A Cumulant Decomposition Reveals Modality-Specific Varentropy Effects on Reading Time

Anonymous ACL submission

Abstract

Prior work reports that sub-Shannon Rényi entropies predict reading times better than Shannon entropy, but does not say why. We close the gap with an identity: in nats, the Taylor expansion of H_α about $\alpha = 1$ has *varentropy* $V = \mathrm{Var}_p(-\ln p(W))$ as its second-order term, so any sub-Shannon order picks up V at leading order. The identity generates three testable cognitive predictions about V, run on four reading-time corpora using GPT-2 small under Pimentel et al.'s (2023) chunked-stride protocol. The HC3 in-sample partial coefficient on V is positive on all three self-paced corpora (NS $\hat{\beta} = +2.39$, $p = 6 \times 10^{-12}$, $n = 10\,215$; UCL SPR $+1.15$, $p = 4 \times 10^{-4}$, $n = 3\,513$; SAP Fillers $+2.51$, $p = 0.08$, $n = 542$), but only NS passes the stricter paired-fold CV criterion ($+0.23 \times 10^{-2}$ nats/obs, $p = 0.002$). V predicts inter-reader RT log-variance on all three SPR corpora ($p \in [10^{-7}, 0.013]$) and fails on every eye-tracking contrast. The surprisal–varentropy interaction is strongest on Provo eye-tracking ($\hat{\beta}_{h \times V} \geq +3.5$, $p \leq 0.027$ on GD, TRT, GPT). Tail-truncated re-extraction localises the signal to the moderate tail of GPT-2's softmax: effects survive top-$p = 0.95$ but largely collapse at top-$k = 100$. Across modalities, varentropy's main-effect and dispersion signatures attach to self-paced reading while its slope-modulation signature attaches to eye-tracking—the sub-Shannon Rényi advantage is explained by a cumulant expansion of surprise whose visibility on reading time depends on what the measurement modality lets readers do.

1 Introduction

Pimentel et al. (2023) report an under-explored empirical fact: contextual *Rényi* entropy of order $\alpha < 1$ predicts reading times better than Shannon entropy. The mathematical reason was left open. We close the gap with an identity, derive three cognitive predictions from it, and run them on four corpora.

The identity. In nats, the Taylor expansion of H_α about $\alpha = 1$ gives

$$\mathrm{H}_\alpha = \mathrm{H}_1 + \tfrac{1}{2}(1-\alpha)\, V + \tfrac{1}{6}(1-\alpha)^2\, \kappa_3 + \cdots, \quad (1)$$

where $V = \mathrm{Var}_p(-\ln p(W))$ is *varentropy*, the variance of surprise, and κ_n is its nth cumulant (Appendix B). The MGF representation of Rényi entropy is itself a known result in information theory (Campbell, 1965; van Erven and Harremoës, 2014); the contribution here is to use it as an explanation of the Pimentel et al. (2023) sub-Shannon advantage and to derive three testable cognitive predictions from it. Any sub-Shannon $\alpha < 1$ picks up V at leading order. The identity does not single out $\alpha = 1/2$. The cognitive claim it motivates—that readers are sensitive to V—is then what our three contrasts on the next page test.

Three contrasts.

- **F1**. Residual $V^{\perp H} = V \mid \mathrm{H}_1$ improves held-out RT prediction.
- **F2**. The interaction $h \times V$ on RT is positive (V amplifies the surprisal slope).
- **F3**. V predicts between-reader RT log-variance.

Headline finding. On *all three* self-paced reading corpora, the HC3 in-sample partial coefficient on V is positive, and the strict CV held-out test passes on Natural Stories ($n = 10{,}215$). On UCL SPR ($n = 3{,}513$) and SAP Fillers ($n = 542$) the CV gain is null even though the HC3 direction is correct; the CV-null cells are exactly the lower-n SPR corpora, consistent with a small partial effect of V on mean SPR RT whose magnitude is too small relative to fold-level noise to translate into held-out predictive gain at smaller n. F3 (varentropy on between-reader RT log-variance) passes on all three SPR corpora and fails on every eye-tracking contrast. F2 (the $h \times V$ interaction) sits on Provo eye-tracking, where V amplifies the surprisal slope

without raising mean dwell. On eye-tracking the two F1 tests *agree negatively*: three of four UCL ET CV residuals are significantly *wrong-signed* and their HC3 coefficients are near zero. The varentropy contribution the cumulant identity extracts from sub-Shannon Rényi entropy attaches differently across modalities: the F1 main effect and F3 dispersion effect are SPR-skewed, while the F2 slope-modulation is an eye-tracking effect.

2 Methods

Corpora. **Natural Stories** (Futrell et al., 2021): 10 English stories, 181 readers, non-cumulative moving-window self-paced reading. After RT filtering ($100 \leq \text{RT} \leq 3000$ ms) and sentence-boundary trimming, $n = 10{,}215$ word tokens. **SAP Fillers** (Huang et al., 2024): the filler subset of the Syntactic Ambiguity Processing Benchmark (40 sentences, 2,000 self-paced readers; `osf.io/b6rqh`); $n = 542$ word tokens after the authors' accuracy/exclusion filters and our sentence-boundary trim. Small n makes SAP a supportive replication rather than a primary test corpus. **UCL** (Frank et al., 2013): 361 isolated English sentences with both self-paced and eye-tracking measurements ($n = 3{,}513$ SPR; $n = 1{,}111$ ET). **Provo** (Luke and Christianson, 2018): 55 short passages, 84 readers, eye-tracking, $n = 2{,}510$ word tokens. We report all four fixation measures (FFD, GD, TRT, GPT) under both word-skipping conventions of Pimentel et al. (2023): *skip-excluded* (drop fixations with RT$= 0$) and *skip2zero* (keep skipped words at RT $= 0$).

Language model and extraction. We follow Pimentel's GPT-2 setup: GPT-2 small via HuggingFace, max chunk length 1022 tokens, BOS prepended and EOS appended to each chunk, sliding-window decoding with STRIDE $= 200$ (drop the first STRIDE-1 predictions of every chunk after the first).[1] Surprisal of each word is summed over its byte-pair subwords; entropy and varentropy are read off the predictive distribution at the first subword (Pimentel's convention). All predictors are z-scored.

Baseline M0. Pimentel's twelve-predictor baseline: surprisal, word length, log-frequency at lags $t, t{-}1, t{-}2, t{-}3$, plus intercept. For each corpus we average RTs across readers (one observation per word token) and fit OLS with Gaussian errors.

