# Effective approximation, low-intensity bounds, and finite-size effects for autocatalytic emergence in reversible polymer networks

September 2026

**Abstract**

The critical-window law for reflexively autocatalytic and food-generated (RAF) sets in the reversible binary-polymer model states that, when the mean number of reactions catalysed per molecule grows like $\lambda n$, the RAF probability converges to $S_\nu(1-\mathrm{e}^{-\lambda})$, where $S_\nu$ is the probability that ordinary reversible closure of the food set is unbounded in an infinite field of independently open reaction channels. The law is known for the ordered split-position catalogue ($\nu = \mathrm{sp}$) and for the commuting-factor quotient catalogue ($\nu = \mathrm{qt}$). It is an existence theorem: it names the limit but says nothing effective about its value, its shape near zero, or what a finite network actually does. This paper answers those three questions.

First, for both profiles we construct a terminating algorithm which, given rational openness $a$ and rational tolerance $\varepsilon$, returns a rational interval of width below $\varepsilon$ containing $S_\nu(a)$, with an explicit approximation error $1/m + (2^{L+1}-2)(1-a^{L+1})^r$ for a computable seed length $L$ and record index $r$, uniform on every openness interval bounded away from zero. The key estimate is a finite-cap record contraction proved without conditioning on survival and without independence between repair attempts. Second, tracking total word length along productive closure histories replaces a coefficient of order $\binom{J}{r}$ over enormous channel caps by $C_r^\nu \le 36^r 2^{r(r-1)/2}$, giving $\log S_\nu(a) \le -\log^2(1/a)/(2\log 2) + O(\log(1/a))$ and exact certificates such as $S_\nu(10^{-20}) < 10^{-629}$. Third, at fixed word length $n$ the probability that some singleton RAF exists is exactly $1-(1-p)^{248}$ (split) or $1-(1-p)^{234}$ (quotient), so the finite RAF probability is linear in $p$ at zero while the limiting profile is flatter than every power; we quantify this order-of-limits effect with rational two-sided predictions and two fully explicit examples at $n = 16$. We also prove strict monotonicity of the profile, an effective global modulus of continuity, terminating inverse-probability brackets, and non-analyticity at zero. The effective evaluator, its error contract with the finite catalytic model, the settled-history structure and all printed arithmetic are checked in Lean 4; the sharper counting theorem, the singleton law and the regularity results have conventional proofs and are labelled as such.

**Keywords.** autocatalytic sets; RAF theory; binary polymer model; critical window; computable analysis; certified numerics; productive histories; finite-size effects; formal verification.

**MSC 2020.** 60K35, 92E20, 05C80, 03D78, 68V20.

## 1 Introduction

A reflexively autocatalytic and food-generated set, or RAF, is a nonempty family of reactions each of whose reactants can be built from a food set using reactions of the family, and each of whose reactions has a catalyst among the molecules so produced. The notion was distilled by Steel [21] from Kauffman's collectively autocatalytic sets [16, 17] and Eigen's hypercycles [5], and has been developed into a structural theory with polynomial-time detection

algorithms [11, 14, 10, 22]; see [8, 13] for surveys. RAF theory is deliberately structural: it asks which reaction families are self-sustaining in principle, and separates that question from kinetics, thermodynamics and molecular inventories. In particular a reaction whose product catalyses it is a singleton RAF, since the catalyst need not be present before the first hypothetical firing. This feature turns out to govern the finite-size behaviour studied below.

The standard random test bed is the binary polymer model [16, 21, 11]: molecules are binary words of length at most $n$, reactions are ligations and their reverse cleavages, the food set consists of the words of length at most two, and each molecule catalyses each reaction independently with a probability chosen so that a molecule catalyses $f_n$ reactions on average. Mossel and Steel [20] showed that $f_n$ growing linearly in $n$ is the relevant scale, and simulations [9, 12, 15] display a steep but smooth rise of the RAF probability within a narrow range of $f_n/n$. Two companion manuscripts identified the limit on this scale. For the ordered *split-position* catalogue [3], and for the *commuting-factor quotient* catalogue used by the public generator of Varanasi and Korenaga [25, 24, 2, 1], if $f_n/n \to \lambda > 0$ then

$$P_n^{\nu}(p_n) \longrightarrow \Theta_{\nu}(\lambda) = S_{\nu}(1 - \mathrm{e}^{-\lambda}), \qquad \nu \in \{\mathrm{sp}, \mathrm{qt}\}, \tag{1}$$

where $S_{\nu}(a)$ is the probability that ordinary reversible closure of the food set is unbounded in an infinite field of reaction channels that are independently open with probability $a$. The profile $S_{\nu}$ is continuous and nondecreasing, $S_{\nu}(0) = 0$, $S_{\nu}(1) = 1$, and $0 < S_{\nu}(a) < 1$ on $(0, 1)$. Thus the whole linear scale is a nontrivial critical window rather than a threshold.

## The questions

The law (1) is an existence theorem. It identifies the limit as a probability in an infinite product space but does not say how to compute $S_{\nu}(a)$ to a prescribed accuracy, how $S_{\nu}$ behaves near $a = 0$, or how a finite network at moderate $n$ relates to the limit. These are different questions with different answers, and we keep them separate throughout:

(Q1) *Computability.* Given rational $a$ and rational $\varepsilon > 0$, does an explicit algorithm terminate with a rational interval of width below $\varepsilon$ that provably contains $S_{\nu}(a)$?

(Q2) *Shape.* What can be proved about $S_{\nu}$ near zero, about strict monotonicity, inverse probabilities and regularity?

(Q3) *Finite size.* What does the finite catalytic model at a given $(n, p)$ actually satisfy, and how does its behaviour near $p = 0$ differ from that of the limit?

An answer to one is not an answer to the others. A certified evaluator can exist while being astronomically expensive; a sharp small-openness bound can be proved without any interior evaluation; and a finite network can be governed at small $p$ by a mechanism that is invisible in the limit.

## Main results

Write $M_n = 2^{n+1} - 2$ for the number of molecules and $R_n^{\nu}$ for the number of channels of cap $n$; let $p$ be the molecule–channel catalytic probability and $a$ the static channel openness.

**(A) Effective approximation and a certified evaluator** (Theorems 3.1 and 3.5). For a computable seed length $L = L_{\nu}(q, m)$ and the caps $D_r = 2^r(L + 2) - L$,

$$0 \le e_{D_r}^{\nu}(a) - S_{\nu}(a) \le \frac{1}{m} + (2^{L+1} - 2)(1 - a^{L+1})^r \qquad (a \in [q, 1]),$$

where $e_B^{\nu}(a)$ is the finite-cap probability that closure generates a word longer than $B$. Since $e_B^{\nu}(a)$ is a finite rational computation, this yields a terminating rational enclosure

algorithm for $S_\nu$ at every rational $a \in [0,1]$, uniform on $[q,1]$. The heart of the proof is a *record contraction* (Lemma 3.3): each escape beyond a cap exposes a fresh record word from which any missing short target can be repaired using at most $\ell+1$ channels of strictly larger product length, and the repair coordinates are disjoint from the exposed prefix. No conditioning on eventual survival and no independence between attempts is used.

**(B) Total-length history bounds** (Theorem 4.3, Corollary 4.5). Along a productive closure history every previously used channel stays settled, so the number of legal next channels is bounded by the square of the number of available types plus the number of cleavage slots, both of which are controlled by the *total word length* of the current set. This gives $S_\nu(a) \le C_r^\nu a^r$ with $C_r^\nu \le 36^r 2^{r(r-1)/2}$ and hence

$$\log S_\nu(a) \le -\frac{\log^2(1/a)}{2\log 2} + O(\log(1/a)),$$

replacing the earlier coefficient $\binom{J_r}{r}$ over a channel cap of size $J_r \approx 2^{2^{r+2}}$. Exact certificates such as $S_\nu(10^{-4}) < 10^{-15}$ and $S_\nu(10^{-20}) < 10^{-629}$ are kernel-checked.

**(C) Exact singleton law at finite size** (Theorem 5.1). For $n \ge 4$ the probability that some singleton RAF exists is exactly $1-(1-p)^{A_\nu}$ with $A_{\mathrm{sp}} = 248$ and $A_{\mathrm{qt}} = 234$, and every RAF without a singleton witness requires at least two positive catalytic coordinates. Consequently $P_n^\nu(p) = A_\nu p + O_n(p^2)$, so the finite network is linear in $p$ at zero while $\Theta_\nu(\lambda)$ is flatter than every power of $\lambda$. The logarithmic rates of the two iterated limits differ (equation (20)).

**(D) Finite predictions without the approximation cap** (Theorem 5.3, Section 5.3). Combining the singleton lower bound with the history bound and the inherited finite catalytic comparison gives

$$1-(1-p)^{A_\nu} \le P_n^\nu(p) \le \min\{1,\, C_r^\nu (M_n p)^r + G_\nu M_{2^{r+1}} p\},$$

which is evaluable in exact rational arithmetic. At $n=16$ and $\lambda = 10^{-4}$ this gives $2.16\cdot 10^{-7} < P_{16}^{\mathrm{sp}} < 4.57\cdot 10^{-5}$ while the limiting profile at the same intensity is below $10^{-15}$.

**(E) Regularity of the profile** (Section 6). $S_\nu$ is strictly increasing on $[0,1]$ with an explicit finite-forcing lower bound on increments; it has an effective global modulus of continuity including at zero; the inverse map $t \mapsto S_\nu^{-1}(t)$ admits a terminating rational bracketing algorithm; and $S_\nu$ is not real-analytic at zero.

