

Entropy to the Mean

Winning Streaks Carry a Rising, Measurable Cost: Evidence from College Basketball and Six Additional Competitive Substrates

A Regression Duel Theory (RDT) White Paper

James Castranova

Independent researcher

Correspondence: jcastra@gmail.com

Preprint. v0.8, August 2026

Supersedes v0.7 (structural predecessor) and v0.6 (evidentiary predecessor).

Abstract

The hot hand literature asks whether success predicts success. This paper asks the inverse: **as a team carries a winning streak into successive games, does the quality of its performance deteriorate as entering streak depth increases, and is that deterioration visible in raw basketball events rather than only in expectation residuals?** The regression frame includes all eligible team-games. Entering streak depth is zero when the team carries no winning streak into the game and positive otherwise; the game that ends a streak remains in the frame at the depth carried into that game.

Using **72,638 Division-I team-game rows across eight college basketball seasons** on a fully repaired panel with point-in-time expectation snapshots from three independent rating systems, we estimate the effect of entering win-streak depth on opponent-adjusted margin residual within team-season fixed effects, with standard errors clustered on team-season. **The gradient is present and adverse against every independent expectation ruler:** KenPom -0.2797 (SE 0.0235), BPI -0.2865 (SE 0.0233), Torvik -0.1041 (SE 0.0217); Mean(3) -0.2234 , Median(3) -0.2432 ; all intervals clear.

The gradient is not a KenPom peculiarity, not explained by simple cumulative-overperformance reversion on the primary KenPom and BPI rulers, and not a within-season calendar artifact. A within-team-season joint model of depth against cumulative streak overperformance leaves depth intact on the two rulers with large gradients (KenPom: 97.9% retained; BPI: 108.7%) while cumulative overperformance collapses to insignificance; **on Torvik the decomposition inverts; the inversion is documented, but its underlying reason remains open.** Adding season $\times$ week fixed effects retains 97.7% (KenPom) to 101.4% (Torvik) of the flagship gradient — the direct rebuttal to the calendar objection.

The paper's central new result, and the reason v0.8 supersedes v0.7 and v0.6, is an expectation-free bridge. On the same games where the residual gradient is measured, apart from trivial channel missingness (16 of 72,666 rows in the reconstructed comparison frame), we remove the expectation systems from the outcome entirely and estimate 18 objective basketball channels on entering streak depth with focal and opponent team-season fixed effects and season $\times$ week fixed effects. **Thirteen of eighteen primary channels clear the standing dark gate; ten move adversely, including aggregate offensive efficiency; three — turnover control, foul rate, and raw fouls — improve.** On the restricted 72,650-row expectation-covered comparison run, twelve of eighteen clear the dark gate with every direction preserved and the major production channels sharpening.

Cross-substrate evidence, filed under the v0.7 seven-substrate architecture at the frozen A–E hierarchy: **tennis** confirms the phase-tier residual under a structure-preserving lifecycle-null (v0.7 §5 verbatim, Grade A); **soccer** replicates enter-depth compression at the goal-differential channel (Grade A); **NHL** discovery-selected `hd_share_tied` returned not-confirmed under a reduced-ceiling reproduction on the frozen 2018-19 through 2025-26 holdout (Grade B under the frozen v0.7 assignment); **CFB** identified a defensive rushing carrier but the sealed 2016-2021 corpus was underpowered for confirmation under the frozen power gate and the play-by-play holdout remains unopened (Grade C); **NFL** identified WR/TE yards-after-catch per reception at depths 1-4 as a predicted-direction discovery candidate but the pre-holdout power gate recovered only 172 of 500 planted worlds against the required 416; the 2021-2025 holdout was never opened (Grade C); **MLB** returned an honest null under a well-powered starting-pitcher process design plus feasibility and boundary results on three other adjudicated designs, and the substrate is closed for RDT testing under the current data modality (**historical v0.7 Grade D provenance; the updated well-powered Layer-1 null is ungraded pending an explicit taxonomy amendment; see §3 and §8.6**).

The duration-burden interpretation is [**supported**]; the compensation reading — that discipline gains reflect active resistance to duration-linked deterioration — is [**supported hypothesis**] with the plain-English framing *teams are trying to escape entropy* preserved alongside; the causal mechanism is [**open**]; the entropic nature of what is being resisted is [**open**].

Keywords: regression to the mean; hot hand; streaks; ordered states; selection bias; within-team identification; entropy; pre-registration; college basketball; multi-substrate.

1. Introduction

Forty years of streak research has been organized around one question: does success predict success? Gilovich, Vallone, and Tversky (1985), building on Tversky and Kahneman (1971), answered no and attributed the perception of momentum to misreading random sequences. Miller and Sanjurjo (2018) showed that the original analysis itself contained a subtle selection bias, and that correcting it revives evidence for a modest hot hand. Bar-Eli, Avugos, and Raab (2006) document how sensitive the debate is to conditioning: the act of selecting sequences bends the statistics of what is selected, and both sides of the hot hand argument have at times been fooled by it.

This paper asks a different question, one the hot hand framing does not reach. It is not whether a winning team is more likely to keep winning. It is whether **the same team, as the winning streak it carries into a game grows older, performs progressively worse relative to expectation, and whether that decay is visible in raw basketball events without any expectation ruler at all.** The unit of interest is every eligible team-game, indexed by the winning-streak depth carried into it (zero when the team carries no winning streak, positive otherwise).

The distinction matters because the two default deflationary accounts attack different objects. Regression to the mean (Galton, 1886) says displacement from equilibrium predicts movement back toward it: a team that has banked unusual overperformance should give some back, in proportion to the size of the bank. It says nothing about duration as such. Selection geometry says conditioning on streak membership mechanically bends conditional averages even for a memoryless process, so within-streak decline proves nothing. A finding that survives as a duration effect must therefore do three things: **beat displacement in a direct horse race, exceed the decline that a memoryless process of the same skill produces under the identical construction, and appear in the raw event record when the expectation ruler is removed.** The program has tested all three: the memoryless-floor result is established on the disclosed pre-repair frame, while the displacement and expectation-free tests are reported on the repaired record. **The historical flagship cleared its own-construction memoryless benchmark on the pre-repair frame; the repaired-panel floor has not yet been refit.** The repaired evidence presented here adds multi-ruler recovery, prospective replication, calendar robustness, and expectation-free channel deformation.

The result, previewed: within team-season, opponent-adjusted margin residual declines by roughly a fifth to a third of a point per game of entering win-streak depth across three independently constructed rating systems. Accumulated overperformance cannot absorb it on the primary rulers. Fatigue, recent form, operational opponent-preparation proxies, and calendar drift cannot absorb it. A loss-streak placebo runs the other way. On the historical pre-repair flagship, an own-construction memoryless null reproduced part of the raw slope mechanically; the excess beyond that floor survived. **In v0.8 that historical falsification result is preserved with its provenance, while the repaired-panel case rests additionally on multi-ruler recovery, prospective replication, calendar robustness, mean-reversion discrimination, and expectation-free channel deformation.** And on essentially the same games, with the expectation systems removed and 18 raw event channels put on the same entering-depth clock, thirteen channels deform at the dark gate and the same duration variable predicts adverse deformation of the underlying basketball performance system.

