Phase Transitions in Combinatory Logic: An Exhaustive Study of Structural Emergence under Basis Extension

Kidan Nelson

February 2026

Abstract

We exhaustively enumerate and reduce all combinatory logic (CL) terms up to size 7 in the standard $\{S, K, I\}$ basis (323,175 terms total) and measure four metrics of structural emergence: normal form compression ratio, sub-expression frequency distribution, motif count and compositional reuse value. We find that the compression ratio $\rho(N)$ decays exponentially as $e^{-0.727N}$ ($R^2 = 1.000$), with a cascade of transitions: compression onset at size 3 (when K begins discarding), divergence onset at size 6 and term explosion at size 7. The normal form multiplicity distribution follows a power law with exponent $\alpha \approx 1.5$, with the identity combinator I as the strongest attractor (6.9% of all normalizing terms at size 7).

We then conduct a systematic *naming experiment*: we add 11 individual motifs and 6 multi-motif combinations as new basis elements and re-enumerate to measure the effect on compression decay. Of the 11 motifs, only two produce a compression advantage above $1.0\times$: $S(SS)$ at $1.68\times$ and SS at $1.61\times$, both slowing the exponential decay rate. The remaining 9 motifs are neutral or harmful. All three K-headed motifs produce identical compression ratios and accelerate decay to $e^{-0.900N}$. SII, the self-application combinator, is S-headed but harmful ($0.44\times$) due to a 17.3% non-normalization rate. Pairing the two best motifs ($SS + S(SS)$) yields $1.95\times$ advantage, but this is sub-multiplicative relative to the product of their individual gains. These results show that the value of a basis extension depends on the specific reduction behavior of the added primitive, not on syntactic properties alone.

1 Introduction

We study the following question: *when a recursive rewriting system is enumerated exhaustively, how does the relationship between syntactic complexity and semantic variety change as terms grow larger?* Specifically, we ask whether naming a frequently occurring sub-expression and adding it to the system's primitives measurably changes this relationship.

Our formal system is combinatory logic (CL), which consists of three primitive combinators S, K, I and a single binary operation (application). Despite this minimal syntax, CL is Turing-complete: every computable function can be expressed as a CL term [Hindley and Seldin, 2008]. This makes it a useful substrate for studying how the space of reachable computations changes under basis extension, because the system is simple enough to enumerate exhaustively at small sizes yet powerful enough to exhibit non-trivial computational phenomena including non-termination.

Our approach is straightforward. We enumerate *every* CL term up to a given size, reduce each to its normal form (or classify it as divergent) and measure structural properties of the resulting population. We then select frequently occurring sub-expressions ("motifs"), add them as new primitives and re-run the survey to measure the effect.

This connects to several lines of prior work:

- **Algorithmic information theory.** The compression ratio measures how many distinct outputs a system produces relative to its input space, which is related to Kolmogorov

complexity [Li and Vitányi, 2019]. Enumeration-based approaches to studying term spaces have been applied in the binary lambda calculus [Tromp, 2014], where similar compression phenomena appear.

- **Logical depth.** The number of reduction steps to reach normal form is a measure of computational content, analogous to Bennett's logical depth [Bennett, 1988].
- **Library learning.** The naming experiment mirrors the abstraction-learning loop in systems like DreamCoder [Ellis et al., 2021], but conducted exhaustively and bottom-up rather than heuristically and top-down. Where DreamCoder uses neural guidance to select which sub-expressions to name, we test all candidates and measure the effect of each.
- **Phase transitions in combinatorics.** The sharp onset of divergence and the exponential decay of compression resemble phase transitions studied in random satisfiability and random graph theory [Mézard and Montanari, 2009].

The paper proceeds as follows. Section 2 defines the formal system and metrics. Section 3 describes the computational methodology. Section 4 presents results for the baseline survey, the naming experiment across 11 motifs, a motif classification analysis and multi-motif combination experiments. Section 5 interprets these findings. Section 6 summarizes results and outlines future work.

2 Background and Definitions

2.1 Combinatory Logic

Definition 1 (CL Term). *A combinatory logic term over a basis $\mathcal{B}$ is defined inductively:*
  1. Every $b \in \mathcal{B}$ is a term (an atom*).*
  2. If M and N are terms, then $(M\ N)$ is a term (application*).*
The standard basis is $\mathcal{B}_0 = \{S, K, I\}$.

Definition 2 (Reduction Rules). *The reduction rules for S, K, I are:*

$$I\ x \rightarrow x \tag{1}$$
$$K\ x\ y \rightarrow x \tag{2}$$
$$S\ x\ y\ z \rightarrow x\ z\ (y\ z) \tag{3}$$

where application is left-associative: $M\ N\ P$ means $((M\ N)\ P)$.

The three combinators have distinct computational roles:
- I is the identity: it passes its argument through unchanged.
- K is the constant-maker (or *discarder*): it takes two arguments and returns the first, discarding the second.
- S is the *distributor*: it takes three arguments and applies the first and second to the third, then applies the results to each other. Crucially, S *duplicates* its third argument.

Definition 3 (Term Size). *The* size *of a CL term, written $|t|$, is the number of combinator leaves (atoms): $|b| = 1$ for $b \in \mathcal{B}$, and $|M\ N| = |M| + |N|$.*

Definition 4 (Enumeration Count). *The number of distinct CL terms of size N over a basis of size k is:*

$$T(N) = C_{N-1} \cdot k^N \tag{4}$$

where $C_n = \frac{1}{n+1}\binom{2n}{n}$ is the n-th Catalan number (counting full binary tree shapes with N leaves) and k^N assigns each leaf to one of k basis elements.

For the standard basis ($k = 3$), this gives:

N	C_{N-1}	3^N	$T(N)$
1	1	3	3
2	1	9	9
3	2	27	54
4	5	81	405
5	14	243	3,402
6	42	729	30,618
7	132	2,187	288,684

2.2 Reduction Strategy

We use *normal-order reduction* (leftmost outermost redex first), which is the standard strategy for computing full normal forms in CL [Hindley and Seldin, 2008]. A term is in *normal form* if no reduction rule applies at any position.

Not all CL terms have normal forms. For example, $\Omega = S\ I\ I\ (S\ I\ I)$ reduces to itself indefinitely. We handle non-termination with two limits:

- **Fuel limit** $L = 10{,}000$: maximum reduction steps before declaring a term *divergent*.
- **Size limit** $M = 50{,}000$: if the term grows beyond M atoms during reduction, we declare it *explosive*.

These limits are chosen to be well beyond what any normalizing term at size ≤ 7 requires (the maximum observed step count for a normalizing term was 23 in the baseline experiment).

2.3 Metrics

We define four metrics for characterizing structural emergence.

Definition 5 (Compression Ratio). *Let $T(N)$ be the total number of CL terms of size N and $D(N)$ the number of* distinct *normal forms among those that normalize. The compression ratio is:*

$$\rho(N) = \frac{D(N)}{T(N)} \tag{5}$$

A value of $\rho(N) = 1$ means every term reduces to a unique normal form (no compression). A value near 0 means many terms collapse to the same structures, which we call attractors.