F1. Held-out 10-fold Δllh in 10^{-2} nats/observation, comparing M0 $+$ H$_1$ $+ V$ to M0 $+$ H$_1$ on the same paired folds. Under OLS the held-out gain isolates exactly $V^{\perp H}$. Significance: paired-permutation p (20 000 sign-flips) and 95% paired-bootstrap CI (20 000 resamples). We also fit the joint model with HC3 robust SE (Long and Ervin, 2000) and report the in-sample $\hat{\beta}_V$. A *pass* requires both $\Delta\text{llh} > 0$ with paired-permutation $p < 0.05$ *and* the bootstrap CI excluding zero.

F2. On M0$+$H$_1 + V$ we add $h \cdot V$ and report the HC3 coefficient and p.

F3. Per word, compute $\log \text{Var}_{\text{readers}}(\text{RT})$ over readers with RT > 0 (drop words with < 5 contributing readers). Regress on M0 $+$H$_1 + V$ with HC3 SE.

Tail-robust varentropy. Varentropy is sensitive to tail estimation because rare-word surprisals contribute quadratically. We recompute V under four truncation schemes (top-k with $k \in \{10, 100\}$; top-p with $p \in \{0.95, 1.0\}$, renormalising the kept mass, minimum two tokens kept) and rerun all three tests. The $p = 1.0$ variant includes all vocabulary tokens and is a numerical sanity check against standard V.

3 Results

Table 1 reports the three tests with standard V.

3.1 F1: residual V on mean RT

F1 passes the strict CV criterion only on Natural Stories: residual V adds $+0.23 \times 10^{-2}$ nats per observation over Shannon entropy alone (paired-permutation $p = 0.002$), with HC3 $\hat{\beta}_V = +2.39$, $p = 5.95 \times 10^{-12}$.

The CV and HC3 tests probe different aspects of the relationship: CV asks if V improves out-of-fold predictions, HC3 asks if a partial linear effect of V exists holding M0$+$H$_1$ fixed. The two diverge in a way consistent with lower power and fold-level noise at smaller n across the SPR corpora. NS ($n = 10{,}215$): both pass. UCL SPR ($n = 3{,}513$): HC3 $\hat{\beta}_V = +1.15$ ($p = 4.3 \times 10^{-4}$), CV $+0.15$ ($p_{\text{perm}} = 0.21$). SAP Fillers ($n = 542$): HC3 $\hat{\beta}_V = +2.51$ ($p = 0.083$), CV -0.08 ($p = 0.89$). **All three SPR corpora have positive in-sample partial V**

[1] See `rycolab/anticipation-on-reading-times`, `src/h01_data/models/gpt2.py`.

Corpus / measure	n	F1 Δllh_V (p_{perm})	F2 $\hat{\beta}_{h\times V}$ (p)	F3 $\hat{\beta}_V$ (p)
Self-paced reading				
Natural Stories (SPR)	10,215	$+\mathbf{0.23}$ ($\mathbf{0.002}$)	$+1.10$ (0.08)	$+\mathbf{0.035}$ ($\mathbf{1.46\times10^{-7}}$)
SAP Fillers (SPR)	542	-0.08 (0.89)	-2.26 (0.36)	$+\mathbf{0.042}$ ($\mathbf{0.013}$)
UCL SPR	3,513	$+0.15$ (0.21)	$+0.45$ (0.20)	$+\mathbf{0.029}$ ($\mathbf{0.008}$)
Eye-tracking				
Provo / GD (skip-excl)	2,510	-0.04 (0.26)	$+\mathbf{3.95}$ ($\mathbf{0.003}$)	-0.02 (0.13)
Provo / GD (skip2zero)	2,510	-0.03 (0.68)	$+\mathbf{3.47}$ ($\mathbf{0.009}$)	-0.02 (0.13)
Provo / FFD (skip-excl)	2,510	$+0.01$ (0.87)	$+0.72$ (0.07)	-0.01 (0.18)
Provo / TRT (skip-excl)	2,510	-0.01 (0.86)	$+\mathbf{5.09}$ ($\mathbf{0.005}$)	-0.02 (0.07)
Provo / GPT (skip-excl)	2,510	$+0.03$ (0.71)	$+\mathbf{5.34}$ ($\mathbf{0.014}$)	-0.01 (0.55)
UCL ET / FFD	1,111	-0.13 ($\mathbf{0.019}^\star$)	-0.25 (0.76)	$+0.03$ (0.22)
UCL ET / GD	1,111	-0.12 ($\mathbf{0.031}^\star$)	-1.30 (0.26)	$+0.02$ (0.50)
UCL ET / RBT	1,111	-0.09 ($\mathbf{0.036}^\star$)	-1.44 (0.31)	$+0.04$ (0.16)
UCL ET / GPT	1,111	-0.06 (0.47)	$+0.41$ (0.88)	$+0.03$ (0.38)

Table 1: F1/F2/F3 with standard varentropy. F1 in 10^{-2} nats/obs, paired-permutation p. F2/F3: HC3 coefficient and p. Bold = pass at $p < 0.05$; $\star$ on F1 marks a CV result with $p < 0.05$ in the *wrong* direction. Provo F3 reported for GD; FFD/TRT/GPT values are similar in direction (all β near zero, $p \geq 0.07$) and are in Appendix C. The n column refers to F1/F2; UCL ET F3 uses $n = 1{,}103$ after the reader-variance filter (words with < 5 contributing readers dropped).

coefficients, and the corpora that fail CV are also the ones with smaller n. With only three SPR corpora we cannot firmly attribute the CV failures to n alone, but the pattern is consistent with V having a partial effect on mean SPR RT that is too small relative to fold-level noise to register as a CV gain at modest sample sizes. We do not claim F1 passes on UCL SPR or SAP under the strict CV criterion.

On eye-tracking the CV and HC3 tests *agree negatively*. Across the eight Provo contrasts Δllh_V is small and centred on zero. On three of four UCL ET contrasts (FFD, GD, RBT) the CV residual is significantly negative ($p_{\text{perm}} \in [0.019, 0.036]$) and the HC3 partial coefficient is near zero or wrong-signed. Adding V on top of H_1 makes held-out ET prediction *worse*, not better. This is qualitatively different from the SPR pattern: ET F1 failures are not an underpowered-CV problem; on the mean-RT contrast both the CV and HC3 tests agree that V does not help. (F2, separately, *is* positive on Provo ET; see §3.2.)

