A NONTRIVIAL CRITICAL WINDOW FOR RAF EMERGENCE IN THE BINARY POLYMER MODEL

AUTHOR NAME TO BE INSERTED

Abstract. We determine the limiting probability that a reflexively autocatalytic and food-generated (RAF) reaction set exists in the uniform reversible binary-polymer model. Molecules are the nonempty binary words of length at most n, the food set consists of the monomers and dimers, reactions are indexed by product word and split position, and every molecule–reaction catalysis coordinate is present independently with probability p_n. Let $f_n = p_n|\mathcal{R}_n|$ be the mean number of reactions catalyzed by one molecule. If $f_n/n \to \lambda \in (0,\infty)$, then the RAF probability converges to

$$\Theta(\lambda) = S\big(1 - e^{-\lambda}\big),$$

where $S(a)$ is the probability that the reaction closure of the food set in an infinite i.i.d. reversible reaction field of openness a is unbounded. We prove that $0 < \Theta(\lambda) \le 1 - e^{-36\lambda} < 1$ for every $\lambda > 0$, that Θ is nondecreasing and continuous, and that $\Theta(\lambda) \to 0$ as $\lambda \downarrow 0$ and $\Theta(\lambda) \to 1$ as $\lambda \to \infty$. Consequently no finite positive constant separates limiting probability zero from limiting probability one inside the linear scaling window, contrary to a conjecture of Hordijk and Steel; the sublinear and superlinear regimes are recovered as corollaries.

The proof combines an exact distributional reduction of catalytic pruning histories, a Peierls-type contour estimate on binary-word geometry that amplifies a fixed finite seed to nearly full molecular mass uniformly in n, a record-word argument showing that an unbounded infinite closure almost surely contains every finite word at the same parameter, and a uniform finite-seed approximation of the survival probability which yields continuity of S. The complete argument, including the geometric appendix, is formalized and machine-checked in Lean 4 over Mathlib.

1. Introduction

Collectively autocatalytic sets were proposed by Kauffman [12, 13], following Eigen's hypercycle [2], as a route by which a random chemistry could bootstrap a self-sustaining metabolism without any single self-replicating molecule. Their modern combinatorial formalization is the notion of a *reflexively autocatalytic and food-generated* set, or RAF, introduced by Steel [16] and developed with Hordijk [7, 6, 9, 17]. A RAF is a nonempty family of reactions which can generate all of its required molecules from an ambient food set and can obtain a catalyst for every reaction from the resulting closure. The two requirements express a form of collective self-support; they impose no order in which catalysts must first be synthesized, and mutual catalytic dependence is permitted. Related but distinct notions of autocatalytic sets appear in the evolutionary graph models of Jain and Krishna [11]; see [4, 10] for surveys.

The binary polymer model [12, 16, 7] provides a finite setting in which molecular diversity, reaction geometry, and random catalysis can be varied independently. Molecules are binary strings of length at most n, reactions are ligations and their reverse cleavages, and each molecule catalyzes each reaction independently with a probability tuned so that a molecule catalyzes on average f_n reactions. Mossel and Steel [14] identified linear growth of f_n in n as the relevant scale for RAF occurrence, and Hordijk, Kauffman and Steel [5] confirmed that scale in simulations across several

Date: September 2026.

2020 *Mathematics Subject Classification.* Primary 60K35; Secondary 92E20, 05C80, 68V20.

Key words and phrases. Autocatalytic sets; RAF theory; binary polymer model; reversible reaction closure; critical window; product measure; Peierls contour; Lean 4; formal verification.

model variants. Simulations of the model exhibit a steep rise in the finite-n RAF probability within a narrow range of f_n/n. Motivated by this, Hordijk and Steel [8, Eq. (21)] proposed a sharper conjecture: there should exist a finite constant $\gamma > 0$ such that, when $f_n \sim \lambda n$, the RAF probability tends to 0 for $\lambda < \gamma$ and to 1 for $\lambda > \gamma$.

This paper gives an exact replacement law. In the fixed-food, uniform, reversible binary-polymer model with split-position reaction identities, the RAF probability converges at every positive intensity, and the limit is a continuous, nondecreasing function

$$\Theta(\lambda) = S(1 - e^{-\lambda})$$

that takes values strictly between zero and one at every finite $\lambda > 0$. The function S is the survival probability of a natural infinite reversible reaction-closure process. Thus the entire linear scale $f_n \asymp n$ is a nontrivial critical window, and the finite step conjecture fails on both of its branches. The result does not contradict the simulation observations that motivated the conjecture: a continuous limiting law is compatible with a steep onset at finite n. What the theorem rules out is the asymptotic step-law interpretation of those observations.

Why the limit is nontrivial. Two elementary facts already show that neither limiting value can be trivial. First, a RAF must contain a reaction that can fire directly from food. With food fixed as the monomers and dimers, there are only 36 such gateway reaction identities, so the mean number of catalytic opportunities at the entrance stays bounded by 36λ as $n \to \infty$, and the probability that all of them are absent stays bounded away from zero (Proposition 4.2). Second, once the closure has escaped the food boundary, the exponentially growing bulk of the polymer universe offers so many alternative reaction paths that an infinite closure survives with positive probability at every positive effective openness (Section 7). Neither fact by itself identifies the limit. The difficulty is that a RAF must retain its own catalysts: catalysts available in the ambient universe may disappear during the pruning that defines the maximal RAF. The proof bridges this gap through complete decreasing-pool pruning histories, whose joint laws admit an exact cardinality replacement (Lemma 5.1), and through a finite-seed approximation of infinite survival that is uniform in the parameter (Theorem 9.1).

Organization. Section 2 fixes the model and states the main theorem. Section 3 introduces the infinite reversible closure process and the survival function S. Section 4 proves the gateway ceiling. Section 5 proves the exact law of catalytic pruning histories, and Section 6 converts it into a static mass barrier that produces a literal RAF. Section 7 proves that a fixed rich seed amplifies to nearly full molecular mass uniformly in n, using a contour estimate whose geometric ingredients are proved in Appendix A. Section 8 proves that an unbounded infinite closure almost surely contains every word, and Section 9 derives the uniform finite-seed approximation and continuity of S. Section 10 assembles the proof of the main theorem. Section 11 discusses consequences, scope, and open questions. Appendix B describes the Lean 4 formalization and gives a concordance between the numbered results and the formal declarations.

Scope. All probability statements use independent molecule–reaction catalysis, with a single catalysis coordinate shared between the two directions of each split-position reaction. The results concern structural RAF existence; they make no claim about reaction rates, concentrations, inhibition, or kinetic persistence.

2. Model and main theorem

2.1. **Molecules, reactions and food.** For $n \geq 1$ let

$$X_n = \bigcup_{j=1}^{n} \{0,1\}^j$$

be the set of nonempty binary words of length at most n, and let $F = X_2$ be the food set, consisting of the two monomers and four dimers. A *reaction* is a pair (w, i) with $w \in X_n$ and $1 \le i < |w|$. Writing $w = uv$ with $|u| = i$, the reaction represents the reversible transformation $u + v \leftrightarrow w$: ligation of u and v to w, and cleavage of w into u and v. Split positions are distinct reaction identities, even when the unordered reactant multisets coincide after forgetting the position. Let $\mathcal{R}_n$ denote the set of reactions. For $n \ge 2$,

$$U_n := |X_n| = 2^{n+1} - 2, \qquad R_n := |\mathcal{R}_n| = \sum_{\ell=2}^{n} (\ell - 1)2^{\ell} = (n-2)2^{n+1} + 4. \tag{2.1}$$

For $n = 5$ these give $U_5 = 62$ and $R_5 = 196$, matching the model sizes reported in [8].

2.2. **Random catalysis.** For every pair $(x, r) \in X_n \times \mathcal{R}_n$, independently set $C(x, r) = 1$ with probability p_n and $C(x, r) = 0$ otherwise. Both directions of a reaction use the same coordinate. The expected number of reactions catalyzed by a molecule is $f_n = p_n R_n$. The food set is deterministic and remains fixed as n increases.

To parametrize by intensity, for real v let

$$p_n(v) = \mathrm{clamp}\Big(\frac{vn}{R_n}\Big), \qquad \mathrm{clamp}(x) = \min(1, \max(0, x)), \tag{2.2}$$

and write $\mathbb{P}_{n,v}$ for the product law with $p_n = p_n(v)$. Thus $v = f_n/n$ whenever $0 \le vn/R_n \le 1$; the clamp only serves to define a measure for every n and every real v, and Lemma 10.4 shows that it is inactive throughout the asymptotic regime of interest.

2.3. **Closure and RAFs.** For $A \subseteq \mathcal{R}_n$ let $\mathcal{C}_n(A; F)$ be the smallest subset of X_n containing F and closed under the following rule: if $(w, i) \in A$ with $w = uv$, $|u| = i$, then $u, v \in \mathcal{C}_n(A; F)$ implies $w \in \mathcal{C}_n(A; F)$, and $w \in \mathcal{C}_n(A; F)$ implies $u, v \in \mathcal{C}_n(A; F)$. Equivalently, $\mathcal{C}_n(A; F)$ is the union of the finite stages obtained by repeatedly adding the outputs of any orientation of any reaction in A whose inputs are already present.

