

Two consecutive productive windows in random polymer reactors: startup, resource bounds, and selection by successful operation

16 September 2026

Abstract

A reflexively autocatalytic and food-generated set certifies that a reaction network can regenerate its catalysts from food. It does not say when the first catalyst molecule appears in a reactor started from food alone, how much material leaves the reactor, or whether stock remains after an operating interval. We study these questions in one explicit random model: the reversible binary polymer model with ordered split identities, the capped, shifted Zipf catalysis law, independent kinetic marks, and a fed, diluted count process with falling-factorial propensities. The operating requirement is a joint event on one uninterrupted trajectory: measured stock at times 1, 100 and 199, nonfood monomer export above $V/10$ in each of the windows $(1, 100]$ and $(100, 199]$, a mass corridor, and a gross supply allowance, where V is the count scale. We prove a positive lower bound for its probability, constructive on an exactly counted source class with arbitrary nonfood background, and a converse; both hold for every Zipf exponent $a > 1$ and every polymer length $n \geq 4$ at a sufficient count scale. Along the critical sequence $a_n = 2 - 2/n$ the probability is $\Theta(2^{-n})$, the limiting prefactors relative to the incidence probability lie between $(6/\pi^2)^6$ and 224, where 224 is the exact number of productive one-channel incidences, and conditioning on success selects such an incidence with probability tending to one. The same estimates bound the fraction of highly reliable environments, the cost of independent screening, and the effect of deleting all catalysis. For every fixed exponent the rarity order is computed exactly; singleton dominance persists for $a \geq 2$ and the argument identifies why heavier tails need a different converse. A stopped first-birth argument gives a necessary startup scale: in food-silent environments the success probability is at most $1 - \exp[-(484/3)\epsilon V]$, so 99% reliability needs $V \geq 1.43 \times 10^7$, while the sufficient scale of the constructive proof, even after an explicit tolerance optimisation, remains near 5×10^{49} ; we explain which single tolerance sets it. Material identities show that the export is net internal synthesis and recover at least $1/12000$ of the supplied monomer. The loss envelopes, potential, startup, interval and noise lemmas are stated with complete hypotheses; the finite core is machine-checked in Lean 4 and every conventional step is labelled.

1 Introduction

Collectively autocatalytic sets were proposed by Kauffman [18] as a way for a random chemistry to sustain itself without a single self-replicating molecule. Their combinatorial formalization is the reflexively autocatalytic and food-generated set, or RAF, of Steel and Hordijk [12, 25, 26]: a set of reactions that generates every molecule it needs from a food set and whose every reaction is catalyzed by a molecule so generated. In the binary polymer model RAFs appear once the mean number of reactions catalyzed per molecule grows linearly in the maximum word length [16, 20],

and heterogeneous, power-law catalysis changes both their appearance and their size [14, 15, 17]. A RAF is a statement about closure. It does not specify when the first catalyst appears, how much material leaves the reactor, or whether anything is left after an operating interval. Those questions need an initial state, a kinetic law, a deadline, a collection rule, and a resource account.

Kinetic studies of autocatalytic sets are as old as the model [8]. Hordijk and Steel [13] simulated the molecular flow of RAF networks and described uncatalyzed seeding of self-catalytic reactions; Giri and Jain [11] separated the catalytic strengths needed to maintain a large-molecule state from those needed to reach it from food; Ravoni [24] studied how composition shapes the dynamics of autocatalytic sets in stochastic network models; the stoichiometric theory of autocatalysis [4, 21, 28] classifies the motifs that can amplify their own species; and seed-dependent activation [22] shows why food-only initiation is a substantive restriction. Two companion papers fixed one random model in which structure and operation can be compared exactly: the capped-Zipf binary polymer model of Hordijk and Steel in a fed, diluted stochastic mass-action reactor. The first [2] proved that RAFs exist with limiting probability $\theta \in (0, 1)$ while one-window productive output has probability of order $p_n \asymp 2^{-n}$, the probability that one specified molecule catalyzes one specified channel. The second [1] proved that, conditional on productive output, the successful environments contain a one-channel RAF with probability tending to one.

This paper asks for more from the same trajectory. The reactor is prepared with food only, and one uninterrupted run must show measured stock at three observation times, export nonfood material in two consecutive windows, stay inside a mass corridor, and consume at most a stated gross supply. No reset, no resampling of the chemistry, and no fresh initial law at the middle observation are allowed. We prove that this joint event has positive probability of a computable order, identify what typical successful environments contain, give a necessary discrete startup scale that the concentration-level argument cannot see, and separate three requirements that a structural certificate conflates: availability of a productive organization, physical creation of its first molecules, and execution of a specified operating task within a resource budget.

Contributions.