3.2 F2: surprisal–varentropy interaction

F2 is strongest and most consistent on Provo eye-tracking. The interaction $h \times V$ is positive with HC3 SE on three of four fixation measures and both skip conventions (GD, TRT, GPT, all $p \leq 0.027$; FFD borderline at $p = 0.07$), with HC3 effect sizes $\hat{\beta}_{h\times V}$ between $+3.5$ and $+5.3$. F2 also appears on Natural Stories at borderline significance ($+1.10$, $p = 0.08$). It is null on UCL self-paced reading and null or wrong-signed on UCL eye-tracking. The interpretation: on Provo, V does not raise mean dwell, but it scales how strongly surprise translates into dwell.

3.3 F3: V on between-reader RT log-variance

F3 passes on *all three* self-paced corpora: Natural Stories ($\hat{\beta}_V = +0.035$, $p = 1.5\times10^{-7}$, larger than the parallel entropy coefficient $\hat{\beta}_{\text{H}} = +0.021$, $p = 0.005$); UCL SPR ($\hat{\beta}_V = +0.029$, $p = 0.008$; $\hat{\beta}_{\text{H}} = -0.011$, $p = 0.43$); and SAP Fillers ($\hat{\beta}_V = +0.042$, $p = 0.013$, $n = 542$). Varentropy is a stronger and more consistent predictor of between-reader RT log-variance than Shannon entropy is: entropy itself reaches significance on Natural Stories but not on UCL SPR or SAP, whereas varentropy is significant on all three. F3 fails on every eye-tracking contrast in both corpora ($\hat{\beta}_V$ between -0.02 and $+0.04$, all $p \geq 0.07$; Appendix C).

Tail location of the signal. Recomputing V under top-k ($k \in \{10, 100\}$) and top-p ($p \in \{0.95, 1.0\}$) truncation (Appendix C, Appendix D) localises where in the predictive distribution most of the varentropy signal lives. The top-$p = 1.0$ sanity-check column reproduces standard V to numerical precision. Beyond the sanity check: every effect we report survives top-$p = 0.95$ except Provo FFD F2 and Provo GPT skip2zero F2 (the latter borderline at $p = 0.051$). NS F1, NS F3, and the Provo F2 effect on GD weaken or collapse at top-$k = 100$ and are fully null at top-$k = 10$. The two F3 effects on smaller SPR corpora are more truncation-robust:

UCL SPR F3 and SAP F3 both survive top-$k = 100$ ($p = 0.010$ and $p = 0.033$) and are marginal at top-$k = 10$ ($p = 0.063$ and $p = 0.065$). The bulk of the varentropy signal therefore sits in the moderate tail of GPT-2's softmax—past the top-100 tokens but inside the 95% probability mass—except on UCL SPR F3 and SAP F3, where some signal remains even when the distribution is aggressively trimmed.

4 Discussion

Where varentropy is and is not visible. Equation (1) converts the Pimentel et al. (2023) sub-Shannon advantage into predictions about specific moments of surprise; it does not single out $\alpha = 1/2$ and does not require F1–F3 to hold on every corpus. The empirical picture is then: across all three self-paced corpora the HC3 in-sample partial coefficient on V is positive; only Natural Stories ($n = 10{,}215$) passes the stricter CV test, with the two smaller-n SPR corpora ($n = 3{,}513$ and $n = 542$) showing positive in-sample direction but CV-null gain; F3 passes on all three SPR corpora; on every eye-tracking F1 contrast both the CV and HC3 mean-RT tests agree that V does not help, with three CV cells significantly wrong-signed. F2, separately, is a positive eye-tracking effect on Provo. The varentropy main-effect and dispersion signatures cluster on self-paced reading; the slope-modulation signature (F2) clusters on Provo eye-tracking. Because Provo and UCL ET use entirely different stimuli and readers, the joint failure across both suggests modality dependence rather than corpus-specific noise. We caution that $\mathrm{H}_{1/2}$ alone is itself a positive predictor of mean RT on some eye-tracking cells (e.g. Provo GPT skip-excl, $+1.00$ over M0, $p = 0.0037$; Appendix D), so the modality split is a fact about varentropy specifically, not about every sub-Shannon information-theoretic quantity.

A regression-distribution mechanism, tested. One hypothesis for the split is that SPR forces every word's processing demand into a single button-press interval, while eye-tracking lets readers distribute demand across refixations and go-past time. We test this directly on Provo by regressing four regression-mediated outcomes on $\mathrm{M0}+\mathrm{H}_1+V+hV$ with HC3 SE (Appendix A). The $h \times V$ interaction is essentially zero on regression-out probability and count ($p > 0.9$) and not significant on go-past$-$GD ($+1.29$ ms, $p = 0.38$). The strict mechanism is not supported; the F2 pattern on Provo is qualitatively what it predicts but is not driven by elevated regression rates that we can detect.

Limitations

GPT-2 small only; replicating with a larger LM (Oh and Linzen, 2025; Grattafiori et al., 2024) is the natural next step. SAP Fillers has only 542 word tokens after sentence-boundary trimming, which limits the power of held-out CV contrasts on that corpus (F1 and F2 are CV-null even though the HC3 in-sample direction is positive for F1). We tested the regression-distribution mechanism on Provo directly and found it not supported (Appendix A); we do not have an alternative mechanism for the SPR–ET split.

Use of AI Assistants

Anthropic Claude was used to assist with prose editing, to scaffold analysis scripts in Python, and to cross-check the cumulant-expansion derivation. All experimental design decisions, statistical methodology, falsification criteria, and final wording are the author's; every number reported in the paper was audited against the underlying CSVs produced by the analysis pipeline.

5 Conclusion

The Taylor expansion of Rényi entropy in nats has varentropy as its second-order term. This identity converts the Pimentel et al. (2023) sub-Shannon advantage into three concrete predictions about varentropy. On four corpora spanning two modalities, F3 (varentropy on between-reader RT log-variance) passes on all three self-paced corpora; F1 (varentropy main effect) passes cleanly on Natural Stories and is in-sample positive elsewhere on SPR; F2 (the $h \times V$ interaction) is an eye-tracking effect on Provo. The varentropy signatures the cumulant identity extracts attach differently across modalities rather than being uniformly absent in either.

References

L. L. Campbell. 1965. A coding theorem and Rényi's entropy. *Information and Control*, 8(4):423–429.

Stefan L. Frank, Irene Fernandez Monsalve, Robin L. Thompson, and Gabriella Vigliocco. 2013. Reading time data for evaluating broad-coverage models of English sentence processing. *Behavior Research Methods*, 45(4):1182–1190.