Two disclosures frame everything that follows. First, this is a theory-first research program with frozen specifications, pre-registered verdict logic, synthetic validation gates before any real-data run, and independent adversarial review of every instrument, with the author as sole adjudicator. Several of the program's own earlier claims were killed by its own nulls, and those kills are reported here as part of the evidence that the surviving results are different in kind. Second, the paper's own prior CBB channel decomposition (v0.6 §8) reported a pace slowdown as part of the behavioral fingerprint. **A dedicated tempo instrument built after v0.6 falsified that claim** (Section 6.7); the pace decomposition is retracted here and does not enter the current channel map.

2. Theory: Regression Duel Theory

Regression Duel Theory (RDT) is named for the direction of its central claim: that much of what the literature files under regression to the mean is not a passive statistical settling but the adverse deformation of sustained ordered com-

petitive states as ordered duration accumulates, with the deformation scaling with how long the order has persisted. The provocation can be stated in four words: **entropy to the mean**. The name marks a hypothesis about direction and form, not a physical identity: no thermodynamic or information-theoretic entropy is computed or claimed to be measured anywhere in this paper, and the firewall of Section 2.3 governs every interpretive sentence. The duel is empirical and runs throughout: duration of order against distance from the mean.

RDT gives the phenomenon it observes a name. **Entropression** is duration-indexed adverse deformation of sustained ordered competitive performance that exceeds the substrate-appropriate null or benchmark for the tested architecture. In college basketball, the substrate-appropriate benchmark is the same-construction memoryless floor; in tennis, a structure-preserving residual null within tournament-structure cells; in soccer, a within-team-season permutation null. The term names the observed form, not its causal mechanism and not a thermodynamic identity. An entropression signal is a substrate-specific statistical registration of that form: a depth coefficient, a channel deformation, a phase-tier rise. The statistics are not entropressing; the ordered performance state undergoes entropression, and the statistics register its signature.

2.1 The law of form

The theory's most compressed observable statement is a single sentence: **a streak becomes more difficult as it ages inside its own lifespan**. This is not the claim that long streaks eventually end (trivially true), and it is not the claim that unusually good measured performance comes back down (regression to the mean, also true and not in dispute). It is the claim that, as an ordered state persists inside its own life, the state carries an increasing duration-memory burden against continuation, measurable as adverse deformation in performance channels while the streak is still alive.

Status: **[supported]** at the phenomenon level under prospectively frozen, substrate-appropriate confirmatory architectures (Sections 4 through 8); **not [established]** as a universal law and **not [established]** as to mechanism.

2.2 Vocabulary

RDT holds that the live ordered competitive performance state carries a duration-memory burden that scales with accumulated ordered duration. *Burden* is descriptive, not causal. Older program vocabulary — regression debt, regression credit, displacement, discharge — remains descriptive shorthand where historically necessary, but none is treated here as an operationally measured reservoir, obligation, or causal mechanism. Words that project directional compulsion onto free processes (*a team is due, the streak must break*) are banned from the program's empirical writing.

2.3 The firewall

The observable layer, everything estimated in this paper, is the **footprint**: compression, channel deformation, duration-indexed decline. The interpretive layer is the author's reading of the recurrence: that sustained competitive order exhibits duration-dependent dissolution. That interpretation remains **[speculative]** as mechanism. Interpretive claims are explicitly labeled and kept separate from empirical results; in v0.8 they appear only in designated interpretive passages, principally §§5.6 and 11. No regression coefficient in this paper is asserted to be a physical quantity, and no thermodynamic identity is claimed.

2.4 What the theory stakes in advance

(P1) Within the same team-season, opponent-adjusted performance declines with entering streak depth. (P2) The decline is indexed to duration beyond what is explained by banked opponent-adjusted overperformance. (P3) The decline is order-specific: a loss-streak placebo does not reproduce it in the same direction. (P4) Expression is substrate-appropriate: where multiple channels are measurable, concordant deformation strengthens the inference. (P5) The decline exceeds what a same-construction memoryless process produces. (P6) The same form recurs in other competitive substrates, expressed through substrate-appropriate clocks and channels rather than through an identical statistic. Sections 5 through 8 test each.

Added in v0.8 (P7). *The decline is not confined to expectation residuals: it is visible in raw event channels of basketball*

performance on the same games, under the same duration variable, with no rating system in the outcome. Section 5.6 tests this.

3. Evidentiary Grading

Substrate results are graded on the frozen v0.7 evidentiary hierarchy (A through E), reproduced here verbatim. **Grade A** is prospective confirmation under a frozen instrument with predeclared verdict logic. **Grade B** is frozen discovery plus a concordant but non-significant real holdout. **Grade C** is discovery-only, or a frozen discovery whose sealed holdout was never opened. **Grade D** is a feasibility failure. **Grade E** is a powered directional contradiction. Per-substrate assignments are in §8.

Null results and boundary results are counted as evidence about the phenomenon, not as failures of the program. **A theory that finds its footprint everywhere it looks deserves suspicion**; MLB (§8) is the clearest example of the program taking a real chance to fail and reporting the null honestly. The MLB starting-pitcher process test's well-powered null is a Layer-1 result the frozen taxonomy does not classify cleanly: it is not a feasibility failure and it is not a directional contradiction. v0.8 preserves MLB's **historical v0.7 Grade D** as provenance, states the updated Layer-1 result as a well-powered null, and assigns **no new Layer-2 grade pending an explicit taxonomy amendment** (a proposed *Grade U* for well-powered nulls would be a program-governance decision made outside this paper).

4. Data

The primary CBB frame is **72,638 Division-I team-game rows across eight seasons on the fully recovered bridge panel**: 2016-17, 2017-18, 2018-19, 2019-20, 2021-22, 2022-23, 2023-24, 2024-25. The 2020-21 season is excluded (COVID irregularity). The 2025-26 season is discussed in §5.7 (prospective replication).

The panel is the fully repaired v2 panel (repository authority w4_rebuild_repaired_v2.json, SHA 50b5b9c5a87d9ecb6b90391f3c40a5e9f1a93515fa5c3081a0e76d4345889db4). The pre-repair panel used in v0.6 contained a duplicate-contest defect and orientation errors that the current panel corrects. **Panel repair attenuated the flagship gradient 18–29% across every lane** (KenPom $-0.3425 \rightarrow -0.2797$, BPI $-0.3498 \rightarrow -0.2865$, Torvik $-0.1456 \rightarrow -0.1041$), and this is disclosed with the same weight §6 gives every falsification battery result. The v0.6 headline of -0.3207 is a pre-repair value and is not reported unqualified anywhere in v0.8.

Two related row counts and their relationship, stated once so it does not recur as ambiguity. The flagship coefficients of §5.1 are estimated on the fully recovered bridge panel of **72,638 rows** (frozen artifact bridge_recovered.json). The calendar audit of §5.3, the horse-race re-certification of §5.2, and the restricted-panel channel comparability run of §5.6 use a **72,666-row** reconstruction of that panel built from the same authority w4_rebuild_repaired_v2.json with the same inclusion rules; the reconstruction reproduces every lane's frozen coefficient to within 0.0005. The 28-row difference is a reconstruction artifact and is disclosed here rather than laundered.