Definition 2.1. A nonempty $A \subseteq \mathcal{R}_n$ is a *RAF* if

(i) (food generation) every reaction in A has at least one complete side contained in $\mathcal{C}_n(A; F)$; and
(ii) (reflexive autocatalysis) every reaction $r \in A$ has a catalyst $x \in \mathcal{C}_n(A; F)$ with $C(x, r) = 1$.

Since a reaction with one complete side in the closure has both sides in the closure, condition (i) may equivalently be read as: both sides of every reaction of A belong to $\mathcal{C}_n(A; F)$.

Let $P_n(v) = \mathbb{P}_{n,v}\{\text{some RAF exists}\}$, and for a sequence (f_n) write $P_n = P_n(f_n/n)$. The event that a RAF exists is a finite Boolean combination of catalysis coordinates and is therefore measurable.

2.4. **The survival function.** Let $X_\infty = \bigcup_{j \ge 1} \{0, 1\}^j$ and let each valid split-position reaction on X_∞ carry an independent Bernoulli mark of parameter $a \in [0, 1]$, with product law $\mathbb{P}_a$. Let $\mathcal{C}_2(\omega)$ be the closure of the food set $F = X_2$ under the open reactions in the mark field ω, defined as in Section 3. The *survival function* is

$$S(a) = \mathbb{P}_a\{\mathcal{C}_2(\omega) \text{ is unbounded in word length}\}. \tag{2.3}$$

Section 3 shows that this event is measurable and that S is nondecreasing.

2.5. **Main results.**

Theorem 2.2 (Critical-window law)**.** *Let (f_n) be a real sequence with $f_n/n \to \lambda \in (0,\infty)$, and set $a_\lambda = 1 - e^{-\lambda}$. Then*

$$P_n \longrightarrow \Theta(\lambda) := S(a_\lambda), \qquad \text{and} \qquad 0 < \Theta(\lambda) \le 1 - e^{-36\lambda} < 1. \tag{2.4}$$

The function Θ is nondecreasing and continuous on $(0,\infty)$, with $\Theta(\lambda) \to 0$ as $\lambda \downarrow 0$ and $\Theta(\lambda) \to 1$ as $\lambda \to \infty$. Moreover, for all sufficiently large n the catalysis parameter is exactly $p_n = f_n/R_n$.

Corollary 2.3 (No finite step threshold)**.** *There is no $\gamma > 0$ such that $P_n(\lambda) \to 0$ for all $\lambda \in (0,\gamma)$ and $P_n(\lambda) \to 1$ for all $\lambda > \gamma$. In fact, for every $\lambda > 0$ the limit $\lim_n P_n(\lambda)$ exists and lies strictly between 0 and 1.*

Corollary 2.4 (Outer scaling regimes)**.** *Let (v_n) be a real sequence. If $v_n \to 0$ then $P_n(v_n) \to 0$. If $v_n \to \infty$ then $P_n(v_n) \to 1$.*

Corollary 2.3 refutes the conjecture of [8, Eq. (21)] for the fixed-food uniform model on both branches: at $\lambda = \gamma/2$ the limit is positive, and at any finite $\lambda > \gamma$ the limit is below one. Together with Corollary 2.4, the three regimes identify the whole linear scale as the nontrivial window, rather than dividing its interior by a second positive constant.

Remark 2.5. The theorem covers arbitrary sequences (f_n) with $f_n/n \to \lambda$, not only $f_n = \lambda n$. It is an exact limit theorem; no finite-size convergence rate is established here, and the proof does not rely on any numerical approximation to Θ.

3. Infinite reversible closure

Give each valid split-position reaction (w,i) on X_∞ an independent Bernoulli mark of parameter $a \in [0,1]$; the resulting field is ω and its product law is $\mathbb{P}_a$. A reaction is *open* when its mark equals 1. For a food cutoff $L \ge 1$ put $F_L = X_L$ and let $\mathcal{C}_L(\omega) \subseteq X_\infty$ be the least set containing F_L and closed under open ligations and open cleavages. Every membership assertion $w \in \mathcal{C}_L(\omega)$ has a finite derivation: a finite sequence of open reaction applications starting from F_L. The *cap* of a derivation is the maximal length of a word occurring in it. For $n \ge L$, let $\mathcal{C}_{n,L}(\omega)$ denote the closure of F_L in which all intermediate molecules are restricted to X_n; it depends only on the marks of reactions with product in X_n.

An unbounded closure is infinite, and a closure contained in some X_K is finite. The reaction index set is countable, and membership of a fixed word in $\mathcal{C}_L(\omega)$ is a countable union of finite-coordinate events; hence the survival event in (2.3) is measurable.

Lemma 3.1 (Finite witnesses and first escape)**.**

(a) Every finite set of words in $\mathcal{C}_L(\omega)$ has a common finite derivation cap.
(b) If $\mathcal{C}_L(\omega)$ contains a word of length greater than $K \ge L$, then some word of length greater than K is generated by a derivation of cap at most $2K$.

Proof. (a) Take the maximum of the finitely many intermediate word lengths in finitely many derivations.

(b) Induct on a derivation of a word longer than K, proving the alternative: either the derivation contains an escape witness of cap at most $2K$, or all of its words have length at most K. Food words have length at most $L \le K$. A cleavage produces words shorter than its input, so it cannot be the first step to cross length K. For a ligation, either one input derivation already supplies a witness, or both input derivations stay within length K; in the latter case the product has length at most $2K$, and if it is longer than K it is itself a witness generated within cap $2K$. This also covers the case in which a short target word is obtained by cleavage from a longer intermediate. □

For $K \geq 2$ let E_K be the event that, starting from F_2, some word of length greater than $K+2$ is generated by a derivation of cap at most $2(K+2)$. By Lemma 3.1(b), E_K coincides with the event that $\mathcal{C}_2(\omega)$ contains a word longer than $K+2$. The events E_K decrease in K, and their intersection is the event that $\mathcal{C}_2(\omega)$ is unbounded. Hence

$$\mathbb{P}_a(E_K) \downarrow S(a) \qquad (K \to \infty). \tag{3.1}$$

Each E_K is determined by the finitely many marks of reactions with product length at most $2(K+2)$, so $\mathbb{P}_a(E_K)$ is a polynomial in a, in particular continuous. Coupling all parameters by common uniform marks shows that $\mathcal{C}_L(\omega)$, E_K and the survival event are increasing in a; thus S is nondecreasing. These finite approximations will always be used with their caps fixed before any parameter limit is taken.

4. The finite food boundary

Call $r \in \mathcal{R}_n$ a *gateway reaction* if at least one of its sides is contained in F. Because a RAF must eventually leave the food set, it must use a gateway.

Lemma 4.1. *Every nonempty food-generated reaction set, and in particular every RAF, contains a gateway reaction.*

Proof. Suppose A is nonempty and contains no gateway reaction. Then no reaction in A has a complete side in F, so the first closure stage adds nothing to F, and by induction $\mathcal{C}_n(A; F) = F$. But food generation requires each reaction of A to have a complete side in $\mathcal{C}_n(A; F) = F$, which would make it a gateway reaction, a contradiction. □

For $n \geq 4$, the gateway reactions with a food side consisting of two food words are indexed exactly by the ordered pairs $(u, v) \in F \times F$: the reaction is $(uv, |u|)$. There are $6^2 = 36$ such identities. A reaction whose product is itself food has both of its factors in F as well, so it already appears in this list and is not counted again.

Proposition 4.2 (Gateway ceiling)**.** *If $f_n/n \to \lambda \in (0, \infty)$, then*

$$\limsup_{n\to\infty} P_n \leq 1 - e^{-36\lambda}. \tag{4.1}$$

Proof. By Lemma 4.1, if every catalysis coordinate (x, r) with r among the 36 gateway identities is absent, no RAF exists. There are $36U_n$ such coordinates, and independence gives

$$P_n \leq 1 - (1 - p_n)^{36U_n}. \tag{4.2}$$

By (2.1), $nU_n/R_n \to 1$, so $p_nU_n = (f_n/n) \cdot nU_n/R_n \to \lambda$ and $p_n \to 0$. Since $\log(1-p_n)/p_n \to -1$, the right side of (4.2) converges to $1 - e^{-36\lambda}$. □

Remark 4.3 (Why the boundary does not disappear). Although the number of molecule types grows exponentially, the number of gateway identities stays fixed, so the total mean number of catalytic opportunities at the entrance tends to 36λ rather than to infinity. This leaves positive extinction probability at every finite intensity, and refutes the upper branch of a step law without any assumption about the subsequent bulk. Four of the 36 identities concatenate two monomers into a dimer that is already food; in the static reaction model only 32 of them enlarge the closure of F. The count 36 is the correct catalytic entrance count for RAF existence, but (4.1) is only a ceiling: subsequent finite traps also contribute to extinction, and the proposition does not identify the leading asymptotics of $1 - \Theta(\lambda)$. An earlier version of this obstruction, with the cruder bound $|R_4| = 68$ on the number of gateway identities, was the first machine-checked refutation of the upper branch and is retained in the formal development (Appendix B).

The main difficulty is the opposite inequality. Catalysts available in the ambient universe may disappear during pruning, so the closure under ambient-catalyzed reactions is only an upper bound for the RAF closure. The next two sections treat this dependence exactly.