Richard Futrell, Edward Gibson, Harry J. Tily, Idan Blank, Anastasia Vishnevetsky, Steven T. Piantadosi, and Evelina Fedorenko. 2021. The Natural Stories corpus: a reading-time corpus of English texts containing rare syntactic constructions. *Language Resources and Evaluation*, 55(1):63–77.

Grattafiori et al. 2024. The Llama 3 herd of models. *arXiv preprint arXiv:2407.21783*.

Kuan-Jung Huang, Suhas Arehalli, Mari Kugemoto, Christian Muxica, Grusha Prasad, Brian Dillon, and Tal Linzen. 2024. Large-scale benchmark yields no evidence that language model surprisal explains syntactic disambiguation difficulty. *Journal of Memory and Language*, 137:104510.

J. Scott Long and Laurie H. Ervin. 2000. Using heteroscedasticity consistent standard errors in the linear regression model. *The American Statistician*, 54(3):217–224.

Steven G. Luke and Kiel Christianson. 2018. The Provo corpus: A large eye-tracking corpus with predictability norms. *Behavior Research Methods*, 50(2):826–833.

Byung-Doh Oh and Tal Linzen. 2025. To model human linguistic prediction, make LLMs less superhuman. *arXiv preprint arXiv:2510.05141*.

Tiago Pimentel, Clara Meister, Ethan G. Wilcox, Roger P. Levy, and Ryan Cotterell. 2023. On the effect of anticipation on reading times. *Transactions of the Association for Computational Linguistics*, 11:1624–1642.

Jeroen S. van Erven and Peter Harremoës. 2014. Rényi divergence and Kullback-Leibler divergence. *IEEE Transactions on Information Theory*, 60(7):3797–3820.

A Mechanism test on Provo

Outcome (per-word reader avg)	$\hat{\beta}_{h\times V}$	SE	p	95% CI
Pr(regression-out)	-0.0003	0.0023	0.91	$[-0.005, +0.004]$
regression-out count	-0.0001	0.0026	0.96	$[-0.005, +0.005]$
go-past − GD (ms)	$+1.29$	1.47	0.38	$[-1.58, +4.16]$
log(go-past − GD + 1)	$+0.022$	0.018	0.24	$[-0.014, +0.057]$

Direct test of the regression-distribution mechanism on Provo. HC3 coefficient on $h \times V$ for four regression-mediated outcomes ($n = 2{,}510$). The strict mechanism predicts positive coefficients; none are significant, and the two discrete regression measures are essentially zero.

B Rényi entropy as the CGF of surprise

All quantities here are in nats; the bits conversion is at the end. Let $I = -\ln p(W)$ be the surprise of a random word and $S(\alpha) = \sum_w p(w)^\alpha$. Rényi entropy in nats is

$$\mathrm{H}_\alpha = \frac{\ln S(\alpha)}{1-\alpha}. \tag{2}$$

Rewrite S as an expectation under p: $S(\alpha) = \mathbb{E}_p[e^{-(\alpha-1)I}]$. With $t = 1-\alpha$, $S(\alpha) = M_I(t)$, the MGF of I, and $\mathrm{H}_\alpha = K_I(t)/t$ where $K_I = \ln M_I$ is the CGF of I. Taylor-expanding K_I about $t = 0$ and dividing by t gives

$$\mathrm{H}_\alpha = \sum_{n=0}^{\infty} \frac{(-1)^n \kappa_{n+1}}{(n+1)!} (\alpha-1)^n, \tag{3}$$

where $\kappa_1 = \mathrm{H}_1$, $\kappa_2 = V$, $\kappa_3, \ldots$ are higher cumulants. The first two terms give Equation (1). In bits, the coefficient of V is $\frac{\ln 2}{2}(1-\alpha)$ when V is measured in bits^2.

Proposition 1 (Rényi = CGF of surprise)**.** *With $t = 1-\alpha$ and $I = -\ln p(W)$, $\mathrm{H}_\alpha = K_I(t)/t$. The Taylor series of $K_I(t)/t$ about $t = 0$ has nth coefficient $\kappa_{n+1}/(n+1)!$.*

C Full F1/F2/F3 table across V variants

Tables 2 to 4 report every (corpus, measure, skip-convention) under every V variant. Cells show the test statistic and its p-value; **bold** = pass at $p < 0.05$ in the predicted direction; <u>underlined</u> = $p < 0.05$ in the wrong direction. The top-$p = 1.0$ column reproduces standard V to numerical precision (sanity check). The machine-readable table is at `results/all_tests_pimentel.csv`.

F1

F2

F3

Reading the tables. The standard-V and top-$p = 1.0$ columns are near-identical for every cell, as required by the sanity check. F1 passes only on Natural Stories standard V and the two top-p variants; CV wrong-signed cells concentrate on UCL ET. F2 on Provo GD/TRT and GPT skip-excl survives top-$p = 0.95$ (Provo GPT skip2zero borderline at $p = 0.051$; Provo FFD does not survive on either skip variant); none of these survives either top-k truncation. F3 on UCL SPR is positive under *every* variant and significant under all except top-$k = 10$ ($p = 0.063$, marginal); F3 on SAP Fillers is positive under *every* variant and significant at standard V, top-$k = 100$, top-$p = 0.95$, and top-$p = 1.0$ ($p \in [0.013, 0.033]$), borderline at top-$k = 10$ ($p = 0.065$); F3 on Natural Stories survives top-p but is borderline-to-null at top-k.