**(F) Formal verification** (Appendix A). The evaluator, the uniform approximation, the record contraction, the seed selector, the finite source contract, the all-powers flatness, the settled-history structure, the generic support union bound and every printed inequality are Lean 4 declarations compiled with warnings promoted to errors and axiom closure {`propext`, `Classical.choice`, `Quot.sound`}. The polymer census behind the sharper coefficient, the singleton semantics, and the regularity results are conventional proofs and are marked as such in the text.

### What is new relative to the companions

We do not reprove the law (1); positivity, continuity, the endpoint values, the supplied-seed mass estimate and the finite catalytic comparison are inherited at their literal source conventions from [3, 1] and are listed in Section 2.5. The companions prove that finite escape probabilities converge to $S_\nu$; they do not give a rate, and their seed lengths are existential. The contribution

here is the effective rate, the sharper shape bound, the exact finite-size law, and the regularity results, together with the separation of the three questions above.

### Organisation

Section 2 fixes both models, the observables and the inherited inputs, and audits the channel counts. Section 3 proves the record contraction and the certified evaluator. Section 4 proves the total-length history bound and its logarithmic corollary. Section 5 proves the singleton law and the finite predictions, with two explicit examples. Section 6 proves strict increase, the modulus, inverse brackets and non-analyticity. Section 7 discusses scope and open problems. Appendix A is the concordance between numbered results and Lean declarations.

## 2 Models, observables, and inherited inputs

### 2.1 Words, channels, closure

Let $X_n = \bigcup_{\ell=1}^{n} \{0,1\}^\ell$ be the nonempty binary words of length at most $n$, $M_n = |X_n| = 2^{n+1}-2$, and $F = X_2$ the six food words. A *split channel* is a pair $(w, i)$ with $1 \le i < |w|$; it represents the reversible reaction $u + v \rightleftharpoons uv = w$ with $|u| = i$, and its ordered factors are part of its identity. Write $\mathcal{R}_n^{\mathrm{sp}}$ for the split channels of cap $n$; for $n \ge 2$,

$$R_n^{\mathrm{sp}} = |\mathcal{R}_n^{\mathrm{sp}}| = \sum_{\ell=2}^{n} (\ell - 1)2^\ell = (n-2)2^{n+1} + 4. \tag{2}$$

The *quotient* model identifies the descriptions $(u, v)$ and $(v, u)$ exactly when $uv = vu$; it never identifies channels with different products. This is the channel catalogue of the generator [24] formalised in [2, 1]. Write $\mathcal{R}_n^{\mathrm{qt}}$ and $R_n^{\mathrm{qt}}$ for the quotient channels and their number; Lemma 2.1 computes $R_n^{\mathrm{qt}}$ exactly. Throughout, $\nu \in \{\mathrm{sp}, \mathrm{qt}\}$ indexes the model, and the inclusions $\mathcal{R}_n^\nu \subseteq \mathcal{R}_{n+1}^\nu$ define the infinite channel set $\mathcal{R}_\infty^\nu$.

For a channel set $A$, the *ordinary reversible closure* $\mathcal{C}_n(A)$ is the smallest subset of $X_n$ containing $F$ such that whenever the channel of $(u, v)$ lies in $A$: if $u, v \in \mathcal{C}_n(A)$ then $uv \in \mathcal{C}_n(A)$, and if $uv \in \mathcal{C}_n(A)$ then $u, v \in \mathcal{C}_n(A)$. Availability is set-valued: a factor occurring twice needs one available type, not two copies. The same channel may be used in either direction. There are no concentration, rate, energy or inventory variables. In the infinite channel set the closure $\mathcal{C}_\infty(A)$ is defined identically without a cap; every membership has a finite derivation.

### 2.2 The finite catalytic model

Let $\xi_{x,r}$, $(x, r) \in X_n \times \mathcal{R}_n^\nu$, be independent Bernoulli variables with parameter $p \in [0, 1]$; $\xi_{x,r} = 1$ means that $x$ catalyses $r$ in both directions, one mark serving ligation and cleavage. A nonempty $A \subseteq \mathcal{R}_n^\nu$ is a *RAF* if every endpoint of every channel in $A$ lies in $\mathcal{C}_n(A)$ and every $r \in A$ has some $x \in \mathcal{C}_n(A)$ with $\xi_{x,r} = 1$. This is the closure-based RAF definition of [11, 9] for reversible reactions; it coincides with the predicate used in the companions. Food generation is a property of the closure, not of a catalysed order of firing, so a product may serve as the catalyst of its own reaction. Let $P_n^\nu(p)$ be the probability that some RAF exists. Under the *canonical normalisation* $p_n = \lambda n / R_n^\nu$ a molecule catalyses $\lambda n$ channels on average, and $\lambda$ is the dimensionless intensity.

### 2.3 The static field and the survival profile

Independently of catalysis, give each channel $j \in \mathcal{R}_\infty^\nu$ an independent Bernoulli$(a)$ *open* mark, with product law $\mathbb{P}_a$. The *survival profile* is

$$S_\nu(a) = \mathbb{P}_a\{\mathcal{C}_\infty(\text{open channels}) \text{ contains words of arbitrarily large length}\}. \tag{3}$$

Table 1: Channel conventions for caps $n \geq 4$. A growth gateway produces a nonfood word from two food words; the RAF gateways are all channels with a food side, including the four food-only channels. The last row is proved in Theorem 5.1.

| Quantity | Split | Quotient |
|---|---|---|
| Food types | 6 | 6 |
| Growth gateways $g_\nu$ | 32 | 30 |
| RAF gateways $G_\nu$ | 36 | 34 |
| Food-only channels | 4 | 4 |
| Singleton witness coordinates $A_\nu$ | 248 | 234 |

For a raw length threshold $B \geq 2$ let $e_B^\nu(a)$ be the probability that the open closure generates a word longer than $B$. A first crossing of length $B$ is a ligation of two available words of length at most $B$, so the event is determined by the channels of cap $2B$; in particular $e_B^\nu(a)$ is a finite rational computation for rational $a$, and $S_\nu(a) = \lim_{B\to\infty} e_B^\nu(a)$ by monotone continuity. (The formal development indexes escape by $K$ with $B = K + 2$; every comparison below is stated in terms of $B$.)

The three parameters used throughout are

$$p \text{ (catalytic probability)}, \qquad q_k(p) = 1 - (1-p)^k \text{ (pool openness)}, \qquad a_\lambda = 1 - \mathrm{e}^{-\lambda}. \quad (4)$$

The pool openness $q_k(p)$ is the probability that at least one of $k$ candidate catalysts marks a given channel; the law (1) arises because $q_{M_n}(\lambda n / R_n^\nu) \to a_\lambda$.

### 2.4 Gateways and the channel census

A channel is a *food gateway* if at least one of its sides is wholly contained in $F$; it is a *growth gateway* if in addition its product is not food. Exactly four channels have all endpoints in $F$ (the ligations $\mathtt{0}+\mathtt{0}$, $\mathtt{0}+\mathtt{1}$, $\mathtt{1}+\mathtt{0}$, $\mathtt{1}+\mathtt{1}$); we call these *food-only*. Table 1 collects the counts.

The growth counts need no network enumeration. There are $6^2 = 36$ ordered ligations of two food words, four of which have a food product, giving 32 split growth gateways. Quotienting merges the two commuting pairs $(\mathtt{0}, \mathtt{00}), (\mathtt{00}, \mathtt{0})$ and $(\mathtt{1}, \mathtt{11}), (\mathtt{11}, \mathtt{1})$; no other pair of distinct food words commutes. Adding the four food-only channels gives $G_{\mathrm{sp}} = 36$ and $G_{\mathrm{qt}} = 34$, the latter being the exact gateway count of [2].

**Lemma 2.1** (Quotient channel count)**.** *Let $\pi_d$ be the number of primitive binary words of length $d$, so that $\pi_d = 2^d - \sum_{k|d,\, k<d} \pi_k$. Then for $n \geq 2$*

$$R_n^{\mathrm{qt}} = R_n^{\mathrm{sp}} - D_n, \qquad D_n = \sum_{d \leq n/3} \pi_d \, \#\{(i,j) : 1 \leq i < j,\ d(i+j) \leq n\}. \quad (5)$$

*In particular $R_{16}^{\mathrm{sp}} = 1\,835\,012$ and $R_{16}^{\mathrm{qt}} = 1\,834\,798$.*

Conventional proof; not formalised.

*Proof.* Two distinct nonempty words $u \neq v$ commute exactly when both are powers of a common primitive word $t$: cancelling the shorter from the longer in $uv = vu$ produces a shorter commuting pair, and iterating terminates at a common root [19, 18]. Writing $u = t^i$, $v = t^j$ with $i \neq j$, the split descriptions $(u, v)$ and $(v, u)$ are distinct and merge into one quotient channel; equal exponents give a single description and no merge. Each unordered pair $\{i, j\}$ with $i < j$ and $d(i+j) \leq n$ therefore deletes exactly one description per primitive $t$ of length $d$, and $d(i+j) \geq 3d$ forces $d \leq n/3$. At $n = 16$ the contributions for $d = 1, \ldots, 5$ are $112, 24, 24, 24, 30$, totalling 214, and (2) gives the stated values. □

The formula agrees with literal enumeration for $n \le 9$ (Appendix A); the quotient loses only $O(n^2 2^{n/3})$ of $(n-2)2^{n+1}$ channels, but, as [1] shows, the two random experiments are genuinely different and neither profile is a relabelling of the other.

### 2.5 Inherited inputs

We use the following established results at their literal source conventions [3, 1].

