leverage early gains into extreme dominance, consistent with very heavy, near-power-law tails in which the top performer or two hold a large share of all outcomes.
- **Broad-Access / low-selectivity competition**. Opportunities are plentiful and weakly selective; success in one event provides little advantage in the next. Outcome distributions are expected to be much thinner, closer to geometric or exponential tails.
- **Rigorous competition under*Relative-Fairness***. Entry is filtered and competition is demanding, but active participants enjoy *adequate repeated opportunities* to deploy and refine their skill. In this regime, heterogeneity in underlying ability and match quality is expressed through repeated, largely independent multiplicative gains, yielding heavy but sub-power-law tails: success accumulates, but outcomes do not collapse into a vanishingly small elite.

Our contribution is to use tail shape as a *diagnostic* of which regime is operative in real competitive systems, and to ask when a particular regime appears and when it breaks. In essence, we offer a conditional theory of Relative-Fairness with a testable signature and a falsifiable set of alternatives. Building on the three-regime lens above, we ask:

RQ1: In elite, freely competing domains where individuals can re-enter competitions and accumulate outcomes over many engagements, which upper-tail regime best describes the data: Runaway cumulative advantage, Broad-Access low-selectivity competition, or rigorous competition under Relative-Fairness?

RQ2: How stable are these regimes over time within a given domain? Do upper tails remain in the same family across institutional eras, or do they shift in response to documented changes in opportunity structure (e.g., funding shocks, professionalization, or wartime asymmetries)?

RQ3: Can upper-tail shape serve as a compact, falsifiable summary of opportunity structure—a way to diagnose when skill has a "fighting chance" to accumulate under Relative-Fairness—rather than as an unexplained empirical regularity?

To answer these questions, we take a *system lens* and analyze three domains that span martial, intellectual, and athletic competition in the modern era: (i) World War II Luftwaffe *fighter aces* (LW), meaning fighter pilots with at least five confirmed aerial victories, measured by total credited victories in one-on-one or few-on-few engagements[5,19–21]; (ii) U.S. biology/biomedical (BIO) and computer science (CS) faculty in research universities competing for research grants from the National Institutes of Health (NIH) and the National Science Foundation (NSF)[22–24]; and (iii) elite athletes representing the United States in Olympic swimming and France in Olympic fencing, constructed from long-run Olympic data and sport-specific studies[4,25–29]. For brevity, we refer to the first domain as *Fighter Aces LW* in tables and figures, and we formalize the ace definition in the Methods. These domains differ sharply in institutional structure and competitive time horizons, and they cover a wide range of stakes—from career-defining but non-lethal successes in grants and Olympic medals to life-or-death stakes in aerial duels, where each engagement can be fatal for the pilot. In our portfolio, Luftwaffe aerial combat thus serves as the lethal endpoint, bracketing the non-mortal but high-stakes grant and Olympic systems. They share three design features that are central for treating cumulative outcomes as valid indicators of underlying competitive performance: (1) entrants are already highly selected in their domain (skilled pilots, investigators, national-team athletes); (2) outcomes are systematically attributed and recorded at the level of individuals using domain-specific scoring procedures (confirmed victories, awarded grants, Olympic medals); and (3) individuals can re-enter competitions and accumulate outcomes over many engagements, seasons, or award cycles, creating adequate repeated opportunities for skill to manifest.

These domains are not arbitrary examples. For the martial domain, WWII aerial combat provides a rare and well-documented example of individually attributed martial success at scale, in which technological asymmetries were not yet overwhelming and where, on the German side, institutional rules did not impose a hard cap on the number of sorties or victories an ace could accumulate[20,30,31]. For the intellectual domain, NIH and NSF grants represent canonical high-stakes competitions for resources and prestige, with detailed records that allow us to track investigators across funding eras and policy shocks[32–35]. For the athletic domain, U.S. swimming and French fencing jointly satisfy demanding design criteria: continuous Olympic participation from 1896 onward, political continuity of the national entity, large and systematically selected national teams, and high-quality individual-level records spanning multiple institutional epochs[25,26,28,29,36,37]. Together, these systems span a broad spectrum of human activity—from intellectual to somatic performance, and from career to mortal stakes—providing a demanding test bed for any cross-domain regularity.

Methodologically, we focus on the *upper tail* of individual outcomes in each domain and time window, and we systematically compare three discrete tail models that operationalize the competitive regimes introduced earlier: a discrete lognormal (dLN) for heavy but sub-power-law tails under Relative-Fairness; a Zipf (discrete power law) for Runaway cumulative advantage; and a shifted geometric for Broad-Access, low-selectivity competition. Rather than fixing a tail cut-off in advance, we scan over retained fractions of the sample (e.g., the top fraction of individuals by outcome) in each domain–period. We then identify a *retained-fraction frontier*, defined as the range of retained fractions for which at least one model provides an adequate fit and the retained sample remains large enough for reliable estimation; within this frontier, we compare the dLN, Zipf, and geometric models on equal footing. This frontier turns the common qualitative observation that many competitive outcome distributions look roughly lognormal into a formal model-comparison framework, in which the dLN is tested directly against Zipf and geometric tails as explicit counterfactuals.

Substantively, our main finding is that in periods where historical and institutional evidence points to stable rules and sustained opportunities for participation, the discrete lognormal model consistently dominates at the frontier. Across very different domains, upper tails are heavy but systematically sub-power-law, consistent with a regime of rigorous competition under Relative-Fairness in which skill has a fighting chance to accumulate. When institutions change sharply, the selected model shifts in theoretically interpretable ways: for example, when funding becomes temporarily much less selective, tails move toward geometric behavior, reflecting broad access with weaker selection; when a small number of athletes or combatants achieve historically exceptional dominance, tails tilt toward Zipf-like heads indicative of strong cumulative advantage. Across domains and periods, we find little evidence that the extreme rich-get-richer regime provides the best description of these competitive systems once we account for adequate repeated opportunities.

In the remainder of the paper, we present the three datasets, introduce the frontier-based model comparison and study designs that link them, and report domain-level and period-specific results. We conclude by showing how tail shape can serve as a compact summary of opportunity structure in competitive human systems, and by considering implications for fairness, institutional design, and the interpretation of exceptional performance.

Results

We begin by summarizing descriptive patterns in each dataset and then compare upper-tail regimes using the retained-fraction frontier analysis (full details under Analytic Methodology in the Methods Section).

Table 1 | Descriptive statistics for the Fighter Aces LW, NSF Grantees CS, NIH Grantees BIO, Olympic Fencing Medalists FRA, Olympic Swimming Medalists US and GBR, and Tennis Top-tier and Challenger Runs data sets

Dataset	*N*	Min	Q1	Median	Mean	Q3	Max	SD	Skewness
Fighter Aces LW	1148	6	9.00	16	30.31	38	353	35.66	3.25
NSF Grantees CS	1323	2	5.00	8	9.77	12	79	7.38	3.24
NIH Grantees BIO	524	2	8.00	16	25.23	35	213	26.78	2.48
Olympic Fencing Medalists FRA	137	2	2.00	3	4.58	6	18	3.24	2.08
Olympic Swimming Medalists US	446	2	3.00	4	5.71	7	77	5.66	5.77
Olympic Swimming Medalists GBR	56	2	2.00	3	3.43	4	17	2.42	3.62
Tennis Top-tier Runs	4547	2	2.00	2	2.79	3	7	1.04	1.28
Tennis Challenger Runs	9457	2	2.00	2	2.83	3	6	1.03	1.00

Skewness values are positive in all cases, indicating right-skewed outcome distributions, with the magnitude of skewness varying across domains.

Descriptive statistics

Table 1 reports descriptive statistics for the eight data sets we analyze: Fighter Aces LW ($N = 1148$), NSF Grantees CS ($N = 1323$), NIH Grantees BIO ($N = 524$), Olympic Fencing Medalists FRA ($N = 137$), Olympic Swimming Medalists US ($N = 446$), Olympic Swimming Medalists GBR ($N = 56$), Tennis Top-tier Runs ($N = 4547$), and Tennis Challenger Runs ($N = 9457$). Across all eight settings, skewness is positive, indicating systematically right-skewed outcome distributions. In most domains the means exceed the medians by a substantial margin, consistent with heavy upper tails; in the tennis runs the gap is more modest, reflecting the fixed ceiling on run length imposed by single-elimination play (each tournament offers only a small number of consecutive wins).

The two Olympic swimming teams (US and GBR) share similar medians for medal scores but differ sharply in dispersion: the US team has a higher mean, a markedly larger upper quartile, and a much larger maximum total, indicating a substantially heavier upper tail despite the same minimum score threshold. The fencing and grant data likewise exhibit pronounced right skew, with means well above medians and wide ranges. The tennis data, which summarize lengths of winning runs in top-tier and Challenger events, show right-skewed but relatively thinner tails, with most runs concentrated at 2–3 wins. The Challenger runs, used as a lower-selectivity comparison, have the smallest skewness in the panel, consistent with their role as a relatively thin-tailed reference case. Overall, these descriptive patterns confirm that our outcome measures are consistently right-skewed, with tail thickness varying across domains.

Exploratory cross-domain comparisons

We investigate separately system-total and generational/era-bounded comparisons.

System-total QQ comparisons. We first use QQ plots as an exploratory tool to compare the full system-total distributions across domains. Figure 1 juxtaposes all pairwise combinations of Fighter Aces LW (LW), NSF Grantees CS (CS), NIH Grantees BIO (BIO), Olympic Swimming Medalists US (USOT_S), and Olympic Fencing Medalists FRA (FRAOT_F). In the lower triangle, QQ plots for each domain pair track their reference lines closely over most of the range, and the upper triangle shows Pearson correlation coefficients of quantiles that are all greater than 0.8. These patterns suggest that, at the system-total level, the domains share broadly similar right-skewed shapes, despite their very different institutional settings.

We emphasize that these QQ comparisons are descriptive: they reveal strong cross-domain concordance in the overall distributional form, but they do not by themselves identify a specific tail family. They motivate the more formal tail modeling that follows.

Generational and era-bounded QQ comparisons. Figure 2 extends the QQ analysis from system totals to entry cohorts and era-bounded windows within each domain. We compare three LW entry cohorts (1940, 1941, 1942), three NIH biology (BIO) and three NSF computing (CS) grant cohorts (1996–2000, 2001–2005, 2006–2010), and four Olympic eras for swimmers and fencers (1896–1912, 1920–1936, 1948–1976, 1980–2016), yielding 17 generational or era-bounded outcome distributions in total. The resulting 17 × 17 matrix contains 136 distinct off-diagonal QQ cells in the lower triangle. Of these, 127 have correlation coefficients above 0.8; the remaining nine lower cells (0.668–0.791) occur primarily in comparisons that involve the smallest dataset, French fencing, once it is split into eras.

Taken together, these generational QQ comparisons suggest that the strong cross-domain distributional similarity observed at the system-total level persists across entry cohorts and historical eras within domains, again at an exploratory and descriptive level rather than as a formal model test.

Tail-frontier model comparisons for system totals

The exploratory QQ results suggest common right-skewed shapes, but they cannot adjudicate among tail regimes. We therefore turn to the retained-fraction frontier analysis described in the Methods Section, which compares three discrete candidates for the upper tail: a discrete lognormal (dLN), a discrete power law (Zipf), and a shifted geometric (Geom). For each system total we scan retained fractions $f \in \{0.20, 0.25, \ldots, 0.80\}$; at each f we choose the smallest integer $k_{\min}(f)$ that retains approximately the top $100f$% of observations, fit all three models on that tail, and obtain Kolmogorov–Smirnov (KS) bootstrap p-values that respect discreteness. We require at least $N_{\text{kept}} \geq 40$ retained observations for stability, and we define the frontier $f^{\star}$ as the largest scanned fraction where the dLN is adequate (KS–bootstrap $p_{dLN} \geq 0.05$) under this size constraint.

Table 2 A summarizes the frontier results for the five primary system totals (LW, BIO, CS, U.S. swimmers, French fencers), and Fig. 3 shows the corresponding frontier scans and dLN fits at $f^{\star}$. In four of the five domains (LW, BIO, CS, FRA) the dLN is adequate at a broad frontier $f^{\star} = 0.80$, while the U.S. swimmers select a somewhat narrower frontier at $f^{\star} = 0.45$. In all five domains the discrete lognormal passes the KS–bootstrap adequacy test at a substantive frontier, and in every case it is the frontier winner at $f^{\star}$ under our rule that models must first be adequate (KS–bootstrap $p \geq 0.05$), and among adequate candidates we select the model with the lowest AIC.

