First-Principles Derivation of the Steiner Sentence Length Distribution

Jon Seymour
jon@wildducktheories.com
Sydney, Australia

July 2026

Abstract

A *Steiner word* (Steiner circuit [1]) is a path segment of an odd Collatz orbit whose compressed mod-8 form matches $(7^*3)?(1 \mid 5)$: a sequence of Syracuse-map applications beginning at any letter in $\{1,3,5,7\}$ and terminating at the terminal letter 1 or 5. A *Steiner sentence* is a string of $k \geq 1$ consecutive Steiner words in which all but the last word terminate at letter 1, and the final word terminates at letter 5; its *length* is the number of words k. We prove that the sentence length distribution is $P(\text{length} = k) = 3^{k-1}/4^k$ for all $k \geq 1$.

The proof models sentences as first-passage times of a Markov chain on $\{1,3,5,7\}$ (mod 8), whose transition matrix is computed row by row from elementary arithmetic. The key structural fact is an invariant $d_3 = d_7$ — equal weight on residue classes 3 and 7 in the non-terminated distribution — which holds unconditionally and forces the surviving mass to decay by exactly 3/4 per word.

The proof requires no ergodic theory and no appeal to the Collatz conjecture. It rests on two arithmetic facts: that $v_2(3n+1)$ is a pointwise constant for $n \equiv 1,3,7 \pmod 8$ (giving exact rational transition probabilities), and that $\gcd(3, 2^j) = 1$ (giving uniform output from $n \equiv 5 \pmod 8$, proved in the companion paper [4]). All results have been machine-verified in Lean 4/Mathlib with no `sorry`.

Appendix B provides a complete empirical validation: 10^5 sentences sampled uniformly over $[1, 10^{15}]$, with χ^2 goodness-of-fit tests. The naive $(1/2)^k$ prediction is rejected with $\chi^2 \approx 10^6$; the formula $3^{k-1}/4^k$ is consistent ($p \approx 0.61$).

Contents

1 Background and Notation

We work throughout with odd positive integers.

Definition 1.1 (Syracuse map)**.** The *Syracuse map* on odd positive integers is

$$S(n) \;=\; \frac{3n+1}{2^{v_2(3n+1)}}, \qquad n \text{ an odd positive integer,}$$

where $v_2(m)$ denotes the 2-adic valuation of m.

Definition 1.2 (Steiner word)**.** A *Steiner word* (or *Steiner circuit* [1, 2]) is a path segment of an odd Collatz orbit whose compressed mod-8 form matches the regular expression $(7^*\,3)?(1 \mid 5)$ [2]: a sequence of Syracuse-map applications whose successive input residues mod 8 spell a string over the alphabet $\{1, 3, 5, 7\}$ of that form, terminating when the input residue is 1 or 5 (the two *terminal letters*). The *entry letter* of a word is the residue modulo 8 of its first node; the *terminal letter* is 1 or 5.

Definition 1.3 (Steiner sentence)**.** A *Steiner sentence* of length $k \geq 1$ is a string of k consecutive Steiner words in a Collatz orbit such that the first $k-1$ words terminate at letter 1 and the final word terminates at letter 5.

Remark 1.4*.* A sentence of length k corresponds to a path of k edges in the overlay tree whose nodes are odd integers $\equiv 5 \pmod 8$ and whose edges are individual Steiner words [2]: consecutive words in the sentence that terminate at letter 1 extend the current edge, while the word terminating at letter 5 completes it. This paper is self-contained; Paper 33 [2] (in preparation) develops the Steiner circuit decomposition, the regular-expression characterisation, and the overlay tree in full generality, but is not required here.

Remark 1.5 (Why Steiner sentences are the right scale)*.* The Steiner sentence decomposition is chosen precisely so that class 5 (mod 8) is *always a sink and never a source* within a sentence. This has a crucial arithmetic consequence.

For $n \equiv 1, 3, 7 \pmod 8$, the 2-adic valuation $v_2(3n+1)$ is a *constant* determined entirely by n mod 8: it equals 2, 1, and 1 respectively (Paper 66 [4], Theorem 2.1). The Syracuse image $S(n)$ mod 8 is therefore completely determined by n mod 8 alone, with exact rational transition probabilities.

For $n \equiv 5 \pmod 8$, by contrast, $v_2(3n+1) = 3 + v_2(3k+2)$ where $n = 8k+5$, and $v_2(3k+2)$ is an unbounded random variable whose expectation requires summing a geometric series (Paper 66). The transition probabilities out of class 5 are still exactly 1/4 each, but they arise from a distribution over infinitely many valuation values, not a single fixed shift.

The Markov chain on entry letters tracks how the entry letter of each word determines the entry letter of the next, via the transition matrix P. Letter 5 is the unique terminal letter that ends a sentence; letter 1 is the other terminal letter, but it begins the next word rather than ending the sentence. Within a sentence the surviving chain operates on entry letters in $\{1, 3, 7\}$: a word with entry letter 5 would have terminated the *previous* sentence, so letter 5 never acts as an entry letter for a non-initial word inside a sentence. For entry letters 1, 3, and 7, the 2-adic valuation $v_2(3n+1)$ is a *constant* determined by the entry letter alone, so every transition probability is an exact rational. The variable-valuation complexity of letter 5 is absorbed at the sentence boundary and plays no role in the counting argument. This is why the sentence length distribution $3^{k-1}/4^k$ admits a clean closed-form proof: at the scale of Steiner sentences (strings of words), the arithmetic is exact.