Definition 6 (Motif Count). *A sub-expression e is a* motif *at size N if e appears as a sub-term of some normal form at size N but e did not appear in any normal form at sizes $< N$. The motif count $\mu(N)$ is the number of such novel sub-expressions.*

Definition 7 (Compositional Reuse Value). *For a candidate sub-expression e appearing in the normal forms at size N, define:*

- *$f(e)$: the number of occurrences of e across all normal forms*
- *$n(e)$: the number of distinct normal forms containing e*
- *$\sigma(e) = f(e) \cdot (|e| - 1)$: the description-length savings if e were named as a new primitive (each occurrence saves $|e| - 1$ atoms)*

This measures whether e is useful as a building block.

Definition 8 (Logical Depth). *The number of reduction steps $s(t)$ to reach normal form (for normalizing terms) is a measure analogous to Bennett's logical depth [Bennett, 1988]: the computational work required to produce a term's normal form. We report the average $\bar{s}(N)$ and maximum $s_{\max}(N)$ across all normalizing terms at each size.*

3   Methodology

3.1   Implementation

The experiment is implemented in Rust (edition 2024), using the `rayon` crate for parallelized reduction. Terms are represented as an algebraic data type using reference-counted pointers (`Arc<Term>`) with cached sizes, ensuring $O(1)$ cloning and $O(1)$ size queries during reduction. This is critical for performance on the S combinator's duplication rule.

All results are stored in a SQLite database with resume support. Each experiment records the basis, fuel limit and size limit as metadata, and each size's reductions are committed in a single transaction before proceeding to the next size.

3.2   Enumeration

For each size N, we enumerate all CL terms by generating every full binary tree shape with N leaves (there are C_{N-1} such shapes) and assigning each leaf to each basis element independently. This produces exactly $T(N) = C_{N-1} \cdot k^N$ terms where $k = |\mathcal{B}|$.

For the extended-basis experiments, k depends on the number of named motifs: $k = 4$ for single motifs (172,032 terms at size 6), $k = 5$ for pairs (656,250 terms at size 6) and $k = 6$ for triples (109,350 terms at size 5). Triples are enumerated only to size 5 to keep the computation tractable.

3.3   Reduction

Each term is reduced independently using normal-order reduction. The reduction engine works by decomposing each term into its *spine*: the chain of left-nested applications leading to the head combinator. It then checks whether the head has enough arguments for its reduction rule to fire.

Specifically, a term t is decomposed as $h\ a_1\ \ldots\ a_n$ where h is the head (an atom) and $a_1, \ldots, a_n$ are the spine arguments. If h has enough arguments:
- I: needs 1 argument, returns a_1 (with remaining args re-applied)
- K: needs 2 arguments, returns a_1 (discarding a_2)
- S: needs 3 arguments, returns $a_1\ a_3\ (a_2\ a_3)$

If the outermost spine has no redex, the engine recurses into sub-terms (left child first, then right) to find the leftmost innermost redex. This ensures normal-order evaluation while still computing *full* normal forms (reducing under lambda-like positions).

3.4   The Naming Experiment

After the baseline survey, we select motifs discovered through the reuse value analysis and run additional experiments. For each motif m:
1. We add m as a new library combinator #0 with its *effective arity*: the number of additional arguments needed before the head combinator of m's expansion can fire. For example, SS has head S which needs 3 arguments, of which 1 (S) is already provided, so the effective arity is 2.
2. We extend the basis: $\mathcal{B}_m = \{S, K, I, \#0\}$.
3. We re-enumerate and reduce all terms of sizes 1–6 using $\mathcal{B}_m$.

When #0 accumulates enough arguments during reduction, it expands to its CL definition and reduces normally. When it has fewer arguments than its arity, it remains in normal form as an opaque primitive.

We test 11 single motifs:
- **S-headed (8 motifs):** SS, SK, SI, $S(SS)$, SII, SKK, $S(KS)$ and $S(KI)$. These have S as the head combinator with varying spine arguments.

- **K-headed (3 motifs):** KS, KK and KI. These have K as the head, making them constant functions that discard their second argument.

We additionally test multi-motif combinations:

- **4 pairs** ($k = 5$ basis, sizes 1–6): $\{SS, S(SS)\}$, $\{SS, KS\}$, $\{SS, SII\}$ and $\{S(SS), SII\}$.
- **2 triples** ($k = 6$ basis, sizes 1–5): $\{SS, S(SS), SII\}$ and $\{SS, SK, KS\}$.

For combinations, each motif receives its own library index (#0, #1, etc.) and the basis is extended accordingly.

4   Results

4.1   Baseline Survey

Table 1 shows the baseline results for the standard SKI basis at sizes 1–7.

Table 1: Baseline survey: SKI basis, sizes 1–7. $T(N)$: total terms; $D(N)$: distinct normal forms; $\rho(N)$: compression ratio; Divg/Expl: divergent/explosive terms; μ: new motifs; $\bar{s}$: average reduction steps.

N	$T(N)$	$D(N)$	$\rho(N)$	Divg	Expl	μ	$\bar{s}$
1	3	3	1.0000	0	0	3	0.00
2	9	9	1.0000	0	0	6	0.33
3	54	30	0.5556	0	0	21	0.72
4	405	108	0.2667	0	0	78	1.19
5	3,402	438	0.1287	0	0	330	1.69
6	30,618	1,920	0.0627	5	0	1,483	2.22
7	288,684	8,706	0.0302	128	44	6,803	2.75

4.1.1   Exponential Compression Decay

The compression ratio $\rho(N)$ decays exponentially for $N \geq 3$:

$$\rho(N) \approx e^{-0.727 \cdot N} \tag{6}$$

with $R^2 = 1.000$ (Figure 1). The per-step decay factor is $e^{-0.727} \approx 0.483$: at each size increment, roughly half the "expressive distinctness" is lost. The half-life is 0.95 size steps.

This decay arises because the total term count grows as $T(N) \sim C_{N-1} \cdot 3^N$ (super-exponential in N), while the distinct normal form count grows at approximately $3.82\times$ per size increment. The normal form count grows fast, but not fast enough to keep up with the combinatorial explosion of syntactically distinct inputs.