- (i) *A finite two-window theorem* (Theorem 4.1). For every $n \geq 4$, every Zipf exponent $a > 1$, and every integer count scale $V \geq 10^{60}(n+1)^2$, the joint event $F_{n,V}$ has probability at least $e_a^6 p_{a,n}(1 - \delta_{n,V}) > 0$ and at most $224p_{a,n} + \bar{r}_{a,n,V}$, with $e_a = 1/\zeta(a)$, $p_{a,n}$ the exact incidence probability, $\delta_{n,V} = 24e^{-cV/n}$ and an explicit residual $\bar{r}$. The lower bound is uniform over every kinetic mark configuration on a source class that leaves every nonfood catalytic row unrestricted. On the event, net internal nonfood synthesis exceeds $V/5$ and the recovered fraction of supplied monomer exceeds $1/12000$.
- (ii) *Rarity and selection along the critical sequence* (Theorem 4.2). At $a_n = 2 - 2/n$ the success probability is $\Theta(2^{-n})$ with $(6/\pi^2)^6 \leq \liminf S_n/p_n \leq \limsup S_n/p_n \leq 224$; conditional on success, a productive one-channel incidence is present with probability tending to one and the constructed food-silent class has posterior at least $(6/\pi^2)^6/224$ in the limit.
- (iii) *A necessary startup scale* (Theorem 4.3). In every food-silent environment, whatever the marks, the success probability is at most $1 - e^{-(484/3)\varepsilon V}$; reliability $1 - \eta$ needs $V \geq \log(1/\eta)/\lambda$ with $\lambda = 484\varepsilon/3$. The proof is a stopped finite-state uniformization with corridor exits sent to a safe absorbing state.
- (iv) *Robustness in the exponent* (Theorem 4.4). For every fixed $a > 1$ the rarity order $S_{a,n} = \Theta(p_{a,n})$ holds with $p_{a,n}$ computed exactly; for $a \geq 2$ and along every family $a_n = 2 - \beta/n$ singleton dominance and the 224 prefactor persist; for $1 < a < 2$ the same-row pair probability is a

positive fraction of $p_{a,n}$, which identifies exactly why the present converse does not extend there.

- (v) *Explicit sufficient scales* (Proposition 4.5). The uniform witness bound needs only $V \geq \max\{2n/d, (n/c) \log(24/\eta)\}$, and an unequal-tolerance evaluation of the six-term noise estimate lowers the 99% witness scale from 8.97×10^{50} to 5.0×10^{49} at $n = 4$ while removing the n -dependence of the dominant term. Both remain impractical, and we show that one absolute tolerance, on the selected product coordinate, sets the scale.
- (vi) *Self-contained kinetic derivations*. The loss coefficients 23 and 25, the potential and its drift, the endpoint constant 54, the startup floor, the interval telescope and the six-term probability estimate are derived or stated with complete hypotheses (Section 5), so that the argument can be read without the formal development.
- (vii) *Formal verification and diagnostics*. The finite core, including the actual-law probability, the exact source integration, the original converse and the signed synthesis account, is a single Lean 4 [6, 27] declaration; the productive census and the startup algebra are separately checked (Appendix B). A matched three-environment example at $n = 4$ (Section 10) illustrates the initiation distinction with exact-algorithm count simulations and deterministic trajectories; it proves nothing and is labelled accordingly.

Three kinds of evidence. Three kinds of statement occur and are kept apart throughout. *Formal results* have a Lean declaration named in Appendix B: the finite two-window core with the containing-rectangle converse coefficient 15480, the census 224 with its predicate equivalence and injection, the food-pair algebra, the finite-kernel survival induction, the logarithmic necessary-scale inequality and the material-ratio inequality. *Conventional results* are proved in full in the text and have no separate formal export: the union bound that sharpens 15480 to 224, the limiting constants, the screening, reliability and transfer corollaries, the identification of the stopped generator in Theorem 4.3, the fixed-exponent asymptotics and the unequal-tolerance evaluation. *Numerical illustrations* supply no premise to any proof. Every printed number is replayed by an exact-arithmetic script distributed with the source.

What is not claimed. The model is schematic chemistry: binary words, ordered split identities, one basal coefficient scale ε , paired reversible coefficients, uniform washout, and a well-mixed vessel. No laboratory molecule is identified and no rate is fitted. The theorem concerns a fixed finite horizon as n and V grow; it asserts no infinite-time persistence, no uniform restart guarantee from an arbitrary state in the stock region, no exact success coefficient, and no equal operational weights for the 224 productive labels. The measured output is aggregate nonfood monomer effluent, not a purified functional product, and gross material recovery is distinct from chemical work or entropy production [23]. The sufficient count scales are conservative by many orders of magnitude; the necessary scale is a genuine lower bound only for the stated class, preparation, coefficients and deadline.

Organization. Section 2 fixes the source and the count process, Section 3 the operating requirement and its material identities. Section 4 states the theorems. Section 5 proves the finite theorem, Section 6 the limits and consequences, Section 7 the startup ceiling, Section 8 the exponent robustness, and Section 9 the explicit scales and the converse obstruction. Section 10 gives the matched example, Section 11 dimensions and scope, and Section 12 the discussion. Appendix A collects the explicit constants and Appendix B the formal concordance.

2 The source and the count process

Everything below is literally the model of the companion papers [1, 2] and of the Lean developments `PowerLawSmallRAF` (source) and `RandomViability` (reactor).