5. Exact laws of catalytic pruning histories

Fix a finite molecule universe X of cardinality U, a finite reaction universe $\mathcal{R}$, and a deterministic monotone operator $\Phi\colon 2^{\mathcal{R}} \to 2^X$ (that is, $A \subseteq A'$ implies $\Phi(A) \subseteq \Phi(A')$). Let $C(x,r)$, $(x,r) \in X \times \mathcal{R}$, be independent Bernoulli(p) variables, and write $q = 1-p$. The *pruning process* is

$$C_0 = X, \qquad A_t = \{r \in \mathcal{R} : \exists x \in C_t,\ C(x,r) = 1\}, \qquad C_{t+1} = \Phi(A_t). \tag{5.1}$$

Since $C_1 \subseteq X = C_0$ and Φ is monotone, induction shows that both (C_t) and (A_t) are decreasing. In the application $\Phi(A) = \mathcal{C}_n(A;F)$, and C_t is the pool of potential catalysts after t rounds of pruning.

Fix an enumeration $X = \{x_1, \dots, x_U\}$ and let $I_k = \{x_1, \dots, x_k\}$. The *canonical process* replaces each pool by an initial segment of the same cardinality:

$$\widehat{C}_0 = X, \qquad \widehat{A}_t = \{r : \exists x \in \widehat{C}_t,\ C(x,r) = 1\}, \qquad \widehat{C}_{t+1} = I_{|\Phi(\widehat{A}_t)|}. \tag{5.2}$$

By the same induction the canonical pools are nested initial segments and the canonical active sets decrease.

Lemma 5.1 (Cardinality replacement)**.** *For every $T \geq 0$, the random sequences $(A_0, \dots, A_T)$ and $(\widehat{A}_0, \dots, \widehat{A}_T)$ have the same law. Consequently the sequences of pool cardinalities $(|C_t|)_{t \leq T}$ and $(|\widehat{C}_t|)_{t \leq T}$ have the same law.*

Proof. Fix a candidate history $A_0^*, \dots, A_T^* \subseteq \mathcal{R}$. The proposed pools $C_0^* = X$, $C_{t+1}^* = \Phi(A_t^*)$ are deterministic. The process follows this history exactly when, for each $t \leq T$, the response at the proposed pool C_t^* equals A_t^*:

$$\{A_t = A_t^*\ \forall t \leq T\} = \{\{r : \exists x \in C_t^*,\ C(x,r) = 1\} = A_t^*\ \forall t \leq T\}.$$

This equivalence is proved successively in t; there is no further random consistency condition. If the proposed pools are not nested the history has probability zero in both processes, so assume $C_0^* \supseteq C_1^* \supseteq \cdots \supseteq C_T^*$.

The right-hand event is an intersection over reaction columns r of events that depend only on the column $(C(x,r))_{x \in X}$, and different columns are independent. Because the pools are nested, the indicator $\mathbf{1}\{\exists x \in C_t^* : C(x,r) = 1\}$ is nonincreasing in t, so a column event has positive probability only if r belongs to A_t^* for $t < d$ and not for $t \geq d$, for some $d \in \{0, \dots, T+1\}$. Its probability is

$$q^{|C_0^*|}\ (d = 0), \qquad q^{|C_d^*|} - q^{|C_{d-1}^*|}\ (1 \leq d \leq T), \qquad 1 - q^{|C_T^*|}\ (d = T+1). \tag{5.3}$$

The middle case says that there is no catalyst in C_d^* but there is one in $C_{d-1}^* \supseteq C_d^*$, which justifies the subtraction. The probability of the history is the product over columns of (5.3), and it depends on the proposed pools only through their cardinalities. The canonical process has proposed pools $I_{|C_t^*|}$, nested initial segments with the same cardinalities, so it assigns the same probability to the same history. Equality on every atom of the history space proves equality of the laws, and the cardinality statement follows because $|C_{t+1}| = |\Phi(A_t)|$ and $|\widehat{C}_{t+1}| = |\Phi(\widehat{A}_t)|$ are the same function of the histories. □

Remark 5.2. Lemma 5.1 is a distributional statement about complete dynamics. It does not condition on a selected terminal pool and then declare its coordinates independent, nor does it identify the original catalysis matrix pathwise with a reaction-percolation field. This distinction is what permits a static comparison in the next section without losing the endogenous catalytic constraint.

6. A static mass barrier produces a RAF

Apply the previous section with $X = X_n$, $\mathcal{R} = \mathcal{R}_n$ and $\Phi(A) = \mathcal{C}_n(A; F)$. In the canonical representation let O_k be the set of reactions catalyzed by at least one molecule of I_k. The indicators $\mathbf{1}\{r \in O_k\}$ are independent across reactions with parameter

$$a_{p,k} = 1 - (1-p)^k, \tag{6.1}$$

so $\Phi(O_k) = \mathcal{C}_n(O_k; F)$ has the law of the restricted closure $\mathcal{C}_{n,2}$ of Section 3 under $\mathbb{P}_{a_{p,k}}$.

Lemma 6.1 (Static barrier)**.** *If $k > |F| = 6$ and $p = p_n$, then*

$$\mathbb{P}_{a_{p,k}}\{|\mathcal{C}_{n,2}| \geq k\} \leq P_n. \tag{6.2}$$

Proof. First, in the canonical process, $|\Phi(O_k)| \geq k$ implies $|\widehat{C}_t| \geq k$ for every t. Indeed $\widehat{C}_0 = X_n \supseteq I_k$; and if $\widehat{C}_t \supseteq I_k$ then $\widehat{A}_t \supseteq O_k$, so $|\widehat{C}_{t+1}| = |\Phi(\widehat{A}_t)| \geq |\Phi(O_k)| \geq k$, whence $\widehat{C}_{t+1} \supseteq I_k$ because it is an initial segment. Let $T = |\mathcal{R}_n|$. A decreasing sequence of subsets of $\mathcal{R}_n$ can make at most T strict deletions, so A_t and C_t are constant for $t \geq T$. By Lemma 5.1, the probability that $|C_T| \geq k$ in the original process equals the probability that $|\widehat{C}_T| \geq k$, which is at least the left side of (6.2).

It remains to show that $|C_T| \geq k > 6$ forces the existence of a RAF. Let $C = C_T$ be the stabilized pool and $A = A_T$ its active reactions, so that $\Phi(A) = C$ and every reaction of A has a catalyst in C. Remove from A every reaction with neither side contained in C, obtaining $A' \subseteq A$. Deletion cannot enlarge the closure, so $\Phi(A') \subseteq C$; conversely each reaction used in a derivation of a molecule of C has an input side contained in C and is therefore retained, so $\Phi(A') = C$. Thus every reaction of A' has a complete side in $\mathcal{C}_n(A'; F) = C$ and a catalyst in C. Since $|C| > 6 = |F|$, some non-food molecule was generated, so A' is nonempty. Hence A' is a RAF. □

Asymptotic use of the barrier. For fixed $m \geq 2$ take $k_n = (m-1)\lfloor U_n/m \rfloor$. Then $k_n > 6$ eventually, and since $p_n U_n \to \lambda$,

$$a_{p_n,k_n} \longrightarrow 1 - e^{-\lambda(1-1/m)}. \tag{6.3}$$

By monotonicity in the openness, any fixed parameter b strictly below the limit in (6.3) can be used in (6.2) for large n. It remains to show that closure from F_2 attains nearly full static mass with a probability approaching the infinite survival probability. This is where the finite-seed amplification theorem enters.

7. Amplification from a fixed rich seed

Theorem 7.1 (Uniform supplied-seed amplification)**.** *For every $a > 0$ and every integer $m \geq 2$ there is a cutoff $L = L(a, m)$ such that, for every $n \geq L$,*

$$\mathbb{P}_{a'}\{m|\mathcal{C}_{n,L}| < (m-1)U_n\} \leq \frac{1}{m} \qquad \text{for all } a' \geq a. \tag{7.1}$$

The cutoff L depends on a and m but not on n. The distinction between a fixed supplied seed F_L and the moving ambient cap n is essential: generation of every fixed word by itself does not imply (7.1).

7.1. The host graph and effective edges.

Fix L and $n \geq L$. Contract all words of length at most L to a single root vertex o; every other vertex is a word s with $L < |s| \leq n$. Let $\ell(s)$ and $r(s)$ denote deletion of the first and of the last bit followed by contraction. The *basic edges* are the indexed pairs $(s, 0)$ and $(s, 1)$ with endpoints $\{s, \ell(s)\}$ and $\{s, r(s)\}$ respectively.

Fix $w \leq L/2$. A basic edge from s to its one-bit truncation at a chosen end is made *effective* by any one of w *detours*: for $j \in \{1, \ldots, w\}$, delete j bits at that end to reach a common descendant s_j; the j-th detour is *open* when the reaction joining s to s_j and (for $j > 1$) the reaction joining the one-bit truncation to s_j are both open. The removed factors have length at most $w \leq L$ and

therefore belong to the supplied food F_L, so an open detour is a genuine reversible path in the chemistry: if either endpoint lies in $\mathcal{C}_{n,L}$ then so does the other. The support of a detour consists of at most two literal reaction identities (Lemma A.6). Effective edges therefore connect vertices lying in a common closure component, and every vertex in the effective component of the root lies in $\mathcal{C}_{n,L}$.

7.2. **Cuts, envelopes and the contour estimate.** A *cut* is the set of basic edges whose endpoints receive different values under a Boolean vertex labelling; a *bond* is an inclusion-minimal nonempty cut. Two basic edges are *diamond-adjacent* if they lie in a common diamond (A.1). Appendix A proves the following deterministic facts.

