The $3^{k-1}/4^k$ Distribution of Steiner Sentence Lengths

Speculative Results Series, Paper 65

Jon Seymour
jon@wildducktheories.com
Sydney, Australia

July 2026

Abstract

We define a *Steiner sentence* as a maximal sequence of consecutive Steiner circuits in a Collatz orbit, from the Syracuse image of one node $\equiv 5 \pmod 8$ to the next. Each Steiner circuit in the sentence is a *Steiner word*. The sentence length k is the number of words.

A naive analysis of the Steiner circuit state machine suggests that the 50/50 split at the 3-node should produce sentence lengths geometrically distributed with parameter 1/2, giving $P(k) = (1/2)^k$. Empirical sampling of 10^5 sentences via uniformly random entry points $n = 8t + 5$ instead yields:

$$P(\text{sentence length} = k) \;=\; \frac{3^{k-1}}{4^k}.$$

We report the full empirical data: the sentence-start distribution (uniform on $\{1, 3, 5, 7\} \pmod 8$, as predicted), the within-sentence Steiner word entry distribution (suppressed at class 5, approximately equal at 1, 3, 7), the sentence length distribution with χ^2 goodness-of-fit tests, and the implied effective termination probability of 1/4 per word. Both the observed distribution and its explanation are stated as conjectures here; both are proved in Paper 67 [2].

Status. Speculative results paper (stream 64+k, $k = 1$). All claims are empirical and supported by χ^2 tests. This paper is superseded by Paper 67 [2], which provides a complete proof and incorporates this paper's experimental record in full.

Contents

1 Background and Notation

We work with odd positive integers throughout. The *2-adic valuation* $v_2(n)$ is the largest power of 2 dividing n.

Definition 1.1 (Syracuse map)**.** The *Syracuse map* $S : \mathbb{O} \to \mathbb{O}$ is

$$S(n) \;=\; \frac{3n+1}{2^{v_2(3n+1)}},$$

which maps odd positive integers to odd positive integers by performing one odd Collatz step followed by all mandatory even halvings.

1.1 Mod-8 Step Taxonomy

Every odd integer n falls into one of four residue classes modulo 8:

n mod 8	$v_2(3n+1)$	Step type	Exit class $S(n)$ mod 8
1	2	OEE	1 (always)
3	1	OE	1 or 5
5	≥ 3	OEEE$^+$	(uniform on $\{1,3,5,7\}$ per step)
7	1	OE	3 or 7

1.2 Steiner Circuits

Definition 1.2 (Steiner circuit)**.** Let $a_0 \in \mathbb{O}$ with $a_0 + 1 = 2^\alpha m$, m odd, $\alpha \geq 1$. The *Steiner circuit* starting at a_0 is the composite of α consecutive Syracuse steps from a_0, producing the output node

$$b \;=\; \frac{3^\alpha m - 1}{2^\beta}, \qquad \beta = v_2(3^\alpha m - 1).$$

The *Steiner parameter* $\alpha = v_2(a_0 + 1)$ counts the Syracuse steps.

Lemma 1.3 (3-node split, proved in Paper 33)**.** *For* $n \equiv 3 \pmod 8$*, write* $n = 8a+3$*. Then* $S(n) = 12a + 5$*, which satisfies* $12a + 5 \equiv 1 \pmod 8$ *iff* $n \equiv 11 \pmod{16}$*, and* $\equiv 5 \pmod 8$ *iff* $n \equiv 3 \pmod{16}$*. The split is exactly* $1/2$ *each, determined by the parity of* a*.*

1.3 Steiner Sentences

Definition 1.4 (Steiner sentence and Steiner word)**.** A *Steiner sentence* is a maximal sequence of consecutive Steiner circuits $W_1, W_2, \ldots, W_k$ such that $W_1, \ldots, W_{k-1}$ each have output $b \not\equiv 5 \pmod 8$ and W_k has output $b \equiv 5 \pmod 8$. Each circuit W_i in the sentence is called a *Steiner word*. The *sentence length* is $k \geq 1$.