2.1 Catalogue, food, and split identities

Fix $n \geq 4$. Let $\mathcal{X}_n$ be the nonempty binary words of length at most n , with $|z|$ the length of z , and let $\mathcal{R}_n$ consist of the ordered split identities $r = (u, v, uv)$ with $|u|, |v| \geq 1$ and $|uv| \leq n$. Counting products of length j with their $j - 1$ splits,

$$X_n = |\mathcal{X}_n| = 2^{n+1} - 2, \quad R_n = |\mathcal{R}_n| = \sum_{j=2}^n (j-1)2^j = (n-2)2^{n+1} + 4. \quad (2.1)$$

The order of the inputs and the split position belong to the identity of a channel even when another channel has the same net stoichiometry; each identity has a ligation direction $u + v \rightarrow uv$ and a cleavage direction $uv \rightarrow u + v$. The food set is $\mathcal{F} = \{0, 1, 00, 01, 10, 11\}$, the six words of length at most two. At $n = 4$ there are 30 molecules and 68 split identities. Replacing ordered splits by unordered input pairs, or merging channels with equal stoichiometry, changes both the source law and the generator; the constants below are for the stated convention.

2.2 The capped, shifted Zipf source

For a real exponent $a > 1$, independently for each molecule $z \in \mathcal{X}_n$ draw an integer $K_z \geq 1$ with

$$\mathbb{P}(K_z = k) = \frac{k^{-a}}{\zeta(a)}, \quad D_z = \min(K_z, R_n) - 1, \quad (2.2)$$

and, conditional on $D_z = d$, let the row $C_z \subseteq \mathcal{R}_n$ of channels catalyzed by z be a uniformly random d -subset. Thus

$$\mathbb{P}(D = d) = \frac{(d+1)^{-a}}{\zeta(a)} \quad (0 \leq d \leq R_n - 2), \quad \mathbb{P}(D = R_n - 1) = \frac{1}{\zeta(a)} \sum_{k \geq R_n} k^{-a}. \quad (2.3)$$

The whole Zipf tail sits at the largest degree; this atom is retained everywhere. An *incidence* is a pair (z, r) with $r \in C_z$; catalyzing a channel catalyzes both of its directions, and food molecules may catalyze. Rows of different molecules are independent, while the assignments within one row are sampled without replacement given a random degree and are strongly dependent: nothing below treats $\{r \in C_z\}_r$ as independent. By exchangeability within a row, every incidence has the same marginal probability, and we write

$$p_{a,n} = \frac{\mathbb{E}D}{R_n}, \quad e_a = \mathbb{P}(D = 0) = \frac{1}{\zeta(a)}, \quad q_{2,a,n} = \mathbb{P}(r, s \in C_z) = \frac{\mathbb{E}[D(D-1)]}{R_n(R_n-1)} \quad (r \neq s). \quad (2.4)$$

Appendix A gives closed forms for the two moments. At $a_n = 2 - 2/n$ we abbreviate $p_n = p_{a_n,n}$ and $e_n = e_{a_n}$.

2.3 Kinetic marks and propensities

Independently of the source, draw marks S_r, B_r, H_{rz} , indexed by channels and by incidences, independently and uniformly from $\{1, \frac{3}{2}, 2\}$, and set

$$b_r = \varepsilon S_r B_r, \quad k_{rz} = 4 S_r H_{rz}, \quad \varepsilon = \frac{1}{500\,000\,000} = 2 \times 10^{-9}, \quad (2.5)$$

so that $\varepsilon \leq b_r \leq 4\varepsilon$ and $4 \leq k_{rz} \leq 16$. The factor S_r is deliberately shared between the basal and the catalytic coefficients of one channel; they are not independent rates. Each reversible pair uses the same coefficient in both directions. The environment $\omega = (C, S, B, H)$ is fixed throughout a run.

For an integer count scale $V > 0$, a reaction with input multiplicities α_z , coefficient κ and total molecularity $m = \sum_z \alpha_z$ has propensity

$$\kappa V^{1-m} \prod_{z \in \mathcal{X}_n} (N_z)_{\alpha_z}, \quad (N)_j = N(N-1) \cdots (N-j+1), \quad (2.6)$$

which vanishes when an input is unavailable; there is no additional factorial divisor. A basal channel has input $u + v$ and output uv ; the catalytic channel of the incidence (z, r) has input $u + v + z$ and output $uv + z$, and repeated substrate and catalyst roles are combined before the falling factorial is taken. The reverse channels use the corresponding reverse complexes. Bounding a falling factorial by the corresponding power is an upper-bound device used for losses, never an identity.

2.4 Preparation, feed, washout, and probability laws

Each food species arrives at rate V , and every molecule z washes out at rate N_z . Initially $N_f(0) = V$ for each $f \in \mathcal{F}$ and every nonfood count is zero, so the initial molecule count is $6V$, the initial monomer mass is $10V$, and the initial nonfood mass is zero. Internal reactions preserve monomer mass, washout decreases it, and the six independent feeds have finitely many arrivals on every bounded interval; hence on every finite horizon the mass, and with it every count, is bounded by the initial mass plus the accumulated input, the reachable state set is finite, and the process is nonexplosive [3, 7]. Here V is a count scale, not a volume in litres and not a molecule total; Section 11 gives its dimensional reading.

Three laws occur. $\mathbb{P}_\omega$ is the trajectory law in a fixed environment ω from the food-only state; $\mu_{a,n}$ is the law of the environment, source and marks together; and $\mathbb{P}$ averages both. For an event F of the trajectory we write $q(\omega) = \mathbb{P}_\omega(F)$ and $\mathbb{P}(F) = \int q \, d\mu_{a,n}$. Conditioning on F under $\mathbb{P}$ reweights environments by $q(\omega)$; selecting environments with $q(\omega) \geq 1 - \eta$ thresholds that function. These are different distributions and are never substituted for one another.