(I1) $S_\nu(0) = 0$, $S_\nu(1) = 1$, $0 < S_\nu(a) < 1$ for $0 < a < 1$, $S_\nu$ is continuous and nondecreasing, and (1) holds for every sequence with $f_n/n \to \lambda > 0$.

(I2) *Supplied-seed mass.* Let $C_* = 2 \cdot 7^{64} \cdot 81$ be the explicit contour constant of the split source. If $m \ge 2$ and $k \ge 1$ satisfy

$$C_*(1-b^2)^k(m(m+1)+1) \le 1, \tag{6}$$

then, at every static openness $a \ge q$ and in every cap $n$, the closure of $F \cup X_L$ with $L = 10k$ occupies fewer than $(m-1)M_n/m$ molecule types with probability at most $1/m$. Here $b = q$ for split channels and $b = q/2$ for quotient channels; the latter uses the at-most-two-to-one channel projection and half-openness domination of [1].

(I3) *Finite catalytic comparison.* For $k_{n,m} = (m-1)\lfloor M_n/m \rfloor > 6$, near-full static closure at openness $q \le q_{k_{n,m}}(p)$ lower-bounds $P_n^\nu(p)$ up to the mass error of (I2); and for $n \ge 4$, $B \ge 2$,

$$P_n^\nu(p) \le e_B^\nu(q_{M_n}(p)) + G_\nu M_B p. \tag{7}$$

Every RAF contains a catalysed food gateway, so the RAF event is contained in the gateway event; the second term is a union bound over catalysts of length at most $B$ for the $G_\nu$ food gateways, and it retains food-only RAFs. The split adapter with the exact constant 36 (rather than the coarse constant 68 of the earlier source) is compiled in this development.

Condition (6) is the contour-ratio form of the supplied-seed estimate: with $z = C_*(1-b^2)^k$ it gives $z/(1-z) \le 1/[m(m+1)]$, which is the hypothesis under which the mass theorem delivers error $1/m$. The nontrivial geometric content is inherited; nothing below re-derives it.

## 3 Effective finite approximation

### 3.1 An executable seed selector

For rational $0 < q \le 1$ and integer $m \ge 2$ let $k_\nu(q,m)$ be the least positive integer satisfying (6) with $b = q$ ($\nu = \mathrm{sp}$) or $b = q/2$ ($\nu = \mathrm{qt}$), and set $L_\nu(q,m) = 10k_\nu(q,m)$. Since $0 \le 1-b^2 < 1$ the search terminates by geometric decay, and each test is an exact rational comparison. This replaces the existential seed length of the companions by a computable one. In this section $L = L_\nu(q,m)$.

Lean: `seedTest`, `seedIndex`, `seedIndex_spec`, `rational_deficit_bound`, `selected_contour_deficit`.

**Theorem 3.1** (Explicit uniform approximation)**.** *Define $D_0 = 2$ and $D_{r+1} = 2D_r + L$, so that $D_r = 2^r(L+2) - L$. For every $a \in [q,1]$ and every integer $r \ge 0$,*

$$0 \le e_{D_r}^\nu(a) - S_\nu(a) \le \frac{1}{m} + (2^{L+1}-2)(1-a^{L+1})^r. \tag{8}$$

Lean: `explicit_split_truncation`, `explicit_quotient_truncation`, `split_uniform_approximation`, `quotient_uniform_approximation`.

The proof occupies the rest of this subsection and the next. Fix the model $\nu$ and the openness $a$; all probabilities are with respect to $\mathbb{P}_a$ on the infinite static field.

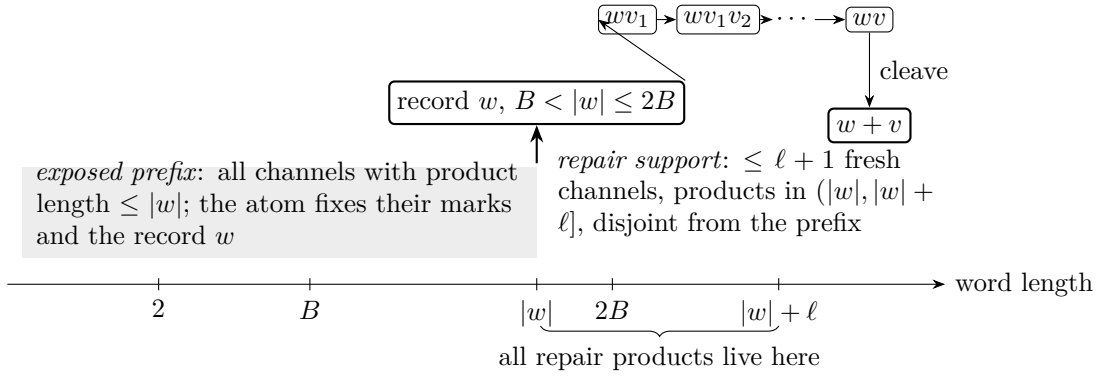


Figure 1: The record-and-repair mechanism of Lemmas 3.2 and 3.3. Conditioning on a finite exposed prefix through the first own-cap record $w$ leaves the repair coordinates untouched, so the conditional probability that they are all open is at least $a^{\ell+1}$ regardless of how many earlier attempts failed.

### 3.2 Records and fresh repairs

For a nonempty target word $v$ of length $\ell$ and a threshold $B \ge 2$ let $F_B(v)$ be the event that the open closure escapes length $B$ (generates a word longer than $B$) but does not generate $v$ using channels of cap $B$ only.

**Lemma 3.2** (Bounded record and repair)**.** *On the event of escape beyond $B \ge 2$ there is a word $w$ with $B < |w| \le 2B$ that is generated using only channels of cap $|w|$. From any such $w$ the target $v$ is generated by cap $|w| + \ell$ as soon as a specified set of at most $\ell + 1$ channels is open, and every one of these channels has product length strictly greater than $|w|$.*

*Proof.* Follow a finite derivation witnessing escape and stop at the first step producing a word longer than $B$. Before that step all available words have length at most $B$, and cleavage cannot increase length, so the crossing step is a ligation of two words of length at most $B$ and its product $w$ has $B < |w| \le 2B$. The derivation up to and including this step uses only products of length at most $|w|$, so $w$ is generated within its own cap. (Formally one proves by induction on derivations that a generated word is either confined to cap $B$ or the derivation contains such an own-cap record.)

Write $v = v_1 \cdots v_\ell$ with each $v_i \in F$. Append the letters successively, $w \to wv_1 \to wv_1v_2 \to \cdots \to wv$, and finally cleave $wv$ into $w$ and $v$. These are at most $\ell + 1$ channel identities (coincidences only reduce the number), each with product length in $(|w|, |w|+\ell]$. In the quotient model, canonicalising a channel preserves its product, so the separation of these repair channels from the cap $|w|$ is preserved; no half-openness comparison is needed for repair marks. □

Lean: `escape_has_bounded_record`, `generated_or_bounded_record`, `finite_record_support`, `quotient_finite_record_support`.

**Lemma 3.3** (Finite record contraction)**.** *For $B \ge 2$ and every nonempty target $v$ of length $\ell$,*

$$\mathbb{P}(F_{2B+\ell}(v)) \le (1 - a^{\ell+1})\,\mathbb{P}(F_B(v)). \tag{9}$$

*Proof.* Partition the escape event beyond $B$ according to the first own-cap record length $M > B$ and the complete configuration of the channels with product length at most $M$. These *prefix atoms* are pairwise disjoint and measurable with respect to finitely many coordinates. On each nonempty atom choose a deterministic record carrier $w$ with $|w| = M$; Lemma 3.2 gives $M \le 2B$.

Intersect the atom with the event that $v$ is absent from the cap-$B$ closure. This event is also determined by the prefix, so the intersection is a prefix event. The repair support of Lemma 3.2 consists of $s \le \ell + 1$ distinct channels with product length above $M$, hence disjoint from the prefix coordinates; conditionally on the prefix its all-open probability is $a^s \ge a^{\ell+1}$. When all repair channels are open, $v$ is generated by cap $M + \ell \le 2B + \ell$. Hence on each intersected atom the measure of continued absence of $v$ at cap $2B + \ell$ is at most $(1 - a^{\ell+1})$ times the measure of the atom.

Finally $F_{2B+\ell}(v)$ is covered by these continued-absence events: escape beyond $2B + \ell$ implies escape beyond $B$, and absence at the larger cap implies absence at cap $B$. Summing over the disjoint atoms, whose intersections with cap-$B$ absence partition $F_B(v)$, gives (9). The argument conditions only on a finite exposed configuration; it neither conditions on eventual survival nor asserts independence between failed attempts. □

Lean: `escaped_missing_contraction`, `quotient_escaped_missing_contraction`, `finite_missing_determined`, `first_record_cap`.

*Remark* 3.4. The contraction is the effective replacement for the record-word argument of the companions, which proves that unbounded closure eventually generates every fixed word but extracts no rate. The rate $1 - a^{\ell+1}$ per doubling of the cap is what makes the seed length below usable in a terminating algorithm. It is also why the approximation is uniform on $[q, 1]$: the factor only improves as $a$ increases.

### 3.3 From target repairs to survival

Iterate (9) along $d_0 = 2$, $d_{j+1} = 2d_j + \ell$; after $r$ steps the residual probability of escaping $d_r$ while missing $v$ at cap $d_r$ is at most $(1-a^{\ell+1})^r$. If $\ell \le L$ then $d_r \le D_r$ and $1-a^{\ell+1} \le 1-a^{L+1}$. A union bound over the $M_L = 2^{L+1} - 2$ words of $X_L$ gives

$$\mathbb{P}(\text{escape beyond } D_r \text{ and fail to acquire } X_L \text{ by cap } D_r) \le M_L(1 - a^{L+1})^r. \tag{10}$$

No independence between the repairs of different targets is used.
Lean: `repeated_record_failure`, `common_cap_target_failure`, `finite_seed_failure`, `exposure_budget`.