Paper 65 [3] was the first attempt at this subject. It observed empirically that $P(\text{sentence length} = k) = \frac{1}{3}(3/4)^k$, equivalently $3^{k-1}/4^k$, and stated the result as a conjecture. The present paper supersedes Paper 65 by providing a complete proof, and incorporates its experimental record in Appendix B.

2 The Syracuse Transition Matrix mod 8

We model the sentence as a Markov chain on the *entry letter* of each successive Steiner word — the residue modulo 8 of the word's first node, a letter in the alphabet $\{1, 3, 5, 7\}$. Each word terminates at a terminal letter (1 or 5); this terminal letter is the entry letter of the *next* word (since the terminal node of one word is the starting node of the next). A sentence ends when a word terminates at letter 5. The transition probability $P[a][b]$ is therefore the natural density of odd integers $n \equiv a \pmod 8$ for which $S(n) \equiv b \pmod 8$:

$$P[a][b] \;=\; \lim_{N\to\infty} \frac{1}{N}\#\{n \le N : n \equiv a \pmod 8,\ S(n) \equiv b \pmod 8\}.$$

All results in this paper are unconditional: they require no assumption beyond this standard density definition and do not depend on the Collatz conjecture.

Theorem 2.1 (Syracuse transition matrix mod 8)**.** *The transition matrix on* $\{1, 3, 5, 7\} \pmod 8$ *is:*

$P =$

	1	3	5	7
1	1/4	1/4	1/4	1/4
3	1/2	0	1/2	0
5	1/4	1/4	1/4	1/4
7	0	1/2	0	1/2

(rows = source letter, columns = target letter).

Proof. We prove each row by direct arithmetic.

Row 1 (uniform). Write $n = 8k + 1$. Then $3n + 1 = 24k + 4 = 4(6k + 1)$. Since $6k + 1$ is always odd, $v_2(3n + 1) = 2$ exactly and $S(n) = 6k + 1$. Now $6k + 1 \pmod 8$: as k runs through $0, 1, 2, 3 \pmod 4$, we get $6k + 1 \equiv 1, 7, 5, 3 \pmod 8$ respectively. Each residue in $\{1, 3, 5, 7\}$ appears with equal density, so $P[1][b] = 1/4$ for all $b \in \{1, 3, 5, 7\}$.

Row 5 (uniform). Write $n = 8k+5$. Then $3n+1 = 8(3k+2)$, so $v_2(3n+1) = 3+v_2(3k+2)$. Write $3k + 2 = 2^m q$ with q odd; then $S(n) = q$. We need $q \pmod 8$ to be uniform on $\{1, 3, 5, 7\}$.

Since $\gcd(3, 2^j) = 1$ for all $j \geq 1$, the map $k \mapsto 3k + 2$ is a bijection on $\mathbb{Z}/2^j\mathbb{Z}$. As k ranges over any 2^{j+3} consecutive integers, the value $3k + 2$ hits each residue class mod 2^{j+3} exactly 2^j times. Consequently the value $S(8k + 5) \pmod 8$ is equidistributed over $\{1, 3, 5, 7\}$ *in natural density*: the arithmetic core is that, as k ranges over $\{0, 1, \ldots, 2^N - 1\}$, exactly 2^{N-j} values satisfy $2^j \mid (3k+2)$ (Paper 66 [4], `uniformity_count`; formalised here as `row5_uniform_dvd`), from which each target residue $b \in \{1, 3, 5, 7\}$ receives limiting density 1/4. Thus $P[5][b] = 1/4$ for all $b \in \{1, 3, 5, 7\}$.

Remark 2.2 (Row 5 is a density, not an exact finite count). Unlike rows 1, 3 and 7 — whose per-residue counts over $\{0, \ldots, 2^N - 1\}$ are *exactly* 2^{N-2} or 2^{N-1} (formalised as `count_row1`, `count_row3_*`, `count_row7_*`) — row 5 holds only as a natural-density limit. The finite count $\#\{k < 2^N : S(8k + 5) \equiv b \pmod 8\}$ is *not* equal to 2^{N-2} in general: at $N = 6$, $b = 1$ the count is 17, not $16 = 2^{6-2}$. This is machine-verified in the companion Lean development by `row5_count_not_exact` (via `native_decide`). The extra, variable number of halvings $v_2(3n + 1) = 3 + v_2(3k + 2)$ from class 5 is exactly what prevents an exact finite count while still forcing uniform limiting density.