3 The operating requirement

Let $M(t) = \sum_z |z| N_z(t)$ be the total monomer mass, $M_{\text{nf}}(t) = \sum_{|z| > 2} |z| N_z(t)$ the nonfood monomer mass, $Q(s, t]$ the monomer equivalents removed by nonfood washouts during $(s, t]$, $A(t)$ the number of food-molecule arrivals by time t , and $B(t)$ their monomer equivalents, so that $A(t) \leq B(t) \leq 2A(t)$. These are gross accumulated inputs, not net changes of inventory. At deterministic observation times the holding-state convention of the jump process agrees almost surely with the usual right-continuous version.

Definition 3.1 (Endpoint stock and the two-window event). A state has *endpoint stock* if

$$M/V \leq 21/2 \quad \text{and} \quad \text{some } z \text{ with } |z| = 4 \text{ satisfies } N_z/V \geq (3 \times 10^{18})^{-1}. \quad (3.1)$$

The event $F_{n,V}$ requires endpoint stock at the times 1, 100 and 199, together with

$$Q(1, 100] > \frac{V}{10}, \quad Q(100, 199] > \frac{V}{10}, \quad M(t) \leq 11V \quad (0 \leq t \leq 199), \quad A(199) \leq 1195V. \quad (3.2)$$

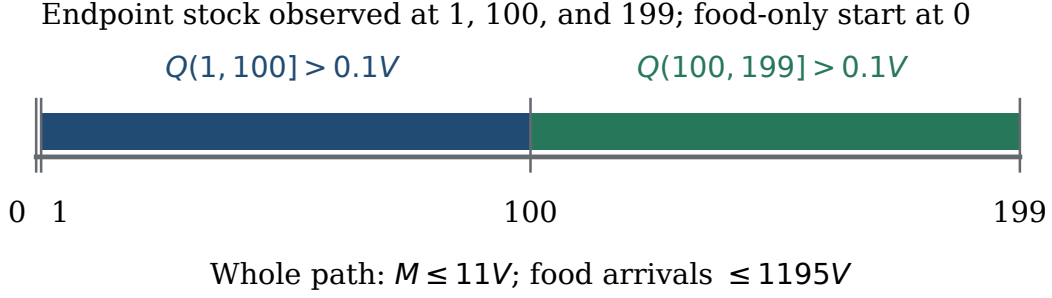


Figure 1: The single-trajectory requirement. The endpoint stock condition is observed at the three marked times; the mass corridor and the arrival allowance cover the whole interval, including startup from the food-only state.

No catalytic label occurs in this definition. The stock condition is existential over individual length-four species, not a pooled sum of all length-four counts, and the qualifying species may differ between the three observation times on a general successful path; the constructive proof tracks one particular species. The reactor is not reset at time 100: there is no resampling of the chemistry, the clock, or the middle state. Observing stock at three times does not assert that stock is positive at every intervening time. Washout is the collection mechanism, and the collected material is aggregate nonfood, not a named sequence and not a purified active product. Figure 1 shows the requirement.

Write $S_{a,n,V} = \mathbb{P}(F_{n,V})$. The one-window event of the companion papers, F^{old} , requires $M(t) \leq 11V$ for $0 \leq t \leq 100$ and $Q(1, 100] > V/10$. Consequently

$$F_{n,V} \subseteq F^{\text{old}}, \quad (3.3)$$

and this exact inclusion is the route by which every inherited converse estimate transfers.

Lemma 3.2 (Material identities). *Let $J(t)$ be the signed change of nonfood monomer mass produced by all internal reactions up to time t , counting reverse reactions with their actual sign. Then for every trajectory and every $t \geq 0$,*

$$M_{\text{nf}}(t) + Q(0, t] = M_{\text{nf}}(0) + J(t) = J(t), \quad 10V + B(t) \text{ is the total monomer supplied by time } t. \quad (3.4)$$

On $F_{n,V}$, therefore, $J(199) > V/5$ and

$$\frac{Q(0, 199]}{10V + B(199)} > \frac{V/5}{10V + 2 \cdot 1195V} = \frac{1}{12000}. \quad (3.5)$$

Proof. Every jump changes M_{nf} by one of three amounts: a food arrival by zero, since foods carry no nonfood mass; a nonfood washout by exactly minus its export weight, which is what Q accumulates; and an internal reaction by its signed nonfood mass change, which is what J accumulates. Summing over the finitely many jumps on $[0, t]$ gives (3.4), and $M_{\text{nf}}(0) = 0$ by the preparation. The supplied monomer is the initial mass $10V$ plus the arrivals $B(t)$. On the event the two windows export more than $V/5$ in total, the inventory $M_{\text{nf}}(199)$ and the startup export $Q(0, 1]$ are nonnegative, so $J(199) > V/5$; and $B(199) \leq 2A(199) \leq 2390V$ gives (3.5). $\square$

The fraction $1/12000 \approx 0.0083\%$ has monomer equivalents in numerator and denominator. It is a conservative material-recovery statement and not an energy efficiency, a purified yield, or a claim about the function of any exported species. Its content is that the export cannot be explained by preloaded nonfood inventory: there is none.