- **Fighter Aces LW**. At $f^{\star} = 0.80$, the tail begins at $k^{\star}_{\min} = 8$ victories with $N^{\star}_{kept} = 970$ pilots. The dLN is adequate ($p_{dLN} \approx 0.124$), whereas Zipf and Geom both have very small p-values ($p_{\text{Zipf}} \approx p_{\text{Geom}} \approx 0.005$) and fail adequacy.
- **NIH Grantees BIO**. BIO totals also select $f^{\star} = 0.80$ (cutoff $k^{\star}_{\min} = 6$, $N^{\star}_{kept} = 432$). The dLN is adequate ($p_{dLN} \approx 0.279$), while Zipf ($p_{\text{Zipf}} \approx 0.005$) and Geom ($p_{\text{Geom}} \approx 0.025$) fall below the adequacy threshold.
- **NSF Grantees CS**. For CS totals, $f^{\star} = 0.80$ with $k^{\star}_{\min} = 5$ and $N^{\star}_{kept} = 1071$; the dLN achieves a high bootstrap p-value

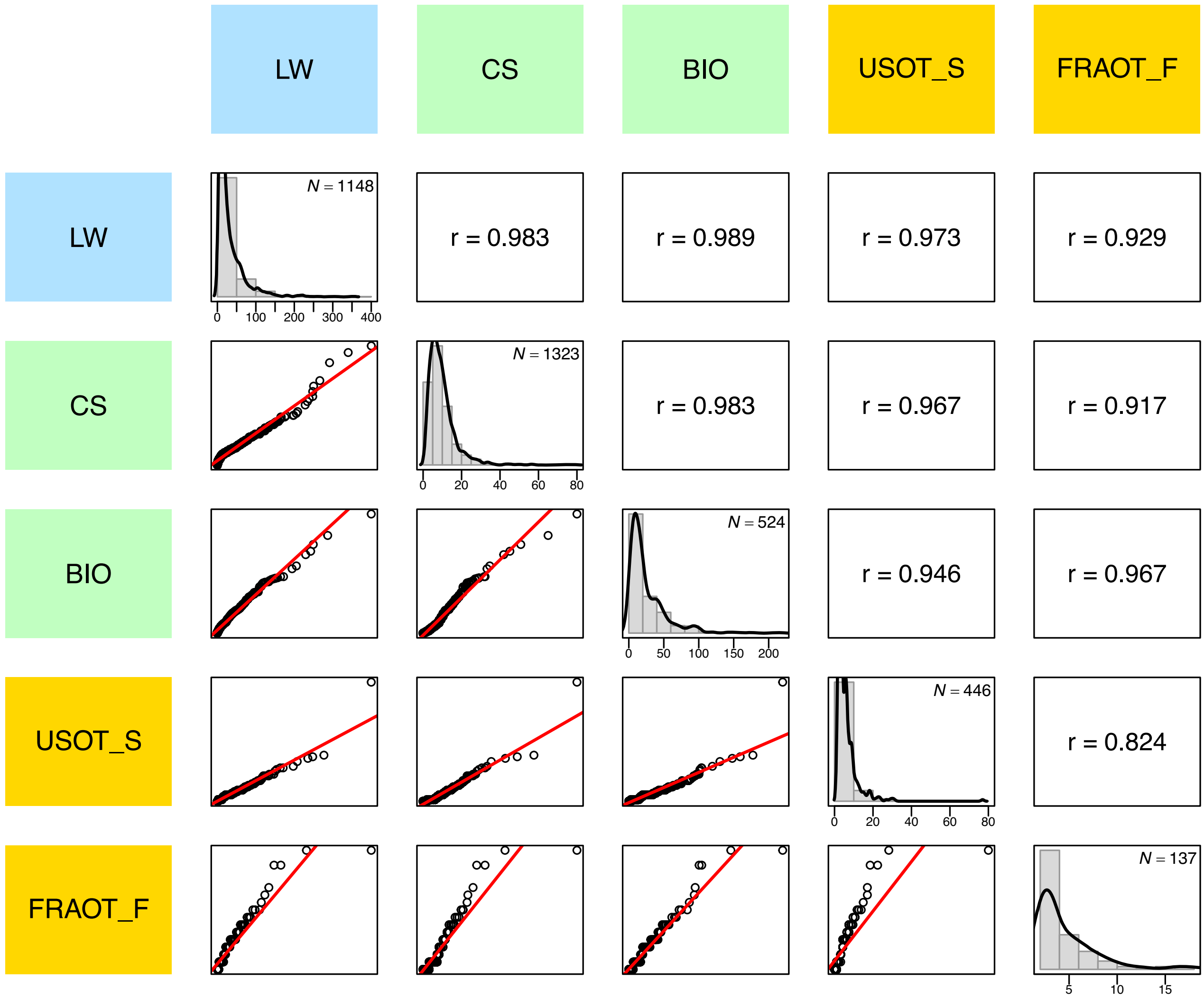


Fig. 1 | Cross-domain QQ comparisons of system-total outcome distributions. Quantile-quantile (QQ) plots compare all pairs of Fighter Aces LW (LW), NSF Grantees CS (CS), NIH Grantees BIO (BIO), U.S. Olympic swimming medalists (USOT_S), and French Olympic fencing medalists (FRAOT_F). The lower triangle shows QQ plots with reference lines; points track the lines closely over most of the range, indicating broadly similar right-skewed distributional shapes across domains. The upper triangle reports Pearson correlations of matched quantiles, with $r > 0.8$ for all pairs, consistent with strong cross-domain concordance at the system-total level.

($p_{dLN} \approx 0.433$). Zipf is clearly inadequate ($p_{\mathrm{Zipf}} \approx 0.005$), whereas Geom is marginally adequate ($p_{\mathrm{Geom}} \approx 0.055$); among adequate candidates, the dLN has the lowest AIC and is therefore selected as the frontier winner.

- **Olympic Swimming Medalists US**. The U.S. swimmers exhibit a somewhat narrower adequate tail, with $f^{\star} = 0.45$ (cutoff $k^{\star}_{\min} = 4$, $N^{\star}_{kept} = 223$). Here the dLN is adequate ($p_{dLN} \approx 0.129$), whereas Zipf again has a very small p-value ($p_{\mathrm{Zipf}} \approx 0.005$) and Geom falls just below our adequacy threshold ($p_{\mathrm{Geom}} \approx 0.045$), making the dLN the only adequate candidate at $f^{\star}$.
- **Olympic Fencing Medalists FRA**. For French fencers, $f^{\star} = 0.80$ (cutoff $k^{\star}_{\min} = 2$, $N^{\star}_{kept} = 137$); the dLN is again adequate ($p_{dLN} \approx 0.224$) and has both the highest bootstrap p-value and the lowest AIC among adequate candidates. Geom is also adequate ($p_{\mathrm{Geom}} \approx 0.134$), whereas Zipf falls just short of adequacy ($p_{\mathrm{Zipf}} \approx 0.045$) at this frontier.

Across these heterogeneous systems, the system-total tails thus share a common pattern: at a nontrivial retained fraction $f^{\star}$, the discrete lognormal provides the best and often the only adequate description of the upper tail, whereas a Zipf tail (strong cumulative advantage) and a shifted geometric tail (near-memoryless chance) are never the frontier winners for these system totals.

Tail-frontier comparisons within cohorts and eras

We next apply the same frontier procedure to within-domain cohorts and period windows: LW pilot entry cohorts by first victory year, BIO/CS grant cohorts by first award period, and era-bounded windows for Olympic swimmers and fencers. Table 2B reports the retained-fraction frontier $f^{\star}$, the corresponding cutoff $k^{\star}_{\min}$, the tail size $N^{\star}_{kept}$, and the KS–bootstrap p-values for dLN, Zipf, and Geom in each cohort/window with $N^{\star}_{kept} \geq 40$. Three broad patterns emerge.

1. **Stability of dLN dominance under Relative-Fairness**. In the majority of LW, BIO, CS, and Olympic windows the discrete lognormal remains both adequate and dominant at $f^{\star}$, with $f^{\star}$ typically at or near 0.80 in BIO and CS clusters and in most LW and US swimming periods. This stability indicates that the dLN tail is not an artifact of pooling across generations: when we re-partition each system into cohorts that share institutional and vintage conditions, the same heavy-but-subpower-law tail form persists wherever competition remains demanding and opportunities are repeated.
2. **Zipf episodes under strong head concentration**. A small number of windows select Zipf at the frontier. For LW, the 1940 entry cohort chooses Zipf rather than dLN at $f^{\star}$, consistent with historical accounts of early-war asymmetries where a pre-war cadre of highly trained pilots confronted less prepared opponents. For U.S.

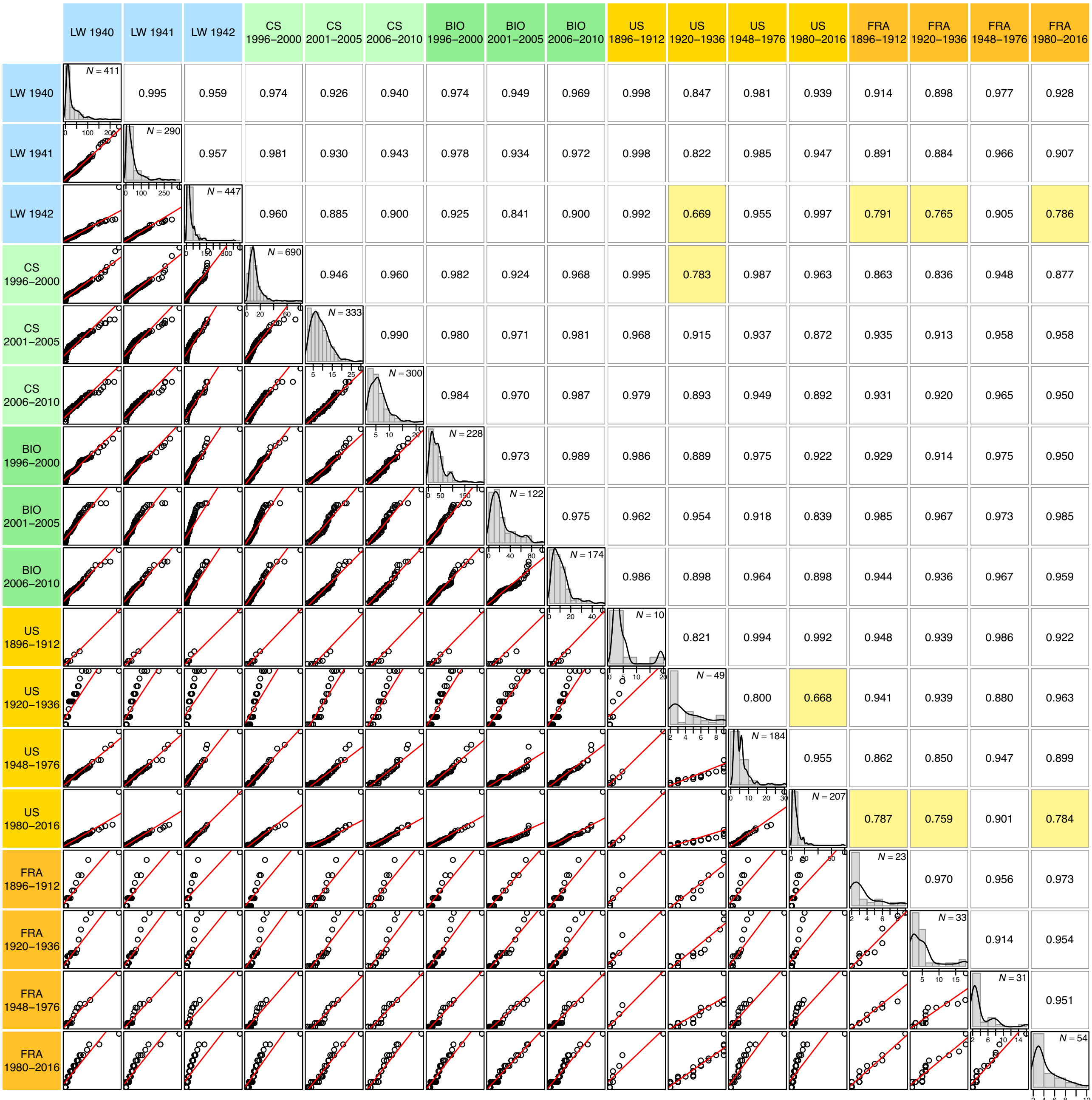


Fig. 2 | Generational and era-bounded QQ comparisons across domains. Each axis indexes one of 17 generational or era-bounded outcome distributions drawn from Fighter Aces LW [LW], NSF Grantees CS [CS], NIH Grantees BIO [BIO], Olympic Swimming Medalists US [USOT_S], and Olympic Fencing Medalists FRA [FRAOT_F]. These include three LW entry cohorts (1940–1942), three CS and three BIO grant cohorts (1996–2000, 2001–2005, 2006–2010), and four Olympic eras for swimmers and fencers (1896–1912, 1920–1936, 1948–1976, 1980–2016). The lower triangle shows QQ plots for all domain--generation pairs, with red reference lines. In most cases the points track the reference lines closely, indicating broadly similar right-skewed distributional shapes across cohorts and eras. The upper triangle reports Pearson correlations between QQ-plot quantiles. Nearly all correlations satisfy $r > 0.8$; only 9 out of 136 cells fall below this threshold (0.668–0.791, yellow background), mainly in comparisons involving the smallest dataset, French fencing.

swimmers, the 1980–2016 window also selects Zipf, reflecting the unusually strong concentration of medals in a small number of multi-Olympiad superstars (most notably Michael Phelps).

3. **Geom episodes under broad access/saturation**. The BIO 1996–2000 grant cohort is distinctive: at its frontier the shifted geometric, not the dLN, provides the best fit, with both models adequate but Geom achieving the lowest AIC value. This cohort coincides with the NIH budget doubling and an unusually permissive funding environment in which many investigators obtained grants. The resulting thinner, near-memoryless tail is consistent with our chance-dominated regime.