Corpus / measure	standard	top-$k = 10$	top-$k = 100$	top-$p = 0.95$	top-$p = 1.0$
NS SPR	$+0.23$ (0.00)	$+0.01$ (0.49)	$+0.02$ (0.06)	$+0.18$ (0.00)	$+0.23$ (0.00)
SAP Fillers	-0.08 (0.89)	$+0.02$ (0.98)	-0.02 (0.99)	-0.23 (0.67)	-0.08 (0.89)
UCL SPR	$+0.15$ (0.21)	$+0.09$ (0.42)	$+0.24$ (0.06)	$+0.12$ (0.27)	$+0.15$ (0.21)
Provo GD (skip-excl)	-0.04 (0.26)	$+0.03$ (0.73)	$+0.02$ (0.88)	$\underline{-0.05}$ (0.00)	-0.04 (0.26)
Provo GD (skip2zero)	-0.02 (0.68)	$\underline{-0.03}$ (0.01)	$\underline{-0.04}$ (0.01)	-0.02 (0.88)	-0.02 (0.68)
Provo FFD (skip-excl)	$+0.01$ (0.87)	$+0.08$ (0.33)	$+0.03$ (0.58)	-0.01 (0.91)	$+0.01$ (0.87)
Provo FFD (skip2zero)	-0.01 (0.54)	-0.02 (0.45)	-0.01 (0.58)	-0.01 (0.62)	-0.01 (0.54)
Provo TRT (skip-excl)	-0.01 (0.86)	-0.00 (0.99)	$+0.02$ (0.87)	-0.03 (0.24)	-0.01 (0.86)
Provo TRT (skip2zero)	$\underline{-0.02}$ (0.01)	$\underline{-0.03}$ (0.00)	-0.03 (0.27)	-0.03 (0.12)	$\underline{-0.02}$ (0.01)
Provo GPT (skip-excl)	$+0.03$ (0.71)	$+0.02$ (0.94)	$+0.08$ (0.66)	-0.02 (0.76)	$+0.03$ (0.71)
Provo GPT (skip2zero)	-0.03 (0.38)	-0.05 (0.58)	-0.03 (0.79)	-0.05 (0.11)	-0.03 (0.38)
UCL ET FFD	$\underline{-0.13}$ (0.02)	$\underline{-0.15}$ (0.00)	$\underline{-0.17}$ (0.02)	$\underline{-0.16}$ (0.00)	$\underline{-0.13}$ (0.02)
UCL ET GD	$\underline{-0.12}$ (0.03)	$\underline{-0.13}$ (0.01)	-0.12 (0.21)	-0.13 (0.06)	$\underline{-0.12}$ (0.03)
UCL ET RBT	$\underline{-0.09}$ (0.04)	$\underline{-0.10}$ (0.03)	-0.10 (0.07)	$\underline{-0.11}$ (0.00)	$\underline{-0.09}$ (0.04)
UCL ET GPT	-0.06 (0.46)	-0.07 (0.40)	$\underline{-0.14}$ (0.03)	-0.07 (0.10)	-0.06 (0.46)

Table 2: F1 held-out Δllh in 10^{-2} nats/obs (with paired-permutation p for F1, HC3 p for F2/F3) under each V variant. **Bold** = pass ($p < 0.05$, predicted direction); <u>underlined</u> = $p < 0.05$ in wrong direction. The top-$p = 1.0$ column reproduces standard V as a sanity check.

Corpus / measure	standard	top-$k = 10$	top-$k = 100$	top-$p = 0.95$	top-$p = 1.0$
NS SPR	$+1.10$ (0.08)	-0.21 (0.74)	-0.00 (1.00)	$+1.02$ (0.13)	$+1.10$ (0.08)
SAP Fillers	-2.26 (0.36)	-0.96 (0.81)	-2.03 (0.60)	-2.40 (0.23)	-2.26 (0.36)
UCL SPR	$+0.44$ (0.19)	$+0.33$ (0.34)	$+0.30$ (0.39)	$+0.21$ (0.53)	$+0.44$ (0.19)
Provo GD (skip-excl)	$+3.95$ (0.00)	-1.56 (0.36)	-0.43 (0.77)	$+4.49$ (0.00)	$+3.95$ (0.00)
Provo GD (skip2zero)	$+3.47$ (0.01)	-1.82 (0.30)	-0.59 (0.69)	$+3.86$ (0.01)	$+3.47$ (0.01)
Provo FFD (skip-excl)	$+0.72$ (0.07)	$+0.27$ (0.52)	$+0.75$ (0.06)	$+0.54$ (0.18)	$+0.72$ (0.07)
Provo FFD (skip2zero)	$+0.30$ (0.57)	$+0.13$ (0.82)	$+0.68$ (0.21)	-0.07 (0.90)	$+0.30$ (0.57)
Provo TRT (skip-excl)	$+5.09$ (0.01)	-2.17 (0.33)	-0.92 (0.64)	$+5.84$ (0.00)	$+5.09$ (0.01)
Provo TRT (skip2zero)	$+4.79$ (0.01)	-2.62 (0.25)	-1.20 (0.55)	$+5.51$ (0.01)	$+4.79$ (0.01)
Provo GPT (skip-excl)	$+5.34$ (0.01)	-0.36 (0.87)	$+1.02$ (0.62)	$+5.06$ (0.02)	$+5.34$ (0.01)
Provo GPT (skip2zero)	$+4.26$ (0.03)	-0.81 (0.70)	$+0.47$ (0.80)	$+3.99$ (0.05)	$+4.26$ (0.03)
UCL ET FFD	-0.25 (0.75)	-0.15 (0.84)	$+0.22$ (0.78)	$+0.00$ (1.00)	-0.25 (0.75)
UCL ET GD	-1.30 (0.26)	-0.92 (0.40)	-0.67 (0.57)	-1.01 (0.38)	-1.30 (0.26)
UCL ET RBT	-1.44 (0.31)	-0.78 (0.56)	-0.43 (0.76)	-1.45 (0.29)	-1.44 (0.31)
UCL ET GPT	$+0.41$ (0.88)	-2.02 (0.46)	-0.93 (0.75)	$+0.65$ (0.80)	$+0.41$ (0.88)

Table 3: F2 interaction HC3 $\hat{\beta}_{h \times V}$ (with paired-permutation p for F1, HC3 p for F2/F3) under each V variant. **Bold** = pass ($p < 0.05$, predicted direction); <u>underlined</u> = $p < 0.05$ in wrong direction. The top-$p = 1.0$ column reproduces standard V as a sanity check.