4 Main results

Fix the split identity $r_* = (00, 11, 0011)$ and let A_n be the source event that all six food rows are empty and $r_* \in C_{0011}$; the other entries of the row C_{0011} and all other nonfood rows are unrestricted. Row independence and exchangeability within a row give

$$\mu_{a,n}(A_n) = e_a^6 p_{a,n}. \quad (4.1)$$

Put

$$d = \frac{1}{6 \times 10^{20}}, \quad c = \frac{d^2}{8 \times 10^7} \approx 3.47 \times 10^{-50}, \quad \delta_{n,V} = 24e^{-cV/n}, \quad V_*(n) = 10^{60}(n+1)^2. \quad (4.2)$$

A *productive incidence* is a pair (r, z) with $r = (u, v, uv)$, $u, v \in \mathcal{F}$, $|uv| > 2$, and $z \in \mathcal{F} \cup \{uv\}$; T_n is the event that some productive incidence is assigned. Let $K_* = 10^{12}$ and let L_n count the assigned incidences whose catalyst has length at most K_* and whose product has length at most $K_* + 2$; this K_* -rectangle is finite and independent of n and is never enumerated. With $c_{\text{old}} = d^2/(4 \times 10^7) = 2c$ define the residual

$$\bar{r}_{a,n,V} = \mathbb{P}_{a,n}(L_n \geq 2) + 8e^{-c_{\text{old}}V/n}. \quad (4.3)$$

Theorem 4.1 (Two consecutive windows). *For every $n \geq 4$, every $a > 1$ and every integer $V \geq V_*(n)$:*

$$\mathbb{P}_\omega(F_{n,V}) \geq 1 - \delta_{n,V} \quad \text{for every } \omega \in A_n \text{ and every mark configuration}, \quad (4.4)$$

$$0 < e_a^6 p_{a,n}(1 - \delta_{n,V}) \leq \mathbb{P}(F_{n,V} \cap A_n) \leq S_{a,n,V}, \quad (4.5)$$

$$S_{a,n,V} \leq \min\{1, 224p_{a,n} + \bar{r}_{a,n,V}\}, \quad (4.6)$$

$$\mathbb{P}(F_{n,V} \cap T_n^c) \leq \bar{r}_{a,n,V}. \quad (4.7)$$

On $F_{n,V}$ the net internal nonfood synthesis through time 199 exceeds $V/5$, and $Q(0, 199]/(10V + B(199)) > 1/12000$.

The formal root uses the containing-rectangle coefficient $15480 = 30 \cdot 516$ in place of 224; the census in Section 5.6 sharpens it. The uniform bound (4.4) is what makes the theorem constructive: it holds pointwise for every environment in A_n , however many assignments the nonfood rows carry.

Now let $a_n = 2 - 2/n$, let $V_n \geq V_*(n)$ be any integer sequence, and write $S_n = S_{a_n, n, V_n}$, $\bar{r}_n = \bar{r}_{a_n, n, V_n}$ and $\gamma = (6/\pi^2)^6$.

Theorem 4.2 (Rarity and conditioning along the critical sequence). *Along every such sequence,*

$$\gamma \leq \liminf_n \frac{S_n}{p_n} \leq \limsup_n \frac{S_n}{p_n} \leq 224, \quad (4.8)$$

$$\liminf_n \mathbb{P}(A_n \mid F_{n,V_n}) \geq \frac{\gamma}{224}, \quad \mathbb{P}(T_n^c \mid F_{n,V_n}) \rightarrow 0, \quad (4.9)$$

$$2^n p_n \rightarrow \frac{9}{2\pi^2}, \quad S_n = \Theta(2^{-n}). \quad (4.10)$$

Numerically $\gamma \approx 0.0504788$ and $\gamma/224 \approx 2.2535 \times 10^{-4}$.

The posterior statement is a liminf bound: it proves a positive asymptotic contribution of the constructed class, not that this class accounts for most successful reactors, and no exact limit of S_n/p_n or of the posterior is asserted.

Theorem 4.3 (Stopped first-birth ceiling). *Let τ be the first time a nonfood molecule exists and σ the first time $M > 11V$. For $n \geq 4$, food-only preparation, empty food rows, and every mark configuration, for every $t \geq 0$,*

$$\mathbb{P}_\omega(\tau \leq t, \tau < \sigma) \leq 1 - e^{-\lambda V t}, \quad \lambda = \frac{484}{3} \varepsilon \approx 3.2267 \times 10^{-7}, \quad (4.11)$$

and consequently $\mathbb{P}_\omega(F_{n,V}) \leq 1 - e^{-\lambda V}$. Reliability $q(\omega) \geq 1 - \eta$, $0 < \eta < 1$, therefore requires $V \geq \log(1/\eta)/\lambda$; at $\eta = 0.01$ this is $V \geq 14\,272\,222$.

Theorem 4.4 (Robustness in the Zipf exponent). *Let $V_n \geq V_*(n)$ be any integer sequence and write $S_{a,n} = S_{a,n,V_n}$.*

(a) (Exact orders.) *As $n \rightarrow \infty$, with $R = R_n \sim 2n2^n$,*

$$p_{a,n} \sim \begin{cases} \frac{R^{1-a}}{(a-1)(2-a)\zeta(a)}, & 1 < a < 2, \\ \frac{\log R}{\zeta(2)R}, & a = 2, \\ \frac{\zeta(a-1)/\zeta(a) - 1}{R}, & a > 2, \end{cases} \quad \text{so} \quad p_{a,n} = \begin{cases} \Theta(n^{1-a}2^{-(a-1)n}), \\ \Theta(2^{-n}), \\ \Theta(n^{-1}2^{-n}). \end{cases} \quad (4.12)$$

In particular $2^n p_{2,n} \rightarrow 3 \log 2 / \pi^2$.