Row 3 (restricted to $\{1, 5\}$). Write $n = 8k+3$. Then $3n+1 = 24k+10 = 2(12k+5)$. Since $12k$ is even, $12k+5$ is always odd, so $v_2(3n+1) = 1$ exactly and $S(n) = 12k+5$. Now $12k+5 \pmod 8$: $12k \equiv 4k \pmod 8$, so $12k + 5 \equiv 4k + 5 \pmod 8$. For k odd: $4k \equiv 4 \pmod 8$, giving $12k+5 \equiv 1 \pmod 8$. For k even: $4k \equiv 0 \pmod 8$, giving $12k+5 \equiv 5 \pmod 8$. Since k is odd and even with equal density, $P[3][1] = P[3][5] = 1/2$ and $P[3][3] = P[3][7] = 0$.

Row 7 (restricted to $\{3, 7\}$). Write $n = 8k + 7$. Then $3n + 1 = 24k + 22 = 2(12k + 11)$. Since $12k + 11$ is always odd, $v_2(3n + 1) = 1$ exactly and $S(n) = 12k + 11$. Now $12k + 11 \pmod 8$: $12k + 11 \equiv 4k + 3 \pmod 8$. For k odd: $4k \equiv 4$, giving $4k + 3 \equiv 7 \pmod 8$. For k even: $4k \equiv 0$, giving $4k + 3 \equiv 3 \pmod 8$. Therefore $P[7][3] = P[7][7] = 1/2$ and $P[7][1] = P[7][5] = 0$. □

Remark 2.3 (Forbidden transitions). The zero entries in P — transitions $3 \to 3$, $3 \to 7$, $7 \to 1$, $7 \to 5$ — are *structurally forbidden* for all n, not just on average. Classes 3 and 7 each map into a strict subset of $\{1, 3, 5, 7\}$. This structural restriction is the engine of the proof.

Remark 2.4 (Machine verification). All results in this theorem have been formally verified in Lean 4/Mathlib with no `sorry` or `admit` and no axioms beyond `propext`, `Classical.choice`, and `Quot.sound`. The key theorems are `S_class1/3/5/7` (arithmetic identities), `S_class1_mod8` (row 1 uniformity by direct mod-4 computation), `count_row1/3/7` (exact finite counts for rows 1, 3, 7), and `row5_count_not_exact` (counterexample for row 5 via `native_decide`: at

$N = 6$, $b \equiv 1$ the count is 17, not 16). The Lean source is in `Paper67.lean` in the paper directory.

Remark 2.5 (The surviving chain has exact arithmetic). Row 5 of P is never used in the proof of Theorem 5.1. Within a sentence, class 5 is a sink: once the chain reaches 5 the sentence ends, and class 5 never acts as a source for subsequent words. The surviving distribution (Definition 3.2) propagates mass only through rows 1, 3, and 7 of P.

For these three rows, all transition probabilities are *exact rational constants* that follow directly from $n \bmod 8$ alone: $v_2(3n+1) = 2$ always for class 1, and $v_2(3n+1) = 1$ always for classes 3 and 7. No geometric series, no expectation, no appeal to the distribution of $v_2(3n+1)$ for class 5 is ever needed. The variable-valuation behaviour of class 5 — which requires an infinite geometric series to characterise (Paper 66 [4]) — is invisible to the surviving chain. This exactness is what makes the sentence decomposition the right scale for counting, and what makes the proof of Theorem 5.1 purely algebraic.

3 The Sentence as a First-Passage Problem

A sentence is the first-passage time of the Markov chain (P, π_0) to state 5, where π_0 is the sentence-start distribution.

Lemma 3.1 (Sentence-start distribution)**.** *The sentence-start distribution is* $\pi_0 = (1/4, 1/4, 1/4, 1/4)$ *— uniform on* $\{1, 3, 5, 7\}$.

Proof. The sentence starts at $b_0 = S(n)$ with $n \equiv 5 \pmod 8$. By Theorem 2.1, row 5 of P is uniform, so $b_0 \bmod 8$ is uniform on $\{1, 3, 5, 7\}$. □

The sentence terminates at the first word where the chain visits state 5. The sentence length k is this first-passage time.

Definition 3.2 (Surviving distribution)**.** Let $d_s^{(k)}$ denote the probability that the chain is at state s after k words without having visited state 5. Formally, $d_s^{(0)} = \pi_0(s)$ and for $k \geq 0$:

$$d_s^{(k+1)} \;=\; \sum_{t \neq 5} d_t^{(k)} P[t][s], \qquad s \neq 5.$$

The *surviving mass* is $M_k = \sum_{s \neq 5} d_s^{(k)}$.

The sentence length distribution follows from the surviving mass: M_{k-1} is the probability of surviving past word $k-1$, and M_k is the probability of surviving past word k, so

$$P(\text{sentence length} = k) \;=\; M_{k-1} - M_k.$$

Expanding: $M_{k-1} - M_k \;=\; \sum_{s \neq 5} d_s^{(k-1)} - \sum_{s \neq 5} d_s^{(k)} \;=\; \sum_{t \neq 5} d_t^{(k-1)} \Big(1 - \sum_{s \neq 5} P[t][s]\Big) \;=\; \sum_{t \neq 5} d_t^{(k-1)} P[t][5]$.

4 The Invariant $d_3 = d_7$

The key structural fact is a symmetry between classes 3 and 7 in the surviving distribution.