In the mod-8 alphabet, a sentence of length k matches the language

$$((7^*3)?1)^{k-1}\,(7^*3)?5.$$

A sentence of length $k = 1$ is a single Steiner word whose output is $\equiv 5 \pmod 8$.

Remark 1.5. Steiner sentences partition the Collatz orbit into segments between consecutive values $\equiv 5 \pmod 8$. The sentence starts at the Syracuse image $b_0 = S(n)$ of a node $n \equiv 5 \pmod 8$ and ends at the next node $\equiv 5 \pmod 8$. The starting node n is the terminal word of the *previous* sentence.

2 The Naive Prediction and Why It Fails

From Lemma 1.3, every Steiner word that visits the 3-node terminates the sentence with probability exactly 1/2. A naive model treats the sentence as a sequence of independent 3-node coin flips, giving:

$$P(\text{sentence length} = k) \;\overset{?}{=}\; \left(\frac{1}{2}\right)^k.$$

This is decisively rejected by the data (Section 4). The structural explanation is provided in Paper 67 [2]: the forbidden transitions $7 \to \{1, 5\}$ (class 7 cannot directly reach classes 1 or 5) reduce the effective per-word termination probability from 1/2 to 1/4.

3 Sampling Methodology

1. Draw t uniformly at random from $[1, T_{\max}]$ with $T_{\max} = 10^{15}$.
2. Set $n = 8t + 5$ (so $n \equiv 5 \pmod 8$).
3. Compute $b_0 = S(n)$, the output of the Steiner circuit starting at n. This is the first Steiner word of the new sentence.
4. Follow consecutive Steiner circuits from b_0, counting words, until the first output $b \equiv 5 \pmod 8$. Record the word count as k.

We collected $N = 10^5$ independent sentences with seed 42 and $T_{\max} = 10^{15}$. The node n is the terminal word of the previous sentence and is not counted. Independence between samples is ensured by the random draw of t; no orbit tails are shared.

4 Empirical Results

4.1 Sentence-start distribution

The first Steiner word of each sentence starts at $b_0 = S(n)$, $n \equiv 5 \pmod 8$. By Paper 67 Theorem 2.1, the Syracuse transition matrix from class 5 is uniform, so b_0 mod 8 should be uniform on $\{1, 3, 5, 7\}$.

b_0 mod 8	Count	Observed fraction	Expected (uniform)
1	25181	0.25181	0.25000
3	25023	0.25023	0.25000
5	24852	0.24852	0.25000
7	24944	0.24944	0.25000

Table 1: Sentence-start distribution: $b_0 = S(n)$, $n \equiv 5 \pmod 8$, $N = 10^5$ samples. The distribution is uniform on $\{1, 3, 5, 7\}$ as predicted.

4.2 Within-sentence Steiner word entry distribution

The entry class of each Steiner word within a sentence determines its termination behaviour. Class 5 words always terminate the sentence, so they can only appear as the final word — their frequency within sentences is structurally suppressed.

Entry mod8	Count	Fraction	Visits 3-node?	Term. prob.
1	125450	0.31333	no	0
3	125119	0.31250	yes	1/2
5	24852	0.06207	no	1 (terminal only)
7	124955	0.31209	yes	1/2
Total	400376			

Table 2: Within-sentence Steiner word entry distribution across all 10^5 sentences (400376 total words). Class 5 appears only as the terminal word, giving frequency ≈1/16 rather than the uniform 1/4. The 3-node fraction (classes 3+7) is 62.46%; the implied per-word termination probability is $0.6246/2 = 0.3123$. The theoretical value is 1/4; the discrepancy is sampling variation.

4.3 Sentence length distribution

The theoretical distribution

$$\boxed{P(\text{sentence length} = k) \;=\; \frac{3^{k-1}}{4^k} \;=\; \frac{1}{3}\left(\frac{3}{4}\right)^k}$$

is a valid probability distribution: $\frac{1}{4}\sum_{k=1}^{\infty}(3/4)^{k-1} = \frac{1}{4} \cdot 4 = 1$.