Opponent adjustment uses point-in-time expected margins from three independently constructed rating systems: Ken Pomeroy's adjusted efficiency margin (kenpom.com), ESPN's Basketball Power Index (BPI), and Bart Torvik's T-Rank. Expected margins are point-in-time snapshots on or before each game date with no future leakage. **The residual is realized margin minus expected margin**, signed to the focal team.

Streak depth (enter_depth) is the count of consecutive Division-I wins the team carries into the game, resetting to zero on any loss. The matched-only definition (D-I opponents) is primary; §5.5 reports the full-schedule alternative and discloses the difference plainly.

5. Core Result — CBB

5.1 The flagship gradient across three independent rulers

Within team-season, with clustered standard errors:

lane	β per game of entering depth	SE	95% CI
KenPom	-0.2797	0.0235	[-0.3258, -0.2335]
BPI	-0.2865	0.0233	[-0.3321, -0.2409]
Torvik	-0.1041	0.0217	[-0.1466, -0.0617]
Mean(3)	-0.2234	0.0223	[-0.2671, -0.1797]
Median(3)	-0.2432	0.0223	[-0.2868, -0.1995]

72,638 rows, 2,582 team-season clusters, all intervals clear. **The gradient is not a KenPom peculiarity.** The three rulers were constructed by three independent teams from partially overlapping inputs, and the two with large gradients (KenPom and BPI) agree to within 3%. Torvik's shallower gradient is the object of §5.4.

5.2 The mean-reversion horse race, re-certified

Simple regression to the mean says the observed decline should be carried by *accumulated overperformance*, not by *duration*. The pre-registered decisive test (v0.6 §6.4) puts both in the same regression and lets them compete. Repeated on the repaired panel:

lane	corr(depth, cum)	depth alone	joint depth	joint cum	depth retained
KenPom	0.716	-0.2802	-0.2743 (t -8.1)	-0.0011 (t -0.3)	97.9%
BPI	0.637	-0.2870	-0.3119 (t -9.8)	+0.0053 (t +1.2)	108.7%
Mean(3)	0.682	-0.2239	-0.1578 (t -5.1)	-0.0141 (t -3.2)	70.5%
Torvik	0.604	-0.1045	+0.0245 (t +0.9)	-0.0323 (t -7.2)	-23.5%

On the two primary rulers, duration survives almost intact and accumulated overperformance collapses.

This establishes that simple cumulative-overperformance reversion is not a sufficient explanation on KenPom and BPI. **On Torvik the horse race inverts** — depth dies, cumulative overperformance survives. That result is not buried and is not explained away. §5.4 documents this as one of two related empirical asymmetries in the Torvik lane (the other is the shallow residual gradient of §5.1) and states plainly what is and is not established about the relationship between them.

5.3 The calendar audit

Within-season calendar drift is a real hazard. Deeper streaks tend to happen later in the season, and if pregame systems are less well-calibrated in November than in February — which is exactly what a decaying preseason prior produces — part of the gradient could be calendar rather than depth. Adding season $\times$ week-of-season fixed effects to the flagship model:

lane	baseline	+ season $\times$ week FE	retained
KenPom	-0.2802	-0.2739	97.7%
BPI	-0.2870	-0.2862	99.7%
Torvik	-0.1045	-0.1060	101.4%
Mean(3)	-0.2239	-0.2220	99.1%
Median(3)	-0.2437	-0.2418	99.2%

Every change is inside a quarter of one standard error. Calendar drift is not the explanation.

5.4 The Torvik asymmetries (empirical measurement, no mechanism claimed)

Torvik exhibits two related empirical asymmetries relative to KenPom and BPI: a shallower residual gradient (37% of the mean of KenPom and BPI in §5.1) and a horse-race inversion (§5.2). Its expected-margin trajectory as a function of entering depth was recovered independently on the repaired panel:

lane	β of E[margin] on entering depth	t
KenPom	+0.0761	+3.67
BPI	+0.0829	+3.78
Torvik	−0.0997	−4.50

KenPom and BPI's expected margins *rise* with entering depth; Torvik's *falls*. **Because realized margin is identical across rulers, differences in expected-margin trajectories translate mechanically, with opposite sign, into differences in residual gradients. This explains the shallower Torvik residual gradient. The alignment with Torvik's horse-race inversion of §5.2 is notable, but whether the same underlying feature explains that inversion remains open.**

Why Torvik's expectation trajectory is different from KenPom's and BPI's is not settled here. A separate documentary architecture audit closed with no established mechanical distinction between the three systems: nobody publicly documents consuming turnover rate, offensive rebounding, foul rate, two-point percentage, assist share, or transition composition as separate rating inputs, and the recency and preseason-persistence disclosures across the three systems are asymmetric in ways that make a clean cross-system comparison unavailable. The paper carries the empirical fact and refrains from a mechanism claim.

The correct public statement is: **Torvik exhibits two related empirical asymmetries: a shallower residual gradient and a horse-race inversion. Its expected-margin trajectory mechanically explains the former; whether the same feature explains the latter remains open.**

5.5 Depth definition — the audit with a recorded wrong prediction

Streak depth can be defined two ways, and the choice is material. The matched-only definition (consecutive D-I wins) is primary and yields the numbers above. The full-schedule definition (all wins including non-D-I) yields a smaller but still adverse gradient. The original audit predicted that counting non-D-I wins toward depth would *strengthen* the effect. It did the opposite. **That wrong directional prediction is logged here rather than buried**, because a program that hides its misses cannot be trusted on its hits. The theory-justified matched-only definition is primary, and that choice was frozen before the out-of-sample test.

5.6 The expectation-free bridge (new in v0.8)

The five preceding subsections all rely on an expectation ruler in the outcome. The concern the paper's own §7.1 anticipates is that a residual, however constructed, is a function of both realized performance and a chosen expectation model. **The bridge test removes the expectation ruler entirely and asks the same duration question of raw basketball events.**

Design. For each team-game, we estimate:

```
channel(i, t) ~ enter_depth(i, t)
               + team-season FE (focal)
               + team-season FE (opponent)
               + neutral-site indicator
               + season × week-of-season FE
```

with standard errors clustered on focal team-season. The 18 primary channels are the objective basketball fingerprint frozen in `objective_channel_inventory_v1_4_FROZEN` and preregistered per `objA_real_census_PREREGISTRATION_v1_0`.

No pregame rating enters any outcome, anywhere. No point pricing, no summation, no attempt to reconstruct the residual gradient. The output is a coefficient per channel.

Standing dark gate. $|t| > 2.807$ two-sided, the program's standing threshold for instrument admissibility across substrates; a channel clears the dark gate only at this threshold. Ordinary 95% is $|t| > 1.96$. The 18-channel preregistration governed which channels enter the map, not a separate real-result significance threshold; the dark gate here is applied ex post as the same standing threshold used elsewhere in the program.