(G1) Every bond is connected in the diamond adjacency, whose degree is at most 9 (Lemma A.2).

(G2) For every actual molecule $s \in X_n$ and every $b \geq 1$, the bonds of size b having s on a *molecularly smaller side* (sides counted by actual molecules, each contracted food word counted separately) number at most A^b, where

$$A = 2 \cdot 7^{64} \cdot 81 \tag{7.2}$$

(Lemma A.4).

(G3) Each literal reaction belongs to the supports of at most three detours; consequently among the wb detours of a size-b bond one can select at least $wb/5$ with pairwise disjoint supports (Lemma A.6).

(G4) If two vertices lie in distinct effective components, some bond separates them and all of its edges are non-effective (Lemma A.5).

Treating the food root as a single vertex of weight one in (G2) would not give the needed estimate; the molecular count is what makes the envelope uniform over all actual molecules including the root's constituents.

Fix a bond of size b and take $w = 5k$. By (G3), at least kb of its detours have pairwise disjoint supports, each support is open with probability at least a^2, and disjoint supports are independent. Hence

$$\mathbb{P}_a\{\text{every detour of the bond fails}\} \leq (1 - a^2)^{kb}. \tag{7.3}$$

Independence is used only after selecting disjoint literal supports. Different bonds may overlap arbitrarily; their probabilities will be added, not multiplied.

Proof of Theorem 7.1. Let $\theta = A(1 - a^2)^k$ and choose k so large that $\theta < 1$; set $w = 5k$ and $L = 10k$, so that $w \leq L/2$. Call a molecule s *bad* if it lies on a molecularly smaller side of some bond all of whose detours fail. By (G2) and (7.3),

$$\mathbb{P}_a\{s \text{ is bad}\} \leq \sum_{b \geq 1} A^b (1 - a^2)^{kb} = \sum_{b \geq 1} \theta^b = \frac{\theta}{1 - \theta} =: \delta. \tag{7.4}$$

If two molecules are both not bad, they lie in the same effective component: otherwise by (G4) a bond with all detours failing separates them, and at least one of them lies on a molecularly smaller side of it and would be bad. Fix a monomer s_0 as root representative. If s_0 is not bad, every molecule that is not bad lies in the effective component of the root and hence in $\mathcal{C}_{n,L}$.

Let B be the number of bad molecules. Linearity of expectation gives $\mathbb{E}_a B \leq U_n \delta$ without any independence between bad events, and the root failure costs at most δ. Markov's inequality yields, for every integer $d > 0$,

$$\mathbb{P}_a\{|\mathcal{C}_{n,L}| + d \leq U_n\} \leq \delta + \frac{U_n \delta}{d}. \tag{7.5}$$

Given m, choose k so large that $\delta \leq 1/[m(m+1)]$ and put $d = \lfloor U_n/m \rfloor + 1$. The event in (7.1) implies a deficit $U_n - |\mathcal{C}_{n,L}| > U_n/m$, hence a deficit of at least d, and $U_n/d \leq m$. So (7.5) bounds

the probability in (7.1) by $(m+1)\delta \le 1/m$. Finally, the failure event is decreasing in the openness (common-mark coupling), so the same L works for every $a' \ge a$. □

7.3. **A positive probability of ignition.**

Proposition 7.2. *$S(a) > 0$ for every $a \in (0,1]$.*

Proof. Let $L = L(a,2)$ from Theorem 7.1. Force the finitely many reactions with product length at most L to be open; there are R_L of them, and successive monomer ligations then generate F_L from F_2. The supplied-seed closure $\mathcal{C}_{n,L}$ does not depend on these coordinates, since both sides of such a reaction already lie in F_L. Hence the event $\{2|\mathcal{C}_{n,L}| \ge U_n\}$ is independent of the forced coordinates, and

$$\mathbb{P}_a\{2|\mathcal{C}_{n,2}| \ge U_n\} \;\ge\; \mathbb{P}_a\{\text{all } R_L \text{ short reactions open}\} \cdot \mathbb{P}_a\{2|\mathcal{C}_{n,L}| \ge U_n\} \;\ge\; \frac{a^{R_L}}{2},$$

uniformly in $n \ge L$, because on the forced event $\mathcal{C}_{n,2} \supseteq \mathcal{C}_{n,L}$. Fix K. Since only finitely many words have length at most $K+2$ and $U_n \to \infty$, half mass forces a word longer than $K+2$ once n is large, and by Lemma 3.1(b) applied inside X_n the escape event E_K occurs. Hence $\mathbb{P}_a(E_K) \ge a^{R_L}/2$ for every K, and (3.1) gives $S(a) \ge a^{R_L}/2 > 0$. □

In particular the infinite reversible closure process has zero critical openness.

8. Record words and same-parameter richness

The lower bound in Section 10 needs more than survival: it needs the closure from F_2 to acquire a prescribed rich finite seed F_L. This section shows that survival already implies acquisition of every finite word, almost surely, at the same parameter.

Lemma 8.1 (Own-cap records)**.** *If $\mathcal{C}_2(\omega)$ is unbounded, then for every $B \ge 2$ there is a word $w \in \mathcal{C}_2(\omega)$ with $|w| > B$ that is generated by a derivation of cap $|w|$.*

Proof. Induct on a finite derivation to prove the alternative: either all words of the derivation have length at most B, or the derivation contains such a record. A food derivation fits within B. At a ligation, retain a record supplied by either input derivation if one exists; otherwise both input derivations fit within B, and then either the product has length at most B and the whole derivation fits, or the product is longer than B and is generated within its own length cap. A cleavage inherits the alternative from the derivation of its input. Apply this to a derivation of any word longer than B. □

Lemma 8.2 (Fresh repair support)**.** *Fix a nonempty target v of length ℓ and a word w. There is a deterministic set of at most $\ell+1$ valid reaction identities, all with product length greater than $|w|$, whose simultaneous opening implies $v \in \mathcal{C}_2(\omega)$ whenever $w \in \mathcal{C}_2(\omega)$. Its all-open probability under $\mathbb{P}_a$ is at least $a^{\ell+1}$.*

Proof. Write $v = v_1 \cdots v_\ell$. Successively append the monomer v_j to $wv_1 \cdots v_{j-1}$ using the split immediately before the appended letter; these are valid ligations with a food factor. Then cleave wv at the split $wv \to w + v$. Every product is strictly longer than w. There are at most $\ell+1$ reaction identities (coincident identities are counted once), and they are open simultaneously with probability at least $a^{\ell+1}$. □

The record condition is what makes these coordinates *fresh*. A word found by an escape search may be shorter than the cap already examined, and appending to it could reuse inspected coordinates. By contrast, the generation of an own-cap record is certified entirely by coordinates with product length at most $|w|$, and every repair product lies strictly above that boundary.

To exploit this probabilistically, fix $B \geq 2$ and let M be the least length of an own-cap record beyond B, that is, the least $M > B$ such that some word of length M is generated within cap M. Whether a given M' has this property depends only on the marks of reactions with product length at most M'. Partition the event $\{M < \infty\}$ according to the value of M and the complete assignment of marks to reactions with product length at most M. There are countably many such *prefix atoms*: finitely many assignments for each finite value of M. Every nonempty atom D fixes a record word w_D of length M_D, and D is determined by the coordinates with product length at most M_D.

Theorem 8.3 (Same-parameter richness)**.** *Let $a > 0$. Almost surely under $\mathbb{P}_a$, either $\mathcal{C}_2(\omega)$ is bounded (hence finite) or $\mathcal{C}_2(\omega) = X_\infty$. In particular, on the survival event every prescribed finite set of words is contained in $\mathcal{C}_2(\omega)$ almost surely, and*

$$S(a) = \mathbb{P}_a\{\mathcal{C}_2(\omega) = X_\infty\}. \tag{8.1}$$

Proof. Fix a nonempty target v and let A_v be the event that $\mathcal{C}_2(\omega)$ is unbounded but $v \notin \mathcal{C}_2(\omega)$. Let C be any event determined by the marks of finitely many reactions, and choose $B \geq 2$ so that all these reactions have product length at most B. Partition $A_v \cap C$ by the prefix atoms D of the first record beyond B; by Lemma 8.1 the atoms cover every unbounded configuration, and they are disjoint.