Let $H_L$ be the event that all of $X_L$ is acquired in the infinite field. The supplied-seed estimate (I2) gives

$$\mathbb{P}(H_L) \le S_\nu(a) + 1/m. \tag{11}$$

Here is the finite-to-infinite step, which is where an effective seed replaces an existential one. Fix a length $A$. On the event that the food closure stays inside $X_A$ it has at most $M_A$ types. If $X_L$ has been acquired by cap $n$, the closure of $F \cup X_L$ in cap $n$ is contained in the food closure. For all large $n$, $(m-1)M_n > mM_A$, so acquisition together with boundedness forces the supplied-seed mass failure, whose probability is at most $1/m$. Letting $n \to \infty$ exhausts the finite witnesses of acquisition and letting $A \to \infty$ exhausts bounded closure. Hence $\mathbb{P}(H_L \cap \{\text{bounded}\}) \le 1/m$, which is (11). Both events live in the same field.
Lean: `effective_seed_probability_le_survival`, `effective_quotient_seed_survival`.

*Proof of Theorem 3.1.* Split the event of escape beyond $D_r$ according to whether $X_L$ has been acquired by cap $D_r$. The failure branch has probability at most (10); the success branch is contained in $H_L$ and has probability at most (11). Survival implies escape beyond every cap, which is the lower inequality. The closed form $D_r + L = 2^r(L + 2)$ follows by induction. □

### 3.4 The certified evaluator

**Theorem 3.5** (Certified rational evaluation)**.** *For either model there is an explicit terminating algorithm which, given rational $a \in [0,1]$ and rational $\varepsilon > 0$, returns a rational interval of width less than $\varepsilon$ containing $S_\nu(a)$. The escape cap computed for a positive rational lower endpoint $q$ controls every real openness $a \geq q$ with the same error budget.*

Lean: `splitCertifiedInterval_correct`, `quotientCertifiedInterval_correct`, `splitTotalInterval_correct`, `quotientTotalInterval_correct`, `splitEscapeEval_correct`, `quotientEscapeEval_correct`, `recordIndex_error`.

*Proof.* At $a = 0$ return $[0,0]$, using $S_\nu(0) = 0$. Otherwise put $m = \lceil 4/\varepsilon \rceil + 2$, so $1/m \leq \varepsilon/4$; compute $L = L_\nu(a,m)$; and let $r$ be the least positive integer with

$$2^{L+1}(1 - a^{L+1})^r \leq \varepsilon/4. \tag{12}$$

This is an exact rational test that terminates by geometric decay, the case $a = 1$ being immediate. Now evaluate $e = e^\nu_{D_r}(a)$ exactly: the event is determined by the finitely many channels of cap $2D_r$; for each Boolean configuration iterate the one-step closure operator $M_{2D_r}$ times, which computes the full finite closure since each round either adds a type or is stable; sum the weights $a^s(1-a)^{J-s}$ over the configurations witnessing escape, where $J$ is the channel cardinality and $s$ the number of open channels. By Theorem 3.1, $[e - \varepsilon/2, e]$ contains $S_\nu(a)$ and has width $\varepsilon/2 < \varepsilon$; it may be intersected with $[0,1]$ without losing containment.

For uniformity compute $L$, $r$ and $D_r$ at the lower endpoint $q$. The seed bound (I2) is uniform for $a \geq q$, and $(1 - a^{L+1})^r \leq (1 - q^{L+1})^r$, so the same budget applies. Every search in this construction terminates by a proof, and the finite-event semantics are proved rather than assumed; no approximation oracle enters. □

*Remark* 3.6 (Cost). Theorem 3.5 is a computability statement in the sense of computable analysis [26], not a practical interior algorithm. At $m = 100$ and $q = 1/2$ the selected seed lengths are $L = 4830$ (split) and $L = 21520$ (quotient); already in the split case a repair error of 0.01 requires a record index with $\log_2 r \approx 4843$, after which the recurrence $D_{r+1} = 2D_r + L$ and the enumeration of channel fields in cap $2D_r$ are applied. These are the costs of the *proof*, not estimates of $S_\nu$. Section 4 obtains useful bounds without constructing any such cap, and Section 5 obtains useful finite predictions without it.

## 4 Total-length history bounds

A *productive history* is an ordered list of distinct channels $r_1, \ldots, r_k$, starting from $C_0 = F$, such that each $r_{j+1}$ is enabled by the current available set $C_j$ (one side of the channel is available) and adds at least one new type; $C_{j+1}$ is $C_j$ together with all endpoints of $r_{j+1}$. Productive histories are the combinatorial skeleton of closure. The following structural lemma is what makes counting them efficient.

**Lemma 4.1** (Settled channels and finite histories)**.** *After every productive step all endpoints of every previously used channel belong to the current set, and the current set is exactly the full closure of $F$ under the channels used so far. If the open closure reaches a word longer than $2^{r+1}$, then there is a productive history of $r$ distinct open channels; in particular survival supplies such a history for every $r$.*

*Proof.* The first assertion holds initially and persists: after firing an enabled channel its previously missing endpoints are available, so it too is settled, and settled channels stay settled when types are added. The new set is closed under every channel used so far, is reachable from

$F$ by those channels, and contains $F$; hence it is their full closure. In particular adjoining a new channel causes no further cascade through earlier channels.

In $j$ steps the maximum word length is at most $2^{j+1}$, since a ligation at most doubles length and a cleavage does not increase it. If the current set omits a target that is reachable using all open channels, then some open channel is productive on the current set: otherwise the set is closed under all open channels and contains $F$, hence contains the target. A previously used channel cannot be productive because it is settled, so the productive channel is new. Repeat $r$ times; the length bound prevents reaching the long target early. In the infinite field the target has a finite derivation, so the whole argument takes place in a finite cap; no infinite scheduling is needed. □

Lean: `settled_insert_closure`, `productive_history_settled`, `productive_history_ordered`, `split_long_target_history`, `quotient_long_target_history`, `split_unbounded_histories`, `quotient_unbounded_histories`, `publication_history_resolution`.

**Lemma 4.2** (Length and cleavage-slot budgets)**.** *For a productive history let* $T_j = \sum_{w \in C_j} |w|$ *and* $Q_j = \sum_{w \in C_j} (|w| - 1)$. *Then for all* $j \geq 0$

$$|C_j| \leq 6 + 2j, \qquad \max_{w \in C_j} |w| \leq 2^{j+1}, \qquad T_j \leq 2^{j+2} + 6, \qquad Q_j \leq 4 \cdot 2^j - j. \tag{13}$$

*Proof.* Initially $(|C_0|, T_0, Q_0) = (6, 10, 4)$ and each step adds at most two types. A productive ligation adds one word of length at most $2^{j+2}$, increasing $T$ by at most $2^{j+2}$ and $Q$ by at most $2^{j+2} - 1$. A productive cleavage of an available product $w$ adds its factors; their total length is $|w| \leq 2^{j+1}$ whether or not one factor was already present or the factors coincide, and at least one factor is new, so $T$ increases by at most $|w|$ and $Q$ by at most $|w| - 1$. Summing $T_{j+1} - T_j \leq 2^{j+2}$ and $Q_{j+1} - Q_j \leq 2^{j+2} - 1$ gives the last two bounds. □

Conventional proof; not formalised.

The point of tracking $T_j$ and $Q_j$ as *sums* rather than as (maximum length)×(cardinality) is the quantitative gain: the number of cleavage candidates is the number of split positions in available words, which is $Q_j$, not $|C_j| \cdot 2^{j+1}$. Neither quantity is a physical mass; both count word types.

Set $C_0^\nu = 1$ and, for $r \geq 1$,

$$b_j = (6 + 2j)^2 + 4 \cdot 2^j - j \quad (j \geq 1), \qquad C_r^\nu = g_\nu \prod_{j=1}^{r-1} b_j, \tag{14}$$

so that $C_1^\nu = g_\nu$, $C_2^{\mathrm{sp}} = 32 \cdot 71 = 2272$ and $C_2^{\mathrm{qt}} = 30 \cdot 71 = 2130$.

**Theorem 4.3** (History upper bound)**.** *For* $a \in [0, 1]$, $r \geq 0$ *and either model,*

$$S_\nu(a) \leq e_{2^{r+1}}^\nu(a) \leq \min\{1,\, C_r^\nu a^r\} \leq \min\{1,\, (36a)^r 2^{r(r-1)/2}\}. \tag{15}$$

*The middle inequality also holds when the channels are independently open with nonidentical probabilities all bounded by* $a$.

*Proof.* From $C_0 = F$ exactly the $g_\nu$ growth gateways are productive: food-only channels add nothing, and a cleavage of a food word yields food. After $j \geq 1$ steps, at most $|C_j|^2$ ordered factor pairs can be ligated and at most $Q_j$ (product, split position) pairs can be cleaved; discarding non-productive candidates or identifying quotient duplicates cannot increase the count. Lemma 4.2 bounds the total by $b_j$. Hence there are at most $C_r^\nu$ legal ordered histories of length $r$, and the family is finite because each prefix has finite branching and the length bound confines every endpoint to a finite word universe.