4.1.2   The Transition Cascade

We observe a structured sequence of transitions in the baseline (Figure 2):

1. **Size 1–2: Trivial regime.** Every term reduces to a distinct normal form ($\rho = 1.0$). No combinator has enough arguments to fire.
2. **Size 3: Compression onset.** The first non-trivial reductions occur. K begins to fire, discarding arguments: $K\,S\,K \to S$, etc. The compression ratio drops to 0.556.
3. **Size 4: Distribution onset.** S begins to fire, distributing and duplicating. The compression ratio drops to 0.267.
4. **Size 5–6: Deepening.** Inner reductions become common (reducing sub-terms that are not at the outermost spine). Average step count increases from 1.19 to 2.22.
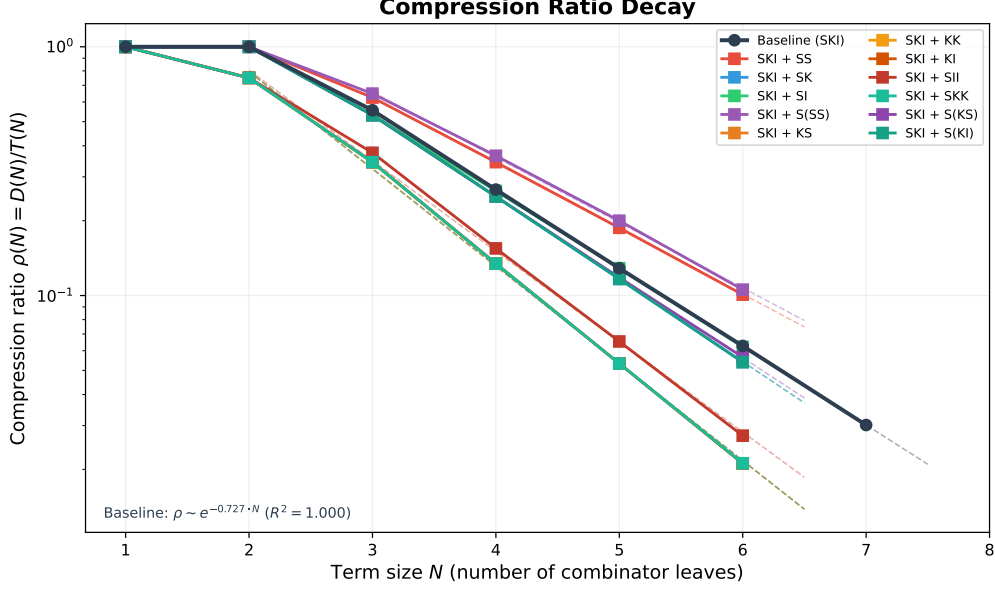
5

Figure 1: Compression ratio $\rho(N)$ vs. term size N for baseline and all single-motif experiments. Dashed lines show exponential fits. The baseline (dark) decays at $e^{-0.727N}$; the best extended basis, $S(SS)$ (purple), decays at $e^{-0.605N}$.

5. **Size 6: Divergence onset.** Five terms fail to normalize within 10,000 steps (0.016% of terms). These are the first terms that loop.
6. **Size 7: Explosion onset.** 44 terms grow beyond 50,000 atoms during reduction. Additionally, 128 terms diverge (0.044% and 0.015% respectively).

4.1.3 Attractor Normal Forms

The distribution of normal form multiplicities (how many input terms reduce to each distinct normal form) follows a power law (Figure 3):

$$\text{multiplicity(rank)} \sim \text{rank}^{-\alpha}, \quad \alpha \approx 1.5 \tag{7}$$

The top attractors at size 7 are shown in Table 2. The three atomic combinators (I, K, S) are the strongest attractors, collectively accounting for 19.7% of all normalizing terms. This is expected, since many terms reduce by discarding or simplifying to atoms. The second tier (the six size-2 normal forms KK, KI, SI, SS, SK, KS) have comparable multiplicities, ranging from 9,201 to 10,629.

4.1.4 Motif Reuse Value

The top motifs by reuse savings at each size are dominated by S-headed sub-expressions (Figure 4). At size 7, the top three motifs are SS, SI and SK, each appearing in over 3,100 distinct normal forms. The reuse savings grow rapidly with size (from 1 at size 2 to over 4,000 at size 7), indicating that the payoff of naming a sub-expression increases as the system grows.

4.2 The Naming Experiment: Single-Motif Results

Table 3 shows the results at size 6 for all 11 single motifs, ranked by compression advantage.

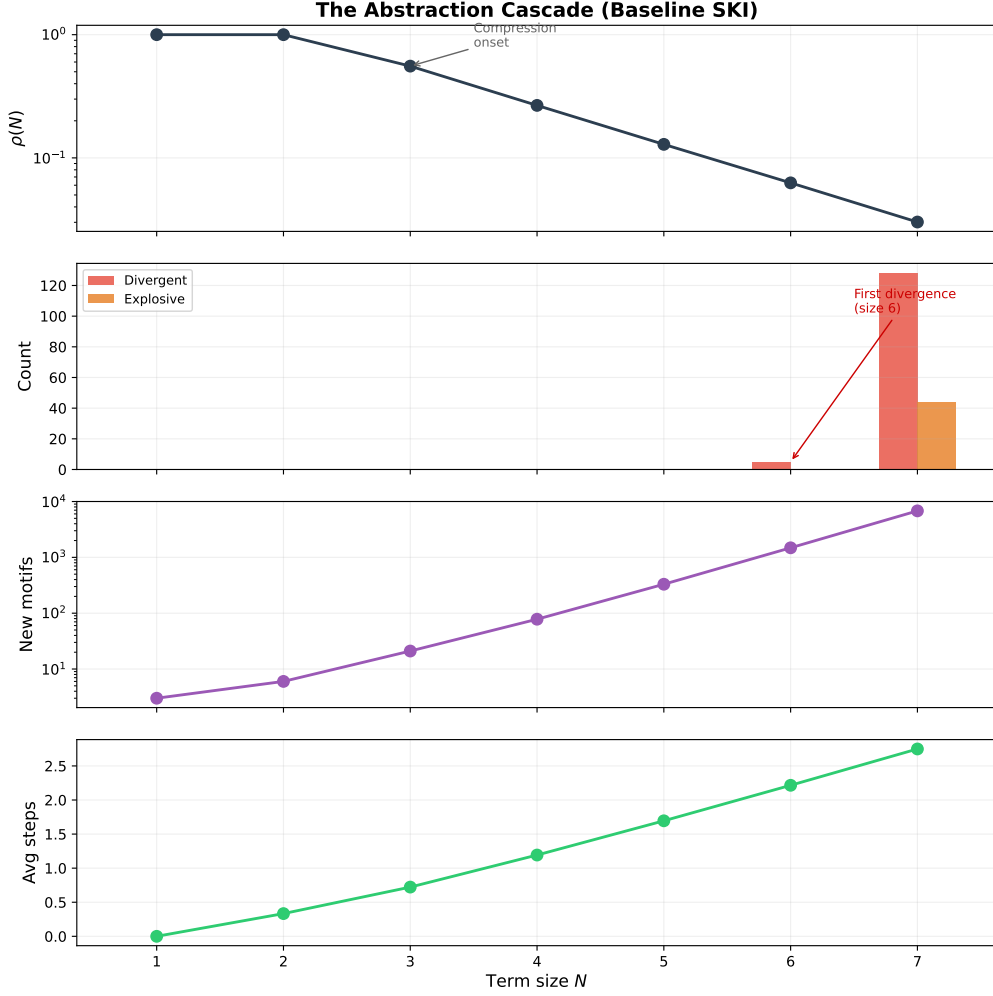The 11 motifs separate into four groups (Figure 5, Table 3):

Figure 2: The transition cascade in the baseline SKI experiment. From top: compression ratio (log scale), divergent/explosive term counts, new motif count (log scale) and average reduction steps.