On an atom D with record w_D of length $M_D > B$, the repair support of Lemma 8.2 for the pair (v, w_D) consists of reactions with product length greater than M_D. It is therefore disjoint from the coordinates defining D (product length at most M_D) and from those defining C (product length at most $B < M_D$). On $A_v \cap C \cap D$ the word w_D is generated but v is not, so the repair must fail; independence of the product measure on the deterministic disjoint support gives

$$\mathbb{P}_a(A_v \cap C \cap D) \leq (1 - a^{|v|+1})\mathbb{P}_a(C \cap D). \tag{8.2}$$

Summing (8.2) over the disjoint atoms incurs no multiplicity and yields

$$\mathbb{P}_a(A_v \cap C) \leq q_v \mathbb{P}_a(C), \qquad q_v = 1 - a^{|v|+1} < 1. \tag{8.3}$$

Finite-coordinate cylinder events form an algebra generating the product σ-algebra, so every measurable event is approximated in symmetric-difference probability by cylinders. Given $\varepsilon > 0$ choose a cylinder C with $\mathbb{P}_a(C \triangle A_v) < \varepsilon$. Then (8.3) gives

$$\mathbb{P}_a(A_v) \leq \mathbb{P}_a(A_v \cap C) + \varepsilon \leq q_v \mathbb{P}_a(C) + \varepsilon \leq q_v \mathbb{P}_a(A_v) + (1 + q_v)\varepsilon.$$

Letting $\varepsilon \downarrow 0$ gives $(1 - q_v)\mathbb{P}_a(A_v) \leq 0$, so $\mathbb{P}_a(A_v) = 0$. There are countably many nonempty words, so almost surely on survival every word is generated. Conversely the complete closure is unbounded, and a bounded closure is contained in some finite X_K. This proves the dichotomy and (8.1). □

Remark 8.4. The proof never conditions on the entire survival event and asserts independence of a selected future. Independence is invoked only on deterministic finite-prefix atoms, and the cylinder approximation supplies the passage to the measurable event A_v.

9. Uniform finite-seed approximation and continuity

Let $G_L(a) = \mathbb{P}_a\{F_L \subseteq \mathcal{C}_2(\omega)\}$ be the probability that the closure from F_2 acquires the seed F_L, and let $G_{n,L}(a)$ be the probability of the same event with all derivations restricted to cap n. By Lemma 3.1(a), monotonicity in the cap, and Theorem 8.3,

$$G_{n,L}(a) \uparrow G_L(a) \quad (n \to \infty), \qquad S(a) \leq G_L(a) \qquad (a > 0). \tag{9.1}$$

Each $G_{n,L}$ is a polynomial in a, hence continuous.

Theorem 9.1 (Uniform approximation)**.** *For every $a_0 > 0$ and integer $m \geq 2$ there is a cutoff L such that, simultaneously for all $a \geq a_0$,*

$$S(a) \leq G_L(a) \leq S(a) + \frac{1}{m}. \tag{9.2}$$

Proof. Choose $L = L(a_0, m)$ from Theorem 7.1; the supplied-seed estimate (7.1) then holds for all $a \geq a_0$. Work in a single field ω. If F_2 generates F_L within cap n, then $\mathcal{C}_{n,2} \supseteq \mathcal{C}_{n,L}$, because the closure operator is monotone and idempotent. Hence

$$\{F_L \subseteq \mathcal{C}_{n,2}\} \subseteq \{m|\mathcal{C}_{n,2}| \geq (m-1)U_n\} \cup \{m|\mathcal{C}_{n,L}| < (m-1)U_n\},$$

and a union bound with (7.1) gives

$$G_{n,L}(a) \leq \mathbb{P}_a\{m|\mathcal{C}_{n,2}| \geq (m-1)U_n\} + \frac{1}{m}. \tag{9.3}$$

This is an event inclusion and a union bound, not a product of dependent probabilities. Fix an escape cutoff K. For n large, the mass event in (9.3) forces a word longer than $K+2$, because $U_n \to \infty$ while only finitely many words have length at most $K+2$; by Lemma 3.1(b) this implies E_K. Thus $G_{n,L}(a) \leq \mathbb{P}_a(E_K) + 1/m$ for large n. Let $n \to \infty$ using (9.1) to get $G_L(a) \leq \mathbb{P}_a(E_K) + 1/m$, and then $K \to \infty$ using (3.1). The lower inequality in (9.2) is (9.1). □

Theorem 9.1 describes an approximate finite ignition event. Reaching F_L is not a deterministic finite certificate of infinite survival; rather, the probability of reaching F_L without surviving is at most $1/m$, uniformly above any fixed positive openness. This is what allows finite information to approximate an infinite event.

Theorem 9.2 (Continuity of survival)**.** *S is nondecreasing on $[0,1]$ and continuous at every $a \in (0,1]$ (with the relative topology at 1). Moreover $S(1) = 1$.*

Proof. Monotonicity was noted in Section 3. Let $a_j \to a \in (0,1]$.

Upper semicontinuity. For each K, $S(a_j) \leq \mathbb{P}_{a_j}(E_K)$ and $\mathbb{P}_a(E_K)$ is continuous in a, so

$$\limsup_{j\to\infty} S(a_j) \leq \mathbb{P}_a(E_K). \tag{9.4}$$

Let $K \to \infty$ and use (3.1).

Lower semicontinuity. Fix $a_0 \in (0, a)$ and $m \geq 2$, and choose L from Theorem 9.1 at a_0. Eventually $a_j \geq a_0$, so $G_{n,L}(a_j) \leq G_L(a_j) \leq S(a_j) + 1/m$ for every n. At fixed n, continuity of $G_{n,L}$ gives

$$G_{n,L}(a) \leq \liminf_{j\to\infty} S(a_j) + \frac{1}{m}. \tag{9.5}$$

Let $n \to \infty$ and use $S(a) \leq G_L(a)$ from (9.1); then let $m \to \infty$.

Endpoint. At $a = 1$ all countably many reactions are open almost surely, and successive monomer ligations generate every word, so $S(1) = 1$. □

The order of choices in (9.5) matters. The seed cutoff is fixed using a positive lower openness; the witness cap n is fixed while the parameter varies; only afterwards is the cap enlarged. No uniform convergence of finite witness probabilities over all parameters is assumed. This is also why monotonicity alone, which permits jumps, would not suffice to identify the critical-window limit.

10. Proof of the main theorem

Throughout this section $\lambda > 0$ is fixed and $a_\lambda = 1 - e^{-\lambda}$. We first treat the exact sequence $f_n = \lambda n$, so that $p_n = p_n(\lambda)$, and then pass to arbitrary sequences.

10.1. **The lower transition bound.**

Proposition 10.1. *If $f_n = \lambda n$ then $\liminf_{n\to\infty} P_n \geq S(a_\lambda)$.*

Proof. Fix $b \in (0, a_\lambda)$ and take $m \geq 2$ so large that

$$b < 1 - e^{-\lambda(1-1/m)}. \tag{10.1}$$

Let $L = L(b, m)$ from Theorem 7.1. The event inclusion behind (9.3), now read as a lower bound on mass, gives

$$\mathbb{P}_b\{m|\mathcal{C}_{n,2}| \geq (m-1)U_n\} \geq G_{n,L}(b) - \frac{1}{m}. \tag{10.2}$$

Let $k_n = (m-1)\lfloor U_n/m \rfloor$. The mass event on the left implies $|\mathcal{C}_{n,2}| \geq k_n$. For large n, (6.3) and (10.1) give $a_{p_n,k_n} \geq b$ while $k_n > 6$. Monotonicity of the closure in the openness and the static barrier (6.2) therefore give

$$P_n \;\geq\; \mathbb{P}_{a_{p_n,k_n}}\{|\mathcal{C}_{n,2}| \geq k_n\} \;\geq\; \mathbb{P}_b\{|\mathcal{C}_{n,2}| \geq k_n\} \;\geq\; G_{n,L}(b) - \frac{1}{m}. \tag{10.3}$$

Let $n \to \infty$ with b, m, L fixed; by (9.1), $\liminf_n P_n \geq S(b) - 1/m$. For fixed $b < a_\lambda$, condition (10.1) holds for every sufficiently large m, so $\liminf_n P_n \geq S(b)$. Finally let $b \uparrow a_\lambda$ and use continuity of S (Theorem 9.2). □

The lower bound preserves the full survival probability. It does not pay a fixed all-open nucleus cost, which would only establish a possibly tiny positive lower bound, and it does not assume that every viable first food reaction ignites the bulk: all finite traps are retained in the survival event whose probability appears in the bound. A retained catalyst fraction $1 - 1/m$ produces effective openness approaching a_λ as m grows; supplied-seed amplification controls the mass deficit at this fraction, and continuity removes the fraction loss.

10.2. **The upper transition bound.**

Proposition 10.2. *If $f_n = \lambda n$ then $\limsup_{n\to\infty} P_n \leq S(a_\lambda)$.*

Proof. Let O be the set of reactions catalyzed by at least one molecule of the full universe X_n. Its indicators are i.i.d. with openness

$$a_n = 1 - (1 - p_n)^{U_n} \longrightarrow a_\lambda.$$

Every RAF A satisfies $A \subseteq O$, so its closure is contained in the ambient closure $\mathcal{C}_n(O; F)$, which has the law of $\mathcal{C}_{n,2}$ under $\mathbb{P}_{a_n}$. This inclusion is used only for an upper bound.

Fix K with $n \geq 2(K+2)$. Either the ambient closure contains a word longer than $K+2$, an event of probability $\mathbb{P}_{a_n}(E_K)$ by Lemma 3.1(b); or the ambient closure is contained in X_{K+2}. In the second case a RAF, if one exists, must by Lemma 4.1 contain a gateway reaction catalyzed by a molecule of the deterministic bounded set X_{K+2}. There are at most 68 gateway identities (every gateway has product length at most 4, and $|\mathcal{R}_4| = 68$), so a union bound over the deterministic finite set of coordinates gives

$$P_n \leq \mathbb{P}_{a_n}(E_K) + 68\,|X_{K+2}|\,p_n. \tag{10.4}$$

No catalytic independence conditional on a bounded selected RAF is needed. At fixed K, continuity of the finite event, $a_n \to a_\lambda$ and $p_n \to 0$ give $\limsup_n P_n \leq \mathbb{P}_{a_\lambda}(E_K)$; now let $K \to \infty$ and use (3.1). □

The two propositions use boundedness in complementary ways. A bounded infinite static closure is a legitimate extinction mechanism, and a RAF trapped in a fixed bounded universe becomes improbable in the finite catalytic source because each required coordinate has probability tending to zero. The comparison uses both facts without conflating the two probability spaces.