- (b) (Rarity order for every exponent.) *For every fixed $a > 1$, $S_{a,n} = \Theta(p_{a,n})$, with $\liminf S_{a,n}/p_{a,n} \geq \zeta(a)^{-6}$; the population of environments with $q(\omega) \geq 1 - \eta$ is also $\Theta(p_{a,n})$ for every fixed $\eta \in (0, 1)$.*
- (c) (Singleton dominance for $a \geq 2$.) *For every fixed $a \geq 2$, $\mathbb{P}(L_n \geq 2) = o(p_{a,n})$; hence $\bar{r}_{a,n,V_n}/p_{a,n} \rightarrow 0$, $\limsup S_{a,n}/p_{a,n} \leq 224$, $\mathbb{P}(T_n^c \mid F_{n,V_n}) \rightarrow 0$ and $\liminf \mathbb{P}(A_n \mid F_{n,V_n}) \geq \zeta(a)^{-6}/224$.*
- (d) (The boundary $1 < a < 2$.) *For fixed $1 < a < 2$, $q_{2,a,n}/p_{a,n} \rightarrow 2(2-a)/(3-a) > 0$ and $\mathbb{P}(L_n \geq 2) \geq q_{2,a,n}$, so $\bar{r}_{a,n,V_n} = \Theta(p_{a,n})$: the residual of Theorem 4.1 is of the same order as the success probability and the conclusions of (c) do not follow from it.*
- (e) (Critical families.) *For fixed real β and $a_n = 2 - \beta/n$,*

$$2^n p_{a_n,n} \rightarrow C(\beta) = \frac{3(2^\beta - 1)}{\beta \pi^2} \quad (\beta \neq 0), \quad C(0) = \frac{3 \log 2}{\pi^2}, \quad \frac{q_{2,a_n,n}}{p_{a_n,n}} \sim \frac{2\beta 2^\beta}{n(2^\beta - 1)} \rightarrow 0, \quad (4.13)$$

with $2/(n \log 2)$ in place of the last expression at $\beta = 0$, and every conclusion of (c) holds along the family with lower prefactor $(6/\pi^2)^6$.

Part (e) at $\beta = 2$ is Theorem 4.2; the point of the theorem is that the exponential rarity scale 2^{-n} is shared by $a = 2$ and by every critical family, that the rarity order survives for every exponent, and that heavier tails fail the converse for an identifiable reason, not by accident of the constants.

Proposition 4.5 (Explicit witness reliability scales). *Let $\omega \in A_n$ with any mark configuration and $0 < \eta < 1$.*

- (a) *If $V \geq \max\{2n/d, (n/c) \log(24/\eta)\}$ then $q(\omega) \geq 1 - \eta$. At $\eta = 0.01$ the second term is $2.2416 \times 10^{50} n$, which is 8.966×10^{50} at $n = 4$, against 2.5×10^{61} under the envelope $V_*(4)$.*

- (b) *Choosing the six noise tolerances of Appendix A separately, at 99% of their permitted limits, the condition*

$$V \geq \max\{5.001 \times 10^{49}, 8.9 \times 10^{11} n\} \quad (4.14)$$

implies $q(\omega) \geq 0.99$. At $n = 4$ this is about eighteen times smaller than (a), the dominant term no longer depends on n , and it is set entirely by the absolute tolerance $(3 \times 10^{18})^{-1}/100$ on the selected product coordinate.

Both scales are pointwise reliability statements for the witness class; neither weakens the hypotheses of the converse or of the asymptotic theorems, and neither is a reactor design. The gap between (4.14) and the necessary scale of Theorem 4.3 is 42 orders of magnitude; Section 9 explains which tolerance is responsible and what a better estimate would have to do.

5 Proof of the finite theorem

We separate the source-independent trajectory argument, valid pointwise on A_n , from source averaging and from the converse. Throughout this section $\omega \in A_n$, the marks are arbitrary, and $x_z = N_z/V$ denotes a normalized count; $u = x_{00}$, $w = x_{11}$, $x = x_{0011}$, $m = M_{\text{nf}}/V$ and $L = M/V$.

5.1 Loss envelopes from the generator

The following derivation produces the constants 23 and 25 directly from the propensities (2.6); it is the source of the drift inequalities that the companion paper [2] states.

Lemma 5.1 (Loss envelopes). *On A_n and on the corridor $L \leq 11$, the coordinate drifts b_u , b_w , b_x of the count process, evaluated at any state and divided by V , satisfy*

$$b_u \geq 1 - u - 23Cu, \quad b_w \geq 1 - w - 23Cw, \quad b_x \geq \varepsilon uw + x(4uw - 1 - 25C), \quad C = 4\varepsilon + \frac{16}{3}m. \quad (5.1)$$

Proof. On A_n every food row is empty, so every active catalyst is nonfood and has length at least three; hence $\sum_{z \notin \mathcal{F}} x_z \leq m/3$. For any channel r , its basal coefficient is at most 4ε and each catalytic coefficient is at most 16, so the total coefficient of the channel, basal plus catalytic contributions weighted by catalyst concentration, is at most $4\varepsilon + 16 \sum_{z \notin \mathcal{F}} x_z \leq C$. Falling factorials are bounded by powers, so the same envelope bounds every count propensity in which the channel appears, including the cases where the catalyst coincides with a substrate.