Result on all team-games (93,113 rows, 2,855 focal team-seasons, 5,412 opponent team-seasons, 174 season-week cells):

category	count	representative channels and coefficients
broad adverse deformation (10 of 18 at dark gate)	10	assist_rate -0.190 (t -7.9), offensive_efficiency -0.146 (t -6.0), eFG% -0.111 (t -7.4), TS% -0.098 (t -7.2), 3P% -0.097 (t -5.1), 2P% -0.090 (t -5.4), TRB% -0.055 (t -5.3), ORB% -0.046 (t -3.0), turnover_points_share -0.092 (t -5.2), TO_per_opp_TOV -0.005 (t -5.0)
adverse, cleared 95% but not dark gate	3	paint_share -0.099 (t -2.7), DRB% -0.033 (t -2.2), fastbreak_share -0.028 (t -2.1)
persistent discipline gains (3 of 18 at dark gate)	3	TOV% -0.033 (t -4.0), foul_rate -0.035 (t -3.7), fouls -0.024 (t -3.5)
null	2	FT% +0.014 (t +0.6), FT_rate -0.025 (t -1.0)

Thirteen of eighteen primary channels clear the dark gate; sixteen of eighteen clear 95%.

Same-population comparability. Restricting the sample to the 72,650 rows that satisfy the flagship residual panel's inclusion rules (all three expectations present, orientation resolvable) leaves the same 2,582 team-season clusters as the reconstruction, 16 rows short of the reconstruction's 72,666. On that exact comparison frame, **twelve of eighteen channels clear the dark gate; every direction survives; the major production channels sharpen** (offensive_efficiency -0.170 t -6.3, eFG% -0.119 t -7.1, TS% -0.109 t -7.2, assist_rate -0.174 t -6.6, 2P% -0.100 t -5.4, 3P% -0.096 t -4.5, TRB% -0.049 t -4.4; TOV% -0.028 t -3.0, foul_rate -0.042 t -4.2, fouls -0.031 t -4.1 all still clear).

The record's sentence:

On the same expectation-covered games, apart from trivial channel missingness, relative-to-expectation performance deteriorates with entering streak depth, and expectation-free physical and event performance exhibits broad adverse deformation with that same depth.

That is the bridge.

Structural asymmetry inside the map. The 13 dark-gate channels show a pronounced control-related asymmetry, sorting by whose realization requires *executing against active resistance* versus *deliberate restraint*. The deteriorating channels require the player to successfully produce an outcome the defense is trying to prevent: shooting conversion, effective and true shooting, offensive efficiency, offensive and total rebounding, assisted-shot production, turnover-generated scoring. The three improving channels are the ones most directly susceptible to deliberate restraint: turnover avoidance, foul-rate avoidance, raw foul count. The continuum is not binary — shot selection is a choice, offensive rebounding depends on effort and positioning, and turnovers and fouls are partly forced by opponents — but the endpoints of the continuum sit at the two poles of the observed split. Under this interpretation,

FT% sits outside the proposed continuum (being neither restraint nor contested execution), and its null result is not a puzzle for the interpretation.

The plain-English hypothesis, preserved: *teams are trying to escape entropy.*

The scientific version, adopted: *the persistent improvement in turnover and foul control is consistent with teams actively resisting the duration-linked deterioration by tightening more directly controllable aspects of play. Whether the process being resisted is entropic remains open.*

The word doing the work is *actively*. The behavioral evidence is consistent with active adaptation; we do not observe intentions.

5.7 Prospective replication and the retracted tempo claim

Prospective 2025-26 replication. The 2025-26 season was staged under the frozen protocol and analyzed once. **The KenPom lane was disqualified before the run:** every available point-in-time KenPom snapshot for 2025-26 carried end-of-season values repeated across dates, and using it would have introduced look-ahead. The prospective run therefore used BPI and Torvik plus their two-way mean. On **11,916 team-game rows and 599 team-season clusters**:

lane	β per game of entering depth	SE	t	95% CI
BPI	-0.2964	0.0638	-4.65	[-0.4214, -0.1714]
Torvik	-0.1301	0.0569	-2.28	[-0.2417, -0.0185]
Mean(2)	-0.2133	0.0597	-3.57	[-0.3302, -0.0963]

All three are adverse and all intervals exclude zero. **The fresh season reproduces the important qualitative ordering** from the pooled eight-season repaired estimate: BPI strong, Torvik adverse but materially shallower. The 2025-26 result carries the KenPom-lane disqualification permanently disclosed on the record; no in-season KenPom snapshot for 2025-26 is used anywhere in this paper.

The pace claim from v0.6 §8, retracted. v0.6 §8 Table 4 reported “*Pace -0.07979 (t -5.8) slows*” and folded that into the behavioral fingerprint. A dedicated tempo instrument (cbb_tempo_adaptation_test_FREEZE_v1_2, SHA 773ed2ab87723e1e5698cb6d8e00abd176c18b573f351f9eedbc4c3387039690) was built and certified at 3 of 500 dark-gate false detections with sd t 0.9992 on 500 disjoint-seed planted-null worlds, with power to detect a slowdown of -0.05 possessions per unit DEPTH_SUM in 100% of worlds. **The real coefficient computed once was +0.004548 possessions per combined entering win, SE 0.008235, t = +0.552, 95% CI [-0.0116, +0.0207], on 50,397 contests.** The preregistered linear slowdown prediction was falsified within calendar stage. **The pace decomposition is retracted** and does not appear in the v0.8 channel map or interpretation. A threshold, late-onset, or otherwise nonlinear tempo response has not been tested; it is unexamined, not rescued.

6. The Falsification Battery

Coefficients below are the depth coefficient after adding the named controls. **Legacy falsification controls are carried from the pre-repair flagship unless a subsection is explicitly identified as a repaired-panel re-certification.** §§6.1, 6.2 and 6.5 are pre-repair historical controls. §6.3 is filed with its own construction. §§6.4 (mean-reversion horse race, §5.2) and 6.8 (calendar audit, §5.3) are genuine repaired-panel re-certifications.

6.1 Fatigue (pre-repair). Adding rest days and games played in the last nine days: depth remains approximately **-0.3062** (pre-repair KenPom lane), and the fatigue proxy itself is null.

6.2 Recent form (pre-repair). Adding the prior-three-game residual: depth remains approximately **-0.2717**, $t \approx -9$ (pre-repair KenPom lane). Recent form carries a small separately significant negative term but does not absorb depth.

6.3 The loss-streak placebo. Entering-loss-streak depth is opposite-signed at approximately **+0.29 ($t \approx +9.3$)** on the pre-repair frame; the memoryless own-construction null reproduces about **+0.25** of that mechanically, leaving an excess of approximately +0.04 that does not clear conventional significance. The placebo is weak asymmetry evidence: the sign structure is consistent with an order-specific burden, but on its own it argues only weakly against a generic any-streak artifact. §6.4 discriminates.

6.4 The mean-reversion horse race. See §5.2. On the repaired panel: KenPom joint depth -0.2743 with cumulative overperformance collapsing to -0.0011 ; BPI joint depth -0.3119 with cumulative $+0.0053$; Torvik joint depth $+0.0245$ with cumulative -0.0323 surviving.