Before continuity is invoked, the same argument already yields the sandwich $S(a_\lambda^-) \le \liminf P_n \le \limsup P_n \le S(a_\lambda)$; the finite-seed argument of Section 9 is exactly what closes the remaining gap.

10.3. **Arbitrary intensities and exact normalization.**

Lemma 10.3. *For each n, $v \mapsto P_n(v)$ is nondecreasing.*

Proof. Couple all catalysis coordinates by common uniform marks. Increasing v increases $p_n(v)$ and adds catalysis coordinates without removing any. Food generation is unaffected and every catalytic witness remains present, so every existing RAF is preserved. □

Lemma 10.4 (Exact source normalization)**.** *If $f_n/n \to \lambda > 0$, then for all sufficiently large n the clamp in* (2.2) *is inactive and $p_n(f_n/n) = f_n/R_n$.*

Proof. Eventually $0 < f_n/n < 2\lambda$. The raw parameter $(f_n/n)n/R_n$ is then nonnegative and bounded by $2\lambda n/R_n$, which tends to zero and is eventually at most one. Hence the clamp is inactive, and for $n > 0$ cancellation gives $(f_n/n)\, n/R_n = f_n/R_n$. □

Proof of Theorem 2.2. Propositions 10.1 and 10.2 give $P_n(\lambda) \to S(a_\lambda) = \Theta(\lambda)$ for the exact sequence $f_n = \lambda n$. By Propositions 7.2 and 4.2, $0 < \Theta(\lambda) \le 1 - e^{-36\lambda} < 1$. Since $\lambda \mapsto a_\lambda$ is increasing and continuous from $(0, \infty)$ onto $(0, 1)$, Theorem 9.2 shows that Θ is nondecreasing and continuous. The gateway ceiling gives $\Theta(\lambda) \to 0$ as $\lambda \downarrow 0$, and since $a_\lambda \uparrow 1$ and S is continuous at 1 with $S(1) = 1$, $\Theta(\lambda) \to 1$ as $\lambda \to \infty$.

Now let (f_n) be arbitrary with $v_n = f_n/n \to \lambda > 0$. For fixed $0 < l < \lambda < h$, eventually $l \le v_n \le h$, so by Lemma 10.3

$$P_n(l) \le P_n(v_n) \le P_n(h). \tag{10.5}$$

The fixed-intensity result gives $\Theta(l) \le \liminf P_n(v_n) \le \limsup P_n(v_n) \le \Theta(h)$; sending $l \uparrow \lambda$ and $h \downarrow \lambda$ and using continuity of Θ proves $P_n \to \Theta(\lambda)$. The final assertion is Lemma 10.4. □

Proof of Corollary 2.3. Suppose $\gamma > 0$ had the step property. Then $P_n(\gamma/2) \to 0$, but by Theorem 2.2 $P_n(\gamma/2) \to \Theta(\gamma/2) > 0$. (Likewise any finite $\lambda > \gamma$ contradicts the upper branch because $\Theta(\lambda) < 1$.) □

Proof of Corollary 2.4. If $v_n \to 0$, then for every fixed $l > 0$ eventually $P_n(v_n) \le P_n(l)$ by Lemma 10.3, so $\limsup P_n(v_n) \le \Theta(l)$; let $l \downarrow 0$. If $v_n \to \infty$, then for every fixed $l > 0$ eventually $P_n(v_n) \ge P_n(l)$, so $\liminf P_n(v_n) \ge \Theta(l)$; let $l \to \infty$. The second case uses the clamp (2.2) if a raw parameter exceeds one; for genuine Bernoulli models with $f_n \le R_n$ no totalization is needed. □

11. Discussion

Interpretation of the scaling function. The map $\lambda \mapsto 1 - e^{-\lambda}$ aggregates the independent catalytic opportunities of a single reaction: the probability that at least one of the U_n molecules catalyzes a fixed reaction tends to a_λ. The function S then measures whether the corresponding infinite reversible closure escapes every finite length cap, and by Theorem 8.3, conditional on escape no word remains permanently absent. The finite catalytic source acquires the same limiting probability through the history barrier of Sections 5–6 and the bounded-core comparison of Proposition 10.2. Entrance and bulk should not be represented as an independent product of probabilities: the food boundary is part of the same infinite closure event that defines S, and the gateway ceiling of Proposition 4.2 is a ceiling rather than a formula for $1 - \Theta$.

Relation to simulations. A continuous nontrivial limiting law is compatible with a steep onset in finite simulations, and the theorem does not estimate the steepness of Θ or the rate at which P_n approaches it. The finite-size observations that motivated [8] therefore need not be in error; it is the asymptotic step-law reading of them that the theorem rules out. The value of Θ at moderate λ, its differentiability, and the asymptotics of $1 - \Theta(\lambda)$ as $\lambda \to \infty$ remain open.

Method. The contour argument of Section 7 is a Peierls-type estimate [15, 3] adapted to a host graph whose vertices are binary words and whose bonds are controlled through diamond parity and end-edit distance. Its distinctive features are the molecular weighting of the contracted food root, which yields a uniform envelope count for every actual molecule, and the strict separation between end-edit walks, used only for counting, and literal reaction detours, used only for chemistry. The record-word argument of Section 8 avoids any conditioning on the survival event by locating a fresh set of coordinates above a stopping cap. The uniformity in Theorem 9.1 is the decisive ingredient: monotonicity alone permits jumps of S, and it is the uniform finite-seed approximation that converts survival into a continuous function of the parameter.

Scope. The scope is deliberately fixed. Different alphabets, food sets of other lengths, quotient reaction identities that merge split positions, dependent or directional catalysis assignments, and inhibition define different models; they may change constants or require new arguments, and the present theorem does not transfer its conclusions to them. The constant 36 refers to the split-position identity convention with $F = X_2$. Nothing here concerns mass-action kinetics, concentrations, or dynamical persistence of a RAF once it exists.

What the proof requires. Only the logical chain displayed above is used: literal reaction geometry, finite-history equality, a static mass barrier, supplied-seed amplification, same-parameter richness, uniform finite-seed approximation, and a limiting squeeze. The formal development described in Appendix B contains reusable infrastructure from proof discovery, but none of its exploratory assertions is a premise of the publication theorem.

Appendix A. Finite word geometry: cuts, diamonds and detours

This appendix proves the deterministic statements (G1)–(G4) used in Section 7. Fix $L \geq 1$ and $n \geq L$. Words of length at most L are contracted to the root vertex o; every other vertex is a word s with $L < |s| \leq n$. The maps ℓ and r (deletion of the first, respectively last, bit followed by contraction) fix o, commute, and decrease the height $\max(|s| - L, 0)$.

A.1. **Diamonds and cuts.** The basic edges are $(s, 0)$ with endpoints $s, \ell(s)$ and $(s, 1)$ with endpoints $s, r(s)$, for non-root s. A Boolean vertex labelling h determines the cut consisting of edges whose endpoint labels differ. For each non-root s, its *diamond* has the four indexed slots

$$(s, 0), \qquad (s, 1), \qquad (r(s), 0), \qquad (\ell(s), 1), \tag{A.1}$$

joining s to $\ell(s)$, s to $r(s)$, $r(s)$ to $\ell(r(s))$ and $\ell(s)$ to $r(\ell(s))$. Since ℓ and r commute, the last two edges share their lower endpoint, so the four slots form the 4-cycle $s, \ell(s), \ell r(s), r(s)$. Slots at the root are empty and are not basic edges. A cut indicator has even parity across each diamond, because each of the four vertex labels appears twice around the cycle. Repeated slots near the contracted food boundary are counted with multiplicity.

Lemma A.1 (Integration)**.** *Any basic-edge indicator with even parity on every diamond is the cut of some vertex labelling.*

Proof. Set the label of o to zero and define the label at s as the parity sum of the indicator along the path of repeated right deletions from s to o. By construction the label difference across each right edge $(s, 1)$ equals its indicator. For a left edge $(s, 0)$, apply diamond parity at s together with the induction hypothesis one height level below: commutation identifies the two lower paths, so the difference across the left edge equals its indicator as well. Induction on height terminates at the root and uses only diamonds inside the finite cap. □

Lemma A.2 (Connected bonds)**.** *Every bond is connected in the graph whose vertices are the basic edges and in which two edges are adjacent when they share a diamond. This adjacency has degree at most* 9*, and the number of connected sets of* b *basic edges containing a specified edge is at most* 81^{b-1}*.*

Proof. Let Q be a connected component of a bond in the diamond adjacency, and restrict the cut indicator to Q. On each diamond either all cut slots are retained or none are, so parity remains even, and Lemma A.1 makes Q a cut. Minimality forces Q to be the whole bond.