Taken together, the cohort and period analyses reinforce the system-total results: under conditions that approximate Relative-Fairness (strong selection, repeated opportunities, no overwhelming structural advantage), dLN tails dominate; when conditions shift toward either extreme head concentration or Broad-Access saturation, the selected tail regime changes in theoretically interpretable ways.

Table 2 | Frontier scan summary

A								
Domain	**N**	**f^***	**$k_{min}(f^*)$**	**$N_{kept}(f^*)$**	**p_{dLN}**	**p_{Zipf}**	**p_{Geom}**	**Winner at f^***
LW	1148	0.80	8	970	0.124	0.005	0.005	dLN
BIO	524	0.80	6	432	0.279	0.005	0.025	dLN
CS	1323	0.80	5	1071	0.433	0.005	0.055	dLN
USOT_S	446	0.45	4	223	0.129	0.005	0.045	dLN
FRAOT_F	137	0.80	2	137	0.224	0.045	0.134	dLN

B									
Domain	**Period**	**N**	**f^***	**$k_{min}(f^*)$**	**$N_{kept}(f^*)$**	**p_{dLN}**	**p_{Zipf}**	**p_{Geom}**	**Winner at f^***
LW	1940	411	0.80	8	339	0.110	0.438	0.005	Zipf
LW	1941	290	0.80	9	240	0.522	0.010	0.020	dLN
LW	1942	447	0.80	8	379	0.771	0.010	0.005	dLN
BIO	1996–2000	228	0.80	12	187	0.433	0.005	0.503	Geom
BIO	2001–2005	122	0.80	7	101	0.463	0.085	0.249	dLN
BIO	2006–2010	174	0.80	4	146	0.746	0.025	0.338	dLN
CS	1996–2000	690	0.80	6	562	0.353	0.005	0.005	dLN
CS	2001–2005	333	0.80	5	267	0.711	0.005	0.025	dLN
CS	2006–2010	300	0.80	3	272	0.667	0.005	0.005	dLN
USOT_S	1896–1912	10	NA	NA	NA	NA	NA	NA	NA
USOT_S	1920–1936	47	0.80	2	47	0.702	0.572	0.692	dLN
USOT_S	1948–1976	184	0.45	4	90	0.134	0.005	0.020	dLN
USOT_S	1980–2016	205	0.80	3	181	0.055	0.065	0.005	Zipf
FRAOT_F	1896–1912	23	NA	NA	NA	NA	NA	NA	NA
FRAOT_F	1920–1936	33	NA	NA	NA	NA	NA	NA	NA
FRAOT_F	1948–1976	31	NA	NA	NA	NA	NA	NA	NA
FRAOT_F	1980–2016	50	0.80	3	42	0.652	0.726	0.309	Zipf

A Domain-level totals. **B** Domain-period windows. For each panel, we report N, the selected frontier f^* (largest retained fraction where dLN is adequate: KS–bootstrap $p \geq 0.05$ with $N_{kept} \geq 40$), the cutoff $k_{min}(f^*)$, the tail size N_{kept}, and KS-bootstrap p-values for dLN, Zipf, and Geom at f^*. Entries shown as NA indicate that no candidate f met the adequacy rule $N_{kept} \geq 40$ and $p_{dLN} \geq 0.05$; in our tables these are predominantly due to insufficient tail size ($N_{kept} < 40$). The "Winner at f^*" column reports, for each row, the adequate model with the lowest AIC at the frontier.

Scale and regime stress tests

To probe the robustness and scope of our framework, we examine three complementary stress tests. Each test perturbs a different aspect of the systems we study—roster size, competitive tier, or access regime over time—and asks whether our tail inferences are stable. First, we use Great Britain Olympic swimmers as a roster-size benchmark for the larger U.S. team. Second, we contrast ATP top-tier and Challenger tennis runs to examine how tail regimes shift across competitive levels. Third, we study the 1996–2000 biology (BIO) cohort of NIH-funded faculty as a high-access, near-null case. Together, these three stress tests clarify when dLN tails persist and when thinner or thicker regimes emerge.

Roster size and mechanical scaling: U.S. vs. GBR swimmers. The Olympic swimming teams of the United States and Great Britain provide a joint stress test of roster size and mechanical scaling. Both programs operate in a highly selective environment, fielding athletes who have already passed multiple performance filters, and both show strongly right-skewed medal distributions (Table 1). In our system-total data, the U.S. team includes 446 swimmers with at least two medal points, whereas the GBR team includes 56 such swimmers. Typical medal totals are modest in both teams (medians of 4 and 3 points for the U.S. and GBR, respectively), and both distributions are clearly right-skewed, but the U.S. distribution exhibits a broader and more extreme upper tail: its third quartile, maximum, standard deviation, and skewness are all larger than those of GBR. Thus the two teams are comparable in competitive level and qualitative shape, but differ in roster size and in the extent of the upper tail. We first ask whether a smaller elite roster on its own supports a dLN tail, and then use the GBR roster size and frontier as a mechanical scaling control by downsampling the U.S. team to the same roster size and tail window.

First, at the system-total level, the GBR medal counts alone support a dLN-dominated tail at a substantive frontier: for a retained fraction with $N_{kept}(f^*) \geq 40$, the dLN provides an adequate KS–bootstrap fit and outperforms both Zipf and Geom in AIC (Fig. 4, panels A–B). This shows that an elite national team with a relatively small roster still exhibits a dLN tail regime under repeated Olympic opportunities.

Second, we use the GBR roster size and frontier as a mechanical scaling control for the U.S. team. To keep the comparison controlled, we first identify the tail cutoff $k^{\star}_{min}$ from the GBR frontier and then hold this cutoff fixed. For each resample, we randomly draw 56 U.S. swimmers (without replacement), apply the GBR cutoff $k^{\star}_{min}$, and fit dLN, Zipf, and Geom to the resulting U.S. tail. Fixing the cutoff in this way avoids a moving target: if we re-estimated k_{min} separately for every small U.S. subsample, the frontier would fluctuate across resamples, confounding regime assessment with noise in the cutoff choice. Panels C and D of Fig. 4 summarize these size-matched fits using AIC differences. For each U.S. resample we compute

$$\Delta AIC_{Zipf-dLN} = AIC_{Zipf} - AIC_{dLN} \tag{1}$$

$$\Delta AIC_{Geom-dLN} = AIC_{Geom} - AIC_{dLN}. \tag{2}$$

Positive values mean that dLN has lower AIC (better fit) than the comparison model. The histograms in panels C and D are concentrated on positive values, indicating that, for a substantial majority of size-matched

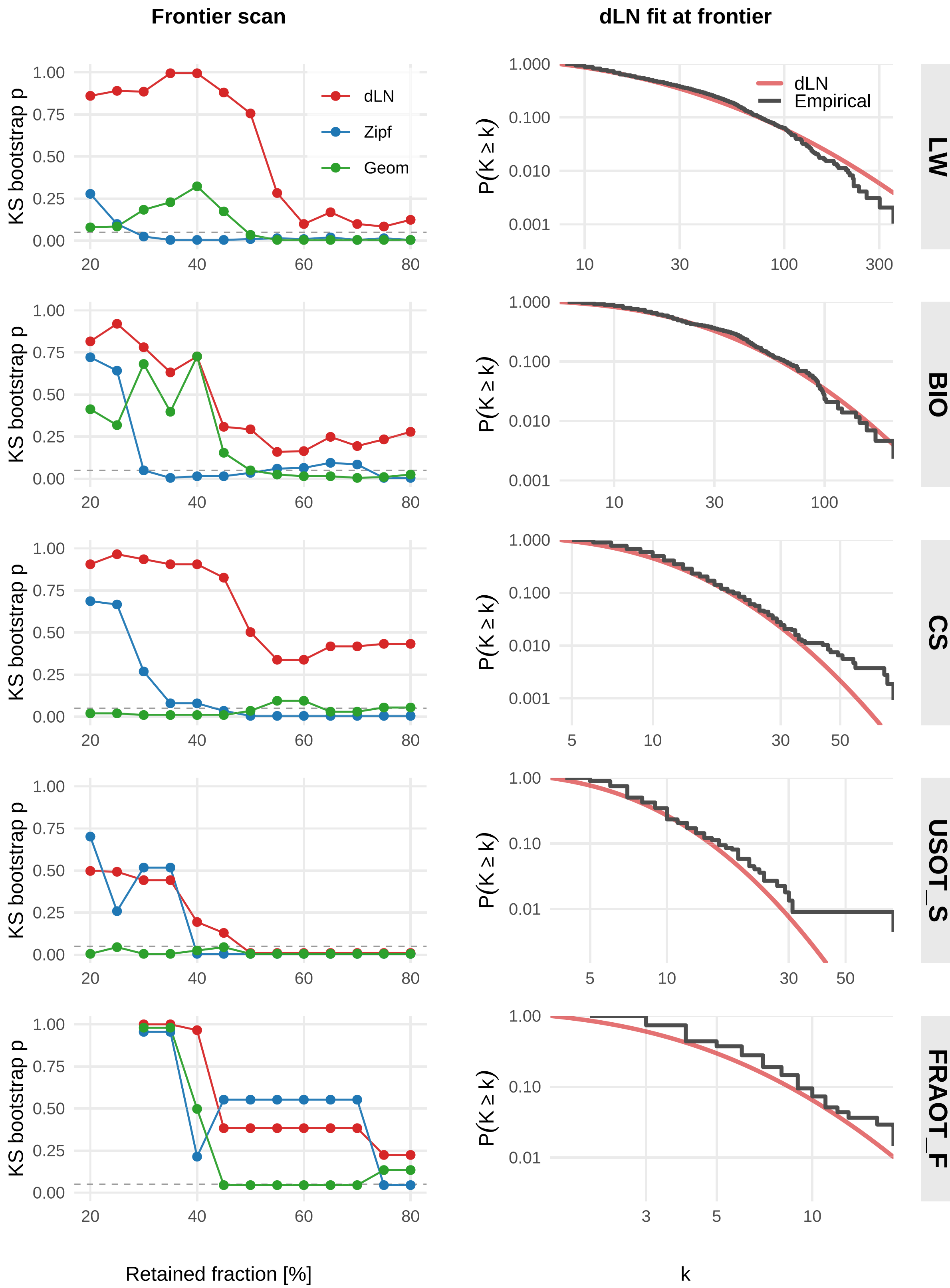


Fig. 3 | Frontier scans and fits by domain (*totals*). Left: KS-bootstrap p-values for dLN (red), Zipf (blue), and Geom (green) as the retained fraction f increases; dashed line at $p = 0.05$. The frontier f^* is the largest f where dLN is adequate with $N_{\text{kept}} \geq 40$. **Right:** empirical complementary cumulative distribution function/CCDF (black) and dLN fit (red) at f^*. Across domains, dLN predominates, consistent with a multiplicative--skill interpretation.