Corpus / measure	standard	top-$k = 10$	top-$k = 100$	top-$p = 0.95$	top-$p = 1.0$
NS SPR	$+0.03$ (1e-07)	$+0.01$ (0.29)	$+0.01$ (0.07)	$+0.03$ (5e-05)	$+0.03$ (1e-07)
SAP Fillers	$+0.04$ (0.01)	$+0.03$ (0.07)	$+0.04$ (0.03)	$+0.04$ (0.03)	$+0.04$ (0.01)
UCL SPR	$+0.03$ (0.01)	$+0.02$ (0.06)	$+0.03$ (0.01)	$+0.03$ (0.01)	$+0.03$ (0.01)
Provo GD (skip-excl)	-0.02 (0.13)	-0.02 (0.17)	-0.02 (0.16)	-0.02 (0.13)	-0.02 (0.13)
Provo GD (skip2zero)	-0.02 (0.13)	-0.02 (0.17)	-0.02 (0.16)	-0.02 (0.13)	-0.02 (0.13)
Provo FFD (skip-excl)	-0.01 (0.18)	-0.01 (0.37)	-0.01 (0.55)	-0.01 (0.20)	-0.01 (0.18)
Provo FFD (skip2zero)	-0.01 (0.18)	-0.01 (0.37)	-0.01 (0.55)	-0.01 (0.20)	-0.01 (0.18)
Provo TRT (skip-excl)	-0.02 (0.07)	-0.01 (0.39)	-0.02 (0.16)	-0.02 (0.12)	-0.02 (0.07)
Provo TRT (skip2zero)	-0.02 (0.07)	-0.01 (0.39)	-0.02 (0.16)	-0.02 (0.12)	-0.02 (0.07)
Provo GPT (skip-excl)	-0.01 (0.55)	$+0.00$ (0.84)	-0.01 (0.65)	$+0.00$ (0.96)	-0.01 (0.55)
Provo GPT (skip2zero)	-0.01 (0.55)	$+0.00$ (0.84)	-0.01 (0.65)	$+0.00$ (0.96)	-0.01 (0.55)
UCL ET FFD	$+0.03$ (0.22)	$+0.01$ (0.63)	$+0.02$ (0.44)	$+0.03$ (0.28)	$+0.03$ (0.22)
UCL ET GD	$+0.02$ (0.50)	$+0.00$ (0.89)	$+0.00$ (0.92)	$+0.01$ (0.67)	$+0.02$ (0.50)
UCL ET RBT	$+0.04$ (0.16)	$+0.02$ (0.52)	$+0.02$ (0.51)	$+0.03$ (0.24)	$+0.04$ (0.16)
UCL ET GPT	$+0.03$ (0.38)	$+0.04$ (0.38)	$+0.02$ (0.72)	$+0.03$ (0.52)	$+0.03$ (0.38)

Table 4: F3 HC3 $\hat{\beta}_V$ in $\log \mathrm{Var}$ (with paired-permutation p for F1, HC3 p for F2/F3) under each V variant. **Bold** = pass ($p < 0.05$, predicted direction); <u>underlined</u> = $p < 0.05$ in wrong direction. The top-$p = 1.0$ column reproduces standard V as a sanity check.

D Held-out Δllh of H, V, $H_{1/2}$ at each truncation

We also compute the held-out 10-fold Δllh of four models against the M0 baseline—$M0 + \mathrm{H}_1$, $M0 + \mathrm{H}_1 + V$, $M0 + \mathrm{H}_{1/2}$, and the joint $M0 + \mathrm{H}_1 + V + \mathrm{H}_{1/2}$—at each truncation, computing *all three quantities* at the same truncation level. The final column reports the paired difference $(M0 + \mathrm{H}_1 + V) - (M0 + \mathrm{H}_{1/2})$, which tests whether the leading-order Taylor truncation suffices.

Natural Stories (SPR)

V variant	+H	+H+V	+$\mathrm{H}_{1/2}$	+H+V+$\mathrm{H}_{1/2}$	(H+V) - $\mathrm{H}_{1/2}$
standard	+0.09 (0.20)	+0.32 (0.00)	+0.36 (0.01)	+0.52 (0.00)	−0.04 (0.47)
top-$k = 10$	+0.01 (0.66)	+0.00 (0.96)	+0.01 (0.62)	−0.01 (0.65)	−0.01 (0.01)
top-$k = 100$	+0.00 (0.87)	+0.01 (0.67)	+0.01 (0.59)	+0.01 (0.71)	−0.00 (0.65)
top-$p = 0.95$	+0.08 (0.24)	+0.27 (0.00)	+0.19 (0.06)	+0.31 (0.00)	+0.08 (0.01)
top-$p = 1.0$	+0.09 (0.20)	+0.32 (0.00)	+0.36 (0.01)	+0.52 (0.00)	−0.04 (0.47)

SAP Fillers (SPR)

V variant	+H	+H+V	+$\mathrm{H}_{1/2}$	+H+V+$\mathrm{H}_{1/2}$	(H+V) - $\mathrm{H}_{1/2}$
standard	−0.07 (0.86)	−0.16 (0.88)	+0.10 (0.87)	+0.06 (0.93)	−0.26 (0.21)
top-$k = 10$	−0.15 (0.36)	−0.10 (0.88)	−0.13 (0.60)	−0.24 (0.74)	+0.04 (0.95)
top-$k = 100$	−0.11 (0.56)	−0.14 (0.99)	−0.07 (0.88)	−0.27 (0.98)	−0.07 (0.97)
top-$p = 0.95$	−0.08 (0.81)	−0.31 (0.74)	−0.06 (0.93)	−0.25 (0.72)	−0.24 (0.39)
top-$p = 1.0$	−0.07 (0.86)	−0.16 (0.88)	+0.10 (0.87)	+0.06 (0.93)	−0.26 (0.21)

UCL SPR

V variant	+H	+H+V	+$\mathrm{H}_{1/2}$	+H+V+$\mathrm{H}_{1/2}$	(H+V) - $\mathrm{H}_{1/2}$
standard	−0.02 (0.26)	+0.12 (0.31)	+0.03 (0.64)	+0.11 (0.38)	+0.09 (0.26)
top-$k = 10$	−0.03 (0.44)	+0.05 (0.48)	−0.03 (0.00)	+0.02 (0.78)	+0.08 (0.30)
top-$k = 100$	−0.02 (0.53)	+0.19 (0.10)	−0.02 (0.30)	+0.16 (0.16)	+0.21 (0.05)
top-$p = 0.95$	−0.03 (0.20)	+0.10 (0.41)	−0.00 (0.98)	+0.04 (0.72)	+0.10 (0.25)
top-$p = 1.0$	−0.02 (0.26)	+0.12 (0.31)	+0.03 (0.64)	+0.11 (0.38)	+0.09 (0.26)

Provo GD (skip-excl)