A species y of normalized concentration y is consumed by ligation in either ordered input position: for each partner z and each of the two orderings the loss propensity over V is at most Cyx_z , and the total over partners is at most $2Cy \sum_z x_z \leq 22Cy$, since $\sum_z x_z \leq L \leq 11$. This count includes ligations of two identical copies, which remove two copies at a propensity bounded by Cy^2 , and it includes the selected channel when y is one of its substrates. Appearances as a catalyst do not consume the species. Cleavage of y removes it at one split position for a dimer and at three positions for a length-four word, at rate at most Cy and $3Cy$ respectively, and this includes the catalyzed reverse of the selected channel, whose catalyst 0011 is nonfood and therefore already inside C . Adding the feed rate 1 and the washout $-y$, and discarding every gain from internal reactions, gives $b_u \geq 1 - u - (22 + 1)Cu$, the same for w , and $b_x \geq -x - (22 + 3)Cx$ plus the retained gains. For x we keep two gains: the basal forward firing of r_* , whose reactants 00 and 11 are distinct so its propensity over V is exactly $b_{r_*}uw \geq \varepsilon uw$, and the self-catalyzed forward firing $00 + 11 + 0011 \rightarrow 2 \cdot 0011$, with three distinct inputs and propensity over V exactly $k_{r_*,0011}uw \geq 4uw$. $\square$

These are envelopes for the full catalogue, not equations for a closed three-species system: arbitrary nonfood background enters only through m , and the corridor bounds its aggregate catalytic influence. On the corridor $C < 59$, which gives the crude startup bounds $b_u \geq 1 - 1358u$, $b_w \geq 1 - 1358w$ and $b_x \geq \varepsilon uw - 1476x$ used in Lemma 5.5.

5.2 The potential

Let $h(s) = (1 - s)_+^2$, $H = h(u) + h(w)$, and

$$\Phi(u, w, x) = \log x - 4H. \quad (5.2)$$

Lemma 5.2 (Drift of the potential). *Let (u, w, x) be differentiable with $u, w \geq 0$, $x > 0$, and derivatives satisfying (5.1) at some $m \geq 0$. Then*

$$\dot{\Phi} \geq 2 - 117C = 2 - 468\varepsilon - 624m. \quad (5.3)$$

Proof. Put $\alpha = (1 - u)_+$ and $\beta = (1 - w)_+$. Since $u \geq 1 - \alpha \geq 0$ and $w \geq 1 - \beta \geq 0$, $uw \geq (1 - \alpha)(1 - \beta) \geq 1 - \alpha - \beta$, and $\alpha \leq 2\alpha^2 + \frac{1}{8}$ because $2(\alpha - \frac{1}{4})^2 \geq 0$; hence $uw \geq \frac{3}{4} - 2H$. Where $u < 1$, $\frac{d}{dt}h(u) = -2\alpha\dot{u} \leq -2\alpha(1 - u - 23Cu) = -2\alpha^2 + 46Cu\alpha \leq -2h(u) + \frac{23}{2}C$, using $u\alpha = u(1 - u) \leq \frac{1}{4}$; where $u \geq 1$ the derivative of $h(u)$ is zero and the same bound holds. Thus $\dot{H} \leq -2H + 23C$. Discarding the nonnegative basal term in b_x/x ,

$$\dot{\Phi} = \frac{\dot{x}}{x} - 4\dot{H} \geq 4uw - 1 - 25C + 8H - 92C \geq 3 - 1 - 117C. \quad \square$$

Remark 5.3 (A sharper elementary inequality). The inequality $uw \geq \frac{3}{4} - 2H$ can be replaced by $uw \geq \frac{4}{5} - 2H$: with α, β as above,

$$(1 - \alpha)(1 - \beta) + 2(\alpha^2 + \beta^2) - \frac{4}{5} = \frac{5}{4}(\alpha + \beta - \frac{2}{5})^2 + \frac{3}{4}(\alpha - \beta)^2 \geq 0,$$

with equality at $u = w = \frac{4}{5}$, so $\dot{\Phi} \geq \frac{11}{5} - 117C$. This is a genuine improvement of (5.3), but the compensated estimate below uses perturbed constants that were established for the weaker bound, and we do not re-derive them here. The improvement would shorten the minimal window length (Section 9) but does nothing for the count scale, which is set by a tolerance and not by the drift.