Winners $(> 1.0\times)$**:** $S(SS)$ **and** SS**.** These two motifs slow compression decay substantially. $S(SS)$ achieves the highest advantage $(1.685\times)$ with a decay rate of $e^{-0.605N}$ compared to the baseline's $e^{-0.727N}$. SS is close behind at $1.608\times$ with decay rate $e^{-0.608N}$. Both are size-2 or size-3 terms whose spine arguments consist entirely of S combinators.

Neutral $(\approx 1.0\times)$**:** SI**.** SI produces $0.992\times$ advantage, essentially matching the baseline. It introduces moderate non-normalization (2.9%) without compensating expressiveness gain.

Mild losers $(0.86–0.90\times)$**:** $S(KS)$**,** $S(KI)$**,** SK**.** These S-headed motifs produce advantages below 1.0 but remain significantly above the bottom tier. They have K in their spine or as their immediate argument, which partially neutralizes the distributing effect of S. None introduce divergence at size 6.

Strong losers $(< 0.44\times)$**:** SII**,** SKK**,** KS**,** KK**,** KI**.** These five motifs all produce compression advantages below $0.44\times$. The three K-headed motifs (KS, KK, KI) are identically harmful: they produce exactly the same 3,636 distinct normal forms and the same compression ratio 0.0211 at every size. This is because Kx for any x is a constant function $(Kx\,y \to x)$, so KS, KK and KI are interchangeable as basis extensions.
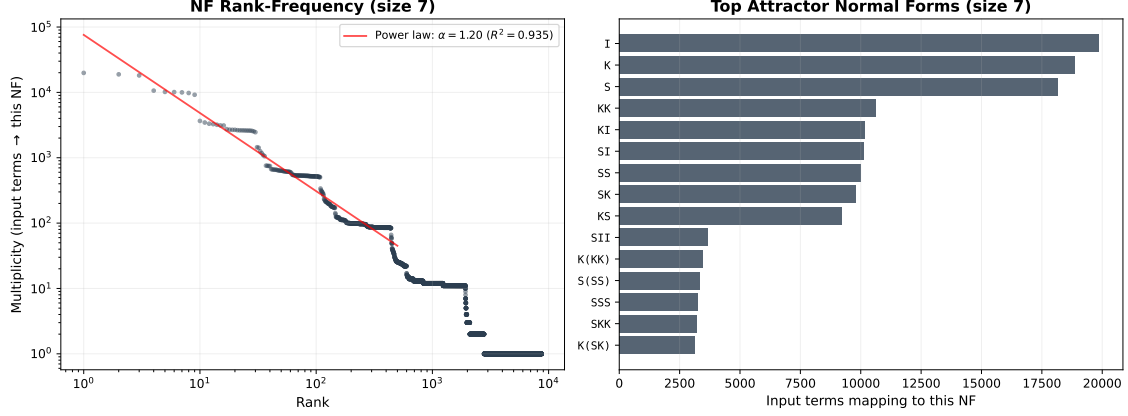
7

Figure 3: Left: rank-frequency plot of NF multiplicities at size 7, showing approximate power-law behavior ($\alpha \approx 1.5$). Right: the 15 most common normal forms.

Table 2: Top 15 attractor normal forms in the baseline at size 7. Multiplicity is the number of distinct size-7 input terms that reduce to this normal form.

Normal Form	Multiplicity	% of normalizing
I	19,864	6.9%
K	18,880	6.5%
S	18,171	6.3%
KK	10,629	3.7%
KI	10,177	3.5%
SI	10,105	3.5%
SS	10,000	3.5%
SK	9,785	3.4%
KS	9,201	3.2%
SII	3,658	1.3%
K(KK)	3,459	1.2%
S(SS)	3,317	1.1%
SSS	3,267	1.1%
SKK	3,197	1.1%
K(SK)	3,124	1.1%

SKK is equally harmful despite being S-headed. It computes the identity function ($SKK\ x \rightarrow x$), which is already available as I in the basis, so naming it is strictly redundant.

SII is the most instructive case. As the self-application combinator ($SII\ x \rightarrow xx$), it is the building block of the diverger $\Omega = SII\ (SII)$. It is S-headed and introduces genuinely new computational behavior (self-application), yet its compression advantage is only $0.436\times$. The reason is clear from the data: 17.3% of terms at size 6 fail to normalize (16.7% divergent, 0.6% explosive). The non-normalization rate overwhelms any expressiveness gain.

4.2.1 Motif Classification

Figure 7 plots each motif's compression advantage against its non-normalizing rate, revealing a structured landscape:

1. **Head combinator determines the floor.** Every K-headed motif achieves advantage $\leq 0.337\times$. No K-headed motif produces more than 3,636 distinct normal forms at size 6. The head combinator is the single strongest predictor of compression advantage.
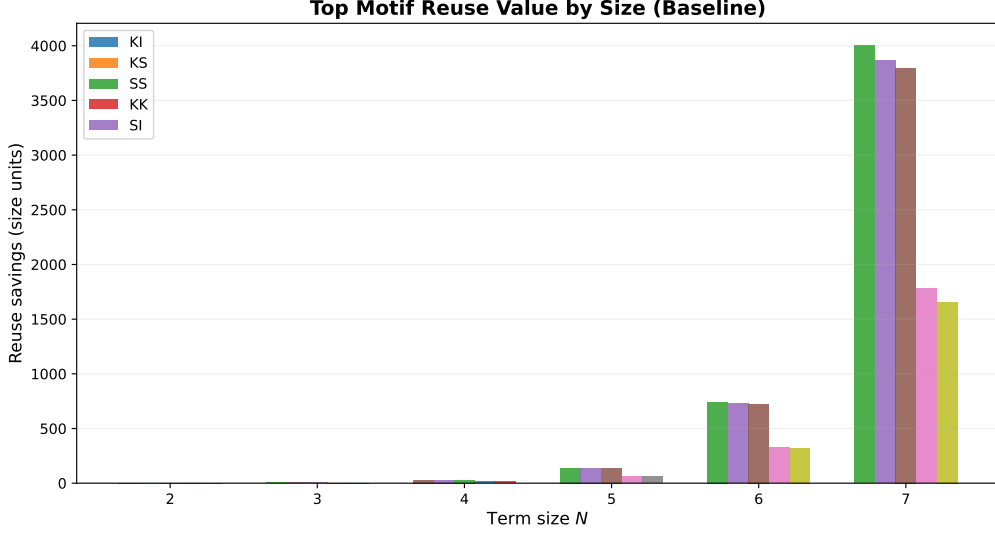
Top Motif Reuse Value by Size (Baseline)

Figure 4: Top 5 motif reuse savings by term size in the baseline experiment. SS, SI and SK consistently dominate.

2. **Spine content determines the ceiling (for S-headed motifs).** Among S-headed motifs, those with S in the spine (SS, $S(SS)$) achieve advantage $> 1.6\times$, while those with K in the spine (SK, $S(KS)$, $S(KI)$, SKK) achieve $< 0.91\times$.