V variant	+H	+H+V	+$\mathrm{H}_{1/2}$	+H+V+$\mathrm{H}_{1/2}$	(H+V) - $\mathrm{H}_{1/2}$
standard	+0.05 (0.56)	+0.01 (0.85)	+0.16 (0.19)	+0.13 (0.24)	−0.14 (0.02)
top-$k = 10$	+0.04 (0.58)	+0.13 (0.48)	+0.12 (0.34)	+0.28 (0.24)	+0.01 (0.93)
top-$k = 100$	+0.02 (0.84)	+0.03 (0.84)	+0.09 (0.48)	+0.12 (0.57)	−0.06 (0.46)
top-$p = 0.95$	+0.04 (0.63)	−0.00 (0.97)	+0.06 (0.47)	−0.02 (0.88)	−0.07 (0.05)
top-$p = 1.0$	+0.05 (0.56)	+0.01 (0.85)	+0.16 (0.19)	+0.13 (0.24)	−0.14 (0.02)

Provo GD (skip2zero)

V variant	+H	+H+V	+$\mathrm{H}_{1/2}$	+H+V+$\mathrm{H}_{1/2}$	(H+V) - $\mathrm{H}_{1/2}$
standard	−0.06 (0.14)	−0.08 (0.10)	−0.05 (0.01)	−0.05 (0.46)	−0.03 (0.47)
top-$k = 10$	−0.05 (0.01)	−0.08 (0.00)	−0.05 (0.02)	−0.04 (0.56)	−0.03 (0.00)
top-$k = 100$	−0.05 (0.45)	−0.09 (0.07)	−0.06 (0.20)	−0.10 (0.16)	−0.03 (0.18)
top-$p = 0.95$	−0.05 (0.17)	−0.07 (0.19)	−0.04 (0.32)	−0.09 (0.15)	−0.03 (0.79)
top-$p = 1.0$	−0.06 (0.14)	−0.08 (0.10)	−0.05 (0.01)	−0.05 (0.46)	−0.03 (0.47)

Provo FFD (skip-excl)

V variant	+H	+H+V	+$\mathrm{H}_{1/2}$	+H+V+$\mathrm{H}_{1/2}$	(H+V) - $\mathrm{H}_{1/2}$
standard	−0.02 (0.29)	−0.01 (0.83)	−0.02 (0.53)	+0.08 (0.58)	+0.01 (0.89)
top-$k = 10$	−0.01 (0.85)	+0.04 (0.64)	−0.04 (0.01)	+0.29 (0.09)	+0.08 (0.38)
top-$k = 100$	+0.00 (1.00)	+0.01 (0.89)	−0.03 (0.12)	+0.12 (0.46)	+0.04 (0.51)
top-$p = 0.95$	−0.02 (0.45)	−0.03 (0.56)	−0.03 (0.05)	−0.06 (0.56)	−0.00 (0.98)
top-$p = 1.0$	−0.02 (0.29)	−0.01 (0.83)	−0.02 (0.53)	+0.08 (0.58)	+0.01 (0.89)

Provo FFD (skip2zero)

V variant	+H	+H+V	+$\mathrm{H}_{1/2}$	+H+V+$\mathrm{H}_{1/2}$	(H+V) - $\mathrm{H}_{1/2}$
standard	+0.27 (0.11)	+0.26 (0.11)	+0.20 (0.17)	+0.24 (0.12)	+0.06 (0.32)
top-$k = 10$	+0.20 (0.03)	+0.21 (0.04)	+0.22 (0.03)	+0.24 (0.04)	−0.01 (0.61)
top-$k = 100$	+0.37 (0.03)	+0.38 (0.02)	+0.39 (0.03)	+0.35 (0.03)	−0.01 (0.68)
top-$p = 0.95$	+0.26 (0.11)	+0.26 (0.11)	+0.28 (0.09)	+0.24 (0.14)	−0.03 (0.37)
top-$p = 1.0$	+0.27 (0.11)	+0.26 (0.11)	+0.20 (0.17)	+0.24 (0.12)	+0.06 (0.32)

Provo TRT (skip-excl)

V variant	+H	+H+V	+$\mathrm{H}_{1/2}$	+H+V+$\mathrm{H}_{1/2}$	(H+V) - $\mathrm{H}_{1/2}$
standard	+0.16 (0.34)	+0.15 (0.37)	+0.28 (0.16)	+0.26 (0.19)	−0.13 (0.03)
top-$k = 10$	+0.06 (0.61)	+0.10 (0.60)	+0.13 (0.44)	+0.22 (0.34)	−0.03 (0.66)
top-$k = 100$	+0.07 (0.65)	+0.08 (0.73)	+0.15 (0.48)	+0.14 (0.55)	−0.06 (0.31)
top-$p = 0.95$	+0.14 (0.36)	+0.12 (0.45)	+0.18 (0.29)	+0.11 (0.45)	−0.06 (0.02)
top-$p = 1.0$	+0.16 (0.34)	+0.15 (0.37)	+0.28 (0.16)	+0.26 (0.19)	−0.13 (0.03)

Provo TRT (skip2zero)

V variant	+H	+H+V	+$\mathrm{H}_{1/2}$	+H+V+$\mathrm{H}_{1/2}$	(H+V) - $\mathrm{H}_{1/2}$
standard	−0.02 (0.69)	−0.04 (0.43)	+0.01 (0.93)	−0.01 (0.92)	−0.05 (0.12)
top-$k = 10$	−0.03 (0.59)	−0.05 (0.42)	−0.03 (0.70)	−0.02 (0.87)	−0.03 (0.02)
top-$k = 100$	−0.05 (0.22)	−0.08 (0.15)	−0.05 (0.44)	−0.09 (0.23)	−0.03 (0.06)
top-$p = 0.95$	−0.02 (0.66)	−0.05 (0.37)	−0.02 (0.69)	−0.06 (0.30)	−0.03 (0.54)
top-$p = 1.0$	−0.02 (0.69)	−0.04 (0.43)	+0.01 (0.93)	−0.01 (0.92)	−0.05 (0.12)

Provo GPT (skip-excl)

V variant	+H	+H+V	+$\mathrm{H}_{1/2}$	+H+V+$\mathrm{H}_{1/2}$	(H+V) - $\mathrm{H}_{1/2}$
standard	+0.72 (0.01)	+0.76 (0.01)	+1.00 (0.00)	+0.98 (0.01)	−0.25 (0.03)
top-$k = 10$	+0.38 (0.11)	+0.45 (0.21)	+0.53 (0.07)	+0.60 (0.18)	−0.08 (0.80)
top-$k = 100$	+0.55 (0.03)	+0.64 (0.09)	+0.78 (0.03)	+0.75 (0.16)	−0.13 (0.20)
top-$p = 0.95$	+0.68 (0.01)	+0.67 (0.01)	+0.75 (0.01)	+0.62 (0.02)	−0.08 (0.24)
top-$p = 1.0$	+0.72 (0.01)	+0.76 (0.01)	+1.00 (0.00)	+0.98 (0.01)	−0.25 (0.03)