6.5 Opponent preparation (the target-game story) (*pre-repair*). The full control adds opponent streak depth, opponent coming off a loss, opponent rest advantage, postseason and road indicators, and own rest and fatigue. Depth coefficient after all these enters: approximately **-0.2893 ($t -10.4$)** on **66,178 pre-repair rows**. The interaction with opponent-off-a-loss is null ($+0.008$), contrary to the target-game prediction that a desperate opponent should amplify the depth burden. The obvious target-game story took a serious hit; the soft-channel remainder (media attention, coaching emphasis, rivalry psychology) is unmeasured and stays [open].

6.6 Depth-definition audit. See §5.5.

6.7 The retracted pace claim. See §5.7.

6.8 The calendar audit. See §5.3. **This is a rebuttal to a real objection and it survives cleanly.**

7. The Null Discipline and the Selection-Geometry Floor

Conditioning on streak membership bends conditional averages even for a memoryless process. To be deep in a win streak, a team had to win the preceding games, and winning is correlated with having played well, so the games at the deep end of a selected streak sit, on average, somewhat closer to the team's own mean than the fortunate games that built the streak. A memoryless coin, flipped at each team's own rate, therefore produces some downward drift across its own selected win streaks. That drift is not a force; it is a property of conditioned averages, and a spreadsheet produces it.

The program's own-construction memoryless null (Track 1) quantifies this floor. Applied to the flagship: a mechanical floor of **-0.2238** on the pre-repair frame, leaving a floored excess of **-0.0969** beyond selection geometry, with 0 of 200 memoryless replicates producing a slope as negative as observed.

Two changes in v0.8 relative to v0.6. First, the floor is filed as an important internal falsification result, not as the paper's headline. The headline is the multi-ruler gradient (§5.1) and the expectation-free bridge (§5.6). Second, the floor and excess are pre-repair numbers; the repaired-panel refit of the floor is deferred to a subsequent memo and is disclosed here as pending rather than quietly used.

The program's record with its own null. A three-point-shooting streak-debt claim was killed by the coin at each team's own rate. An MLB starter-ramp signal was retired against a within-pitcher reorder null. An MLB team-game margin channel returned the compression footprint null three carriers deep. These kills are reported because they are the evidence that the surviving results are different in kind: the surviving results are the ones that beat the same instrument that correctly destroyed the weaker claims.

8. Cross-Substrate Evidence

A duration burden that existed only in one sport, under one construction, would invite the suspicion that it is a construction. The program tests substrate-appropriate instruments elsewhere, under one principle: each substrate expresses load through its own clock and channels, and the law-of-form standard does not require the same statistic across substrates, only the same form.

8.1 Tennis — Grade A confirmation

Tennis is a one-person ordered state with no team-season baseline. The paper's clock for tennis is relative position inside the streak (early, middle, and late thirds of streaks of length six or more). The score-thirds lifecycle-null instrument (freeze v0.1.4) tests whether score-channel within-streak aging survives residualization against tournament structure. The verdict-bearing null (Null B) estimates the expected average number of sets played per match from a coarse structural cell defined by tournament round, playing surface, era bin, best-of format, tournament level, and pre-match player and opponent strength bands, with a frozen collapse order and a minimum cell count of 30. The verdict statistic is the late-minus-early slope on the residual across the streak thirds. Within-cell residual permutation preserves late-round difficulty and erases only the within-streak aging path.

Synthetic validation passed both required fixtures before real-data contact, including the critical over-control safety case: in a structure-only world with no true within-streak aging but a raw late-minus-early slope of +0.671 driven by later matches being tougher rounds, Null B correctly did not light up (residual -0.002 , permutation fraction 0.85). The real run walked 131,187 matches and identified 2,825 eligible streaks (length ≥ 6) comprising 24,350 eligible matches.

Real-run result:

Null	Observed slope	Null median	Null band	Fraction $\geq$ observed	Role
Null B (verdict)	+0.0496	+0.0083	$[-0.0031, +0.0201]$	0.0000	verdict-bearing
Null A (sanity)	+0.2958	-0.0077	—	0.0000	diagnostic

Zero of 2,000 structure-preserving permutations reached or exceeded the observed residual slope of +0.0496; the observed slope exceeded the null median by +0.0412 and lay above the null band's upper bound of +0.0201. **Grade A** — prospective confirmation under a frozen structure-preserving lifecycle-null.

Public claim ceiling for tennis (filed): *"In tennis, within-streak score-channel aging was prospectively confirmed within a frozen structure-preserving lifecycle-null instrument. Later-phase wins required more sets even after accounting for tournament round, surface, era, format, tournament level, and pre-match strength structure. The result confirms the tennis lifecycle-aging form; it does not by itself establish universality, identify entropy as the causal mechanism, or establish every tennis compression channel."*

8.2 Soccer — Grade A replication in a different geometry

An English Premier League instrument was frozen to the basketball estimand: enter-depth on consecutive league wins, team-season fixed effects, opponent strength, home, and rest as covariates, significance assessed only against a within-team-season permutation null. The construction locally replicated enter-depth compression across shot, shots-on-target, and goal channels; a lower-division extension returned a within-football channel-shift result in which the goal-difference channel is robust and process channels are not. Soccer's parity structure differs from college basketball's extreme inequality, and the substrate-clock principle predicts exactly this kind of channel-routed, medium-strength expression rather than a copy of the basketball ladder. **Grade A** — prospective third-substrate confirmation under the frozen v0.7 assignment.

8.3 NHL — Grade B, reduced-ceiling holdout reproduction

The NHL discovery lane identified `hd_share_tied` (tied-state high-danger share) as the primary carrier compressing under accumulated winning-streak duration. Discovery corpus 2010-11 through 2017-18: 18,742 team-games, 9,252 positive-depth; discovery slope adverse in all eight seasons. Frozen holdout 2018-19 through 2025-26: 19,562 team-games, 9,654 positive-depth. **The pre-registered one-shot reproduction returned observed slope -0.002335 against 400 own-coin null draws with 154 exceedances; corrected empirical $p = 0.386534$.** Frozen decision rule was PASS iff 0 of 400 exceedances. **The primary carrier did not clear the frozen gate.**

The reproduction is filed under a **reduced evidentiary ceiling** owing to a documented governance defect in the opener chain (see `nhl_governance_memorial_v0_1.md` for the full defect ledger and byte-identity chain), and it is assigned **Grade B** under the frozen v0.7 hierarchy — a real-holdout run disclosed under a defined ceiling, with the discovery-selected primary carrier not surviving the frozen gate. The Priority 2 volume-migration hypothesis remains prospectively sealed. The program pre-registered a hard gate, executed it, and reported the failure.

8.4 CFB — Grade C, feasibility-blocked discovery candidate

The college-football lane identified `def_line_yards_allowed` as a coherent defensive-rushing carrier, supported by `def_stuff_rate` and `def_ypc_allowed`; the pre-designated offensive rushing-decomposition tier was null and the tested bilateral garbage-time explanation did not account for the discovery pattern. **The sealed 2016-2021 held-out corpus was then found underpowered for confirmation under the frozen power gate.** No 2016-2021 play-by-play channel outcomes were opened. **The recovery threshold was not lowered and the sealed outcomes remain unopened.** Grade C — feasibility-blocked discovery candidate, sealed window preserved.