3. **Self-application is costly.** SII introduces the highest non-normalization rate (17.3%) and falls well below the other S-headed motifs despite being S-headed with no K in its spine.

4.2.2   K-Headed Equivalence and the SKK Coincidence

A striking feature of the data is the exact equivalence of the three K-headed motifs. At every size from 1 to 6, the experiments KS, KK and KI produce identical distinct NF counts (e.g. 3,636 at size 6) and identical compression ratios. This is not an artifact: since $Kx\ y \to x$ for any x, the term Kx with one remaining argument is a constant function returning x. Adding KS, KK or KI as a basis element adds a one-argument function that always returns S, K or I respectively. These are structurally equivalent as basis extensions: each adds a single constant to the reachable computations.

Similarly, SKK produces compression behavior nearly identical to the K-headed motifs (decay rate -0.900 vs. -0.900) because SKK computes the identity, duplicating the role of I.

4.3   The Divergence–Expressiveness Correlation

The most expressive motifs also introduce the most divergence and explosion. The full data at size 6:

- $S(SS)$: 1.8% divergent + 6.8% explosive = 8.6% non-normalizing, advantage $1.685\times$
- SS: 0.4% + 1.1% = 1.5%, advantage $1.608\times$
- SI: 1.1% + 1.8% = 2.9%, advantage $0.992\times$
- SII: 16.7% + 0.6% = 17.3%, advantage $0.436\times$
- SK, $S(KS)$, $S(KI)$, SKK: 0.0%, advantage 0.86–0.90$\times$
- KS, KK, KI: 0.0%, advantage $0.337\times$

The motifs with zero non-normalization are exactly those that provide no compression advantage or are actively harmful. This confirms a basic property of Turing-complete systems: extending the basis with combinators that increase computational reach necessarily introduces

Table 3: All 11 single motifs at $N = 6$, ranked by compression advantage $\rho_{\text{ext}}/\rho_{\text{base}}$. Head: head combinator of the motif.

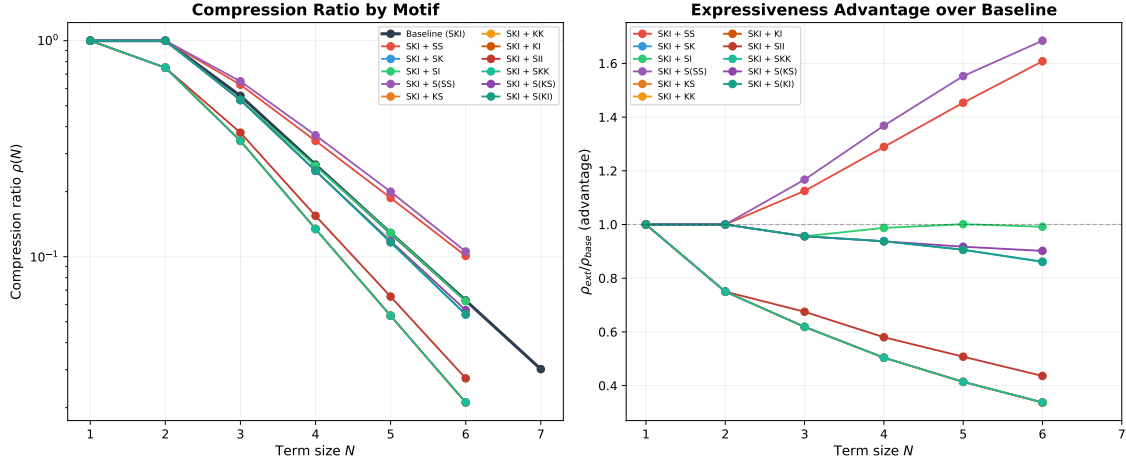| Motif	Head	$D(6)$	$\rho(6)$	Advantage	Divg%	Expl%
Baseline (SKI)	–	1,920	0.0627	$1.000\times$	0.0%	0.0%
$S(SS)$	S	18,174	0.1056	$1.685\times$	1.8%	6.8%
SS	S	17,346	0.1008	$1.608\times$	0.4%	1.1%
SI	S	10,697	0.0622	$0.992\times$	1.1%	1.8%
$S(KS)$	S	9,728	0.0565	$0.902\times$	0.0%	0.0%
$S(KI)$	S	9,294	0.0540	$0.862\times$	0.0%	0.0%
SK	S	9,284	0.0540	$0.861\times$	0.0%	0.0%
SII	S	4,703	0.0273	$0.436\times$	16.7%	0.6%
SKK	S	3,642	0.0212	$0.338\times$	0.0%	0.0%
KS	K	3,636	0.0211	$0.337\times$	0.0%	0.0%
KK	K	3,636	0.0211	$0.337\times$	0.0%	0.0%
KI	K	3,636	0.0211	$0.337\times$	0.0%	0.0%

Figure 5: Left: compression ratio by size for all single-motif experiments. Right: compression advantage ($\rho_{\text{ext}}/\rho_{\text{base}}$) vs. size. Values above 1.0 indicate the extended basis produces more distinct normal forms per input term.

the possibility of non-termination. The converse does not hold uniformly, however: SII introduces the most non-termination but is not the most expressive. High non-normalization is necessary but not sufficient for high compression advantage.

4.4   Combination Experiments

To test whether motif benefits compound, we run 6 combination experiments: 4 pairs at $k = 5$ (size 6) and 2 triples at $k = 6$ (size 5). Table 4 shows the results.

Three patterns emerge from the combination data:

The best pair is sub-multiplicative. $\{SS, S(SS)\}$ achieves $1.950\times$ advantage, the highest of any experiment. However, the product of the individual advantages is $1.608 \times 1.685 = 2.710$, so the combination achieves only 72% of the expected value. Both motifs expand the same region of computational space (S-distributing with S-rich spines), producing diminishing returns when combined.
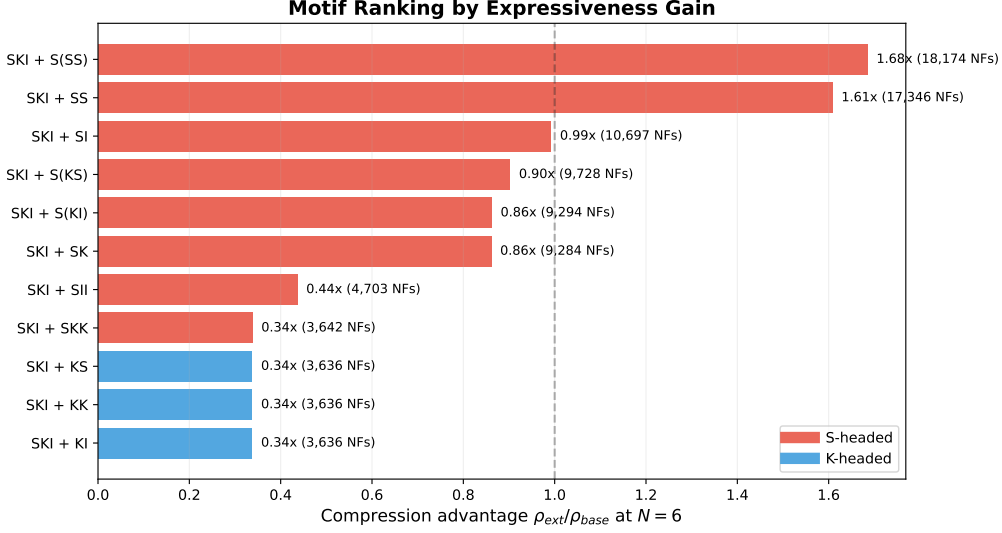
Figure 6: All 11 single motifs ranked by compression advantage at $N = 6$. Red bars: S-headed motifs. Blue bars: K-headed motifs.