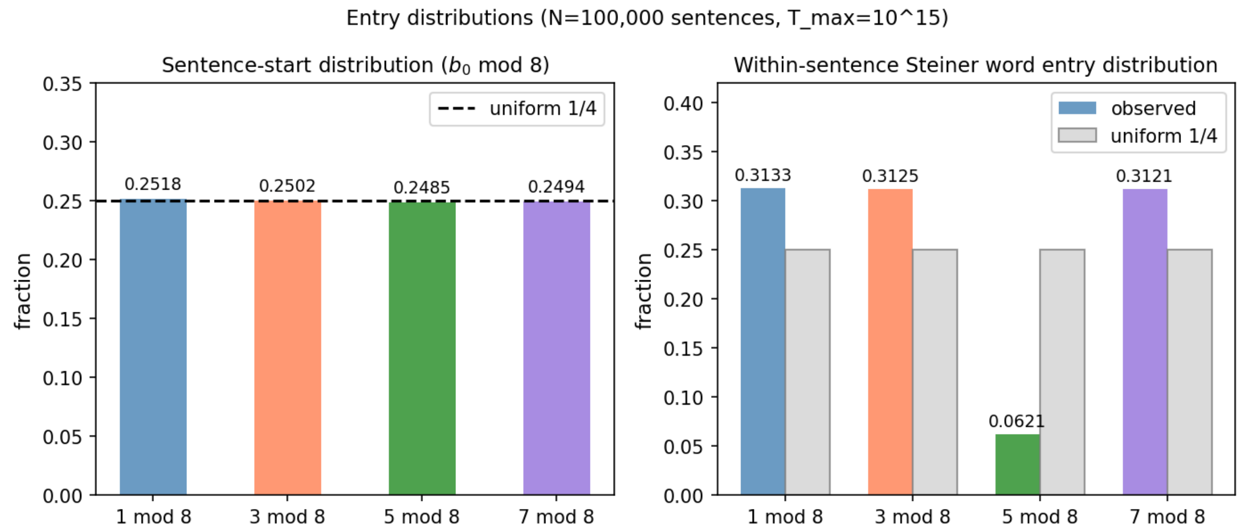

Figure 1: Left: sentence-start distribution (b_0 mod 8) against the uniform 1/4 baseline — confirmed uniform. Right: within-sentence Steiner word entry distribution against the uniform 1/4 baseline — class 5 is suppressed to ≈6.2% (structurally, it can only appear as the terminal word), while classes 1, 3, 7 each carry ≈31%.

4.4 Statistical tests

A χ^2 goodness-of-fit test (15 bins plus tail) on $N = 10^5$ sentences:

- Naive $(1/2)^k$: $\chi^2 \approx 1{,}022{,}029$, $p \approx 0$ — **rejected**.
- Theory $3^{k-1}/4^k$: $\chi^2 \approx 12.9$, $p \approx 0.61$ — consistent.

4.5 The effective termination probability

The 3/4 decay rate implies each Steiner word independently terminates the sentence with probability 1/4. Given the 3-node termination probabilities:

- Class 1 words: termination probability 0.
- Class 3 and 7 words: termination probability 1/2 each.
- Class 5 words: termination probability 1 (terminal only).

The effective rate 1/4 constrains the within-sentence entry distribution. With class 5 appearing only terminally, the within-sentence distribution over $\{1, 3, 7\}$ must satisfy:

$$0 \cdot f_1 + \tfrac{1}{2} \cdot f_3 + \tfrac{1}{2} \cdot f_7 \ = \ \tfrac{1}{4},$$

i.e. $f_3 + f_7 = 1/2$ among the non-5 words. Empirically $f_3 + f_7 \approx 0.625$ of all words, which (excluding terminal class 5 words) gives $\approx 0.625/0.938 \approx 0.666$, consistent with the within-$\{1, 3, 7\}$ constraint.

Paper 67 [2] proves this exactly: the invariant $d_3 = d_7$ in the surviving word distribution makes the effective termination probability exactly 1/4 at every step, independent of d_1.

k	Observed fraction	Naive $(1/2)^k$	Theory $3^{k-1}/4^k$
1	0.25180	0.50000	0.25000
2	0.18683	0.25000	0.18750
3	0.13931	0.12500	0.14063
4	0.10574	0.06250	0.10547
5	0.07844	0.03125	0.07910
6	0.05958	0.01563	0.05933
7	0.04490	0.00781	0.04449
8	0.03342	0.00391	0.03337
9	0.02490	0.00195	0.02503
10	0.01866	0.00098	0.01877

Table 3: Empirical Steiner sentence length distribution ($N = 10^5$, $T_{\max} = 10^{15}$) vs. the naive $(1/2)^k$ prediction and the theoretical $3^{k-1}/4^k$ (proved in Paper 67 [2]).