8.5 NFL — Grade C, feasibility-blocked discovery candidate

The NFL lane identified **WR/TE yards after catch per reception at entering-depth 1-4** as the predicted-direction discovery candidate on the 2002-2020 discovery corpus. The pre-holdout power gate required 416 of 500 planted worlds to be recovered; **the actual power measurement recovered 172 of 500.** Under the frozen protocol the 2021-2025 real holdout was therefore never opened and remains sealed. Layer-1 verdict: feasibility-blocked discovery candidate. Grade C.

8.6 MLB — historical Grade D provenance; updated well-powered null ungraded pending taxonomy amendment

Major League Baseball is the program's hardest substrate and its clearest demonstration that the pipeline reports what it finds. Four adjudicated designs, 2015-2024 excluding 2020:

1. **Starting-pitcher process test** (K-BB% at team win-streak depth 5+ vs 2-4, pitcher-season FE): well-powered null. 1,626 five-plus starts, 742 episodes; primary coef -0.208 , one-sided p 0.279; secondaries and losing-streak diagnostic null.
2. **Pitcher opposite-tail matched-cell test**: structurally infeasible at the frozen 250-per-tail floor. Not a null; a feasibility boundary.
3. **Team offensive-opportunity opposite-tail test**: primary not supported with the estimate opposite the frozen direction (+1.27 pp RISP share in deep-win tails, p 0.90); panel not supported.
4. **HR localization discriminator** (win 6+ vs win 2-4): non-pass (coef -0.030 , p 0.386).

No tested MLB design detected the RDT footprint, including one design with genuine statistical power (branch 1) and one that failed at feasibility before outcomes (branch 2). The MLB substrate is **closed for RDT testing** under the current data modality; reopening requires either a new modality that defeats the identified confounds (pitch-level leverage controls at scale, or event-chain baserunning data) or a substrate-clock argument specifying why a different duration construction should express where these did not. **Historical grade: D (feasibility provenance in the frozen v0.7 ledger).** **Current Layer-1 status: well-powered null on the starting-pitcher process design, plus feasibility and boundary results on the others.** **Revised Layer-2 grade: not assigned, pending an explicit program taxonomy amendment (see §3).** Filed as substrate boundary, not evidence against the phenomenon where it is established.

Baseball was the theory's best chance to fail quietly and it took it. If we had found the same signal there that we found in basketball, you should have trusted us less.

8.7 Cross-substrate synthesis, at the evidentiary hierarchy

substrate	grade	status
CBB	A (established checkpoint)	multi-ruler gradient fully re-certified on the repaired panel; expectation-free bridge independently established on the repaired/event-linked frame
Tennis	A	prospective confirmation under structure-preserving lifecycle-null
Soccer	A	prospective third-substrate confirmation under permutation null in a different geometry
NHL	B	real-holdout reproduction under reduced ceiling; discovery-selected primary carrier did not clear the frozen gate
NFL	C	feasibility-blocked discovery candidate; 2021-2025 holdout sealed
CFB	C	feasibility-blocked discovery candidate; 2016-2021 holdout sealed
MLB	historical D (provenance); current Layer-2 not assigned	well-powered null on the primary; feasibility and boundary results on three others; substrate closed pending taxonomy amendment

The recurrence is asymmetric. CBB, tennis and soccer confirm; NHL did not clear its frozen gate; NFL and CFB were sealed at the power boundary before the holdout could open; MLB honestly does not detect. **The asymmetric record is inconsistent with a trivial one-size-fits-all positive artifact, but it does not by itself establish the law-of-form interpretation.** Cross-substrate variation is compatible with a substrate-appropriate law of form and with substrate-varying construction artifacts alike; separating those two requires the identification tests reserved to §10 and the still-open pre-registered composition tests.

9. Alternative Explanations and the Identification Frontier

This section states each surviving rival account, what has already constrained it, and what remains open.

9.1 Mean reversion. Constrained decisively on the primary rulers by the horse race of §5.2: on KenPom and BPI, banked cumulative overperformance collapses to insignificance in the joint model while duration survives. On Torvik the horse race inverts; §5.4 reports the trajectory that mechanically explains the shallower Torvik residual gradient and states that whether the same feature explains the horse-race inversion remains open. Simple cumulative-overperformance reversion is not a sufficient explanation on the two primary rulers.

9.2 Selection geometry. Quantified by the own-construction memoryless null (§7): a mechanical floor of -0.2238 on the pre-repair frame, leaving a floored excess of -0.0969 with zero of 200 memoryless replicates producing a slope as negative as observed. The floor is a floor the excess is measured *from*, not a rival explanation of the excess. The repaired-panel floor has not yet been refit and this is disclosed in §7 and §11.

9.3 Inflated-entry / entry-selection. Constrained by the stratification test of the earlier flagship, with 4,044 genuinely cold entries all classifying correctly and depth slopes present in every entry stratum (cold -1.2752 , neutral -1.2152 , hot -1.0035 ; all $t < -10$). A local-quality peak can explain a starting altitude but not a progressively deepening within-episode depth slope.

9.4 Composition across streak lengths — unresolved at confirmatory grade. Two prospectively governed confirmatory instruments reached the validation stage but failed their frozen sensitivity gates before a verdict-bearing real-data null was authorized: **FW2 at 74 of 500** planted worlds recovered against a required 416 of 500, and its

successor **FW3 at 371 of 500** at the frozen MEI of -0.10 . Both failures were filed as **power hard stops**, not as evidence for or against the CBB effect; the frozen FW3 order expressly says the 371/500 failure makes no inference concerning RDT. A downstream FW3-D exploratory result exists in the record but its status is **discovery-only / non-confirmatory** and it does not enter the main proof narrative.

9.5 Fatigue, schedule, opponent preparation. Constrained as reported in §6. The soft, unmeasured preparation channels (media attention, coaching emphasis, rivalry psychology) remain the principal conventional residual; the reported controls have materially constrained the principal conventional confounds.

9.6 True quality drifting upward during the streak. A rival in which the team is genuinely improving as the streak ages is addressed by argument rather than by an empirical test here. A team genuinely ascending in true quality with each win should post *widening* opponent-adjusted margins, the opposite of the observed compression and channel signature (tightening on controllable channels, eroding on less controllable ones).

9.7 Within-season calendar drift — closed as of §5.3. Adding season $\times$ week fixed effects to the flagship retains 97.7%–101.4% of the gradient across every lane. The calendar objection is closed.

The soft, unmeasured preparation channels remain the principal conventional residual. A critic who wishes to retreat to a bare *maybe-it-is-fatigue* must identify a specific defect in the reported controls or a materially different uncaptured mechanism; the reported controls have materially constrained the principal conventional confounds.

10. What Would Falsify This

A theory that cannot state its own kill conditions is not offering evidence. The following outcomes, each tied to a frozen or standing instrument, would force retraction or material rewriting.