By Lemma 4.1, escape beyond $2^{r+1}$ is covered by the events that all $r$ channels of some history in this deterministic family are open. A fixed history has $r$ distinct channel identities, so its cylinder event has probability $a^r$; for heterogeneous independent marks it is the product of the $r$ marginals, at most $a^r$. The union bound over the family gives the middle inequality. The histories are enumerated before probabilities are assigned; no claim is made that an adaptively chosen next mark is an unconditioned Bernoulli variable.

For the closed envelope, $(6+2j)^2 \le 32 \cdot 2^j$ for $j \ge 1$ (equality at $j = 1$, and $(8+2j)^2 \le 2(6+2j)^2$ for $j \ge 1$ by expanding), so $b_j \le 36 \cdot 2^j$ and $g_\nu \le 36$; multiplying gives $C_r^\nu \le 36^r 2^{r(r-1)/2}$. The case $r = 0$ is the trivial bound one. □

Conventional proof; Lean-checked components: the generic union-bound step `history_support_union_bound`, the coefficient envelope `historyBudget_envelope`, and the history extraction of Lemma 4.1; the polymer census bounding the branching by $b_j$ is conventional.

*Remark* 4.4 (Comparison with the sparse-support bound). The companions' all-powers flatness rests on the observation that $r$ open channels confine closure to length $2^{r+1}$, so escape requires $r$ open channels among the $J_r = R^{\mathrm{sp}}_{2(2^{r+1}+2)}$ channels of the witness cap, giving $S_\nu(a) \le \binom{J_r}{r} a^r$. This is valid, Lean-checked, and numerically useless beyond very small $r$: at $r = 2$ the coefficient is $\binom{J_2}{2} = 712\,483\,666\,919\,430$ against $C_2^{\mathrm{sp}} = 2272$, and $J_r$ grows like $2^{2^{r+2}}$. The settled-history structure removes the dependence on the cap entirely. Figure 2 compares the two envelopes.

Lean: `split_sparse_real`, `quotient_sparse_real`, `split_second_coefficient`, `worked_low_openness_bound`.

**Corollary 4.5** (Logarithmic shape and flatness)**.** *Put* $x = \log(1/a)$, $h = \log 2$ *and* $y = x - \log 36 + h/2$. *For* $0 < a < 1$ *with* $y \ge 0$,

$$\log S_\nu(a) \le -\frac{y^2}{2h} + \frac{h}{8} = -\frac{\log^2(1/a)}{2\log 2} + O(\log(1/a)). \tag{16}$$

*In particular* $S_\nu(a) = o(a^d)$ *for every fixed* $d$, *and* $\Theta_\nu(\lambda) = S_\nu(a_\lambda) = o(\lambda^d)$ *as* $\lambda \downarrow 0$, *with the same logarithmic expansion in* $\log(1/\lambda)$.

*Proof.* The logarithm of the last expression in (15) is $hr^2/2 - yr$. Choose $r$ to be a nearest nonnegative integer to $y/h$; then $|r - y/h| \le 1/2$ and completing the square gives (16). The logarithm is defined because $S_\nu(a) > 0$ by (I1). Flatness follows directly from (15) with $r = d + 1$. Since $0 < a_\lambda \le \lambda$, the intensity statement follows from the same fixed-power bound, and $\log(1/a_\lambda) = \log(1/\lambda) + O(\lambda)$ transfers the expansion. □

Conventional proof; Lean-checked components: the all-powers flatness statements `split_low_openness_flatness`, `quotient_low_openness_flatness`, `split_low_intensity_flatness`, `quotient_low_intensity_flatness` (via the sparse coefficient); the logarithmic optimisation is conventional.

The estimate (16) is one-sided. It is not an asymptotic equivalence and no optimality is claimed for the coefficient $1/(2\log 2)$; matching lower bounds are an open problem (Section 7). Lean: `history_certificate_003`, `history_certificate_004`, `history_certificate_006`, `history_certificate_020`, `publication_small_openness_arithmetic`.

**Proposition 4.6** (Independence is essential)**.** *Let one Bernoulli(a) variable open every channel simultaneously. Then every channel is open with marginal probability a, yet the survival probability equals a exactly.*

*Proof.* When all channels are open the closure generates every word; when all are closed it stays at $F$. □

Conventional proof; not formalised.

Marginal openness bounds alone therefore cannot imply any all-powers estimate. The heterogeneous independent extension in Theorem 4.3 accommodates missing or less probable channels; correlated catalytic mechanisms would need a separate dependence theorem.

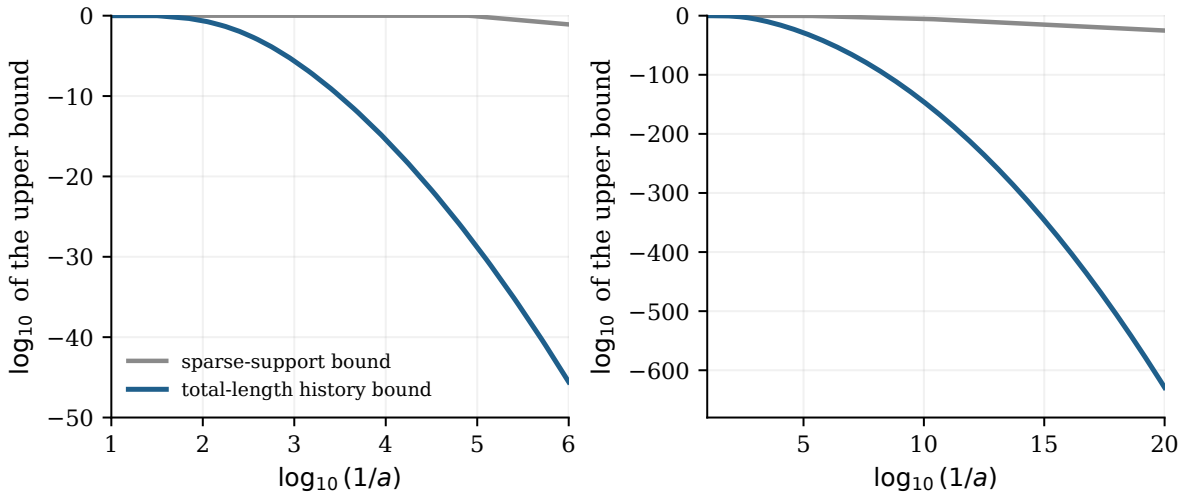


Figure 2: Upper bounds on the split profile, not a plot of $S_{\mathrm{sp}}$. Grey: the sparse-support envelope $\min_{1\le r\le 10}\binom{J_r}{r}a^r$ of Remark 4.4. Blue: the total-length history envelope $\min_{0\le r\le 80} C_r^{\mathrm{sp}} a^r$ of Theorem 4.3. Both are clipped at one; the left panel enlarges the moderate range. Logarithms are floating-point renderings; the certificates in Table 2 are exact.

Table 2: Exact history-product certificates $C_r^{\mathrm{sp}} a^r < 10^{-t}$, valid for both models since $C_r^{\mathrm{qt}} \le C_r^{\mathrm{sp}}$. The indices $r$ need not be optimal. The integer/rational comparisons are kernel-checked; their probabilistic meaning uses Theorem 4.3.

| Openness $a$ | History length $r$ | $\log_{10} C_r^{\mathrm{sp}}$ (approx.) | Certified consequence |
|---|---|---|---|
| $10^{-3}$ | 8 | 18.354 | $S_\nu(a) < 10^{-5}$ |
| $10^{-4}$ | 12 | 32.578 | $S_\nu(a) < 10^{-15}$ |
| $10^{-6}$ | 18 | 62.431 | $S_\nu(a) < 10^{-45}$ |
| $10^{-20}$ | 65 | 670.814 | $S_\nu(a) < 10^{-629}$ |

# 5 Finite-network contributions and predictions

## 5.1 Singleton RAFs and the first-order law

Write $Q_n^\nu = M_n R_n^\nu$ for the number of catalytic coordinates of cap $n$.

**Theorem 5.1** (Exact singleton contribution)**.** *For $n \ge 4$ let $A_{\mathrm{sp}} = 248$ and $A_{\mathrm{qt}} = 234$. Then*

$$\mathbb{P}_p(\textit{some singleton RAF exists}) = 1 - (1-p)^{A_\nu}, \tag{17}$$

$$0 \le P_n^\nu(p) - \left[1 - (1-p)^{A_\nu}\right] \le \binom{Q_n^\nu}{2} p^2. \tag{18}$$

*Consequently, for each fixed $n \ge 4$,*

$$P_n^\nu(p) = A_\nu p + O_n(p^2), \qquad P_n^\nu(\lambda n/R_n^\nu) = \frac{A_\nu n}{R_n^\nu}\,\lambda + O_n(\lambda^2). \tag{19}$$

*Proof.* A single channel $\{r\}$ is food-generated exactly when $r$ is a food gateway: if neither side of $r$ lies in $F$ the channel can never fire alone, so $\mathcal{C}_n(\{r\}) = F$ contains no endpoint of $r$ outside $F$. A food-only channel has $\mathcal{C}_n(\{r\}) = F$, so it has six candidate catalysts. A growth gateway $r$ with product $w \notin F$ has $\mathcal{C}_n(\{r\}) = F \cup \{w\}$, so it has seven candidate catalysts, the product $w$ among them; the structural RAF definition allows a product to catalyse its

own reaction, and a kinetic start-up restriction would define a different model. Hence the set of single catalytic coordinates $(x, r)$ certifying a singleton RAF has exactly $4 \cdot 6 + 7g_\nu = A_\nu$ elements. The coordinate includes the channel identity, so witness sets of different channels are disjoint even when they share a catalyst molecule, and independence gives (17).