A basic edge occurs in at most three diamonds: once as an upper edge of its own word's diamond, and at most twice as a lower edge, because a non-root word has at most two one-bit preimages for either deletion direction. Each diamond contributes at most three neighbours, so the degree is at most 9. A rooted depth-first traversal of a spanning tree of a connected b-set uses $2(b-1)$ steps, each with at most 9 choices, giving at most $9^{2(b-1)} = 81^{b-1}$ connected sets containing a specified edge. □

A.2. **Locating a bond near a molecule.** Let *end-edit distance* count insertions or deletions of one bit at either end of a word, allowing passage through the empty word. Each word has at most six one-step neighbours, so a ball of radius ρ contains at most 7^ρ words (allow an idle step in a code of length ρ). These walks are only a counting device; they need not be chemical derivations.

If two basic edges share a diamond, their product words differ by a bounded number of end edits. Connecting the edges of a size-b bond by at most $b-1$ diamond steps yields an end-edit walk of length at most $4(b-1)$ between any two boundary products. Consequently a substring of one boundary product that is protected by $4b$ bits on both sides persists as a substring of every boundary product.

Lemma A.3 (Molecular anchoring)**.** *Fix a bond of size* b *and any edge* e *of it with product word* s_e*. Count the two sides of the bond by actual molecules of* X_n*, each contracted food word counted separately. Every molecule on a smaller side has end-edit distance at most* $64b$ *from* s_e*.*

Proof. Short-product case: $|s_e| \le 8b$. Put $M = |s_e| + 4b \le 12b$; every boundary product has length at most M. If $M \ge n$ every molecule has length at most M. If $M < n$, all words in the length band $[M, n]$ receive the same cut label: a one-bit shift at length M is an insertion to length $M+1$ followed by a deletion, and neither step crosses an edge whose product has length at most M; repeated shifts connect all length-M words, and truncation connects the rest of the band. The high band contains more than half of the actual molecules, so its label is not the label of a smaller side, and every molecule on a smaller side has length at most M. Delete it to the empty word and insert s_e: at most $M + |s_e| \le 20b < 64b$ end edits.

Long-product case: $|s_e| > 8b$. Remove $\rho = 4b$ bits from each end of s_e to leave a nonempty core z. By the locality observation, z is a substring of every boundary product, and every boundary product has length at most $|z| + 3\rho$. A word avoiding z cannot cross the bond during right deletion to food, so every word on the non-food side contains z. Above all boundary products the high-band connectivity applies, and a constant word opposite to the first bit of z avoids z; hence the entire high band carries the food label, and every non-food-side word has length at most $|z| + 3\rho$.

If the non-food side is a smaller side: a non-food-side word and s_e both contain z; delete their exterior bits to z and insert the exterior bits of s_e. The exterior lengths are at most 3ρ and 2ρ, so the distance is at most $5\rho = 20b$.

If the food side is a smaller side: for any nonempty core z, the number of words containing z with length at most $|z| + k$ is at most

$$\sum_{j=0}^{k} (j+1)2^j < 2^{2(k+1)}, \tag{A.2}$$

by choosing one occurrence of z, its left padding length, and the j exterior bits (counting all occurrences only overcounts, and the bound is independent of $|z|$). The non-food side has at least $U_n/2$ molecules and is contained in the set counted by (A.2) with $k = 3\rho$, so $2^{n+1} - 2 = U_n \leq 2 \cdot 2^{2(3\rho+1)}$, which forces $n \leq 2(3\rho + 1)$. Also $|s_e| \leq n$. Any molecule can therefore be connected to s_e through the empty word in at most $2n \leq 12\rho + 4$ end edits, which is at most $64b$ since $\rho = 4b$ and $b \geq 1$. □

Lemma A.4 (Uniform contour envelope)**.** *For every molecule $s \in X_n$ and every $b \geq 1$, the bonds of size b having s on a molecularly smaller side number at most A^b, with A as in* (7.2)*.*

Proof. By Lemma A.3 the product of any edge of such a bond lies within end-edit radius $64b$ of s, so there are at most 7^{64b} candidate product words and at most two edge directions for an anchor edge. For each anchor edge, Lemma A.2 gives at most 81^{b-1} connected sets of size b containing it. Hence the number of bonds is at most

$$2 \cdot 7^{64b} \cdot 81^{b-1} \leq (2 \cdot 7^{64} \cdot 81)^b = A^b.$$

The argument applies to molecules of the contracted food as well: their weight has not been replaced by the weight of one graph vertex, because the smaller-side comparison used U_n throughout. This is why the envelope controls a uniform bad-event marginal for every actual molecule, including the root representative. □

Lemma A.5 (Separation)**.** *If two vertices lie in distinct effective components, then some bond separates them and every edge of that bond is non-effective.*

Proof. Contract each effective component of the basic graph to a point. The two vertices become distinct points of a connected quotient graph. Choose an inclusion-minimal cut separating them in the quotient and lift it back to the basic graph. Both sides are connected in the quotient, so the lifted cut is a bond of the basic graph, and all of its edges join distinct effective components and are therefore non-effective. □

A.3. **Literal detours.** Fix a basic edge joining a word s to the word obtained by deleting one bit at a chosen end, and $w \leq L/2$. For $j \in \{1, \ldots, w\}$, deleting j bits at that end yields a common descendant s_j. The j-th detour support consists of the reaction joining s to s_j (removed factor of length j) and, for $j > 1$, the reaction joining the one-bit truncation to s_j (removed factor of length $j - 1$). If the one-bit truncation is contracted food, the descendant is food too and the second reaction is unnecessary. Both removed factors have length at most $w \leq L$ and belong to the supplied food, so an open support gives genuine reversible reachability using only reactions with product in X_n.

Lemma A.6 (Bounded support overlap)**.** *Each detour support contains at most two literal reaction identities, and each literal reaction belongs to at most three supports over all edges and all j. Consequently among the wb detours of a size-b bond there are at least $wb/5$ with pairwise disjoint supports.*

Proof. A short reaction is identified by its product, the deletion end, and the removed factor length. It is the upper reaction of at most one edge (the edge at its product). If it is a lower reaction, its owner is a one-bit preimage of the product on that side, and there are at most two such preimages. This accounts for at most three supports. The two end descriptions cannot identify the same split when the product is non-food: that would make both factor lengths at most w, hence the product length at most $2w \leq L$, and such products are contracted. Food contractions only remove lower support coordinates and do not increase incidence.

A support meets at most four other supports, since each of its at most two reactions is used by at most two further supports. Greedily choose a support, discard it and its at most four neighbours, and repeat; at least $wb/5$ pairwise disjoint supports remain. □

With $w = 5k$, Lemma A.6 gives at least kb disjoint supports, each open with probability at least a^2, which is (7.3). Only literal reaction identities enter the probability calculation. End-edit walks used for anchoring do not assert open chemistry; conversely, detours used for chemical reachability explicitly supply their short factors. Keeping the two roles separate avoids importing a graph path as an unsupported chemical derivation.

Appendix B. Formal verification in Lean 4

The complete argument is formalized in Lean 4 [1] over Mathlib [18]. The development defines the concrete model of Section 2 literally (binary words, split-position reactions, the food set X_2, reversible closure, the RAF predicate, and the product Bernoulli measure on catalysis coordinates), defines the infinite product measure of Section 3, and proves Theorem 2.2 and its corollaries as theorems whose statements quantify over the actual RAF-existence event and the actual infinite survival event. No exploratory assumption from proof discovery enters as a premise.

B.1. **Root declarations.** All declarations below live in the namespace `HordijkSteelThreshold`, with the model definitions in `RAF.Concrete` and `RAF.Polymer`; module names refer to files under `proofs/HordijkSteelThreshold/` of the source repository [repository URL to be inserted]. The publication interface is the following declaration (`PublicationResolution.lean`), shown in ASCII notation:

```
theorem publication_resolution {f : Nat -> Real} {lambda : Real}
    (hlambda : 0 < lambda)
    (hf : Tendsto (fun n => f n / (n : Real)) atTop (nhds lambda)) :
    Tendsto (fun n => rafProbability n (f n / (n : Real))) atTop
      (nhds ((infiniteStaticMeasure (transitionOpenness lambda hlambda)
        {field | ReversibleUnbounded 2 field}).toReal)) /\
    0 < transitionLimit lambda hlambda /\
    transitionLimit lambda hlambda < 1 /\
    Continuous (fun x : {x : Real // 0 < x} =>
      transitionLimit x.val x.property) /\
    (forall x y : Real, forall hx : 0 < x, forall hy : 0 < y,
      x <= y -> transitionLimit x hx <= transitionLimit y hy)
```

Here `rafProbability n v` is $P_n(v)$ with the clamped parameter (2.2), `transitionOpenness lambda` is $1 - e^{-\lambda}$ as an element of the unit interval, `infiniteStaticMeasure a` is the product measure $\mathbb{P}_a$ on mark fields, `ReversibleUnbounded 2` is the event that $\mathcal{C}_2(\omega)$ is unbounded, and `transitionLimit lambda` unfolds definitionally to the real value of $S(1-e^{-\lambda})$. The mathematical root `full_corrected_transition` (`CriticalWindowSource.lean`) adds the eventual identity $p_n = f_n/R_n$ of Lemma 10.4 and the gateway ceiling $\Theta(\lambda) \le 1 - e^{-36\lambda}$. The refutation of the step conjecture is the separately authenticated declaration

```
theorem no_finite_step_threshold :
    Not (Exists fun gamma : Real => 0 < gamma /\
      (forall l : Real, 0 < l -> l < gamma ->
         Tendsto (fun n => rafProbability n l) atTop (nhds 0)) /\
      (forall l : Real, gamma < l ->
         Tendsto (fun n => rafProbability n l) atTop (nhds 1)))
```

in `TransitionConsequences.lean`, and the outer regimes of Corollary 2.4 are `sublinear_regime` and `superlinear_regime` in `OuterScalingRegimes.lean`. The earlier gateway-only refutation of the upper branch is the declaration `hordijkSteel_fixedFood_uniform_upper_false` in the module `RAF.Conjecture.ConcreteRefutation`.