Table 4: Combination experiment results. Expected advantage is the product of individual motif advantages (independence assumption). Ratio is actual/expected.

Combination	N	k	$D(N)$	$\rho(N)$	Adv.	Exp.	Ratio
$\{SS, S(SS)\}$	6	5	80,252	0.1223	$1.950\times$	$2.710\times$	0.72
$\{SS, KS\}$	6	5	26,990	0.0411	$0.656\times$	$0.542\times$	1.21
$\{SS, SII\}$	6	5	29,291	0.0446	$0.712\times$	$0.701\times$	1.02
$\{S(SS), SII\}$	6	5	29,258	0.0446	$0.711\times$	$0.734\times$	0.97
$\{SS, S(SS), SII\}$	5	6	12,932	0.1188	$0.923\times$	$1.147\times$	0.80
$\{SS, SK, KS\}$	5	6	9,498	0.0872	$0.678\times$	$0.546\times$	1.24

Harmful motifs are less harmful than expected when paired with strong motifs. $\{SS, KS\}$ achieves $0.656\times$ advantage, which is 1.21 times the expected product ($0.542\times$). Similarly, $\{SS, SK, KS\}$ achieves $0.678\times$ vs. an expected $0.546\times$ (ratio 1.24). The strong motif (SS) partially compensates for the weak ones.

SII drags down any combination. All three combinations involving SII underperform: $\{SS, SII\}$ at $0.712\times$, $\{S(SS), SII\}$ at $0.711\times$ and $\{SS, S(SS), SII\}$ at $0.923\times$. The non-normalization rate of SII propagates through the combination: $\{SS, SII\}$ has 18.8% non-normalizing terms and $\{S(SS), SII\}$ has 27.4%.

5   Discussion

5.1   Compression Decay is Universal but Rate-Variable

Every experiment in our study exhibits exponential compression decay with $R^2 \geq 0.996$. The decay rate, however, varies substantially: from -0.572 for the best combination ($\{SS, S(SS)\}$) to -0.900 for the worst single motifs (KS, KK, KI, SKK). The baseline sits at -0.727. This suggests that exponential compression decay is a structural property of CL enumeration (driven by the super-exponential growth of the input space), while the specific rate is determined by the basis.
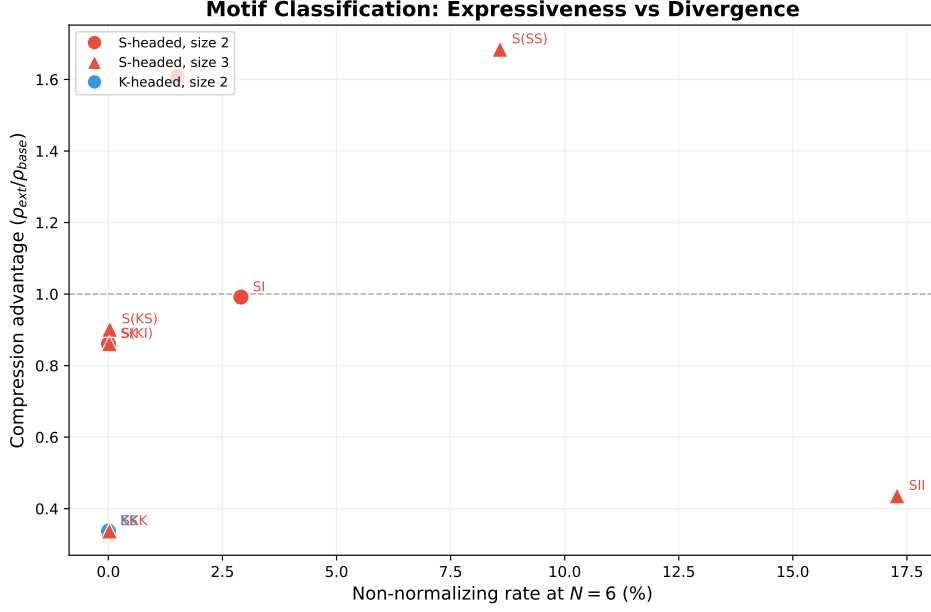
Figure 7: Motif classification scatter: compression advantage vs. non-normalizing rate at $N = 6$. Red: S-headed motifs. Blue: K-headed. Circles: size-2 motifs. Triangles: size-3.

A natural question is whether the decay rate has a lower bound. Our data show diminishing returns: $S(SS)$ achieves -0.605, the combination $\{SS, S(SS)\}$ achieves -0.572, and the triple $\{SS, S(SS), SII\}$ achieves -0.653 (worse than the pair, due to SII's divergence cost). Whether iterating the naming process can push the decay rate arbitrarily close to zero, or whether there is a fundamental floor, remains open.

5.2 What Determines the Value of a Basis Extension

Comparing all 11 single motifs reveals that three properties predict whether naming a motif improves or harms compression:

1. **Spine content (strongest predictor).** Among S-headed motifs, those whose spine consists entirely of S combinators (SS, $S(SS)$) are the only ones with advantage > 1.0. Introducing K into the spine (SK, $S(KS)$, $S(KI)$, SKK) drops the advantage to 0.34–0.90$\times$. The spine determines what the S combinator distributes when it fires: S-rich spines distribute further combinatorial structure, while K-rich spines distribute discarders.

2. **Head combinator (floor predictor).** All K-headed motifs (KS, KK, KI) produce advantages $\leq 0.337\times$. The head combinator sets a hard floor on expressiveness because K discards its second argument, collapsing the space of reachable outputs.

3. **Non-normalization rate (ceiling predictor).** SII demonstrates that even a motif with favorable structural properties (S-headed, no K in spine) can fail if it causes too many terms to diverge. Its 17.3% non-normalization rate at size 6 overwhelms the computational reach it provides.

This yields a practical test for motif selection: a good basis extension is one whose head combinator is S, whose spine arguments are S-rich and whose effective arity does not create excessive self-application.