Lemma 4.1 (Invariant)**.** *$d_3^{(k)} = d_7^{(k)}$ for all $k \geq 0$.*

Proof. By Definition 3.2, the surviving distribution sums only over sources $t \neq 5$; class 5 contributes nothing because once the chain reaches 5 the sentence terminates (the mass is absorbed, not propagated). The transition update for states 3 and 7 is therefore:

$$\begin{aligned} d_3^{(k+1)} &= d_1^{(k)} \cdot P[1][3] + d_3^{(k)} \cdot P[3][3] + d_7^{(k)} \cdot P[7][3] \\ &= \tfrac{1}{4}\, d_1^{(k)} + 0 \cdot d_3^{(k)} + \tfrac{1}{2}\, d_7^{(k)}, \\ d_7^{(k+1)} &= d_1^{(k)} \cdot P[1][7] + d_3^{(k)} \cdot P[3][7] + d_7^{(k)} \cdot P[7][7] \\ &= \tfrac{1}{4}\, d_1^{(k)} + 0 \cdot d_3^{(k)} + \tfrac{1}{2}\, d_7^{(k)}. \end{aligned}$$

The two expressions are identical: $d_3^{(k+1)} = d_7^{(k+1)}$ for any $d_1^{(k)}, d_3^{(k)}, d_7^{(k)}$. By induction, since $d_3^{(0)} = d_7^{(0)} = 1/4$ (Lemma 3.1), $d_3^{(k)} = d_7^{(k)}$ for all $k \geq 0$. □

Remark 4.2 (Structural source of the invariant)*.* The invariant holds because columns 3 and 7 of P, restricted to rows $\{1, 3, 7\}$, are *identical*:

$$P[\cdot][3]\big|_{\{1,3,7\}} \;=\; P[\cdot][7]\big|_{\{1,3,7\}} \;=\; \left(\tfrac{1}{4},\, 0,\, \tfrac{1}{2}\right).$$

This column symmetry is a direct consequence of Theorem 2.1:

- *From class 1*: uniform output gives equal weight 1/4 to both 3 and 7.
- *From class 3*: forbidden to reach 3 or 7 (both have weight 0).
- *From class 7*: reaches 3 and 7 with equal probability 1/2 each.

The invariant is unconditional — it holds for *any* value of d_1, not just at the fixed point of the chain.

5 Main Theorem

Theorem 5.1 (Sentence length distribution)**.** *For any $k \geq 1$:*

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

Equivalently, $P(\textit{sentence length} = k) = \frac{1}{3}\left(\frac{3}{4}\right)^k$.

Proof. We use the splitting: $P(\text{sentence length} = 1) = \pi_0(5) = 1/4$ (sentences that start at $b_0 \equiv 5$ terminate immediately). For $k \geq 2$, the sentence does not terminate at word 1 (probability 3/4) and then terminates at word $k-1$ from the conditional start, i.e. $P(\text{length} = k) = \frac{3}{4} \cdot g_{k-1}$, where g_j is the first-passage probability to 5 starting from the conditional distribution $\pi_0(\cdot \mid \cdot \neq 5) = (1/3, 1/3, 1/3)$ uniform on $\{1, 3, 7\}$.

It suffices to show $g_j = (3/4)^{j-1}/4$ for all $j \geq 1$, which follows if the surviving mass M_j starting from uniform $\{1, 3, 7\}$ satisfies $M_j = (3/4)^j$.

Step 1: $d_3 = d_7$ is preserved. The conditional start has $d_3^{(0)} = d_7^{(0)} = 1/3$, so Lemma 4.1 applies: $d_3^{(k)} = d_7^{(k)}$ for all $k \geq 0$.

Step 2: Surviving mass decays by exactly 3/4. From Definition 3.2, $M_{k+1} = \sum_{s \neq 5} d_s^{(k+1)} = \sum_{s \neq 5} \sum_{t \neq 5} d_t^{(k)} P[t][s] = \sum_{t \neq 5} d_t^{(k)} \sum_{s \neq 5} P[t][s] = \sum_{t \neq 5} d_t^{(k)} (1 - P[t][5])$, giving:

$$M_{k+1} \;=\; d_1^{(k)}\big(1 - P[1][5]\big) + d_3^{(k)}\big(1 - P[3][5]\big) + d_7^{(k)}\big(1 - P[7][5]\big).$$

Substituting $P[1][5] = 1/4$, $P[3][5] = 1/2$, $P[7][5] = 0$:

$$M_{k+1} \;=\; \tfrac{3}{4}\, d_1^{(k)} + \tfrac{1}{2}\, d_3^{(k)} + d_7^{(k)}.$$

We claim $M_{k+1} = \frac{3}{4} M_k = \frac{3}{4}(d_1^{(k)} + d_3^{(k)} + d_7^{(k)})$. This requires:

$$\tfrac{3}{4}\, d_1^{(k)} + \tfrac{1}{2}\, d_3^{(k)} + d_7^{(k)} \;=\; \tfrac{3}{4}\, d_1^{(k)} + \tfrac{3}{4}\, d_3^{(k)} + \tfrac{3}{4}\, d_7^{(k)},$$

which simplifies to $\frac{1}{2} d_3^{(k)} + d_7^{(k)} = \frac{3}{4}(d_3^{(k)} + d_7^{(k)})$, i.e. $d_7^{(k)} = d_3^{(k)}$. This is exactly Lemma 4.1.