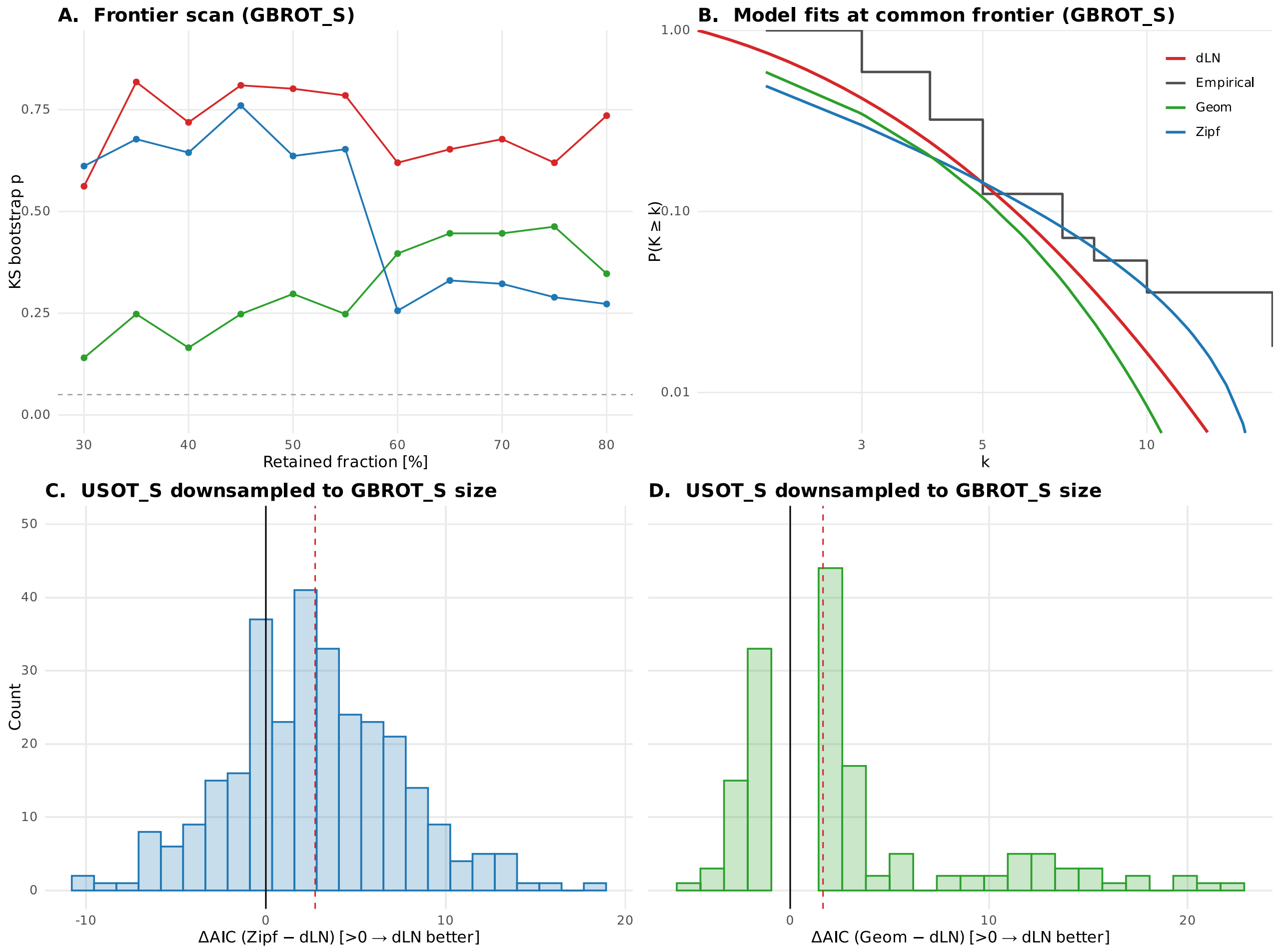


Fig. 4 | Roster-size benchmark and mechanical scaling control for Olympic swimmers. **A** KS-bootstrap p-values for dLN, Zipf, and Geom fits to GBR swimmers (56 athletes with at least two medal points) across retained fractions f; the frontier cutoff $k^{\star}_{\min}$ (with $N_{\text{kept}}(f^{\star}) \geq 40$) marks a tail region where dLN is adequate and AIC-preferred. **B** Complementary cumulative distribution of GBR medal totals with dLN, Zipf, and Geom fits at this cutoff. **C**, **D** Histograms of AIC differences from repeated downsampling of the U.S. swimmers (446 athletes with at least two medal points) to the GBR roster size, using the same cutoff $k^{\star}_{\min}$: $\Delta AIC_{Zipf-dLN} = AIC_{Zipf} - AIC_{dLN}$ in (**C**) and $\Delta AIC_{Geom-dLN} = AIC_{Geom} - AIC_{dLN}$ in (**D**). Positive values indicate that dLN has lower AIC than the comparison model; their predominance shows that, even under size- and cutoff-matched conditions, the dLN tail regime remains preferred for both national teams.

resamples, dLN beats Zipf and Geom when the U.S. team is evaluated at the GBR frontier. In other words, even when we constrain the larger U.S. system to a GBR-sized roster and to the GBR-defined tail window, the dLN regime remains dominant.

Taken together, these results support our broader claim that in selective, high-quality competitive environments with repeated opportunities, the upper tails of performance concentrate around a dLN regime. The observed dLN tails for Olympic swimmers are not artifacts of the larger support range or roster size of the U.S. team, but reflect stable accumulation dynamics shared across elite national programs.

Competition quality: top-tier vs. Challenger tennis. Professional men's tennis offers a within-domain contrast in which the competition format is held fixed while competitive level and selectivity differ. Both the Association of Tennis Professionals (ATP) Tour (top-tier) and the ATP Challenger Tour (developmental circuit one tier below) use single-elimination draws, but the former concentrates stronger players and tighter entry filters. For each match, we record the length of a player's winning run (consecutive wins within an event) and aggregate these into the Tennis Top-tier Runs and Tennis Challenger Runs distributions.

Descriptively, both strata are modestly right-skewed (Table 1). Most runs are short (2–3 wins), and the upper limits are moderate (7 wins in top-tier events, 6 in Challenger), with skewness smaller in the Challenger tier. This suggests that both levels operate in a relatively thin-tailed regime compared to our other domains, with heavier tails at the very top of the ATP Tour.

Frontier scans make the contrast more precise (Fig. 5). In top-tier events, the dLN remains adequate over a broad range of retained fractions and yields a frontier at $f^{\star} \approx 0.45$ with $n_{\text{kept}} \approx 2125$ runs; Zipf and Geom are rejected across this range. In Challenger events, dLN is still the winning model, but its adequate region recedes to a much smaller frontier at $f^{\star} \approx 0.20$ with a comparable number of retained runs; again, Zipf and Geom are rejected at $f^{\star}$. Thus, as we move from the highest level of the professional circuit to a second-tier developmental circuit, while holding the competition format constant, the system remains in the dLN regime but the admissible frontier shrinks.

This graded pattern is consistent with our interpretation of competition quality. In a more selective, higher-intensity environment (ATP Tour), heavy-but-subpower-law tails extend over a larger portion of the distribution. In a somewhat less selective, developmental environment (Challenger), the same dLN regime applies, but only over a narrower upper slice of outcomes.

Saturation and a near-null regime: BIO 1996–2000. The BIO 1996–2000 cohort provides our clearest empirical glimpse of a high-access, near-null regime within the main corpus. In the tennis analyses

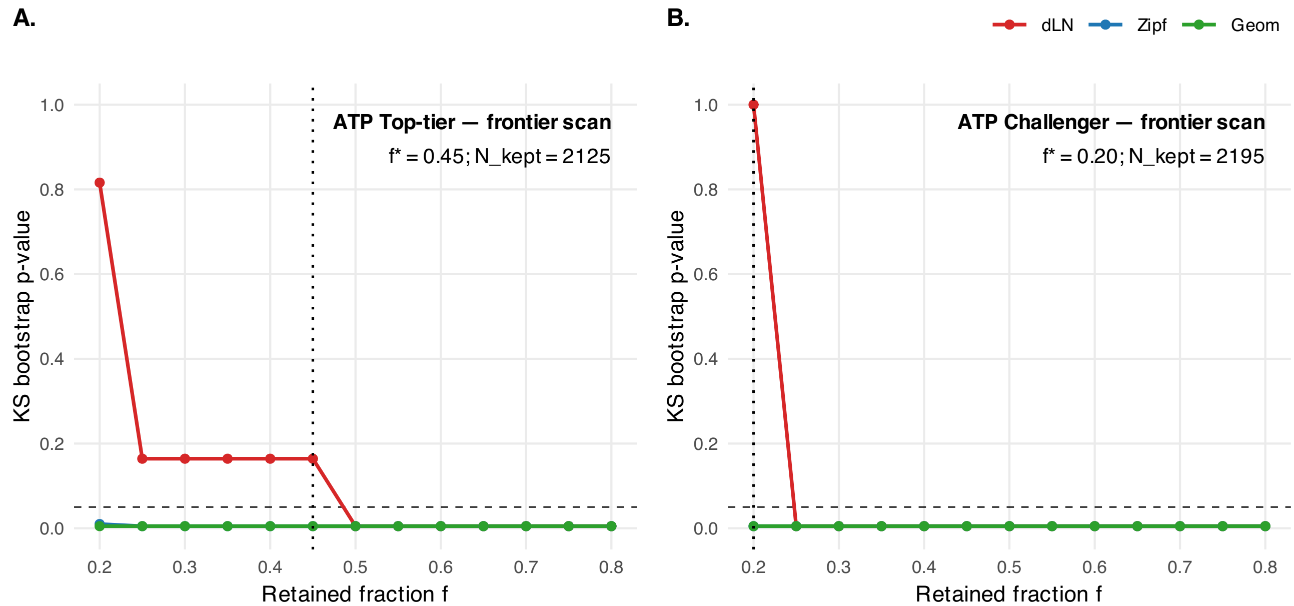


Fig. 5 | Frontier scans for winning-run lengths in professional men's tennis. Each panel shows KS--bootstrap p-values for dLN, Zipf, and Geom fits as a function of retained fraction f, with the selected frontier f^* marked by a vertical dotted line. **A** Top-tier stratum (Grand Slam tournaments and ATP Tour events: ATP 250, ATP 500, and ATP Masters 1000); dLN remains adequate over a broad range of f, and Zipf and Geom are rejected at the frontier. **B** Challenger stratum (ATP Challenger Tour events); dLN is still the preferred model, but its adequate region is confined to smaller f, and Zipf and Geom again fail the KS--bootstrap test at f^*. Together, the scans show that both tiers inhabit a dLN regime, with the admissible tail range shrinking as competition quality is relaxed.

above, we saw that moving from the ATP Top-tier to the Challenger circuit—while holding the competition format fixed—keeps the system in a dLN regime but compresses the admissible frontier as competition quality and selectivity are relaxed. The biology (BIO) cohort of NIH-funded faculty around 2000 represents a much more dramatic shift. During this period the NIH budget approximately doubled[32], sharply increasing available funds and temporarily relaxing competition for major awards. This expansion broadened access: many more investigators were able to clear the funding threshold, and the marginal difficulty of securing additional awards was substantially reduced, weakening success-dependent gating in a way that can approach a chance-dominated (near-memoryless) limit.

At the retained-fraction frontier, this change in selectivity is reflected in a shift toward the geometric regime. For BIO 1996–2000, the frontier scan identifies $f^* = 0.80$ with cutoff $k_{\min}(f^\star) = 12$ and $N_{\text{kept}} = 187$ investigators (Table 2). Both dLN and Geom provide adequate KS–bootstrap fits at this frontier ($p_{dLN} = 0.433$, $p_{\text{Geom}} = 0.503$), but, according to our selection rule, the shifted geometric attains the lowest AIC and is therefore the winner at f^*. In contrast, all other BIO and CS period windows in Table 2 select dLN as the frontier winner, with Zipf and Geom either rejected or clearly disfavored in AIC. Within our three-model family, BIO 1996–2000 is thus the only major grant cohort whose upper tail is best described by the thinnest candidate, Geom, rather than by dLN.

We interpret this as the broad-access route to a near-null regime (Geom) within our framework: broad access can weaken success-dependent gating enough to approach an approximately memoryless, chance-dominated limit. When budget expansion makes it much easier for investigators to obtain repeated awards, individual trajectories become closer to a sequence of nearly independent trials with elevated success probability, and the tail is well captured by a shifted geometric distribution. As competition tightens again in later periods, the system returns to the dLN regime (BIO 2001–2005 and 2006–2010), in line with our broader claim that strongly selective, repeated competitions tend to generate heavy-but-subpower-law tails, whereas pronounced reductions in selectivity push outcomes toward a geometric, near-null limit.

Relative-Fairness across systems

Our results so far suggest that discrete lognormal tails arise robustly when competition is demanding and repeated, but not fully dominated by structural advantages or institutional instability. We refer to this regime as one of *Relative-Fairness*: entry is filtered and opportunities are scarce, yet skill still has a fighting chance to accumulate over many trials. Throughout the paper we therefore focus on systems that are both honestly competitive and, over the observation window, reasonably stable in their institutional conditions. Episodes where competition is disrupted by multifactorial collapse—for example, by simultaneous shocks to training, resources, and operational control—are treated as boundary cases rather than as core evidence, and we do not attempt to model them with our tail-frontier procedure. Here we probe how much latitude the Relative-Fairness regime can accommodate, using three complementary lenses: global competition in Olympic swimming, institutional concentration in science funding (including a transient budget shock), and a boundary case in late–World War II aerial combat that we discuss qualitatively.

Global benchmarks: U.S. swimmers vs. the rest of the world. The first question is whether dLN tails for U.S. Olympic swimmers are compatible with a genuinely demanding competitive field. Figure 6 plots the share of individual-event medals captured by the United States versus the rest of the world by Olympiad, separately for bronze, silver, and gold. The underlying medal records are drawn from the long-run Olympic results compiled in the Kaggle Olympics online repository[25]. Although U.S. swimmers are formidable when compared to any single nation, they almost never dominate the global medal pool. Across most Games the rest of the world secures the majority of medals at each level, and the U.S. share typically hovers between about 20% and 60%, with only a few Olympiads where it briefly approaches 70–80%. From the perspective of a U.S. athlete, the effective competition set is therefore "the world," not a handful of weaker programs. That both the U.S. and GBR system-total distributions support dLN tails under this sustained global pressure, based on the same underlying Olympic data[25], reinforces our interpretation of the dLN regime as compatible with tough but broadly fair