Provo GPT (skip2zero)

V variant	+H	+H+V	+$\mathrm{H}_{1/2}$	+H+V+$\mathrm{H}_{1/2}$	(H+V) - $\mathrm{H}_{1/2}$
standard	+0.24 (0.11)	+0.20 (0.21)	+0.31 (0.09)	+0.27 (0.18)	−0.11 (0.14)
top-$k = 10$	+0.08 (0.53)	+0.04 (0.76)	+0.12 (0.44)	+0.08 (0.70)	−0.07 (0.01)
top-$k = 100$	+0.12 (0.39)	+0.09 (0.58)	+0.16 (0.35)	+0.06 (0.80)	−0.07 (0.07)
top-$p = 0.95$	+0.22 (0.11)	+0.18 (0.23)	+0.23 (0.12)	+0.13 (0.41)	−0.05 (0.33)
top-$p = 1.0$	+0.24 (0.11)	+0.20 (0.21)	+0.31 (0.09)	+0.27 (0.18)	−0.11 (0.14)

UCL ET / FFD

V variant	+H	+H+V	+$\mathrm{H}_{1/2}$	+H+V+$\mathrm{H}_{1/2}$	(H+V) - $\mathrm{H}_{1/2}$
standard	+0.20 (0.36)	+0.07 (0.73)	+0.01 (0.98)	−0.03 (0.90)	+0.06 (0.71)
top-$k = 10$	+0.14 (0.35)	−0.02 (0.88)	+0.15 (0.37)	−0.09 (0.64)	−0.17 (0.01)
top-$k = 100$	+0.17 (0.36)	−0.00 (1.00)	+0.15 (0.53)	−0.11 (0.54)	−0.15 (0.11)
top-$p = 0.95$	+0.21 (0.36)	+0.05 (0.80)	+0.10 (0.60)	+0.03 (0.88)	−0.05 (0.61)
top-$p = 1.0$	+0.20 (0.36)	+0.07 (0.73)	+0.01 (0.98)	−0.03 (0.90)	+0.06 (0.71)

UCL ET / GD

V variant	+H	+H + V	+$H_{1/2}$	+H+V+$H_{1/2}$	(H+V) - $H_{1/2}$
standard	+0.30 (0.15)	+0.18 (0.38)	+0.14 (0.23)	+0.08 (0.78)	+0.04 (0.83)
top-$k=10$	+0.15 (0.48)	+0.09 (0.71)	+0.19 (0.42)	+0.05 (0.87)	−0.11 (0.06)
top-$k=100$	+0.18 (0.44)	+0.07 (0.77)	+0.20 (0.46)	−0.03 (0.93)	−0.12 (0.09)
top-$p=0.95$	+0.30 (0.16)	+0.17 (0.39)	+0.24 (0.14)	+0.22 (0.40)	−0.07 (0.39)
top-$p=1.0$	+0.30 (0.15)	+0.18 (0.38)	+0.14 (0.23)	+0.08 (0.78)	+0.04 (0.83)

UCL ET / RBT

V variant	+H	+H + V	+$H_{1/2}$	+H+V+$H_{1/2}$	(H+V) - $H_{1/2}$
standard	+0.22 (0.31)	+0.12 (0.56)	+0.04 (0.68)	+0.03 (0.94)	+0.09 (0.67)
top-$k=10$	+0.13 (0.52)	+0.05 (0.80)	+0.15 (0.51)	+0.04 (0.88)	−0.10 (0.05)
top-$k=100$	+0.12 (0.58)	+0.03 (0.93)	+0.13 (0.61)	−0.06 (0.81)	−0.10 (0.10)
top-$p=0.95$	+0.23 (0.31)	+0.12 (0.58)	+0.13 (0.35)	+0.19 (0.53)	−0.02 (0.89)
top-$p=1.0$	+0.22 (0.31)	+0.12 (0.56)	+0.04 (0.68)	+0.03 (0.94)	+0.09 (0.67)

UCL ET / GPT

V variant	+H	+H + V	+$H_{1/2}$	+H+V+$H_{1/2}$	(H+V) - $H_{1/2}$
standard	+0.23 (0.49)	+0.16 (0.54)	−0.08 (0.47)	+0.08 (0.80)	+0.24 (0.32)
top-$k=10$	+0.25 (0.50)	+0.02 (0.95)	+0.20 (0.63)	+0.01 (0.98)	−0.18 (0.13)
top-$k=100$	+0.22 (0.53)	+0.02 (0.91)	+0.16 (0.73)	−0.07 (0.86)	−0.13 (0.34)
top-$p=0.95$	+0.24 (0.45)	+0.17 (0.56)	+0.05 (0.80)	+0.12 (0.58)	+0.12 (0.31)
top-$p=1.0$	+0.23 (0.49)	+0.16 (0.54)	−0.08 (0.47)	+0.08 (0.80)	+0.24 (0.32)

Reading the head-to-head tables. Three patterns. (1) On Natural Stories at standard / top-$p = 1.0$, $\mathrm{H}_1 + V$ and $\mathrm{H}_{1/2}$ are statistically interchangeable ($p_{\text{perm}} = 0.47$); the joint model wins at $\Delta\text{llh} = +0.52$, $p = 0.002$. At top-$p = 0.95$, $\mathrm{H}_1 + V$ *beats* $\mathrm{H}_{1/2}$ ($+0.08$, $p = 0.012$): the leading-order truncation suffices once the most pathological tail tokens are removed. At top-$k=100$ no individual model beats M0 at $p < 0.05$; at top-$k = 10$ a tiny but significant wrong-signed $(\mathrm{H}_1+V)-\mathrm{H}_{1/2}$ gap appears (-0.006, $p = 0.005$), of negligible practical magnitude. (2) On Provo GD at standard, $\mathrm{H}_{1/2}$ *beats* $\mathrm{H}_1 + V$ (-0.14, $p = 0.02$): on this corpus the higher cumulants in $\mathrm{H}_{1/2}$ carry signal the two-term truncation drops. At top-$p=0.95$ the gap shrinks to -0.07 (still $p = 0.046$), and disappears under both top-k cuts. (3) On UCL eye-tracking, every contrast is null or wrong-signed at every truncation.