- 1. Composition across streak lengths — pending taxonomy path.** Composition remains unresolved at confirmatory grade. Two prospectively governed confirmatory instruments reached the validation stage but hard-stopped on power (FW2 74/500, FW3 371/500 against required 416/500) before a verdict-bearing real-data null was authorized. **A sufficiently powered confirmatory instrument that fails to find the predicted adverse within-episode position ordering would materially lower the identification claim and leave pure level-composition compatible with the flagship result.** The current standing is *unresolved*, not *unfavorable*.
- 2. A refined own-construction null reproducing the full raw slope.** If a better memoryless null collapses the floored excess of §7 toward zero, the finding reverts to selection geometry, exactly as the program's own three-point-shooting and starter-ramp claims did.
- 3. Failure of prospective replication on future sealed seasons.** The first fresh-season replication, 2025-26, passed (§5.7). Subsequent prospectively frozen seasons remain a direct falsification path.
- 4. A better stratifier reversing the within-stratum ladder.** The inflated-entry stratification of §9 is defined on the program's chosen pre-streak local-quality measure; a rival account is free to propose a materially better stratifier, and if the resulting within-stratum ladder reverses, the entry-selection objection reopens.
- 5. A better ruler with a shallower gradient.** If a subsequently developed rating system that consumes the fingerprint channels documented in §5.6 produces a gradient closer to Torvik's (-0.10) than to KenPom/BPI (-0.28), the interpretation of the multi-ruler agreement of §5.1 shifts materially.
- 6. New in v0.8 — the expectation-free bridge failing on a subsequent cohort.** The bridge's core is *broad expectation-free adverse deformation of the production channels* under entering depth. If a future season's flagship-frame channel map does not reproduce that broad production deterioration, the bridge itself is not stable and the claim narrows. Failure to reproduce the two discipline gains (TOV%, foul rate) would weaken the control-asymmetry / active-resistance hypothesis of §5.6 and §11 but would not by itself weaken the bridge, because the discipline gains are a supported structural interpretation of the pattern, not a required part of the law.

The program has already retired claims under conditions like item 2, so these are not hypothetical standards.

11. Discussion and Interpretation Ceiling

What is on the table after the falsification battery, the floor, the calendar audit, the multi-ruler recovery, the horse-race re-certification with disclosed Torvik inversion, and the expectation-free bridge is a specific, bounded object: **within the same team-season, sustained winning order carries an opponent-adjusted performance cost that scales with the duration of the order, exceeded the own-construction memoryless benchmark on the pre-repair Track 1 frame, cannot be absorbed by accumulated displacement on the primary KenPom and BPI rulers, is not explained by fatigue, operational opponent preparation, calendar drift, or a target-game interaction, runs opposite in sign for losing order, expresses through distributed and directional channels, recurs in form on substrate-appropriate clocks in tennis and soccer, returns honest nulls in baseball while returning a non-clearing holdout in hockey under a reduced ceiling and feasibility boundaries in American football (NFL and CFB), and — the paper’s central new result — appears as broad expectation-free deformation of the underlying basketball performance system on the same games under the same duration variable.**

The order-specific duration-burden interpretation is **[supported]**. RDT calls this duration-indexed deformation of sustained ordered performance *entropression*; the measured coefficients and channel deformations are its substrate-specific signals; its causal generator remains **[open]**.

The compensation reading is [supported hypothesis], not established. The persistent improvement in turnover and foul control is consistent with teams actively resisting the duration-linked deterioration by tightening more directly controllable aspects of play. Whether the process being resisted is entropic remains **[open]**. In plain English: *teams are trying to escape entropy*. The plain-English framing is preserved because it captures what the pattern suggests; it is not the empirical conclusion.

Interpretation ceiling. The author’s causal position is that the recurring form reflects a general dissolutive pressure that sustained order accumulates against its own continuation, and the program’s name for that inferred pressure is entropic in spirit. That position is tagged **[speculative]** as mechanism. This paper measures the footprint; it does not measure the force, it asserts no thermodynamic identity, and nothing in Sections 3 through 10 depends on the interpretation. Whether the burden is law-like or is a deep emergent property of competitive systems (adaptation, information leakage, psychological load, or channels not yet named) is **[open]**, and may not be resolvable from box scores.

Two practical implications stand regardless of interpretation. For forecasting, the rating-lag sign reversal (a within-team sign of -0.28 becomes a cross-team sign of the opposite direction when team-season identity is not conditioned on) is a warning to any system that evaluates streaking teams with lagged ratings across teams. For the streak literature, the paper’s separation of duration from displacement, its subtraction of an own-construction selection floor, and — new in v0.8 — its expectation-free bridge on the same games are portable disciplines: the hot hand debate spent decades on exactly the selection subtleties that the floor and the bridge make explicit.

One methodological carry-forward for future substrate work, on the record. *Future substrate work must allow for and test the possibility that an adaptive system’s response to duration-linked deterioration becomes part of the observed fingerprint; it must not assume either a clean unopposed baseline or that a compensatory response necessarily exists.* If entropression is real and general, different systems may resist differently, and some may not exhibit an identifiable resistance response at all. That has to be discovered, not required.

12. Claim-Status Ledger

#	Claim	Status	Basis
1	Within team-season, opponent-adjusted margin declines with entering win-streak depth on the primary rulers	[established] (checkpoint frame)	§5.1, §5.7
2	The gradient is not a KenPom peculiarity: KenPom, BPI, and Mean/Median composites agree; Torvik registers the same direction at smaller magnitude	[established]	§5.1, §5.4
3	Duration, not displacement, carries the burden on the primary rulers	[supported] via re-certification of the frozen horse-race specification on the repaired panel (Torvik inversion recorded); the re-certification is a robustness run against a re-surfaced objection, not a fresh prospective preregistration	§5.2, §6.4
4	The raw slope contains a mechanical selection-geometry component; the claim is the excess beyond the own-construction null	[established] (Track 1 verified: floor -0.2238 , excess -0.0969 , 0/200 replicates more negative; pre-repair frame)	§7
5	The gradient survives within-season calendar drift	[established] via a narrow re-certification run performed after the calendar objection was surfaced by the tempo test; not prospective, and disclosed as such: season $\times$ week FE retains 97.7%–101.4% of the flagship gradient	§5.3, §6.8
6	The gradient survives fatigue, recent form, and operational opponent-preparation controls; loss-streak placebo runs opposite in sign; channel expression is distributed and directional	[supported]	§5.6, §6
7	The Torvik expectation trajectory rises with depth for KenPom and BPI and falls for Torvik, mechanically explaining the shallower Torvik residual gradient; whether the same feature explains the horse-race inversion remains open	[established, empirical] for the trajectory measurement and its mechanical relationship to the residual gradient; [open] for the horse-race inversion mechanism	§5.4
8	Why Torvik's expectation trajectory differs from KenPom's and BPI's	[open]	documentary architecture audit