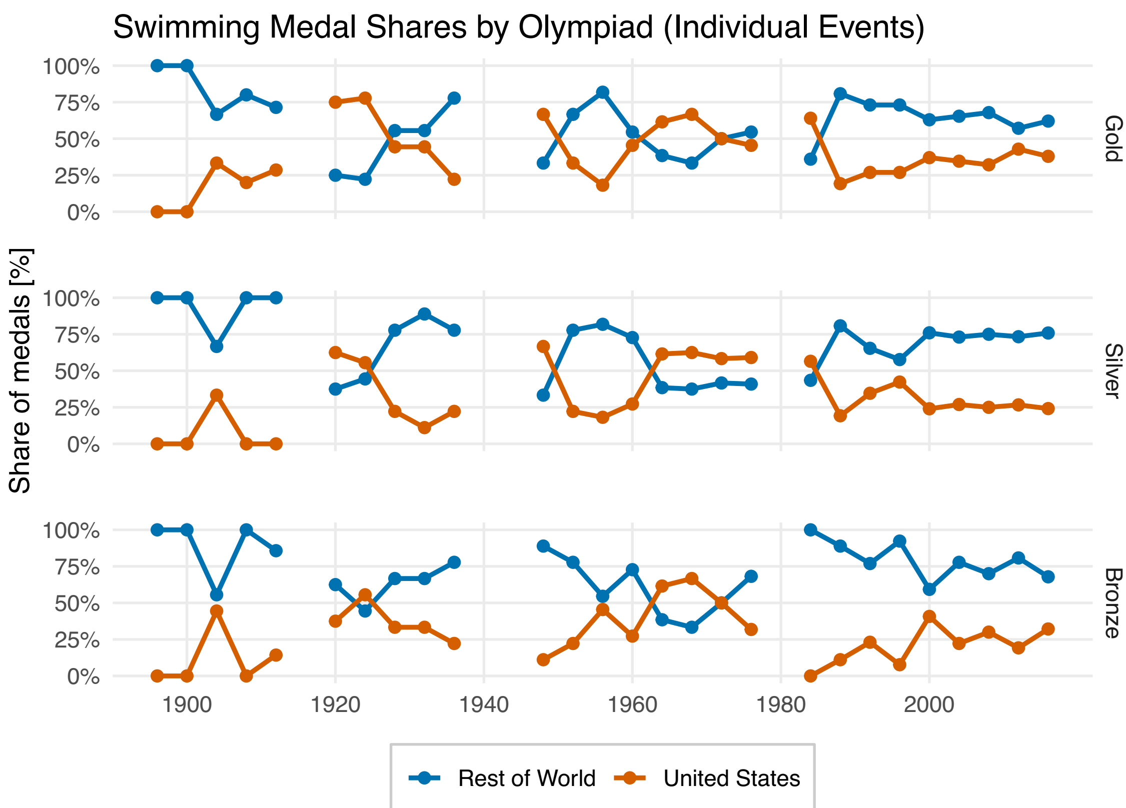


Fig. 6 | Swimming medal shares by Olympiad for individual events. Each panel shows the percentage of gold (top), silver (middle), and bronze (bottom) medals won by U.S. swimmers (red) and by all other countries combined (blue) over time. U.S. swimmers are consistently strong but rarely dominant: in most Games the rest of the world captures the majority of medals at each level. This supports our claim that the dLN tails observed for U.S. swimmers arise in a sustained, globally competitive field rather than in a weakly contested environment. Gaps in the series correspond to Olympiads that were canceled or boycotted (e.g., during World Wars I and II and the 1980 Moscow Games), for which no medals were awarded. Medal data are drawn from the Kaggle Olympics repository[25].

competition among highly selected entrants, rather than with a weakly contested or structurally rigged environment.

Shades of fairness: departmental concentration in BIO vs. CS. Relative-Fairness does not require a perfectly level playing field: structural advantages can be present as long as they do not completely overwhelm repeated competition. To probe this latitude, we examine how strongly awards are concentrated by institutional rank in NIH BIO and NSF CS funding. For these analyses we use the investigator-level grant histories assembled by Petersen et al.[22], which include rankings of the investigators' home departments in their respective fields. Figure 7 plots the cumulative share of awards as a function of department rank (1 = highest), along with Lorenz curves and Gini coefficients. Both fields exhibit substantial head concentration, but BIO is more unequal: the top 10 and top 20 departments capture a larger share of awards than their CS counterparts, and the BIO Gini coefficient (0.40) exceeds that of CS (0.37). This reflects structural differences such as higher fixed costs and infrastructure demands in wet-lab biology, which amplify the advantages of major research hubs.

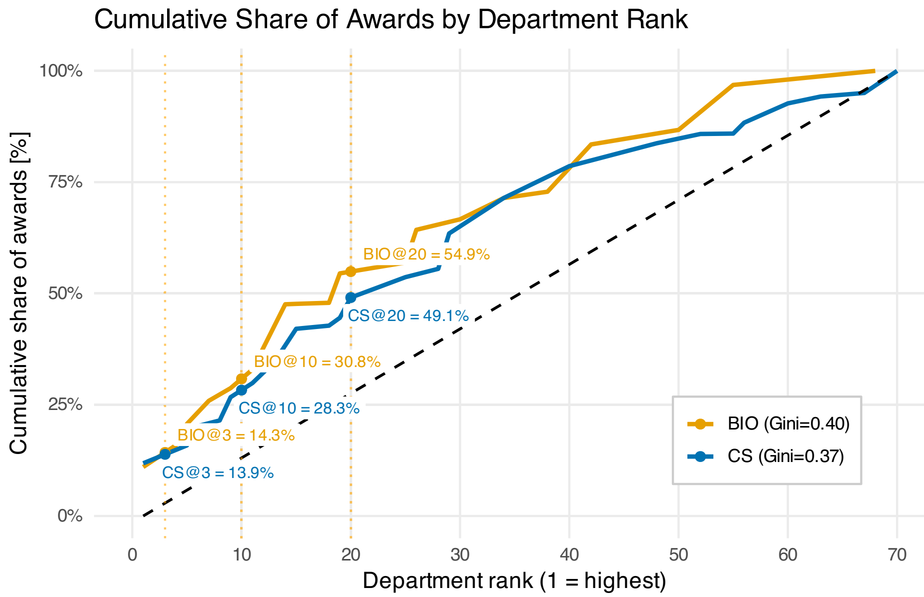
At the same time, the BIO concentration is far from absolute. Even in BIO, a substantial fraction of awards flows to mid-ranked and lower-ranked departments, and top institutions do not monopolize the system. Thus, being based at a major department confers a clear advantage, but not an overwhelming one: investigators at less prestigious institutions still have repeated opportunities to compete successfully. Consistent with this, the frontier scans in Table 2 show that dLN remains the dominant tail model across BIO and CS system totals and cohorts, apart from a well-identified transient disturbance in BIO around the turn of the century.

That disturbance is the doubling of the NIH budget circa 2000, documented by Park[32] and by Petersen et al.[22], which temporarily altered the effective supply of awards without changing the underlying structure of scientific competition. In the 1996–2000 BIO entry cohort, this single, system-wide budget shock produces a thinner, near-geometric tail: our frontier procedure selects Geom as the winning model under adequate KS–bootstrap support. As funding conditions stabilize after the expansion, the selected tail regime reverts to dLN in subsequent BIO cohorts (Table 2B). This pattern illustrates the kind of transient, largely mono-factor perturbation that our framework is designed to accommodate: the system experiences a temporary deviation from the dLN regime but returns to Relative-Fairness once conditions settle.

Boundary of Relative-Fairness: LW 1943–1944. Finally, we consider a case that lies at the boundary of our framework rather than within it. The LW pilots who entered combat in 1943 and 1944 fall outside our main analysis window, and we deliberately do not include these cohorts in our tail-frontier scans. As summarized in Fig. 8, historical sources emphasize that from 1943 onward the Luftwaffe experienced a sharp and sustained deterioration in pilot training hours[20], mounting operational strain as it fought on multiple fronts, escalating losses, and growing shortages of fuel and materiel[38]. Conditions changed from month to month without a realistic prospect of stabilization. Under such circumstances, survival and opportunity were driven increasingly by structural constraints and sheer attrition rather than by accumulated skill expressed over many sorties.

On these historical grounds, we treat the 1943–1944 cohorts as an example of a highly dynamic, multifactorially collapsing system–exactly the kind of setting that our Relative-Fairness framework is not designed to handle. Our focus in this paper is on systems where competition remains

Fig. 7 | Cumulative share of awards by department rank for NIH BIO and NSF CS, using the investigator-level data and departmental rankings from Petersen et al.[22]. Curves show the fraction of total awards captured by the top *k* departments as a function of *k*, with the dashed line indicating perfect equality. BIO exhibits stronger head concentration than CS (Gini 0.40 vs. 0.37), with top-ranked departments capturing a larger share of awards at each cutoff. Despite these differences in institutional inequality—and a transient budget shock in BIO around 2000[32]—both fields predominantly select dLN as the preferred tail model in our frontier scans (Table 2), illustrating the latitude of the Relative-Fairness regime.



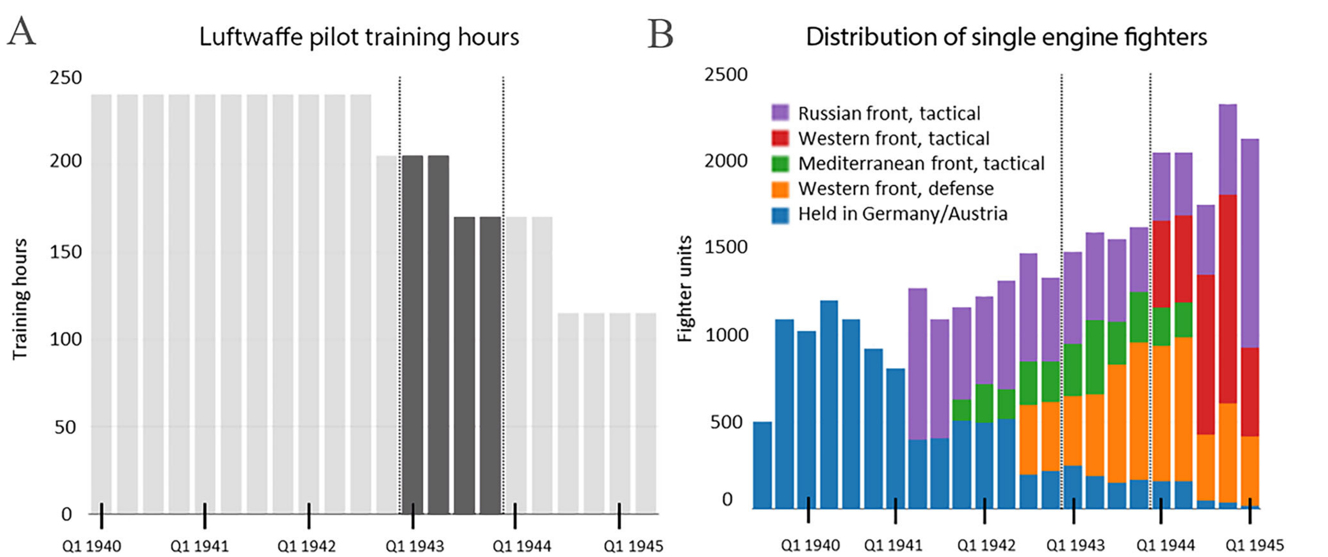


Fig. 8 | Evolution of training and operational strains in the Luftwaffe during World War II. **A** Luftwaffe pilot training hours over the course of World War II. The graph is based on data reported by Williamson Murray in *Strategy for Defeat*[20]. Training hours decline sharply in 1943, 1944, and 1945. **B** Distribution of Luftwaffe single-engine fighters across different fronts during World War II. The graph, based on data reported in the *United States Strategic Bombing Survey and Air Force Doctrine*[38], illustrates the growing operational strain from participating in multiple theaters of war after 1942.

honestly demanding and institutionally coherent over time, and where at most a transient, largely mono-factor disturbance (such as the NIH budget doubling) perturbs otherwise stable conditions. By contrast, the late-war LW environment involves permanent and multi-causal breakdown. We therefore view it as a qualitatively informative boundary case, but we leave formal tail modeling of such regimes to future work that can take their instability as a starting point rather than as a complication.

Taken together, these comparisons support a coherent picture. Under demanding but relatively fair and institutionally stable competition—from globally contested Olympic swimming to scientific grant races with varying degrees of institutional concentration—the upper tails of success typically concentrate around a discrete lognormal. Within this Relative-Fairness regime, our tail-frontier procedure not only recovers dLN as the prevailing pattern, but also detects transient, non-catastrophic departures to thinner (Geom) or heavier (Zipf) tails when conditions experience a temporary, largely mono-factor shock, such as the NIH budget doubling in BIO. As those conditions normalize, the inferred tails revert toward dLN. By design, however, we do not apply this framework to settings undergoing multifactorial collapse, such as the late-war LW cohorts, where the underlying competitive conditions change rapidly and permanently; such regimes fall outside the scope of the present study.

Discussion

Across three very different arenas—martial, intellectual, and athletic—we have shown that once competition is both selective and repeated, the upper tails of individual success tend to concentrate around a discrete lognormal regime. This pattern is not a trivial consequence of pooling across eras or of one particular domain. It appears in system totals for WWII Luftwaffe pilots, U.S. NIH/NSF grantees in biology and computer science, and U.S. and French Olympic medalists, and it reappears when we partition the same systems into exposure-aligned cohorts and era-bounded windows. Within the modeling family we study, a discrete lognormal tail is therefore not an incidental fit but a robust *signature* of how skill accumulates when entrants are already highly selected and enjoy adequate repeated opportunities under reasonably stable rules. The few cases where geometric or Zipf tails are selected arise in clearly identifiable institutional episodes, rather than as random anomalies.

Taken together, these findings address the three questions posed in the Introduction. First, across elite domains with repeated opportunities, dLN tails systematically outperform geometric and Zipf alternatives at the retained-fraction frontier, indicating that runaway cumulative advantage and near-memoryless luck are not the prevailing regimes (RQ1). Second, when we re-partition systems into entry cohorts and era-bounded windows, the same dLN regime persists except in periods with documented institutional shocks, showing that tail shape is largely stable but sensitive to identifiable changes in opportunity structure (RQ2). Third, episodes in which geometric or Zipf tails are selected—such as the NIH budget-doubling cohort in BIO or the historically exceptional dominance of early-war Luftwaffe pilots and modern swimming superstars—demonstrate that shifts in upper-tail shape provide a compact, falsifiable signal of movement toward luck-heavy or dominance-driven regimes (RQ3).