If at most one coordinate is positive and a RAF $A$ exists, every channel of $A$ must be catalysed through that one coordinate, so $A$ is a singleton. Hence the event that a RAF exists but no singleton witness is present is contained in the event that at least two coordinates are positive, whose probability is at most $\binom{Q_n^\nu}{2}p^2$ by the union bound over pairs. This gives (18), and expanding the polynomial $1-(1-p)^{A_\nu}$ at $p=0$ gives (19). □

Conventional proof; not formalised.

The census $A_{\mathrm{sp}} = 248$, $A_{\mathrm{qt}} = 234$ is confirmed by literal enumeration at $n = 4$ (Appendix A); the universal statement for all $n \geq 4$ is the argument above.

**Corollary 5.2** (Order of limits)**.** *For each fixed $n \geq 4$,*

$$\lim_{\lambda \downarrow 0} \frac{\log P_n^\nu(\lambda n / R_n^\nu)}{\log \lambda} = 1, \qquad \lim_{\lambda \downarrow 0} \frac{\log \Theta_\nu(\lambda)}{\log \lambda} = +\infty. \tag{20}$$

*Moreover, conditional on the food-inclusive gateway event (some food gateway catalysed by some molecule), whose probability is $1-(1-p)^{G_\nu M_n}$, the RAF probability at fixed $n$ tends to $A_\nu/(G_\nu M_n)$ as $p \downarrow 0$, whereas the corresponding limiting ratio $\Theta_\nu(\lambda)/(1-(1-a_\lambda)^{G_\nu})$ is still $o(\lambda^d)$ for every $d$.*

*Proof.* The first logarithm is $\log \lambda$ plus a constant plus $o(1)$ by (19). For the second, Theorem 4.3 with $a_\lambda \leq \lambda$ gives $\log \Theta_\nu(\lambda) \leq \log C_r^\nu + r \log \lambda$ for every $r$, and positivity makes the quotient well defined; dividing by $\log \lambda < 0$ gives $\liminf \geq r$ for every $r$. Every RAF contains a catalysed food gateway (Section 2.5), so the conditional statement follows from (19); in the limit the denominator is at least $a_\lambda$, so at most one power is lost. □

Conventional proof; Lean-checked components: `conditional_intensity_flatness`, `split_conditional_intensity_flatness`, `quotient_conditional_intensity_flatness`; the finite-$n$ limits are conventional.

The unscaled probabilities vanish in both iterated limits; what differs is the rate. This is the precise sense in which a fixed finite network and the critical-window limit disagree at small intensity, and it is entirely due to singleton witnesses, which have no analogue in the infinite static field.

## 5.2 A finite error contract

The effective evaluator also calibrates a finite catalytic source, even in regimes where the history bound is not small. Let $h_{n,L}^\nu(q)$ be the static probability that the food closure in cap $n$ contains $X_L$, with $L \leq n$. Choose $m \geq 2$, $k_{n,m} > 6$, rational $0 < q \leq q_{k_{n,m}}(p)$, and the computed seed length $L = L_\nu(q, m)$. In the same field, acquisition of $X_L$ and success of supplied-seed mass imply near-full food closure, so (I3) gives

$$h_{n,L}^\nu(q) - \frac{1}{m} \;\leq\; P_n^\nu(p) \;\leq\; e_B^\nu(q_{M_n}(p)) + G_\nu M_B p. \tag{21}$$

Both finite probabilities are exactly rational for rational inputs. If $[l_P, u_P]$ is such a source enclosure and $[l_S, u_S]$ a profile enclosure from Theorem 3.5, then

$$|P_n^\nu(p) - S_\nu(a)| \leq \max\{u_P - l_S,\, u_S - l_P\}. \tag{22}$$

Lean: `splitSourceInterval_correct`, `quotientSourceInterval_correct`, `interval_comparison_error`, `split_source_profile_error`, `quotient_source_profile_error`, `canonical_raf_le_finite_escape_36`.

Table 3: Finite-source terms at $n = 16$, $\lambda = 10^{-4}$, each model with its own normalisation. The three probability entries are rounded displays of exact fractions, not outward endpoints; the rigorous outer intervals are (24). There is no sampling uncertainty anywhere.

| Model | $R_{16}^{\nu}$ | $C_2^{\nu}$ | $\Lambda_\nu$ | $U_{\text{hist}}$ | $U_{\text{short}}$ |
|---|---|---|---|---|---|
| split | 1 835 012 | 2 272 | $2.162383 \cdot 10^{-7}$ | $2.967407 \cdot 10^{-5}$ | $1.600861 \cdot 10^{-5}$ |
| quotient | 1 834 798 | 2 130 | $2.040551 \cdot 10^{-7}$ | $2.782593 \cdot 10^{-5}$ | $1.512101 \cdot 10^{-5}$ |

This contract retains the actual input $p$; it is not a cap-only convergence rate for an unspecified sequence $f_n/n \to \lambda$, and cannot be, since a convergent sequence may sit one unit from its limit at any prescribed finite index.

**Theorem 5.3** (History-based finite prediction)**.** *Let $n \geq 4$, $r \geq 1$ and $B = 2^{r+1}$ with $2B \leq n$. For every legal catalytic probability $p$,*

$$1 - (1-p)^{A_\nu} \;\leq\; P_n^{\nu}(p) \;\leq\; \min\{1,\, C_r^{\nu}(M_n p)^r + G_\nu M_B p\}. \tag{23}$$

*Proof.* The lower bound is (17). For the upper bound apply the finite-escape form $e_{2^{r+1}}^{\nu}(a) \leq C_r^{\nu} a^r$ of Theorem 4.3 inside (7) with $a = q_{M_n}(p) \leq M_n p$, the last inequality being the union bound over catalyst opportunities. The estimate remains valid when $M_n p > 1$, though clipping at one may then be uninformative. The cap restriction $2B \leq n$ keeps the escape event inside the ambient word universe and suffices for the examples. □

Conventional proof; not formalised.

Increasing $r$ improves the history term at small openness but lets $M_{2^{r+1}} p$ grow doubly exponentially; a useful choice controls the sum.

### 5.3 Two exact examples at cap sixteen

Take $n = 16$, $\lambda = 10^{-4}$, $r = 2$ and $B = 8$, so $M_n = 131\,070$, $M_B = 510$, $2B = n$, $C_2^{\text{sp}} = 2272$ and $C_2^{\text{qt}} = 2130$; the channel counts are those of Lemma 2.1. Each model uses its own normalisation $p = 16/(10^4 R_{16}^{\nu})$, that is $p = 1/1\,146\,882\,500$ (split) and $p = 1/1\,146\,748\,750$ (quotient). Bonferroni's inequality gives the computable lower bound $\Lambda_\nu = A_\nu p - \binom{A_\nu}{2} p^2 \leq 1 - (1-p)^{A_\nu}$; write $U_{\text{hist}} = C_2^{\nu}(M_n p)^2$ and $U_{\text{short}} = G_\nu M_B p$. Table 3 displays the three terms.

Exact integer cross-multiplication proves

$$2.16 \cdot 10^{-7} < P_{16}^{\text{sp}}\Big(\tfrac{16}{10^4 \cdot 1835012}\Big) < 4.57 \cdot 10^{-5}, \qquad 2.04 \cdot 10^{-7} < P_{16}^{\text{qt}}\Big(\tfrac{16}{10^4 \cdot 1834798}\Big) < 4.31 \cdot 10^{-5}. \tag{24}$$

Lean: `split_finite_example_arithmetic`, `quotient_finite_example_arithmetic`, `split_history_coefficient_two`, `quotient_history_coefficient_two`.

The arithmetic is kernel-checked separately from the conventional singleton and history semantics that give it meaning.

At the same intensity, $a_\lambda < 10^{-4}$, so Table 2 gives $\Theta_\nu(10^{-4}) < 10^{-15}$. The finite probability is already above $2 \cdot 10^{-7}$ because of singleton witnesses, more than eight orders of magnitude above the limit. This is a genuine source-versus-profile comparison with no Monte Carlo error and no claim that cap sixteen approximates the limit well; it is the concrete face of Corollary 5.2.

## 6 Regularity of the profile

The results of this section are conventional; they use the inherited estimates and the theorems above at their stated scopes. Throughout, $g = g_\nu$ and $G = G_\nu$.

**Proposition 6.1** (Strict increase by finite forcing). *For $0 < a < b < 1$ and either model there is an explicit finite channel set $T$, with $|T| \le (L+1)M_L$ for a computable $L \ge 4$, such that*

$$S_\nu(b) - S_\nu(a) \;\ge\; \tfrac{1}{2}\,(b-a)(1-b)^{g-1}b^{|T|} > 0. \tag{25}$$

*Together with (I1), $S_\nu$ is strictly increasing on $[0,1]$ and maps $[0,1]$ bijectively onto itself.*

*Proof.* Couple all thresholds by independent uniform variables $U_j$ on $[0,1]$, channel $j$ being open at openness $t$ when $U_j \le t$; then survival at $a$ implies survival at $b$, so $S_\nu(b) - S_\nu(a)$ is the probability of surviving at $b$ but not at $a$. Select the growth gateway $\mathtt{0+00} \rightleftharpoons \mathtt{000}$ with carrier $w = \mathtt{000}$, and let $\Gamma$ be the event that its variable lies in $(a,b]$ while the variables of the other $g-1$ growth gateways exceed $b$. On $\Gamma$ no growth gateway is open at $a$, so the closure at $a$ is $F$ and survival at $a$ fails; at $b$ the selected gateway is open and $w$ is available.