B.2. **Concordance.** Tables 1 and 2 list, for each numbered result of the paper, the principal Lean declarations that establish it. Several constants and inequalities in the displayed proofs are elementary specializations of the listed declarations rather than separately named theorems; the table records the correspondence at the level of the mathematical result.

Paper result	Lean module and declaration(s)
Counts (2.1)	`PolymerCounts`; `RAF.Polymer.Counts`
Definition 2.1, P_n	`RAF.Concrete.SeedGateway.IsRevRAF`; `RAF.Concrete.Events.HasRAFEvent`, `rafProbability`
Survival (2.3)	`SprinkledSeedLowerBound.staticSurvival`; `ReversibleMeasurableEvents`
Lemma 3.1, (3.1)	`ReversibleFiniteCertificates`; `ReversibleEscape.reversible_escape_iff_cap`; `ReversibleEscapeProbability.finite_static_escape_probability_tendsto`
Lemma 4.1	`RAF.Concrete.SeedGateway.polymer_raf_implies_seedOpen`
Proposition 4.2	`GatewayProbability.rafProbability_limsup_le_gateway_ceiling`, `card_seedCoord_exact`, `seedClosedProbability_tendsto_exp`
Lemma 5.1	`ScalarHistoryDistribution.activeTrace_map_eq_scalar`; `NestedColumnHistory`; `HistoryProductLaw`
Lemma 6.1	`SourceStaticRAFBound.canonical_raf_ge_static_barrier`; `SourceTerminalRAF`; `PeelingStabilization`
Limit (6.3)	`TransitionLowerBound.source_probability_eventually_above`
Theorem 7.1	`StaticBulk.static_bulk_near_full`; `WordContourProbability.staticReaction_contour_failure`; `WordRootedMass.measure_rootedMolecularWords_deficit`; `StaticClosureMass`
Proposition 7.2	`TransitionConsequences.staticSurvival_pos`; `StaticFoodTwoBulk`
Lemmas 8.1, 8.2	`ReversibleRecordWords.unbounded_record_words`; `RecordTargetTrial.record_target_support`
Theorem 8.3	`SameParameterRichness.same_parameter_target_ae`, `unbounded_missing_target_null`; `InfiniteClosureDichotomy.survival_iff_complete_ae`, `finite_or_complete_ae`, `staticSurvival_eq_complete`
(9.1)	`InfiniteStaticLaw.finite_static_seed_probability_tendsto`; `SameParameterRichness.staticSurvival_le_same_parameter_seed`
Theorem 9.1	`SeedSurvivalApproximation.seed_probability_le_survival_add_error`; `InfiniteClosureDichotomy.uniform_seed_survival_approximation`

TABLE 1. Concordance between the numbered results of Sections 2–9 and the Lean declarations.

The formal lower bound was first compiled with an independent-sprinkling intermediate; the same-parameter richness theorem now available in the dependency chain replaces that intermediate in the prose of Proposition 10.1. Both routes yield the same limiting inequalities, and the displayed proof needs no sprinkling lemma.

B.3. **Verification record.** The root declarations were checked with Lean `v4.30.0` against a pinned Mathlib revision (`c5ea0035`) with warnings treated as errors; `sorry` and `admit` are rejected. The dependency closure of `publication_resolution` comprises 151 local modules, which were rebuilt from source in an isolated cache separate from any shared proof cache. The axiom report for every root declaration lists exactly `propext`, `Classical.choice` and `Quot.sound`, the standard classical and quotient principles used throughout Mathlib; no project-specific axiom occurs. Source hashes, compiled-artifact hashes, exported declaration interfaces and the axiom reports are recorded in a machine-readable verification manifest distributed with the source. Kernel verification establishes the encoded declarations; source-model fidelity additionally rests on the definitions of Section 2, which are part of the stated theorem surface.

Paper result	Lean module and declaration(s)
Theorem 9.2	`StaticSurvivalContinuity.staticSurvival_tendsto`; `StaticSurvivalLeftContinuity.staticSurvival_le_left`; `OuterScalingRegimes.staticSurvival_one`
Proposition 10.1	`TransitionLowerBound.transition_liminf_lower`
Proposition 10.2	`TransitionUpperBound.transition_limsup_le_finite_escape`, `transition_limsup_upper`
Fixed-intensity limit	`CorrectedTransition.rafProbability_tendsto_transitionLimit`, `corrected_transition`
Lemma 10.3, (10.5)	`CanonicalParameterMonotonicity`; `VariableIntensityTransition.variable_intensity_transition`, `critical_window_transition`
Lemma 10.4	`CriticalWindowSource.critical_window_parameter_exact`
Theorem 2.2	`CriticalWindowSource.full_corrected_transition`; `PublicationResolution.publication_resolution`; `TransitionConsequences.continuous_transitionLimit`, `transitionLimit_mono`, `transitionLimit_mem_Ioo`, `transitionLimit_zero`
Corollary 2.3	`TransitionConsequences.no_finite_step_threshold`
Corollary 2.4	`OuterScalingRegimes.sublinear_regime`, `superlinear_regime`, `transitionLimit_infinity`
Lemma A.1	`WordDiamondCuts.wordDiamond_integrate`
Lemma A.2	`WordBondConnected.wordBond_diamond_connected`; `WordDiamondDegree`; `WordContourCounting`
Lemma A.3	`WordMolecularAnchors.wordBond_molecular_anchor_radius`
Lemma A.4	`WordMoleculeContourCount.wordMoleculeContourEnvelope_card`, `wordBond_mem_molecule_envelope`
Lemma A.5	`FiniteBondSeparation`
Lemma A.6	`WordDetourLiteralPaths.wordDetourSource_reach_parent`; `WordDetourIncidence`; `DisjointDetourSelection.exists_disjoint_detours`

TABLE 2. Concordance (continued): Sections 9–10, the main results, and Appendix A.

REFERENCES

1. Leonardo de Moura and Sebastian Ullrich, *The Lean 4 theorem prover and programming language*, Automated Deduction – CADE 28, Lecture Notes in Computer Science, vol. 12699, Springer, 2021, pp. 625–635.
2. Manfred Eigen, *Selforganization of matter and the evolution of biological macromolecules*, Naturwissenschaften **58** (1971), no. 10, 465–523.
3. Geoffrey Grimmett, *Percolation*, 2nd ed., Grundlehren der mathematischen Wissenschaften, vol. 321, Springer, Berlin, 1999.
4. Wim Hordijk, Jotun Hein, and Mike Steel, *Autocatalytic sets and the origin of life*, Entropy **12** (2010), no. 7, 1733–1742.
5. 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** (2011), no. 5, 3085–3101.
6. Wim Hordijk, Joshua I. Smith, and Mike Steel, *Algorithms for detecting and analysing autocatalytic sets*, Algorithms for Molecular Biology **10** (2015), 15.
7. Wim Hordijk and Mike Steel, *Detecting autocatalytic, self-sustaining sets in chemical reaction systems*, Journal of Theoretical Biology **227** (2004), no. 4, 451–461.
8. ———, *Autocatalytic sets in polymer networks with variable catalysis distributions*, Journal of Mathematical Chemistry **54** (2016), no. 10, 1997–2021.
9. ———, *Chasing the tail: The emergence of autocatalytic networks*, Biosystems **152** (2017), 1–10.
10. ———, *Autocatalytic networks at the basis of life's origin and organization*, Life **8** (2018), no. 4, 62.
11. Sanjay Jain and Sandeep Krishna, *Autocatalytic sets and the growth of complexity in an evolutionary model*, Physical Review Letters **81** (1998), no. 25, 5684–5687.
12. Stuart A. Kauffman, *Autocatalytic sets of proteins*, Journal of Theoretical Biology **119** (1986), no. 1, 1–24.
13. ———, *The origins of order: Self-organization and selection in evolution*, Oxford University Press, New York, 1993.

14. Elchanan Mossel and Mike Steel, *Random biochemical networks: the probability of self-sustaining autocatalysis*, Journal of Theoretical Biology **233** (2005), no. 3, 327–336.
15. Rudolf Peierls, *On Ising's model of ferromagnetism*, Mathematical Proceedings of the Cambridge Philosophical Society **32** (1936), no. 3, 477–481.
16. Mike Steel, *The emergence of a self-catalysing structure in abstract origin-of-life studies*, Applied Mathematics Letters **13** (2000), no. 3, 91–95.
17. Mike Steel, Wim Hordijk, and Joana C. Xavier, *Autocatalytic networks in biology: structural theory and algorithms*, Journal of the Royal Society Interface **16** (2019), no. 151, 20180808.
18. The mathlib Community, *The Lean mathematical library*, Proceedings of the 9th ACM SIGPLAN International Conference on Certified Programs and Proofs (CPP 2020), ACM, 2020, pp. 367–381.

Affiliation to be inserted
Email address: `email to be inserted`