Our framework casts this signature as a regime of *Relative-Fairness*. The term is deliberately modest. We do not assume a level playing field, nor do we claim that success is proportional to merit in any simple sense. Instead, Relative-Fairness describes a conditional regime in which (i) entry is filtered so that participants start from a high baseline of preparation, (ii) outcomes are determined through many individual-level contests or evaluations, and (iii) no single structural advantage or shock persistently overwhelms the competitive process. Under these conditions, heterogeneous abilities and contexts are expressed through many partly independent chances to gain or lose ground, generating heavy but sub-power-law tails well captured by a discrete lognormal. In this sense, dLN tails mark systems in which skill has a fighting chance to accumulate without either collapsing into a single dominant winner or dissolving into near-memoryless luck.

The stress tests sharpen this interpretation by showing how tail regimes respond when we perturb opportunity structure. In professional tennis, we hold rules and scoring fixed while moving from the ATP Tour to the Challenger circuit. Both tiers retain dLN tails, but the adequate frontier shrinks in the lower-tier environment, consistent with a weaker expression of skill under looser selection. In NIH funding, the entry of the 1996–2000 BIO cohort coincides with the onset and unfolding of the NIH budget doubling, which provides a more dramatic intervention. For this cohort, the upper tail shifts toward the shifted geometric regime, our near-null for Broad-Access, low-selectivity competition: many more investigators clear the funding threshold, and the effective hazard of an additional award approaches a memoryless process. As the funding environment tightens again, later biology cohorts return to a dLN regime. On the other side of the spectrum, we see localized Zipf episodes in early-war Luftwaffe pilots and in modern U.S. swimmers and French fencers, where historically exceptional individuals accrue a very large share of victories or medals. Taken together, these contrasts show that tail shape tracks a continuum from chance-heavy to dominance-heavy regimes, with Relative-Fairness and its dLN tails occupying a broad, interior band rather than a fragile knife edge.

A key concern in any tail analysis is whether apparent regularities are driven by sample size or mechanical scaling. Our comparisons between U.S. and Great Britain Olympic swimmers address this directly. Great Britain fields a much smaller but still elite roster; its medal totals alone support a substantive dLN tail, indicating that lognormal-like behavior does not require very large systems. When we repeatedly downsample the larger U.S. team down to the British roster size and evaluate the tails at the British frontier, the discrete lognormal remains preferred in most resamples. This suggests that the dLN regime for Olympic swimming is not an artifact of a particularly deep U.S. bench or of a longer upper support, but reflects accumulation dynamics common to elite national programs facing a strong global field.

Relative-Fairness is also compatible with pronounced but finite structural inequalities. The grant data make this clear. NIH BIO awards are more heavily concentrated in top-ranked departments than NSF computing awards, reflecting the infrastructural advantages of large biomedical hubs. Yet both systems predominantly select dLN tails at their frontiers, apart from the transient budget-doubling episode in BIO. In other words, institutional inequality and geographic concentration do not automatically push a system into a runaway cumulative advantage regime. They shift the playing field and matter greatly for who gets to compete and under what conditions, but as long as a broad stratum of investigators repeatedly participates in high-stakes competitions, the aggregated outcomes can still inhabit a Relative-Fairness regime rather than a Zipf-like head dominated by a handful of institutions or individuals.

At the same time, our design draws a line around systems where the very idea of repeated, comparable opportunities breaks down. Late-war Luftwaffe cohorts (1943–1944) are the clearest example. Historical accounts emphasize concurrent collapses in pilot training, fuel availability, command coherence, and survival prospects, with conditions changing from month to month and no prospect of return to equilibrium. Under such multifactorial and effectively irreversible breakdown, survival and exposure cease to be interpretable as the outcome of a stable competitive process, and our tail-frontier framework is no longer appropriate. Rather than forcing a fit in such settings, we treat them as boundary cases that mark where Relative-Fairness—and with it the dLN regime—stops being a meaningful benchmark.

Beyond organizing existing data, these results have implications for how institutions might monitor and steer competitive systems. First, tail shape offers a compact, falsifiable summary of opportunity structure: a predominately geometric tail in a nominally elite competition is an early warning that outcomes are approaching a near-null regime in which repeated opportunities carry little informational content about skill, whereas a persistent Zipf regime can flag extreme concentration that may be driven more by structural lock-in than by performance differences alone. Second, changes in tail regime across cohorts or policy eras can help separate transient shocks from deeper institutional shifts. In our setting, the NIH budget doubling temporarily thins the tail but does not permanently alter the long-run regime once resources stabilize. Similar diagnostics could be applied to other funding systems, promotion schemes, or tournament structures as they undergo reforms.

Our study has several limitations that also point toward next steps. We focus on three high-resolution domains with unusually good individual-level records and relatively well-documented institutional settings. We do not model exposure directly: sorties, proposal submissions, and training loads remain latent, and thus enter only through their aggregate impact on tails. Our model family is deliberately minimal—discrete lognormal, Zipf, and shifted geometric—chosen to represent three stylized regimes rather than to exhaust all plausible tail forms. Other distributions or mixtures might better capture specific systems or intermediate cases, and micro-level mechanisms that combine learning, selection, and network effects could generate richer dynamics than those we infer from aggregate tails alone.

Future work can extend the present framework along several directions. On the empirical side, the frontier methodology could be applied to creative industries, online platforms, or other sports and funding schemes, probing whether Relative-Fairness and dLN tails appear wherever competition is selective, repeated, and institutionally coherent. On the theoretical side, agent-based or stochastic process models that explicitly couple heterogeneous skill, evolving opportunity sets, and institutional rules could clarify the conditions under which different tail regimes emerge and how they respond to targeted interventions. Finally, a richer treatment of fairness would connect tail diagnostics to questions of access and diversity: who reaches the starting line of these competitions, whose careers are prematurely truncated, and which institutional designs enlarge the region in which skill can accumulate without collapsing into either near-random churn or entrenched dominance.

In sum, our findings suggest that heavy but sub-power-law tails are a common outcome of honestly demanding, relatively fair competition among already-selected participants, and that departures toward thinner or heavier tails are informative about how institutions broaden, ration, or distort access to opportunity. Tail shape alone cannot adjudicate all questions of justice or merit, but it offers a concise, quantitative lens on the health of competitive systems and on the conditions under which exceptional human performance arises.

Methods

Construction of Datasets

In the following, we provide definitional and key design details about the study's multiple datasets.

Construction of the Luftwaffe fighter ace dataset. The term *flying ace* or *fighter ace* was coined by French newspapers during World War I and subsequently was used to refer to exceptional fighter pilots[19]. Here we define as an ace any Luftwaffe fighter pilot who scored at least five aerial victories during his participation in the war. This number is the threshold marking exceptional performance according to widely accepted norms of the period[21].

In our data, an *aerial victory* is defined as a *credited* enemy aircraft destroyed in air-to-air combat, as recorded in wartime rosters and claims in the historical sources. This is the only measure that is consistently available at pilot level across our observation window and matches the unit used in the monographs and archives that underpin the dataset. We do not observe sortie-level exposure, mission type, or task difficulty, and thus cannot adjust victories for mission severity; we acknowledge this as a limitation and focus on the *distributional form* of tallies in the upper tail.

We chose to study the victory records of Luftwaffe aces rather than Allied aces because the former typically participated in the war for longer, offering the opportunity for extended observations. Importantly, once Luftwaffe pilots entered frontline service, they generally remained in combat until they were captured, incapacitated, or killed[21]. By contrast, successful Allied pilots (more than five victories) were often promoted to non-fighting positions or reassigned as instructors to pilot schools[39]. These institutional practices likely reduced their combat mortality but also imposed a hard institutional ceiling on their combat exposure, preventing them from developing their full fighting potential in air-to-air engagements. In our framework, Luftwaffe aces thus provide a more uniform and extended exposure to repeated trials, subject to the natural limits of survival and the course of the war.

We constructed the Luftwaffe fighter ace dataset from two groups of sources: (a) monographs and (b) archives. We consulted six monographs dedicated to the Luftwaffe and its aces during World War II[21,30,40–43]. From these we compiled a list of all aces and recorded the number of victories credited to each in every monograph. In the event of full agreement among sources, we considered the reported number of victories as a potentially reliable figure. We then checked these figures against the corresponding entries in an online archival source[44], which served as a final validation step. Generally, the sources were either in full agreement or had relatively small disagreements. In the latter case, we followed a reconciliation process based on majority rule: we selected the number of victories supported by the majority of sources. Details of this reconciliation procedure are provided in the Supplementary Materials.

The resulting dataset of Luftwaffe aces includes the records of $N = 1148$ fighter pilots. Additionally, we constructed three generational datasets, each holding the records of a distinct entry generation of Luftwaffe aces: those who started their combat career in 1940, in 1941, or in 1942, respectively. We chose a yearly decomposition because this largely aligned with the conscription and buildup practices of the German armed forces at the time. We excluded the war years 1939 and 1945 because they were incomplete calendar years. We also excluded the years 1943 and 1944, when Luftwaffe pilot training deteriorated dramatically (Fig. 8A) and the overall balance of the air war shifted under heavy attrition and multi-front pressures (Fig. 8B). These institutional and operational disruptions violate a key requirement of our comparative design, namely that within an analysis window the rules, training standards, and deployment practices are broadly stable so that pilots face a reasonably consistent opportunity structure. The three entry cohorts for 1940, 1941, and 1942 therefore form a set of *candidate* analysis windows under broadly stable institutional conditions. Together they include the full ace dataset used in our main analysis, where the retained-fraction frontier is used to classify each cohort into Runaway, Relative-Fairness, or Broad-Access regimes.

Construction of the NSF/NIH grantee dataset. The grantee dataset includes National Science Foundation (NSF) and National Institutes of Health (NIH) grant records of computing and biology faculty, respectively, in U.S. research universities. Following Petersen et al. [22] we use the term "biology" for departments that in practice span core biomedical and life-science fields (e.g., molecular and cell biology, biochemistry, related programs); for brevity we refer to this NIH-funded domain as BIO. The analytic focus on NSF grant competitions for computing faculty and NIH grant competitions for biology/biomedical faculty reflects the fact that these agencies are the primary funders of the respective disciplines.

We constructed the grantee dataset starting from the list of computing and biology faculty reported in the article *Cross-disciplinary evolution of the genomics revolution* by Petersen et al. [22] That study identifies $N = 2077$ biology and $N = 2133$ computing faculty in 155 top-ranked U.S. biology and computing departments with PhD programs. Using these rosters as a sampling frame, we queried the NIH RePORTER repository[23] for NIH grants and the NSF Award Search database[24] for NSF grants, retrieving records for the listed biology and computing faculty, respectively. From these rosters, our analytic datasets retain investigators whose first grant award occurs no later than 2010 (to ensure follow-up through 2016) and who accumulated at least two grant awards over the 1996–2016 window. Applying these criteria yields $N = 524$ NIH-funded biology/biomedical investigators (BIO) and $N = 1323$ NSF-funded computing investigators (CS), matching the totals reported in Table 1 and Table 2A.

The performance period under consideration is from 1996 to 2016, capturing the early part of the 21st century. Over this interval, competition norms and formal review procedures remained broadly similar, but funding conditions were not entirely stationary: NSF success rates in computing were relatively stable, whereas NIH underwent a well-known budget-doubling phase around 1998-2003 that temporarily relaxed funding constraints and raised biology success rates[32,33]. The unit of performance is a grant award: if a biologist received five NIH awards during the performance period, their score would be 5. We work at a medium resolution in terms of investigator roles: each named investigator on an award (PI, co-PI, or multi-PI) accrues one unit of credit. NSF and NIH have gradually expanded the use of co-PI and multi-PI arrangements over this period, and the prevalence and interpretation of these roles varies across programs and years. As a result, there is no simple, time-consistent basis for assigning different numerical weights to PI versus co-PI or multi-PI status, and we do not attempt to impose such differential shares.

We do not observe the number of *submitted* proposals nor other latent drivers of grant-seeking behavior. Our analysis is therefore explicitly conditional on realized grant wins per named investigator, and differences in submission intensity or institutional context enter as unobserved heterogeneity rather than as modeled covariates. To ensure that we have adequate repeated opportunities to observe competitive performance, we focus on faculty who received at least one grant award by 2010. Faculty whose first grant award occurred after 2010 are excluded, as they are likely to be relatively new investigators for whom the 1996–2016 window may not capture their full competitive potential.

The stopping rule at 2010, rather than at a later date, is motivated by established criteria of academic performance. In the U.S., tenure decisions in which grant awards play an important role are typically based on a 5-year performance period[45], which we treat as the minimum horizon for evaluating research careers. Investigators who first appear in the grant record in

Table 3 | Countries and sports at the 1896 Athens Olympic Games

Participating countries (Athens 1896)
Australia
Austria
Bulgaria
Chile
Denmark
France
Germany
Great Britain
Greece
Hungary
Italy
Sweden
Switzerland
United States
Sports on the 1896 Olympic program
Athletics (track and field)
Cycling
Swimming
Gymnastics
Weightlifting
Wrestling (Greco–Roman)
Shooting
Tennis
Fencing

Thetop panel lists the 14 participating nations; the bottom panel lists the nine official sports on the 1896 program.