#	Claim	Status	Basis
9	The same form recurs in tennis under a structure-preserving lifecycle-null and in soccer under a permutation null in a different geometry	[supported]	§8.1, §8.2
10	No tested MLB design detected the RDT footprint; the NHL discovery-selected carrier did not clear its frozen holdout gate under a reduced ceiling; NFL and CFB were sealed at the power boundary before the holdout could open	[established] for each substrate as filed; [open] as to whether alternative constructions could express in the substrates that did not confirm	§8.3, §8.4, §8.5, §8.6
11	Within-streak ordering is not a composition artifact	[open] — unresolved at confirmatory grade: FW2 (74/500) and FW3 (371/500 at MEI −0.10) hard-stopped on frozen sensitivity gates before a verdict-bearing real-data null could be authorized; both are power hard stops, not inference concerning RDT; FW3-D remains discovery-only / non-confirmatory	§9, §10
12	The depth slope does not require inflated entry to exist	[supported] (stratification run, all four strata negative including 4,044 cold entries)	§9
13	New in v0.8. On the same expectation-covered games, apart from trivial channel missingness, raw event performance channels broadly deform adversely with entering streak depth, while turnover and foul control persistently improve	[established]	§5.6
14	New in v0.8. The persistent improvement in turnover and foul control is consistent with teams actively resisting the duration-linked deterioration by tightening more directly controllable aspects of play	[supported hypothesis]	§5.6, §11

#	Claim	Status	Basis
15	v0.6's pace-slowdown claim	[retracted] — falsified by a dedicated tempo instrument, +0.005 possessions per combined entering win, t +0.55, CI excludes any linear slowdown steeper than −0.012	§5.7
16	The pressure is a general dissolutive force on sustained order	[speculative] as mechanism	§11

Every substantive claim above is tagged, and nothing tagged higher than its ledger row appears anywhere in this paper.

13. Reproducibility, Governance, and Disclosures

Pre-registration and post-hoc robustness — the two are kept separate. The v0.6 out-of-sample specification was locked in writing and hashed before the 2021-22 season was loaded (hash cb04061f7e1a00d5); the tempo test was frozen and hashed before its real coefficient was computed (freeze SHA 773ed2ab87723e1e5698cb6d8e00abd176c18b573f3). The flagship channel inventory was preregistered per objA_real_census_PREREGISTRATION_v1_0 and the frozen inventory is objective_channel_inventory_v1_4_FROZEN. **The calendar audit (cbb_calendar_audit_v1_0) and the horse-race re-certification (cbb_horserace_recertification_v1_0) are narrow robustness runs performed after their respective objections were surfaced,** not prospective preregistrations; their value comes from reproducing the frozen baseline coefficient on the reconstructed panel and from the tightness of their specifications (season × week FE only; the frozen v0.6 horse-race specification only), not from anti-forking provenance. Both are filed with that provenance explicit.

Panel authority. All CBB numbers in this paper except explicitly labeled pre-repair values are on w4_rebuild_repaired_v2.json (SHA 50b5b9c5a87d9ecb6b90391f3c40a5e9f1a93515fa5c3081a0e76d4345889db4). The 18–29% attenuation between the pre-repair and repaired panels across every lane is disclosed in §4 with the same weight §6.6 gives every falsification battery result.

Standing governance sequence. Specification freeze, independent adversarial cross-review, synthetic validation against planted ground truth with gating fixtures, separately authorized execution, and separately authorized null runs. No construction changes are permitted after synthetic output unless a gate fails and the instrument returns to specification review; no metric selection is permitted after a null.

Adversarial review. Every instrument in this program was reviewed adversarially by two independent large-language-model systems operating under different vendors, one as builder and auditor and one as peer reviewer, with disagreements treated as calibration signals and the author as sole adjudicator of every canon decision. The author further discloses that AI systems assisted in instrument construction, drafting, and audit under that structure; all analytical decisions, adjudications, and this paper's claims are the author's.

Data sources, by layer. The **expectation-free channel layer** of §5.6 uses a single homogeneous **ESPN raw-event summary corpus** (55,323 events; every counted event/score/date drawn from the ESPN summary payload keyed by stable ESPN event id), with no Torvik field, no legacy box field, and no cross-source patch entering Object A; that homogeneity is central to why the expectation-free bridge is independently valuable. The **expectation-adjusted flagship layer** of §§5.1–5.5 uses the panel authority w4_rebuild_repaired_v2.json for schedules and outcomes, with point-in-time expected-margin snapshots from kenpom.com (subscription), ESPN BPI, and barttorvik.com, taken with no future leakage; those inputs are separate from the ESPN raw-event summary corpus that carries

the bridge. Tennis uses the public Sackmann ATP match dataset (1968-2023 fork). Soccer uses public football-data.co.uk league files. NHL uses fastRhockey and hockeyR. CFB uses collegefootballdata.com play-by-play. NFL uses nflverse play-by-play. MLB uses public Retrosheet and Statcast data as pre-registered per branch.

References

Cited works

- Bar-Eli, M., Avugos, S., and Raab, M. (2006). Twenty years of hot hand research: Review and critique. *Psychology of Sport and Exercise*, 7(6), 525–553.
- Galton, F. (1886). Regression towards mediocrity in hereditary stature. *Journal of the Anthropological Institute of Great Britain and Ireland*, 15, 246–263.
- Gilovich, T., Vallone, R., and Tversky, A. (1985). The hot hand in basketball: On the misperception of random sequences. *Cognitive Psychology*, 17(3), 295–314.
- Lopez, M. J., Matthews, G. J., and Baumer, B. S. (2018). How often does the best team win? A unified approach to understanding randomness in North American sport. *Annals of Applied Statistics*, 12(4), 2483–2516.
- Miller, J. B., and Sanjurjo, A. (2018). Surprised by the hot hand fallacy? A truth in the law of small numbers. *Econometrica*, 86(6), 2019–2047.
- Tversky, A., and Kahneman, D. (1971). Belief in the law of small numbers. *Psychological Bulletin*, 76(2), 105–110.

Data sources

- Pomeroy, K. Adjusted efficiency ratings and archives (kenpom.com, subscription; point-in-time snapshots).
- ESPN. Basketball Power Index (BPI) point-in-time snapshots.
- ESPN. College basketball raw-event summary corpus (55,323 events; keyed by stable ESPN event id).
- Bart Torvik. T-Rank (barttorvik.com) point-in-time snapshots.
- Sackmann, J. ATP match results dataset (tennis_atp), community fork covering 1968–2023.
- football-data.co.uk. English league match data files.
- fastRhockey / hockeyR. NHL play-by-play and season files.
- nflverse. NFL play-by-play (nflfastR / nflreadr).
- CollegeFootballData (CFBD). FBS play-by-play and schedule feeds (collegefootballdata.com).
- Retrosheet. Major League Baseball event files (retrosheet.org).
- Statcast / MLB Advanced Media. Pitch-level and event-chain data as pre-registered per branch.

Market prices appear nowhere in the doctrine or the analysis. Artifact inventories, frozen specifications, synthetic-validation reports, and regression outputs are maintained in the program record and are available to reviewers on request.

Preprint v0.8. Not peer-reviewed. Comments to jcastra@gmail.com.