5.3 Scaling Effects Across Naming Experiments

The compression advantage of the best motifs (SS, $S(SS)$) grows with term size: at size 3, both are near 1.1$\times$; by size 6, they reach 1.6–1.7$\times$ (Figure 5, right panel). This growth means
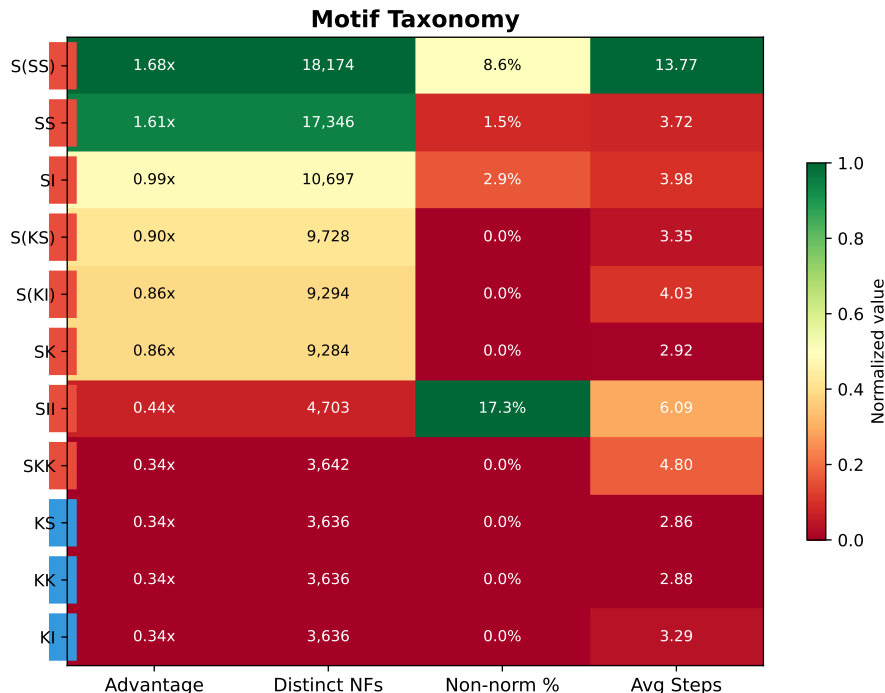
Figure 8: Motif taxonomy heatmap: all 11 single motifs (rows) across 4 metrics (columns). Red sidebar indicates S-headed; blue indicates K-headed.

that the benefit of having named a useful sub-expression *compounds as the system scales*. Each additional unit of syntactic complexity produces more distinct outputs in the extended basis than in the standard basis, and this gap widens at larger sizes.

The combination $\{SS, S(SS)\}$ extends this further, reaching $1.95\times$ at size 6. However, the sub-multiplicative interaction (72% of expected) suggests that the two motifs overlap in the computational territory they open up. A combination of motifs that expand orthogonal regions of the computation space might achieve closer to multiplicative gains.

5.4 Connection to Logical Depth

The average reduction steps $\overline{s}(N)$ increase monotonically, from 0 at size 1 to 2.75 at size 7 in the baseline. However, the maximum steps $s_{\max}$ reveal more: $s_{\max}(7) = 23$ in the baseline, but $s_{\max}(6) = 9{,}174$ for $S(SS)$ and $s_{\max}(6) = 2{,}292$ for SS. These are terms that undergo long chains of rewriting before settling into a normal form. In Bennett's terminology [Bennett, 1988], they have high logical depth: they are "computationally deep" outputs that require significant work to derive.

The extended bases amplify logical depth because the named motifs create more opportunities for S to fire, leading to longer reduction chains. This is another aspect of the divergence–expressiveness correlation: more computational reach means more computation happens, producing both more distinct normal forms and more non-termination.

5.5 Limitations

1. **Size ceiling.** Our exhaustive enumeration reaches size 7 for the baseline (288,684 terms), size 6 for single-motif and pair experiments and size 5 for triple combinations. Larger sizes would strengthen the exponential decay fits and might reveal higher-order transitions.

2. **Fuel sensitivity.** Terms classified as "divergent" might eventually normalize with a higher fuel limit. Our fuel of 10,000 is generous relative to the baseline (max normalizing steps
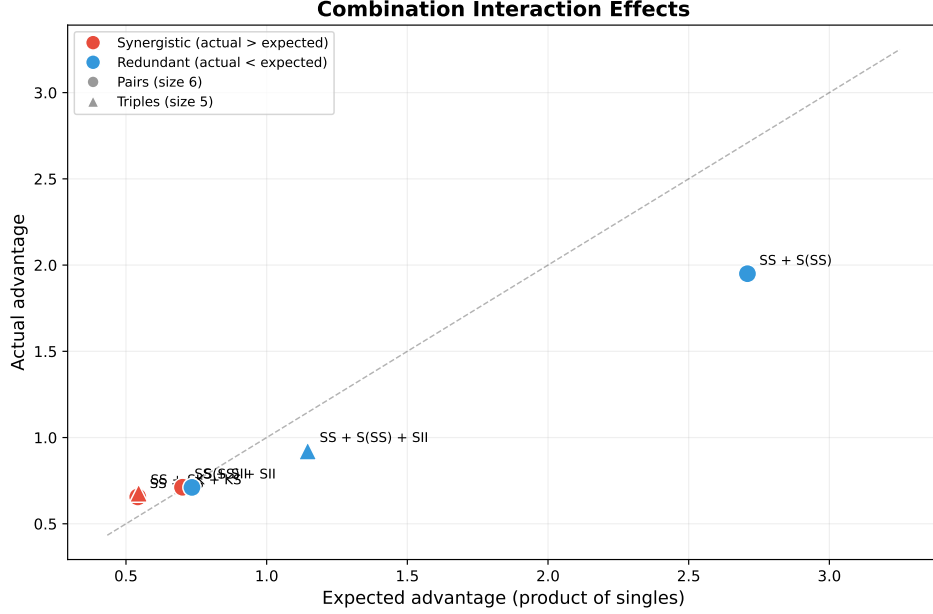
Combination Interaction Effects

Figure 9: Combination interaction effects. Each point is a multi-motif experiment; x-axis shows the expected advantage (product of individual motif advantages), y-axis shows the actual measured advantage. Points above the diagonal indicate super-multiplicative interaction; below indicates sub-multiplicative.

$= 23$) but the extended bases push this much higher (up to 9,174 for $S(SS)$), suggesting the classification boundary may shift at larger sizes.

3. **Basis size confound.** The combination experiments increase the basis size k, which independently increases the enumeration space. A $k = 5$ experiment at size 6 enumerates $3.8\times$ more terms than $k = 4$. The compression ratio already accounts for this (it divides by total terms), but it means the "search space" interpretation differs across experiments.

4. **Syntactic identity only.** We identify normal forms by syntactic equality. A richer analysis would classify them by computational behavior (e.g. identifying which normal forms act as projections, composition operators, etc.), which might reveal structure not visible in our metrics.

6 Conclusion

We have conducted an exhaustive empirical study of compression and structural emergence in combinatory logic, testing 11 individual motifs and 6 multi-motif combinations as basis extensions. Our main findings are:

1. **Exponential compression decay.** The compression ratio $\rho(N)$ decays as $e^{-0.727N}$ in the SKI basis ($R^2 = 1.000$), meaning the fraction of input terms producing unique normal forms roughly halves with every size increment.