1996–2000 are typically observed for well over a decade and, in many cases, for most of their active grant-seeking careers by 2016. Those whose first grant appears in 2001–2005 or 2006-2010 are followed forward for roughly 11–15 and 6–10 years, respectively, often spanning the period in which they progress from junior to tenured positions. We do not reconstruct individual CVs or institution-specific tenure clocks; such differences in career timing enter as unobserved heterogeneity rather than as an explicit stratification.

As with the other domains, we organize the grant data in two ways. In the intergenerational case (the main grantee dataset), the competitor set consists of all faculty who secured at least one grant between 1996 and 2010 and are then followed through 2016; we further filter out investigators who have only a single award in their tally, treating a lone grant as nominal performance over such long windows. In the generational design, we group investigators by the period of their first award (1996–2000, 2001–2005, or 2006–2010), yielding three academic generations in each field (BIO: N = 228, 122, and 174; CS: N = 690, 333, and 300), whose sums equal the system totals above. These entry-cohort subsets are used to study how tail behavior evolves across vintages with different exposure horizons and, in biology, between the NIH budget-doubling era (when the 1996–2000 cohort enters) and the subsequent post-doubling funding regime. The doubling episode provides a quasi-natural experiment in funding stringency: grant competitions became temporarily less selective, while review norms, career structures, and field definitions remained broadly comparable.

Construction of the Olympic winners dataset. The Olympic winners dataset includes detailed records of Olympic medalists in the U.S. swimming and French fencing cohorts from the 1896 Olympiad in Athens up to the 2016 Olympiad in Rio de Janeiro. We chose to analyze U.S. swimmers and French fencers because these country-sport pairs jointly satisfy several design criteria:

- continuity from the inaugural modern Olympics (1896) onward, allowing multiple historical periods to be formed under broadly comparable rules and scoring conventions;
- political stability of the national entities across the study window, avoiding complications from state dissolution and re-formation (e.g., the split of Germany into FRG and GDR and later reunification);
- sufficiently large and regularly selected national teams to support reliable tail estimation at the individual level; and
- somatically and technically coherent disciplines in which athletes typically contest multiple events within a Games (e.g., multiple distances or weapons), yielding many repeated opportunities within a well-defined performance domain.

By somatic and technical coherence, we mean that the events within a discipline draw on closely related physical and technical capacities. For example, freestyle and butterfly swimming share a common training base, and many elite swimmers contest both strokes and multiple distances. By contrast, Olympic athletics (track and field) aggregates highly heterogeneous events such as sprinting and shot put, which rely on very different body types, motor patterns, and training regimes; no athlete seriously competes in both at the medal-winning level. Using such heterogeneous bundles as a single *sport* would blur the performance domain and complicate any interpretation of repeated opportunities.

Table 3 summarizes the countries that participated in the 1896 Athens Games and the individual sports on the original Olympic program. Under our continuity and stability criteria, this already leaves only a small set of plausible country-sport pairs that could be followed from 1896 to 2016. Many otherwise promising combinations either fail on roster size (too few medalists per Games) or on somatic/technical coherence (heterogeneous event portfolios as in athletics). Under these joint constraints, U.S.

swimming and French fencing emerge as natural choices: both have long-standing traditions of excellence and deep rosters, but they represent distinct sport types (aquatic vs. fencing) and institutional contexts[26,27].

A further reason to focus on specific country-sport pairs rather than pooling all countries within a sport is that national programs differ markedly in depth, infrastructure, and continuity. Our interest is in competitors who enter the Olympics from systems with a stable and consistently high preparation standard. U.S. swimmers and French fencers typically arrive as part of such well-established programs and then compete against the rest of the world, which collectively provides formidable opposition but whose national programs vary widely in depth, resources, and continuity. Following a single country with a strong, stable tradition in each sport allows us to treat the national team as a relatively homogeneous source of well-prepared entrants, while still situating them within the full global Olympic competition.

To differentiate the competitive value of the three types of Olympic medals, we assign 3 score points to gold medals, 2 score points to silver medals, and 1 score point to bronze medals. We exclude athletes who won only a bronze medal during their career, as we consider this nominal performance in the context of the Olympic Games. We also do not count team medals, as our study focuses on individual performance; for example, relay events such as 4 × 100 freestyle are excluded from consideration.

As with the other two main datasets, we organize the Olympic winners dataset in two different ways: intergenerational and generational. In the intergenerational case, the competitor set within each cohort (U.S. swimmers or French fencers) comprises all medal winners in the Olympiads from 1896 to 2016, subject to the conditions described above. Each competitor is represented by their total score, computed from their medal-winning record according to the weighting scheme outlined. The intergenerational set lists 446 U.S. swimmer medalists and 137 French fencer medalists with scores ≥ 2.

In the generational design, we include the same athletes (U.S. swimmers and French fencers with scores ≥ 2) but grouped into four epochs: Pre-WWI (1896–1912), Pre-WWII (1920–1936), Cold War (1948–1976), and Turn-of-the-Century (1980–2016). These epochs were selected with two criteria in mind. First, they must provide sufficient time for athletes to develop and exhibit their full potential, given the four-year Olympic cadence and the finite length of athletic careers. Second, they must be delimited by major historical disruptions that sharply reduce the possibility that medal-winning athletes from one epoch can continue into the next, and that coincide with shifts in sports science, equipment, and institutional arrangements[46]. World War I and World War II canceled the 1916, 1940, and 1944 Games, effectively forcing generational resets in Olympic participation. For U.S. swimmers, the U.S.-led boycott of the 1980 Moscow Games removed an entire Olympic cycle of opportunities and marks a structural break between the Cold War and Turn-of-the-Century epochs. Given biological limits on peak performance and the four-year spacing of Olympiads, these breaks make cross-epoch medal careers rare, while each epoch remains long enough for entrants to realize their performance potential under a characteristic technological and institutional regime.

Construction of the tennis and Great Britain swimmers comparative and control datasets. In addition to the three primary domains described above, we construct two auxiliary datasets used as comparative and control cases: a professional tennis dataset that provides a within-domain continuum in competitive rigor under fixed rules and format, and a Great Britain swimmers dataset that serves as a size-matched downsampling control for U.S. swimmers.

Tennis dataset. The tennis dataset is built from modern men's professional singles competition on the Association of Tennis Professionals (ATP) circuit[47]. For our purposes, the men's circuit has two main competitive levels: the *ATP Tour*, which is the primary top-tier circuit, and the *ATP Challenger Tour*, which is the developmental circuit one tier below. We treat the four Grand Slam tournaments (Australian Open, Roland Garros, Wimbledon, and the U.S. Open) together with ATP Tour events (ATP 250, ATP 500, and ATP Masters 1000) as a single *Top-tier* stratum, and ATP Challenger events as a separate *Challenger* stratum.

We focus on recent seasons for which electronic records are complete and formats are stable, covering 2016–2019 and 2021–2023 (2020 is omitted because of extensive pandemic-related disruptions to the calendar and draws). In all of these tournaments, players enter a main-draw singles bracket (after any qualifying rounds, which we exclude from our analysis) and then play a standard single-elimination event: each match is played until one player wins; the winner advances to the next round and the loser exits the tournament. For each player in each main-draw event and year, we define an *event-level run* as the number of main-draw matches that player wins in that tournament before their first loss or the title, yielding an integer from 0 (first-round loss) up to the number of rounds in the draw (typically 5-7 for champions, depending on draw size).

The unit of analysis in our tennis study is therefore a *player-event-year* run, not a career. To align with our Olympic scoring analysis and focus on the competitive tail, the frontier-scan analysis is restricted to runs with at least two main-draw wins in a given event (wins_in_event ≥ 2). In our data, the Top-tier stratum includes $N_{\mathrm{Top}} = 4,547$ such player-event runs and the Challenger stratum includes $N_{\mathrm{Chal}} = 9,457$. We pool these runs across seasons within each stratum and analyze the resulting distributions separately for the Top-tier and Challenger circuits. Because the rules, scoring, draw structure, and exposure (one observation per player-event) are essentially identical across the two strata, while the average competitive caliber is higher on the Top-tier circuit than on the Challenger circuit, tennis provides a controlled within-domain comparison: we can examine how our tail regimes evolve as competition quality is relaxed toward a more luck-dominated regime, without changing the underlying format.

Great Britain swimmers dataset. The Great Britain swimmers dataset is derived from the same Olympic sources used for the U.S. swimmers cohort[25,27]. We restrict the Olympic winners dataset to athletes who represented Great Britain in individual swimming events over the 1896-2016 period and apply exactly the same scoring scheme (3 points for gold, 2 points for silver, 1 point for bronze) and inclusion criteria (total scores ≥ 2; exclusion of relay and other team events). The resulting dataset contains a smaller but still high-caliber set of British swimmers with individual Olympic success: Great Britain has a long-standing tradition of excellence in Olympic swimming, so this cohort remains firmly within the elite range even though its roster is noticeably smaller than that of the United States. In our data, the intergenerational U.S. swimmers cohort includes $N_{\mathrm{US}} = 446$ athletes with scores ≥ 2, whereas the Great Britain cohort includes $N_{\mathrm{GBR}} = 56$.

Because both national teams operate at a similarly elite level, the Great Britain swimmers cohort serves two roles in our design. First, it provides an independent comparison case: we can examine whether a different but approximately comparable Olympic swimming program, albeit with a much smaller roster, exhibits tail behavior consistent with what we observe for U.S. swimmers. Second, its roster size offers a natural benchmark for sample-size robustness checks. Great Britain's rosters lie near the lower bound of what is adequate for tail analysis, whereas U.S. rosters are substantially larger. In our control analyses, we therefore repeatedly downsample the U.S. swimmers cohort to match the Great Britain roster size and re-run the retained-fraction frontier procedure on each resample. This two-step use of the Great Britain cohort—first as an independent elite reference, then as a roster-size target for resampling—allows us to assess both cross-program consistency and the sensitivity of our U.S. results to sample-size and range effects, helping to rule out mechanical scaling as a primary explanation for the observed tail patterns.

Analytic methodology

In the following, we describe important details of the methods we used to operationalize the study.

System-total analyses. We treat each domain's outcome distribution, aggregated over the full observation window and across all entry cohorts or eras, as a *system total*. For Luftwaffe pilots this is the distribution of confirmed aerial victories accumulated by each pilot over World War II; for BIO/CS investigators it is the distribution of grant awards accumulated over each investigator's observed career within our database window; and for Olympic swimmers and fencers it is the distribution of cumulative medals per athlete across all Games in the study period. These system totals therefore provide an across-history baseline for each domain, pooling individuals from different vintages and institutional conditions.

We compare three canonical candidates for the right tail: (i) a discrete lognormal (**dLN**), representing a heavy but sub–power-law tail that arises naturally from compounded but noisy multiplicative growth; (ii) a discrete power law or Zipf form (**Zipf**), often used as an idealized representation of very heavy winner-take-all tails; and (iii) a shifted geometric (**Geom**), representing a much thinner, approximately memoryless accumulation regime. Each model is fitted by maximum likelihood to integer-valued outcomes.

We do not fix the tail region in advance. Instead, we perform a *frontier scan* over retained fractions of the sample. For each domain, we order individuals by their outcome (victories, grants, medals) and consider a grid of retained fractions $f \in \{0.20, 0.25, \ldots, 0.80\}$. For each f, we choose the smallest integer threshold $k_{\min}(f)$ such that retaining all observations with outcome $\geq k_{\min}(f)$ yields a retained fraction as close as possible to f from above, given the discreteness and ties in the data. We then fit dLN, Zipf, and Geom to the observations with $k \geq k_{\min}(f)$.

For goodness-of-fit we use a Kolmogorov–Smirnov (KS) statistic with a parametric bootstrap. For each model–fraction pair we generate many synthetic samples under the fitted model and recompute the KS distance, obtaining a bootstrap p-value $p_{\text{model}}(f)$. This provides an adequacy curve for each model as a function of the retained fraction. We also compute the Akaike Information Criterion (AIC) for each fit.

To summarize each domain, we select a *frontier* retained fraction f^* as the largest scanned f for which the dLN model is adequate (KS–bootstrap $p_{dLN}(f^*) \geq 0.05$) and the retained sample size satisfies $N_{\text{kept}}(f^*) \geq 40$. At that frontier, we compare the dLN, Zipf, and Geom models; among the models that are adequate, we designate the one with the lowest AIC as the *frontier winner* for that domain.

We also implement a control analysis to address potential mechanical scaling artifacts arising from roster-size differences. In particular, U.S. Olympic swimmers form a much larger cohort than British swimmers. To check whether differences in model support between these cohorts could be driven by sample-size or range effects, we first estimate the retained-fraction frontier for the Great Britain swimmers and record the corresponding tail cutoff $k^\star_{\min}$. We then repeatedly downsample the U.S. swimmer cohort to match the British roster size, and for each resample we fix the tail at this same cutoff $k^\star_{\min}$ and fit dLN, Zipf, and Geom to the resulting U.S. tail. For each resample, we compute AIC differences between dLN and the alternative models, summarizing how often dLN remains preferred under size- and cutoff-matched conditions.