5 Open Problems

The following conjectures, stated here from empirical evidence, are proved in Paper 67 [2].

Conjecture 5.1 (Sentence length distribution)**.** *For Steiner sentences sampled by the protocol of Section 3 with* $T_{\max} \to \infty$*:*

$$P(\textit{sentence length} = k) = \frac{3^{k-1}}{4^k}, \qquad k = 1, 2, 3, \ldots$$

Conjecture 5.2 (Within-sentence invariant)**.** *The within-sentence Steiner word entry distribution satisfies* $d_3 = d_7$ *at every step, making the effective per-word termination probability exactly* $1/4$*.*

Both are proved in Paper 67 from the exact Syracuse transition matrix modulo 8 and the column symmetry of that matrix restricted to states $\{1, 3, 7\}$.

Acknowledgements

This paper is part of the speculative results stream (64+k) of the Collatz programme. Paper 33 provides the Steiner circuit decomposition and the 3-node lemma. Paper 67 provides the proof.

References

[1] J. Seymour, *A Regular Expression Language for the Collatz Graph* (Working Paper, Paper 33), 2026.

[2] J. Seymour, *First-Principles Derivation of the Steiner Sentence Length Distribution* (Working Paper, Paper 67), 2026.

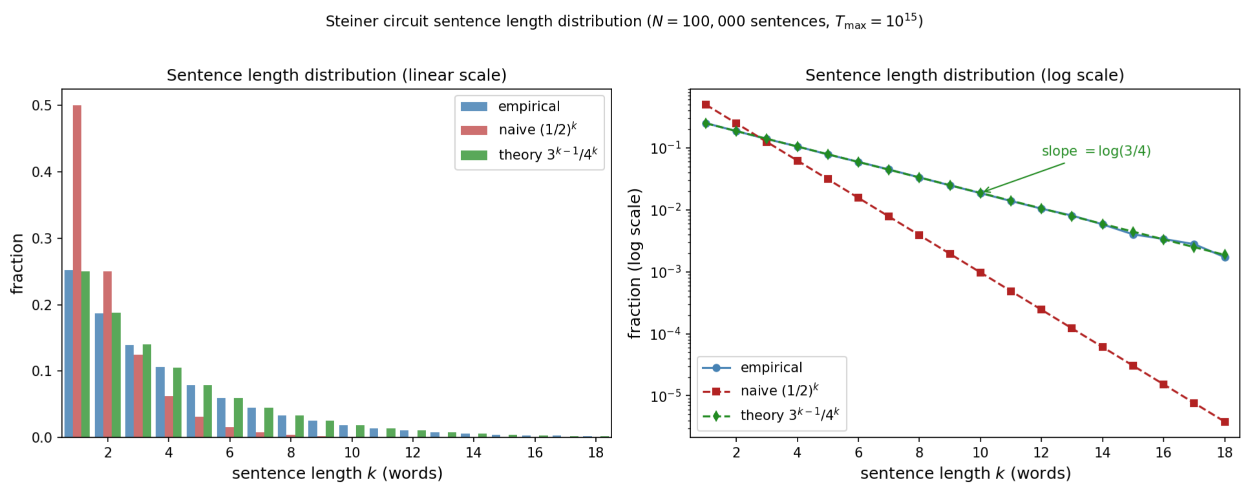

Figure 2: Steiner sentence length distribution ($N = 10^5$, $T_{\max} = 10^{15}$): empirical (blue) vs. naive $(1/2)^k$ (red) and theoretical $3^{k-1}/4^k$ (green), on linear scale (left) and log scale (right). The log-scale plot confirms a geometric distribution; the slope matches $\log(3/4)$, not $\log(1/2)$.