Step 3: Initial condition. At $k = 0$, $M_0 = 1$ (the full conditional mass on $\{1, 3, 7\}$). By induction, $M_k = (3/4)^k$ for all $k \geq 0$.

Conclusion. $g_j = M_{j-1} - M_j = (3/4)^{j-1} - (3/4)^j = (3/4)^{j-1}(1/4) = 3^{j-1}/4^j$. Therefore:

$$P(\text{length} = 1) \;=\; \tfrac{1}{4} \;=\; \tfrac{3^0}{4^1},$$

$$P(\text{length} = k) \;=\; \tfrac{3}{4} \cdot g_{k-1} \;=\; \tfrac{3}{4} \cdot \frac{3^{k-2}}{4^{k-1}} \;=\; \frac{3^{k-1}}{4^k}, \qquad k \geq 2.$$

Both cases give $3^{k-1}/4^k$, completing the proof. □

Corollary 5.2 (Normalisation)**.** $\sum_{k=1}^{\infty} P(\textit{sentence length} = k) = 1.$

Proof. $\sum_{k=1}^{\infty} \frac{3^{k-1}}{4^k} = \frac{1}{4} \sum_{k=1}^{\infty} \left(\frac{3}{4}\right)^{k-1} = \frac{1}{4} \cdot \frac{1}{1 - 3/4} = 1.$ □

Remark 5.3 (Machine verification of the Markov chain algebra). The Markov chain results — Lemma 4.1 (`invariant_d3_d7`), $M_k = (3/4)^k$ (`Mmass_eq`), $g_j = 3^{j-1}/4^j$ (`g_eq`), $P(\text{length} = k) = 3^{k-1}/4^k$ (`Plen_eq`), and the normalisation identity (`normalization`) — have all been formally verified in Lean 4/Mathlib with no `sorry`, no `admit`, and no axioms beyond the standard Lean kernel. The verification operates on exact rational arithmetic throughout; no real-analysis machinery is required. The Lean source is in `Paper67.lean` in the paper directory.

6 Discussion

6.1 The role of class 5

The formula $3^{k-1}/4^k$ is the geometric distribution with success probability 1/4, as if at each word the chain draws uniformly from $\{1, 3, 5, 7\}$ and terminates on drawing 5. This appearance is striking because class 5 plays three distinct and mutually reinforcing roles in the sentence structure.

Class 5 as the termination state. A sentence terminates exactly when a word's terminal letter is 5. Since letter 5 is the unique absorbing boundary of the chain (once the chain's terminal letter is 5, the sentence ends), letter 5 is structurally absent as an entry letter for all intermediate words. Empirically, letter 5 accounts for only $\approx 6.2\%$ of within-sentence word entries (Appendix B, Table 2) — one entry per sentence, uniformly distributed across sentence lengths — rather than the 25% it would contribute if words were unrestricted.

Class 5 as the uniform source. The sentence starts at $b_0 = S(n)$ with $n \equiv 5 \pmod 8$. By row 5 of the transition matrix (Theorem 2.1), letter 5 maps uniformly over $\{1, 3, 5, 7\}$: any of the four letters is equally likely as the entry letter of the first word. In particular, exactly 1/4 of sentences have $b_0 \equiv 5$, meaning length $k = 1$. This is the correct $P(k = 1) = 1/4$, not the naive 1/2. Class 5 is the unique class for which $v_2(3n + 1) \geq 3$ always (proved in Paper 66 [4]), which forces the extra halvings that spread the output uniformly.

Class 5 as the structural asymmetry. The transition $7 \to 5$ is structurally forbidden: $P[7][5] = 0$. Class 5 is therefore unreachable *directly* in one word from class 7, while it is reachable from classes 1 (with probability 1/4) and 3 (with probability 1/2). This asymmetry — class 5 reachable from some classes but not others — is precisely what forces the invariant $d_3 = d_7$: nodes in class 7 can only reach class 5 via a detour through class 3 or class 7, never directly. The proof in Section 4 shows that this forced detour maintains equal weight on classes 3 and 7 throughout the chain, and Section 5 shows that this equal weight is the algebraic condition that makes the effective termination probability exactly 1/4 at every step, regardless of the current value of d_1.

6.2 Connection to Paper 66

Paper 66 [4] proves that $E[v_2(3n+1) \mid n \equiv r \pmod 8]$ equals exactly 2, 1, 4, 1 for $r = 1, 3, 5, 7$ respectively. The two papers are complementary: Paper 66 establishes what happens *at* the class 5 boundary (a geometrically distributed valuation requiring an infinite series for its expectation), while the present paper shows that by choosing Steiner sentences as the unit of analysis, the class 5 boundary need never be crossed from the inside — it only needs to be *entered*.