2. **A structured cascade.** The transition from trivial to complex behavior is not a single event but a cascade: discarding ($N = 3$) $\rightarrow$ distribution ($N = 4$) $\rightarrow$ inner reduction ($N = 5$) $\rightarrow$ divergence ($N = 6$) $\rightarrow$ explosion ($N = 7$).

3. **Basis extension can slow or accelerate decay.** Of 11 motifs tested, 2 slow compression decay (best: $S(SS)$ at $1.68\times$ advantage), 1 is neutral and 8 accelerate it (worst:

$KS/KK/KI$ at 0.34×). The benefit of the two best motifs grows with system size, reaching 1.6–1.7× at size 6.

4. **Spine content predicts motif value.** Among S-headed motifs, those with S-only spines (SS, $S(SS)$) are the only winners. Introducing K into the spine drops the advantage below 1.0. All K-headed motifs produce identical, harmful results.

5. **Non-normalization constrains expressiveness.** SII, the self-application combinator, is S-headed with no K in its spine yet achieves only 0.44× advantage due to 17.3% non-normalization. High non-normalization is a necessary accompaniment of the best motifs ($S(SS)$: 8.6%, SS: 1.5%) but SII shows it can also dominate.

6. **Combination effects are sub-multiplicative for strong pairs.** The best pair $\{SS, S(SS)\}$ achieves 1.95× (72% of the expected product). Harmful motifs are less harmful than expected when paired with strong motifs (super-multiplicative ratios of 1.21–1.24).

6.1 Future Work

Several directions follow naturally:

- **Iterative naming.** Run the naming experiment in a loop: add the best motif, re-survey, add the next best motif. Does the compression advantage compound indefinitely, or is there a fixed point?
- **Optimal motif selection.** The spine-content predictor established here could be tested as a heuristic for selecting motifs without exhaustive evaluation.
- **Larger sizes.** Extend to size 8+ using sampling rather than exhaustive enumeration to test whether the exponential decay rates and the motif ranking remain stable.
- **Semantic classification.** Classify normal forms by computational behavior rather than syntactic identity. For instance, identifying which normal forms are projections ($\lambda xy.x$, $\lambda xy.y$), which are composition operators, etc.
- **Comparison to library learning.** The bottom-up motif evaluation performed here is complementary to DreamCoder's top-down, neural-guided library learning [Ellis et al., 2021]. A direct comparison on shared benchmarks would clarify the relative strengths of exhaustive vs. heuristic approaches.

References

C. H. Bennett. Logical depth and physical complexity. In R. Herken, editor, *The Universal Turing Machine: A Half-Century Survey*, pages 227–257. Oxford University Press, 1988.

K. Ellis, C. Wong, M. Nye, M. Sable-Meyer, L. Morales, L. Hewitt, L. Cary, A. Solar-Lezama, and J. B. Tenenbaum. DreamCoder: building libraries of compositional abstractions. *Proceedings of the ACM SIGPLAN Conference on Programming Language Design and Implementation (PLDI)*, 2021.

J. R. Hindley and J. P. Seldin. *Lambda-Calculus and Combinators: An Introduction*. Cambridge University Press, 2nd edition, 2008.

M. Li and P. Vitányi. *An Introduction to Kolmogorov Complexity and Its Applications*. Springer, 4th edition, 2019.

M. Mézard and A. Montanari. *Information, Physics, and Computation*. Oxford University Press, 2009.

J. Tromp. Binary lambda calculus and combinatory logic. In C. Calude, editor, *Randomness and Complexity, from Leibniz to Chaitin*, pages 237–260. World Scientific, 2014.

A Reduction Step Distribution

Figure 10 shows the distribution of reduction steps for normalizing terms at size 7 in the baseline. The distribution is concentrated at low step counts (mean $\bar{s} = 2.75$, max $= 23$), indicating that most terms normalize quickly. The right tail consists of terms requiring chains of S-reductions.
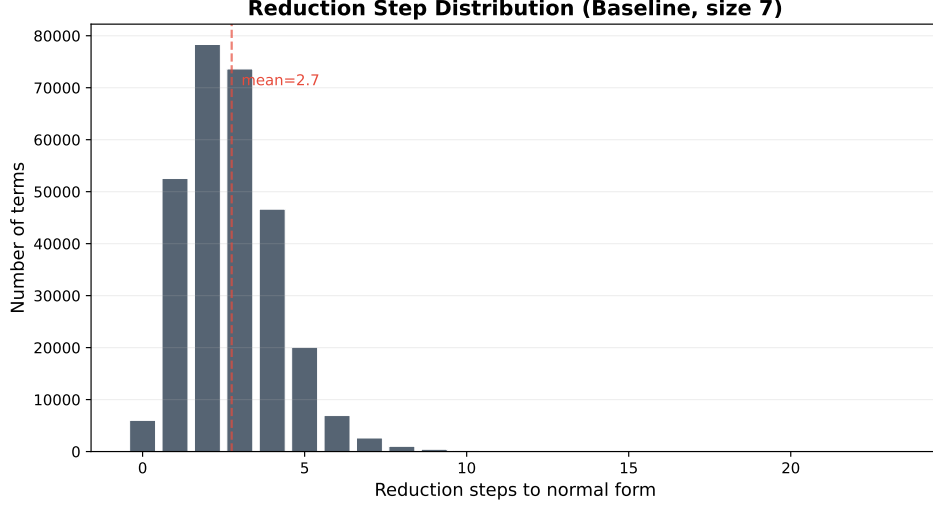

Figure 10: Distribution of reduction steps for normalizing terms at size 7 (baseline).

B Full Exponential Decay Fits

Table 5 shows the exponential decay parameters for all single-motif experiments. Note the exact equivalence of KS, KK, KI and the near-equivalence of SKK; and the exact equivalence of SK and $S(KI)$.

Table 5: Exponential decay parameters $\rho(N) \approx A \cdot e^{bN}$ for each single-motif experiment. Experiments are grouped by decay rate.

Experiment	Decay rate b	Per-step factor e^b	R^2
SKI + $S(SS)$	-0.605	0.546	0.999
SKI + SS	-0.608	0.544	1.000
SKI + SI	-0.715	0.489	1.000
Baseline (SKI)	-0.727	0.483	1.000
SKI + $S(KS)$	-0.747	0.474	1.000
SKI + SK	-0.762	0.467	1.000
SKI + $S(KI)$	-0.762	0.467	1.000
SKI + SII	-0.837	0.433	0.998
SKI + KS	-0.900	0.407	0.999
SKI + KK	-0.900	0.407	0.999
SKI + KI	-0.900	0.407	0.999
SKI + SKK	-0.900	0.407	0.999

C   Combination Experiment Decay Rates

Table 6: Exponential decay parameters for combination experiments.

Experiment	Decay rate b	Per-step factor e^b	R^2
$\{SS, S(SS)\}$	-0.572	0.564	0.999
$\{SS, S(SS), SII\}$	-0.653	0.521	0.996
$\{SS, SII\}$	-0.727	0.483	0.998
$\{S(SS), SII\}$	-0.729	0.482	0.997
$\{SS, KS\}$	-0.748	0.473	0.999
$\{SS, SK, KS\}$	-0.756	0.470	0.998