Generational and period-bounded analyses. To examine how tail behavior varies across historical conditions within each domain, we repeat the frontier analysis in exposure-aligned strata that hold institutional conditions approximately constant while allowing career timing and length to vary. In all such strata we apply exactly the same retained-fraction frontier procedure as for system totals (same grid of f, same KS-bootstrap adequacy criterion $p_{\text{model}}(f) \geq 0.05$, the same $N_{\text{kept}}(f) \geq 40$ requirement, and AIC-based model comparison).

Entry–cohort, forward follow (LW; BIO/CS). For Luftwaffe pilots and BIO/CS investigators we define cohorts by year of first success (first aerial victory or first grant award) and follow individuals forward to exit or to the end of the observation window. Pre-entry exposure is zero by construction, and cohort membership aligns participants who faced similar institutional and technological conditions. Within each cohort that yields at least $N_{\text{kept}}(f) \geq 40$ retained observations for some f, we apply the frontier scan and classify the tail using the dLN, Zipf, and Geom models.

Era-bounded windows (Athletes: U.S. swimmers; French fencers). For Olympic swimmers and fencers we define era-specific windows based on major historical and institutional breaks in the Games (for example, pre–World War II Olympiads, the Cold War, and the post-Cold War period). Within each era we aggregate each athlete's medal counts over the Games in that window and then run the same frontier pipeline, using the common adequacy and sample-size criteria to identify frontier winners.

Tail regimes and theoretical null. The frontier procedure above treats the discrete lognormal (dLN), Zipf, and shifted geometric (Geom) distributions as competing statistical models for the upper tail. We also use these three candidates as proxies for distinct *tail regimes* linked to underlying competitive conditions, which lets us interpret the frontier winner at f^* as evidence about how far a system lies from a luck-heavy baseline. Conceptually, we map the three distributions onto the following regimes:

- *Luck-heavy / approximately memoryless (null).* When outcomes are shaped largely by luck under repeated but only weakly sorted opportunities, the conditional hazard of an additional success is roughly constant from trial to trial. Standard counting-process arguments then yield geometric or negative-binomial–like counts with an exponential tail. In this setting the shifted geometric (Geom) serves as a *theoretical null* for a Broad-Access, low-selectivity regime: exposure may vary, but conditional on continued participation the increment from one success to the next is close to memoryless. By contrast, a perfectly uniform distribution would correspond to a single-shot context in which individuals receive at most one success and exposure does not vary; this lies outside the repeated-opportunity setting we analyze here and so we do not treat the uniform as a practical null.
- *Runaway cumulative advantage*. Under strong success-begets-success mechanisms (e.g., Matthew-effect or preferential-attachment dynamics), early gains substantially increase the probability of future gains. This produces extremely heavy, near-scale-free tails in which a small minority accumulate a disproportionate share of outcomes. We use the discrete power law (Zipf) as an idealized representation of this regime.
- *Relative-Fairness (target regime)*. When heterogeneous skill is expressed through many largely independent opportunities, and no single non-skill factor persistently dominates outcomes, success accumulates approximately as noisy multiplicative growth. This yields heavy but *sub–power-law* tails, for which the discrete lognormal (dLN) is a natural proxy. We interpret this regime as one of *Relative-Fairness*: competition is selective and exposure is ample, so skill has repeated chances to accumulate, but luck remains present and careers are finite.

Classification rule at the frontier. For each domain and stratum (system totals, entry cohorts, era-bounded windows), the frontier procedure yields a retained fraction f^*, a corresponding threshold $k_{\min}(f^*)$, and KS–bootstrap p-values and AIC scores for dLN, Zipf, and Geom. We classify the tail regime at that frontier as follows:

1. If Geom is adequate ($p_{\text{Geom}}(f^*) \geq 0.05$) and preferred by AIC among adequate models, we label the stratum as *luck-heavy / geometric-like* and treat this as evidence for the theoretical null.
2. If dLN is adequate and preferred by AIC among adequate models, we label the stratum as lying in the *Relative-Fairness* regime.
3. If Zipf is adequate and preferred by AIC among adequate models, we label the stratum as exhibiting *runaway cumulative advantage*.

Strata in which no model is adequate at f^* are treated as inconclusive for regime classification.

Movement toward the null. This setup also lets us operationalize movement along a continuum from Relative-Fairness toward the luck-heavy null. We use two signatures: (i) changes in the frontier winner (dLN, Zipf, or Geom) across strata that differ in access, selectivity, or technological constraints; and (ii) changes in the breadth of the adequate tail, measured by the largest retained fraction f^* at which dLN remains adequate. Holding the estimation and adequacy criteria fixed, a receding dLN frontier (smaller f^*) under weaker selection or broader access indicates that the range of outcomes well described by multiplicative skill accumulation is shrinking, in the direction of a geometric-like, approximately memoryless baseline. We later use this regime classification to interpret within-domain contrasts and to connect observed tail behavior to documented changes in competitive conditions.

Data availability
The aggregated data and derived distributions used in this study are available at https://github.com/UH-ACDC/npjcomplexity-competitive-performance. At https://github.com/UH-ACDC/npjcomplexity-competitive-performance.

Code availability
All analysis and figure-generation code is available at https://github.com/UH-ACDC/npjcomplexity-competitive-performance.

Received: 7 May 2025; Accepted: 9 February 2026;
Published online: 16 March 2026

References
1. Barabási, A.-L. The origin of bursts and heavy tails in human dynamics. *Nature* **435**, 207–211 (2005).
2. Perc, M. Zipf's law and log-normal distributions in measures of scientific output across fields and institutions: 40 years of Slovenia's research as an example. *J. Informetr.* **4**, 358–364 (2010).
3. Yair, G., Gueta, N. & Davidovitch, N. The law of limited excellence: Publication productivity of Israel prize laureates in the life and exact sciences. *Scientometrics* **113**, 299–311 (2017).
4. Bernard, A. B. & Busse, M. R. Who wins the Olympic games: Economic resources and medal totals. *Rev. Econ. Stat.* **86**, 413–417 (2004).
5. Simkin, M. V. & Roychowdhury, V. P. Theory of aces: High score by skill or luck? *J. Math. Sociol.* **32**, 129–141 (2008).
6. Merton, R. K. The Matthew effect in science. *Science* **159**, 56–63 (1968).
7. White, C. S., James, K., Burke, L. A. & Allen, R. S. What makes a "research star"? factors influencing the research productivity of business faculty. *Int. J. Product. Perform. Manag.* **61**, 584–602 (2012).
8. Zucker, L. G. & Darby, M. R. Star scientists and institutional transformation: Patterns of invention and innovation in the formation of the biotechnology industry. *Proc. Natl. Acad. Sci.* **93**, 12709–12716 (1996).
9. Stanley, M. H. R. et al. Scaling behaviour in the growth of companies. *Nature* **379**, 804–806 (1996).
10. Cristelli, M., Batty, M. & Pietronero, L. There is more than a power law in Zipf. *Sci. Rep.* **2**, 812 (2012).
11. Felin, T., Foss, N. J. & Ployhart, R. E. The microfoundations movement in strategy and organization theory. *Acad. Manag. Ann.* **8**, 575–632 (2014).
12. Fink, G., Dauber, D. & Yolles, M. Performance heterogeneity in organizations: Toward a micro-foundations perspective. *Strategic Organ.* **15**, 227–248 (2017).
13. Fonti, F. et al. Using sports data to advance management research: A review and a guide for future studies. *J. Manag.* **49**, 325–362 (2023).
14. Makino, S. & Chan, C. M. Skew and heavy-tail effects on firm performance. *Strategic Manag. J.* **38**, 1721–1740 (2017).
15. Becker, G. S. & Murphy, K. M. The division of labor, coordination costs, and knowledge. *Q. J. Econ.* **107**, 1137–1160 (1992).
16. Jovanovic, B. & Nyarko, Y. Learning by doing and the choice of technology. *Econometrica* **64**, 1299–1310 (1997).
17. Leahey, E. Not by productivity alone: How visibility and specialization contribute to academic earnings. *Am. Sociological Rev.* **72**, 533–561 (2007).
18. Jara-Figueroa, C., Jun, B., Glaeser, E. & Hidalgo, C. A. Knowledge in the successful career move: The role of knowledge in the mobility of inventors. *Nat. Commun.* **9**, 1–7 (2018).
19. Haack, H.*Pilots in the First World War: How the Phenomenon of the Flying Ace Impacted the War Effort*. Ph.D. thesis, University of Iowa (2021).
20. Murray, W.*Strategy for Defeat: The Luftwaffe, 1933–1945* (Air University Press, Maxwell Air Force Base, AL, 1983).
21. Spick, M.*Luftwaffe Fighter Aces* (Frontline Books, 2011).
22. Petersen, A. M., Majeti, D., Kwon, K., Ahmed, M. E. & Pavlidis, I. Cross-disciplinary evolution of the genomics revolution. *Sci. Adv.* **4**, eaat4211 (2018).
23. NIH RePORTER. https://reporter.nih.gov Accessed on April 14, 2025 (2025).
24. NSF Award Search Web API. https://catalog.data.gov/dataset/nsf-award-search-web-api-3f6f4 Accessed on April 14, 2025 (2025).
25. Kaggle. 120 years of Olympic history: Athletes and results. https://www.kaggle.com/datasets/heesoo37/120-years-of-olympic-history-athletes-and-results Accessed 16 November 2025 (2023).
26. Baginska, J., Piepiora, P., Piepiora, Z. & Witkowski, K. The geography of Olympic combat sports – part two: Boxing, fencing, modern pentathlon, wrestling. *Arch. Budo* **18**, 327–340 (2022).
27. Trewin, C. B., Hopkins, W. G. & Pyne, D. B. Relationship between world ranking and Olympic performance of swimmers. *J. Sports Sci.* **22**, 339–345 (2004).
28. Allen, S. V. & Hopkins, W. G. Age of peak competitive performance of elite athletes: A systematic review and meta-analysis. *Sports Med.* **45**, 1431–1441 (2015).
29. Roi, G. S. & Bianchedi, D. The science of fencing: Implications for performance and injury prevention. *Sports Med.* **38**, 465–481 (2008).
30. Caldwell, D. & Muller, R.*The Luftwaffe over Germany: Defense of the Reich* (Frontline Books, 2014).
31. Uziel, D.*Arming the Luftwaffe: The German Aviation Industry in World War II* (McFarland, 2011).
32. Park, J., Blume-Kohout, M., Krestel, R., Nalisnick, E. T. & Smyth, P. Analyzing NIH funding patterns over time with statistical text analysis. In *AAAI Workshop: Scholarly Big Data* (2016).
33. Rissler, L. J., Hale, K. L., Joffe, N. R. & Caruso, N. M. Gender differences in grant submissions across science and engineering fields at the NSF. *BioScience* **70**, 814–820 (2020).
34. Lane, J. N. et al. Conservatism gets funded? a field experiment on the role of negative information in novel project evaluation. *Manag. Sci.* **68**, 4478–4495 (2022).
35. Boudreau, K. J., Guinan, E. C., Lakhani, K. R. & Riedl, C. Looking across and looking beyond the knowledge frontier: Intellectual distance, novelty, and resource allocation in science. *Manag. Sci.* **62**, 2765–2783 (2016).
36. International Olympic Committee. The modern Olympic games caught in the upheavals of history. https://www.olympics.com/ioc/news/the-modern-games-caught-in-the-upheavals-of-history Accessed 16 November 2025 (2020).
37. U.S. Department of State, Office of the Historian. The Olympic boycott 1980. https://history.state.gov/milestones/1977-1980/olympic Documents on Diplomacy: Resources. Accessed 16 November 2025 (1980).
38. *The United States Strategic Bombing Survey: Summary Report (European War) September 30, 1945*, vol. 1 (US Government Printing Office, 1945).
39. Spick, M.*Allied Fighter Aces: The Air Combat Tactics and Techniques of World War II* (Greenhill, 1997).

40. Morgan, H. & Weal, J.*German Jet Aces of World War II* (Osprey Publishing, 1998).
41. Shores, C. F.*Air Aces* (Presidio Press, 1983).
42. Musciano, W. A.*Messerschmitt Aces* (Arco Pub., 1982).
43. Toliver, R. F. & Constable, T. J.*Horrido! Fighter Aces of the Luftwaffe* (Bantam War Books, 1979).
44. Aircrew Remembered. https://aircrewremembered.com Accessed on April 14, 2025 (2025).
45. Akita, K. et al. "Publish or Perish": analysis of tenure in U.S. universities. *Journal of the School of Foreign Studies, Aichi Prefectural University: Area Studies and International Relations* 205–219 (2022).
46. Taha, Z. et al. An overview of sports engineering: History, impact and research. *Mov., Health Exerc.* **2**, 1–14 (2013).
47. ATP Tour. Results archive. https://www.atptour.com/en/scores/results-archive Accessed: 16 November 2025 (2023).

Acknowledgements
This research was funded in part through the Eckhard-Pfeiffer Distinguished Professorship Fund at the University of Houston and in part through the NSF AccelNet award 2412731.

Author contributions
V.Z. collected data and co-developed methods, P.T. co-developed methods, I.P. designed research, co-developed methods, and wrote the manuscript. All authors reviewed the manuscript.

Competing interests
The authors declare no competing interests.

Additional information
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s44260-026-00078-y.

Correspondence and requests for materials should be addressed to Ioannis Pavlidis.