More precisely, the 3/4 decay rate follows from three facts, each drawing on the same underlying arithmetic:

- The 1/4 probability of immediate termination ($b_0 \equiv 5$) uses row 5 of P exactly once — at the sentence start. Class 5 is the unique class for which $v_2(3n + 1) \geq 3$ always (Theorem 2.1(3) of Paper 66 [4]), forcing the extra halvings that make its output uniform. This is the only point where the geometric-series character of class 5 is relevant.

- The 1/2 termination probability from class 3 comes from the mod-16 split of Theorem 2.1: $v_2(3n+1) = 1$ exactly for all $n \equiv 3 \pmod 8$, so $S(n)$ mod 8 is determined by n mod 16 alone, with no random valuation.

- The invariant $d_3 = d_7$ — and hence the 3/4 decay — involves classes 1, 3, and 7 only, all of which have constant valuations. The proof is purely algebraic: no expectation, no series.

Paper 66 handles the hard case (class 5, variable valuation). Paper 67 sidesteps it by construction.

6.3 Why $3^{k-1}/4^k$ rather than $(1/2)^k$

The naive prediction $(1/2)^k$ arises from treating each step as an independent trial that terminates whenever the current word is in class 3 (with probability 1/2, since $P[3][5] = 1/2$) or class 7 (also 1/2 by this model). If the within-sentence distribution were uniform over $\{1, 3, 5, 7\}$, the effective per-step termination probability would be:

$$\tfrac{1}{4} \cdot 0 \;+\; \tfrac{1}{4} \cdot \tfrac{1}{2} \;+\; \tfrac{1}{4} \cdot 1 \;+\; \tfrac{1}{4} \cdot \tfrac{1}{2} \;=\; \tfrac{1}{2},$$

where the terms correspond to classes 1 (termination probability 0 in this model), 3 (1/2), 5 (terminal, probability 1), and 7 (1/2). This model is wrong in two ways. First, class 1 *can* reach class 5 directly: $P[1][5] = 1/4$, so the naive assignment of 0 to class 1 understates termination from class 1. Second and more importantly, class 7 *cannot* reach class 5 *directly*: $P[7][5] = 0$, so the naive assignment of 1/2 to class 7 overstates it. The net effect of these two corrections is a reduction in the effective termination rate from 1/2 to 1/4. The structural reason is the forbidden transition $7 \to 5$: a word with entry letter 7 must first transit to letter 3 or 7 (remaining in the growth letters), where it may then terminate at letter 5 on a subsequent word. The invariant $d_3 = d_7$ (Lemma 4.1) shows that these two errors cancel exactly and yield an effective rate of precisely 1/4, independent of d_1.

Acknowledgements

This paper forms part of a programme of incremental research on the Collatz conjecture. The proof structure was found through numerical exploration in `notebooks/steiner-branch-model.ipynb`. All main results have been machine-verified in Lean 4/Mathlib; the Lean source file `Paper67.lean` is included in the arXiv submission. Proof structure and exposition were developed with AI assistance (Claude, Anthropic; Gemini, Google; Aristotle, Harmonic); all mathematical content has been independently verified.

Appendix A: The Steiner Language Hierarchy

The objects studied in this paper — sentences, words, letters — belong to a hierarchy of structures that together form a language-theoretic description of Collatz orbits. This appendix defines the hierarchy from the bottom up and connects it to the Collatz conjecture. The terminology is specific to this body of work; the foundational combinatorial infrastructure is developed in Paper 33 [2].

The Steiner Alphabet and Letters

The *Steiner alphabet* is the set $\{1, 3, 5, 7\}$ of odd residues modulo 8. Each element of the alphabet is a *Steiner letter*, encoding the step type of a single application of the Syracuse map:

Letter	Input class	$v_2(3n+1)$
1	$n \equiv 1 \pmod 8$	2 (OEE step)
3	$n \equiv 3 \pmod 8$	1 (OE step)
5	$n \equiv 5 \pmod 8$	≥ 3 (OEEE$^+$ step)
7	$n \equiv 7 \pmod 8$	1 (OE step)

Letters 1, 3, and 7 have *constant* 2-adic valuations determined entirely by the residue class. Letter 5 has a *variable* valuation $v_2(3n+1) = 3 + v_2(3k+2)$ (where $n = 8k+5$) that depends on k and is unbounded; its statistical behaviour requires an infinite geometric series and is studied in Paper 66 [4]. The two *terminal letters* are 1 and 5: these are the only letters at which a Steiner word can end.

Steiner Words

A *Steiner word* (or *Steiner circuit* [1, 2]) is a path segment of an odd Collatz orbit whose sequence of input residues mod 8 matches the regular expression

$$(7^*\,3)?(1 \mid 5).$$

That is: zero or more letters 7, optionally followed by one letter 3, followed by exactly one terminal letter (1 or 5). Every Steiner word terminates at a terminal letter; words terminating at letter 1 are *OEE-terminal* and words terminating at letter 5 are *OEEE$^+$-terminal*. The *entry letter* of a word is its first letter (the residue mod 8 of its starting node); the *terminal letter* is its last letter (1 or 5).