Choose $L \ge 4$ such that, at openness $b$, the closure of $F \cup X_L$ is unbounded with probability at least $1/2$; this follows from (I2) with $m = 2$ by the finite-to-infinite argument of Section 3, using $L = L_\nu(b, 2)$. Let $T$ be the union over $v \in X_L$ of the repair supports of Lemma 3.2 for the carrier $w$, duplicates removed; $|T| \le (L+1)M_L$. Every channel of $T$ has product length above three and a nonfood factor, so $T$ contains no growth gateway in either convention. Let $E$ be the event that the closure of $F \cup X_L$ at openness $b$ is unbounded. Since every growth gateway has all its endpoints in $X_4 \subseteq X_L$, $E$ does not depend on the growth-gateway variables, and $E$ is increasing in the remaining marks.

On $\Gamma \cap \{T \text{ open at } b\} \cap E$ the closure at $b$ contains $w$, hence $X_L$, hence the unbounded closure of $F \cup X_L$; so survival at $b$ holds. Now $\Gamma$ is independent of $(T, E)$ with $\mathbb{P}(\Gamma) = (b-a)(1-b)^{g-1}$, and $E$ is an increasing event whose conditional probability given that the finite set $T$ is open is at least its unconditional probability (condition on all other coordinates; an increasing function of the $T$-coordinates is at least its average at the all-open point; this is a special case of Harris' inequality [7, 6]). Hence

$$S_\nu(b) - S_\nu(a) \;\ge\; \mathbb{P}(\Gamma)\,\mathbb{P}(T \text{ open})\,\mathbb{P}(E \mid T \text{ open}) \;\ge\; (b-a)(1-b)^{g-1}b^{|T|} \cdot \tfrac{1}{2}.$$

At the endpoints, $S_\nu(0) = 0 < S_\nu(b)$ by positivity and $S_\nu(a) \le 1 - (1-a)^g < 1 = S_\nu(1)$ since some growth gateway must open. □

Conventional proof; not formalised.

**Proposition 6.2** (Effective global modulus of continuity). *Given rational $\varepsilon > 0$ one can compute $\delta > 0$ such that $|S_\nu(s) - S_\nu(t)| \le \varepsilon$ whenever $s, t \in [0,1]$ and $|s - t| \le \delta$.*

*Proof.* Let $C = g$, so $S_\nu(t) \le 1 - (1-t)^g \le Ct$. Put $q = \min\{1/4, \varepsilon/(8C)\}$ and run Theorem 3.5 at lower endpoint $q$ with tolerance $\varepsilon/2$; this yields a cap with $0 \le e(t) - S_\nu(t) \le \varepsilon/4$ for all $t \ge q$, where $e(t) = e^{\nu}_{D_r}(t)$ has $J$ channel coordinates. Set $\delta = \min\{q, \varepsilon/(4\max\{1, J\})\}$. By the uniform coupling, $|e(s) - e(t)| \le J|s-t|$: the finite escape indicator can change only if some coordinate lies between $s$ and $t$. If $s, t \ge q$ and $|s-t| \le \delta$ then $|S_\nu(s) - S_\nu(t)| \le \varepsilon/4 + \varepsilon/4 + \varepsilon/4$. If one of them is below $q$, both are at most $2q$, and monotonicity with nonnegativity give $|S_\nu(s) - S_\nu(t)| \le S_\nu(2q) \le 2Cq \le \varepsilon/4$. □

Conventional proof; not formalised.

The companions prove continuity of $S_\nu$ on $(0,1]$ and at 0; the point here is that the modulus is computable, without any lower bound on the openness actually of interest.

**Proposition 6.3** (Inverse brackets). *Given rational $0 < t < 1$ and rational $\varepsilon > 0$, there is a terminating algorithm returning rationals $u < v$ with $v - u < \varepsilon$ and $u < S_\nu^{-1}(t) < v$. The endpoints $t = 0$ and $t = 1$ are exact. The inverse critical intensity $\lambda(t) = -\log(1 - S_\nu^{-1}(t))$ is enclosed to any rational tolerance by the rational logarithm bounds of Remark 6.5.*

*Proof.* By Proposition 6.1 and (I1), $S_\nu$ is a continuous bijection of $[0,1]$, so $x = S_\nu^{-1}(t) \in (0,1)$ exists. Dovetail over all rational pairs $0 \le u < v \le 1$ with $v-u < \varepsilon$ and all evaluator tolerances, and stop when certified enclosures show $\overline{S_\nu(u)} < t < \underline{S_\nu(v)}$. Any pair with $u < x < v$ has $S_\nu(u) < t < S_\nu(v)$ by strictness; both gaps are positive, so sufficiently accurate enclosures certify them and the search terminates. The stopping test uses strict rational inequalities, so the pathological equality case of midpoint bisection never arises. □

Conventional proof; not formalised.

**Proposition 6.4** (Non-analyticity at zero)**.** *Neither $S_\nu$ nor $\Theta_\nu$ agrees with a convergent power series on any right neighbourhood of* 0*.*

*Proof.* If $S_\nu(a) = \sum_k c_k a^k$ near 0 with some $c_k \neq 0$, let $d$ be the least such index; then $S_\nu(a)/a^d \to c_d \neq 0$, contradicting Corollary 4.5. If every $c_k$ vanishes, $S_\nu$ vanishes near 0, contradicting positivity. The same argument applies to $\Theta_\nu(\lambda) = S_\nu(a_\lambda)$. □

Conventional proof; not formalised.

Flatness does not by itself establish that derivatives of every order exist; no $C^\infty$ claim is made.

*Remark* 6.5 (Rational parameter conversion and normalisation drift). The three parameters of (4) are related by exponentials and logarithms, which are not rational. For rational $z \ge 0$ the truncated exponential series with a tail bound $2z^{N+1}/(N+1)!$ once $N+2 \ge 2z$ gives rational enclosures of $1 - \mathrm{e}^{-z}$ of any width; for rational $0 \le a < 1$ the series $\sum_{k \le N} a^k/k$ with remainder at most $a^{N+1}/((N+1)(1-a))$ encloses $-\log(1-a)$. Both searches terminate by geometric decay and both are implemented in the replay script accompanying the paper. Under the canonical normalisation, $0 \le nM_n/R_n^{\mathrm{sp}} - 1 \le 2/(n-2)$ and $D_n/R_n^{\mathrm{sp}} \le 2^{-\lfloor n/2 \rfloor}$, so $|q_{M_n}(p_n) - a_\lambda|$ is bounded by explicit envelopes decreasing to zero, uniformly for $\lambda$ in a bounded interval; these envelopes are what a user of Theorem 5.3 needs to relate a finite $p$ to a nominal intensity. Details are routine and are omitted here.

# 7 Discussion

## Three observables

The paper separates three quantities that are easily conflated: the probability that a finite random network of cap $n$ contains some RAF; the probability that a static closure escapes an observed length range; and the limiting survival profile. They agree in the critical-window limit and disagree at small intensity by an unbounded number of orders of magnitude (Section 5.3). Any report of finite-size RAF probabilities near the lower edge of the window should therefore state the singleton contribution separately, since it dominates there and has no counterpart in the limit. Reaction quotienting changes the random experiment, not merely a normalisation: the two models have different gateway counts, different singleton coefficients and different profiles [1].

## Scope

The results concern structural autocatalysis under independent homogeneous molecule–channel catalysis with one mark per channel. Proposition 4.6 shows that independence cannot be weakened to a marginal condition; the heterogeneous independent extension of Theorem 4.3 is the strongest dependence relaxation proved here. The probabilities are not concentration thresholds, productive fluxes, persistence probabilities or predictions about viable protocells; finite inventories, inhibition [14], energetic feasibility and degradation would change the model. The examples are exact random-network calculations, not calibrated biochemical measurements.

### Open problems

1. *Matching lower bounds.* Corollary 4.5 bounds $\log S_\nu(a)$ above by $-\log^2(1/a)/(2\log 2) + O(\log(1/a))$. A matching lower bound, or a proof that the true rate is different, would determine the low-intensity asymptotics of the critical window.
2. *Cheaper seeds.* The seed lengths of Remark 3.6 come from the inherited contour constant $C_* = 2 \cdot 7^{64} \cdot 81$. Any improvement of the supplied-seed estimate, or a repair procedure acquiring a smaller effective seed, propagates directly into Theorems 3.1 and 3.5.
3. *Interior evaluation.* Theorem 3.5 settles computability but not feasibility. An algorithm evaluating $S_\nu(a)$ to a few digits at moderate $a$ would require a genuinely different idea, such as a sufficient statistic for closure states that supports a scalar recursion.
4. *Dependence.* A theorem covering correlated catalytic mechanisms with controlled dependence would extend the shape results to more realistic catalysis models [12].

None of these is needed for the results proved above.

### Data and code availability

The replay script `check_paper.py` accompanying the source recomputes every printed constant in exact rational arithmetic, from the channel census to the certificates of Table 2 and the examples of Section 5.3, and renders Figure 2. The Lean sources and verification receipts are described in Appendix A.

## A Formal verification and reproducibility

The formal development is written in Lean 4 [4] over mathlib [23]. Its namespace is `RAFCriticalWindowQuantitative`. It imports the split source development of [3], namespace `HordijkSteelThreshold`, and the quotient development of [1], namespace `RAFReactionQuotient`. Every module was compiled with Lean 4.30.0 with warnings promoted to errors, without `sorry` and without `native_decide`; the axiom report of every requested declaration is {`propext`, `Classical.choice`, `Quot.sound`}. No probability oracle, convergence rate or validity assumption is a premise.