Every Collatz orbit decomposes uniquely into a sequence of Steiner words (Proposition 2.1 of Paper 33 [2]).

Remark A.1 (Word vs. circuit)*.* Steiner [1] introduced the term *circuit* for these path segments, and that term remains standard in the Collatz literature. We prefer *word* here because the present framework embeds circuits into a richer linguistic hierarchy: words compose into sentences, sentences into volumes, and volumes into libraries. Within this hierarchy, *word* is the natural term — it carries the right connotations of a discrete unit that can be concatenated and counted, and it aligns with standard formal-language usage where words are strings over an alphabet. The term *circuit* is retained as a synonym (and in the compound *Steiner circuit*) where it aids clarity or connection to prior work.

Steiner Sentences

A *Steiner sentence* of length $k \geq 1$ is a string of k consecutive Steiner words in a Collatz orbit such that the first $k-1$ words are OEE-terminal (terminal letter 1) and the k-th word is OEEE$^+$-terminal (terminal letter 5).

Equivalently, a sentence is a path of k edges in the $5 \pmod 8$ *overlay tree*: the directed graph whose nodes are odd integers $\equiv 5 \pmod 8$ and whose edges are individual Steiner words [2]. The sentence length distribution $P(\text{length} = k) = 3^{k-1}/4^k$ is the main result of this paper (Theorem 5.1).

Higher-Level Structures

The hierarchy does not end at sentences. Future work in this programme will use the terms *paragraph*, *page*, *chapter*, *volume*, and *library* to name successively larger graph-theoretic structures built from sentences, each relevant to the analysis of Collatz orbits at a different scale. Their precise definitions are deferred to companion paper [2], where the full graph-theoretic infrastructure is developed.

The sentence length distribution proved in this paper (Theorem 5.1) is unconditional: the proof rests on arithmetic facts about $v_2(3n+1)$ that hold for every odd positive integer, and does not depend on the Collatz conjecture or on the structure of any higher-level objects.

Appendix B: Empirical Validation

This appendix records the complete numerical experiment establishing the sentence length distribution. The experiment was first reported in Paper 65 [3] and is reproduced here in full as the empirical record accompanying this paper's proof.

Sampling Protocol

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 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.

$N = 10^5$ independent sentences were collected with seed 42. Independence between samples is ensured by the random draw of t; no orbit tails are shared.

Sentence-start Distribution

By Theorem 2.1, row 5 of P is uniform, so $b_0 = S(n) \bmod 8$ 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 by Theorem 2.1.

Within-sentence Steiner Word Entry Distribution

The entry letter of each Steiner word within a sentence determines its termination behaviour. Words with entry letter 5 always terminate the sentence, so they appear only as the final word — their within-sentence frequency is structurally suppressed to approximately $1/(k+1)$ for a sentence of length k, averaging to $\approx 1/16$ empirically.

Entry mod8	Count	Fraction	Can reach class 5 directly?	Term. prob.
1	125450	0.31333	yes ($P_{15} = 1/4$)	1/4
3	125119	0.31250	yes ($P_{35} = 1/2$)	1/2
5	24852	0.06207	— (terminal only)	1
7	124955	0.31209	no ($P_{75} = 0$)	0
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 ≈6.2% rather than the uniform 25%. The last column gives the probability of a word terminating at letter 5 given its entry letter ($P_{15} = 1/4$, $P_{35} = 1/2$, $P_{75} = 0$) — these are the exact entries of column 5 of the transition matrix P (Theorem 2.1). The invariant $d_3 = d_7$ (Lemma 4.1) makes the effective per-word termination probability exactly 1/4 (Theorem 5.1).

Sentence Length Distribution

Statistical Tests

A χ^2 goodness-of-fit test on 15 length bins plus a tail bin, using $N = 10^5$ sentences:

Model	χ^2	p-value	Verdict
Naive $(1/2)^k$	1,022,029	≈ 0	**Rejected**
Theory $3^{k-1}/4^k$	12.9	0.61	Consistent