The compiled root is

`RAFCriticalWindowQuantitative.quantitative_critical_window_resolution`

which packages, for both models: the enclosure property of the total rational evaluator; the uniform escape error above a rational lower endpoint; the all-powers bound with the sparse coefficient; the intensity flatness; the source error contract with actual catalytic inputs; the extraction of productive histories from long words; and the symbolic worked bound $S_{\mathrm{sp}}(a) < 8 \cdot 10^{-26}$ for $a \le 10^{-20}$. Its source hash is `33564f5d...0a6b2` (SHA-256 of `Resolution.lean`) and the frozen mathlib manifest hash is `a8df0b60...9fac`. The publication endpoints `publication_history_resolution` and `publication_small_openness_arithmetic` import this root unchanged and add the ordered-history bridges and the four certificates of Table 2; the largest, $C_{65}^{\mathrm{sp}} < 10^{671}$, is an exact integer comparison that required raising Lean's exponentiation threshold and no other change.

Remaining formalisation obligations, not claimed here, are the total-length census of Lemma 4.2 and the full polymer instance of Theorem 4.3, the singleton semantics of Theorem 5.1 with their composition in Theorem 5.3, and Section 6. These are *formalisation* tasks, not open mathematics. Table 4 lists the evidence result by result.

Table 4: Concordance between the numbered results and their evidence. "Lean" means the statement, at the stated scope, is a compiled declaration; "conventional" means a proof in this paper.

| Result | Status | Declarations and scope |
|---|---|---|
| Lemma 2.1 | conventional | Formula checked against enumeration for $n \leq 9$ by the replay script. |
| Seed selector | Lean | `seedIndex_spec`, `rational_deficit_bound`, `selected_contour_deficit`. |
| Theorem 3.1 | Lean | `explicit_split_truncation`, `explicit_quotient_truncation`, `split_uniform_approximation`, `quotient_uniform_approximation`; escape indexed by $K = B - 2$. |
| Lemma 3.2 | Lean | `escape_has_bounded_record`, `finite_record_support`, `quotient_finite_record_support`. |
| Lemma 3.3 | Lean | `escaped_missing_contraction`, `quotient_escaped_missing_contraction`: finite prefix atoms, no conditioning on survival. |
| Eqs. (10)–(11) | Lean | `finite_seed_failure`, `effective_seed_probability_le_survival`, `effective_quotient_seed_survival`. |
| Theorem 3.5 | Lean | `splitCertifiedInterval_correct`, `quotientCertifiedInterval_correct`, `splitTotalInterval_correct`, `quotientTotalInterval_correct`; the clipping to $[0, 1]$ is a remark. |
| Lemma 4.1 | Lean | `productive_history_settled`, `settled_insert_closure`, `productive_history_ordered`, `split_unbounded_histories`, `quotient_unbounded_histories`. |
| Lemma 4.2 | conventional | Exhaustive census to depth two in the replay script; `historyBudget_envelope` checks $b_j \leq 36 \cdot 2^j$. |
| Theorem 4.3 | mixed | Generic step `history_support_union_bound` takes the family count and coverage as premises; the polymer census is conventional. The sparse-coefficient version is fully Lean: `split_sparse_real`, `quotient_sparse_real`. |
| Corollary 4.5 | mixed | Flatness: `split_low_openness_flatness`, `split_low_intensity_flatness` and quotient counterparts; logarithmic optimisation conventional. |
| Table 2 | Lean (arithmetic) | `history_certificate_003`, `_004`, `_006`, `_020`; probabilistic meaning via Theorem 4.3. |
| Theorem 5.1, Cor. 5.2 | conventional | Census confirmed at $n = 4$ by enumeration; conditional flatness Lean: `conditional_intensity_flatness`. |
| Eqs. (21)–(22) | Lean | `splitSourceInterval_correct`, `quotientSourceInterval_correct`, `interval_comparison_error`, `canonical_raf_le_finite_escape_36`. |
| Theorem 5.3 | conventional | Composition of Theorems 4.3 and 5.1 with (7). |
| Eq. (24) | Lean (arithmetic) | `split_finite_example_arithmetic`, `quotient_finite_example_arithmetic`, `split_history_coefficient_two`, `quotient_history_coefficient_two`. |
| Section 6 | conventional | Propositions 6.1–6.4 and Remark 6.5; rational conversion implemented in the replay script. |

## References


[1] [Author name to be inserted]. Critical-window emergence of autocatalytic sets under reaction-channel quotienting. Companion manuscript, 2026. Commuting-factor quotient of reaction channels; proves the quotient critical-window law $S_{\mathrm{qt}}(1 - e^{-\lambda})$, the half-openness domination of the split supplied-seed estimate, and the bounded-catalyst gateway comparison, 2026.

[2] [Author name to be inserted]. Exact overlap partitions and gateway scaling in Kauffman RAF networks. Companion manuscript, 2026. Defines the commuting-factor quotient of reaction channels and proves the exact 34-channel food gateway count, 2026.

[3] [Author name to be inserted]. A nontrivial critical window for RAF emergence in the binary polymer model. Companion manuscript, 2026. Split-position reaction catalogue; proves that the RAF probability converges to $S_{\mathrm{sp}}(1 - e^{-\lambda})$ when $f_n/n \to \lambda$, with $0 < S_{\mathrm{sp}}(1 - e^{-\lambda}) \leq 1 - e^{-36\lambda}$, and supplies the supplied-seed mass estimate and the finite catalytic comparison used here, 2026.

[4] Leonardo de Moura and Sebastian Ullrich. The Lean 4 theorem prover and programming language. In *Automated Deduction – CADE 28*, volume 12699 of *Lecture Notes in Computer Science*, pages 625–635. Springer, 2021.

[5] Manfred Eigen. Selforganization of matter and the evolution of biological macromolecules. *Naturwissenschaften*, 58(10):465–523, 1971.

[6] Geoffrey Grimmett. *Percolation*, volume 321 of *Grundlehren der mathematischen Wissenschaften*. Springer, Berlin, 2nd edition, 1999.

[7] Theodore E. Harris. A lower bound for the critical probability in a certain percolation process. *Mathematical Proceedings of the Cambridge Philosophical Society*, 56(1):13–20, 1960.

[8] Wim Hordijk, Jotun Hein, and Mike Steel. Autocatalytic sets and the origin of life. *Entropy*, 12(7):1733–1742, 2010.

[9] Wim Hordijk, Stuart A. Kauffman, and Mike Steel. Required levels of catalysis for emergence of autocatalytic sets in models of chemical reaction systems. *International Journal of Molecular Sciences*, 12(5):3085–3101, 2011.

[10] Wim Hordijk, Joshua I. Smith, and Mike Steel. Algorithms for detecting and analysing autocatalytic sets. *Algorithms for Molecular Biology*, 10:15, 2015.

[11] Wim Hordijk and Mike Steel. Detecting autocatalytic, self-sustaining sets in chemical reaction systems. *Journal of Theoretical Biology*, 227(4):451–461, 2004.

[12] Wim Hordijk and Mike Steel. Autocatalytic sets in polymer networks with variable catalysis distributions. *Journal of Mathematical Chemistry*, 54(10):1997–2021, 2016.

[13] Wim Hordijk and Mike Steel. Chasing the tail: The emergence of autocatalytic networks. *Biosystems*, 152:1–10, 2017.

[14] Wim Hordijk, Mike Steel, and Stuart Kauffman. The structure of autocatalytic sets: Evolvability, enablement, and emergence. *Acta Biotheoretica*, 60(4):379–392, 2012.

[15] Wim Hordijk, Mike Steel, and Stuart A. Kauffman. Molecular diversity required for the formation of autocatalytic sets. *Life*, 9(1):23, 2019.

[16] Stuart A. Kauffman. Autocatalytic sets of proteins. *Journal of Theoretical Biology*, 119(1):1–24, 1986.

[17] Stuart A. Kauffman. *The Origins of Order: Self-Organization and Selection in Evolution*. Oxford University Press, New York, 1993.

[18] M. Lothaire. *Combinatorics on Words*. Cambridge Mathematical Library. Cambridge University Press, Cambridge, 1997.

[19] Roger C. Lyndon and Marcel-Paul Schützenberger. The equation $a^m = b^n c^p$ in a free group. *Michigan Mathematical Journal*, 9(4):289–298, 1962.

[20] Elchanan Mossel and Mike Steel. Random biochemical networks: the probability of self-sustaining autocatalysis. *Journal of Theoretical Biology*, 233(3):327–336, 2005.

[21] Mike Steel. The emergence of a self-catalysing structure in abstract origin-of-life studies. *Applied Mathematics Letters*, 13(3):91–95, 2000.

[22] Mike Steel, Wim Hordijk, and Joana C. Xavier. Autocatalytic networks in biology: structural theory and algorithms. *Journal of the Royal Society Interface*, 16(151):20180808, 2019.

[23] The mathlib Community. The Lean mathematical library. In *Proceedings of the 9th ACM SIGPLAN International Conference on Certified Programs and Proofs (CPP 2020)*, pages 367–381. ACM, 2020.

[24] Vijay Varanasi. Emergence-of-Autocatalysis-in-Prebiotic-Reaction-Networks: public source and data repository. https://github.com/VJ-Varanasi/Emergence-of-Autocatalysis-in-Prebiotic-Reaction-Networks, 2026. Reaction generator and catalysis sampler, as retrieved 1 September 2026.

[25] Vijay Varanasi and Jun Korenaga. Emergence of autocatalysis in prebiotic reaction networks. *Physical Review E*, 113:044304, 2026.

[26] Klaus Weihrauch. *Computable Analysis: An Introduction*. Texts in Theoretical Computer Science. Springer, Berlin, 2000.