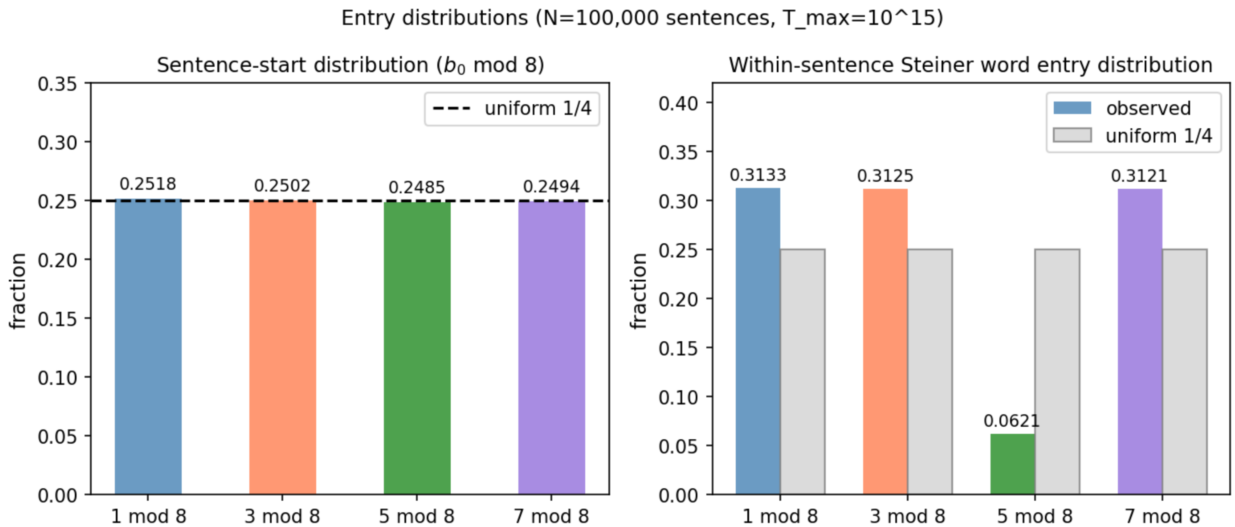

Figure 1: Left: sentence-start distribution (b_0 mod 8) against the uniform 1/4 baseline — confirmed uniform (Table 1). Right: within-sentence Steiner word entry distribution against the uniform 1/4 baseline. Class 5 is structurally suppressed to ≈6.2% (it can only appear as the terminal word); classes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

The naive model is rejected with overwhelming confidence. The theoretical formula $3^{k-1}/4^k$ is consistent with the data at all tested lengths.

References

[1] R. P. Steiner, A theorem on the Syracuse problem, *Proceedings of the 7th Manitoba Conference on Numerical Mathematics and Computing*, Winnipeg, 1977, pp. 553–559.

[2] J. Seymour, *A Regular Expression Language for the Collatz Graph*, in preparation, 2026.

[3] J. Seymour, *The* $\frac{1}{3}(\frac{3}{4})^k$ *Distribution of Steiner Sentence Lengths* (superseded by the present paper), 2026.

[4] J. Seymour, *Exact 2-adic Valuations of* $3n+1$ *and the Geometric Mean Contraction Rate of the Syracuse Map* (companion paper, submitted simultaneously), 2026.

[5] J. C. Lagarias (ed.), *The Ultimate Challenge: The* $3x+1$ *Problem*, American Mathematical Society, Providence, RI, 2010.

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 $3^{k-1}/4^k$ (Theorem 5.1). The naive prediction overestimates $P(k = 1)$ by a factor of 2 and decays far too fast for $k \geq 3$.

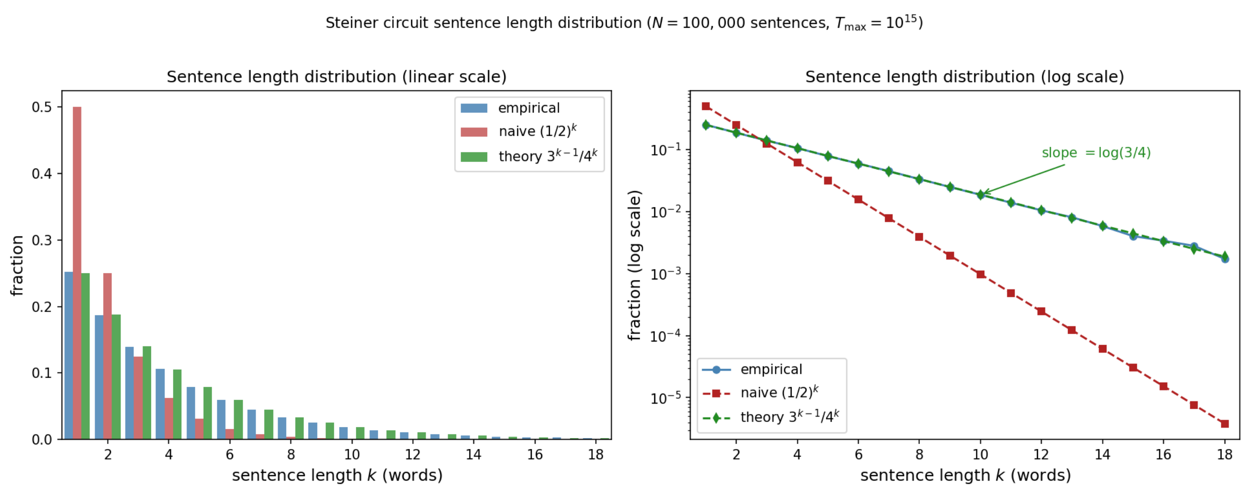

Figure 2: Steiner sentence length distribution ($N = 10^5$, $T_{\max} = 10^{15}$): empirical (blue bars / circles) vs. naive $(1/2)^k$ (red) and theoretical $3^{k-1}/4^k$ (green), on linear scale (left) and log scale (right). On the log scale, the empirical slope matches $\log(3/4)$ precisely — not $\log(1/2)$. The naive prediction is dramatically refuted at $k = 1$ (predicts 50%; observed 25%) and for $k \geq 3$ (predicts too-fast decay).