5.3 The count process: compensated coordinates and the interval telescope

The jump process has holding intervals $[t_k, t_{k+1})$ on which the state $N^{(k)}$ is constant. For a coordinate y write its compensator path $\tilde{y}(t) = y(0) + \int_0^t b_y(N(s)) ds$, a continuous piecewise-linear function whose slope on the k -th holding interval is the drift $b_y(N^{(k)})$, and its noise $y(t) - \tilde{y}(t)$, a martingale. The *corrected potential* $\Phi_{\text{corr}}(t) = \Phi(\tilde{u}(t), \tilde{w}(t), \tilde{x}(t))$ is evaluated along the compensator paths, so it is continuous across jumps and differentiable inside holding intervals. A *coordinate noise bound* with tolerance η on horizon T is the statement that $|y(t) - \tilde{y}(t)| \leq \eta$ for all $t < T$ up to the first exit from the corridor; a *mass noise bound* and *reward noise bounds* for the export and gross-feed counters are defined in the same way, with the reward compensators $\int m dt$ and $6t$. The tolerances used are

$$\eta_M = \frac{1}{4}, \quad \eta_F = 10^{-5} (u, w), \quad \eta_P = \frac{1}{3 \times 10^{20}} (x), \quad \eta_E = \frac{1}{40} (Q/V), \quad \eta_B = 1 (A/V). \quad (5.4)$$

The product tolerance is one percent of the stock floor $(3 \times 10^{18})^{-1}$; this ratio, not the drift constants, is what later sets the count scale.

Lemma 5.4 (Mass localization). *From the food-only state, on the mass noise bound with tolerance $\frac{1}{4}$ on horizon $T = 200$, $M(t) \leq 11V$ for all $t < 200$, and $M(t) \leq \frac{21}{2}V$ at the three observation times.*

Proof. The mass drift is $b_M/V = 10 - L$, so $\tilde{L}(t) = 10$ for all t from $L(0) = 10$, and $|L - \tilde{L}| \leq \frac{1}{4}$ gives $L \leq 10.25$. The tolerance $\frac{1}{4}$ is composed from a nonfood-mass tolerance $\frac{1}{8}$ and a tolerance $\frac{1}{80}$ on each of the six food coordinates, whose mass weights sum to 10; this composition is the origin of the coefficients 2 and 12 in the first two terms of (5.10). $\square$

Lemma 5.5 (Startup floor). *Let $\omega \in A_n$ with any marks, start from the food-only state, and suppose that on horizon $T \leq 200$ the mass noise bound and the three coordinate noise bounds (5.4) for u , w and x hold. Then at every time $1 \leq t < T$,*

$$u, w \geq \frac{1}{1358} - 2 \times 10^{-5}, \quad x \geq (3 \times 10^{18})^{-1}. \quad (5.5)$$

Proof sketch; the checked statement is `physical_startup_product_from_noise`. By Lemma 5.4 the corridor holds on the horizon, so the crude bounds $b_u \geq 1 - 1358u$ apply on every holding interval. The compensator $\tilde{u}$ then satisfies $\tilde{u}(t) \geq \frac{1}{1358}$ for all t from $\tilde{u}(0) = 1$, and the noise bound transfers this to u up to η_F ; the same holds for w . With $u, w \geq \frac{1}{1358} - 2 \cdot 10^{-5}$, the product drift satisfies $b_x \geq \varepsilon uw - 1476x \geq 6.9 \times 10^{-19} - 1476x$, so $\tilde{x}(t) \geq \frac{6.9 \times 10^{-19}}{1476}(1 - e^{-1476t})$, which exceeds $2 \cdot (3 \times 10^{18})^{-1}$ at $t = 1$ and remains above it afterwards because the drift is positive below that level; the product noise η_P is one percent of the floor. $\square$

Lemma 5.6 (Gain on a holding interval). *Under the hypotheses of Lemma 5.5, on a holding interval with state N of mass at most $11V$ and $x \geq (3 \times 10^{18})^{-1}$, for $0 \leq a \leq b$ inside the interval and before T ,*

$$\Phi_{\text{corr}}(b) - \Phi_{\text{corr}}(a) \geq (b - a)(1.9 - 480\varepsilon - 640m(N)). \quad (5.6)$$

Proof sketch; the checked statement is `physical_holding_potential_gain`. Inside the interval the compensator paths are linear with slopes given by the drifts at N , which satisfy (5.1) in terms of the actual concentrations at N . The corrected coordinates differ from the actual ones by at most η_F , η_F and η_P , and η_P is at most one percent of x . Lemma 5.2 applied to the compensator paths with these perturbations, and with $\frac{3}{4} - 2H$ evaluated at the corrected foods, yields (5.3) with the constants 2, 468, 624 degraded to 1.9, 480, 640. $\square$

Lemma 5.7 (Endpoint penalty). *Under the same hypotheses, if $x \geq (3 \times 10^{18})^{-1}$ at the two times $a \leq b < T$ and the mass at time b is at most $11V$, then $\Phi_{\text{corr}}(b) - \Phi_{\text{corr}}(a) < 54$.*

Proof. At time b , $4x \leq L \leq 11$ gives $\tilde{x} \leq \frac{11}{4} + \eta_P$ and $H \geq 0$, so $\Phi_{\text{corr}}(b) \leq \log(\frac{11}{4} + \eta_P) < 1.02$. At time a , $\tilde{x} \geq (3 \times 10^{18})^{-1} - \eta_P$ and each corrected food is at least $-\eta_F$, so $H \leq 2(1 + \eta_F)^2$ and $\Phi_{\text{corr}}(a) > \log(0.99 \cdot (3 \times 10^{18})^{-1}) - 8.0002 > -50.56$. The difference is below 51.6; the constant 54 is the rounded value used in the checked statement `physical_potential_endpoint_penalty`. $\square$

Proposition 5.8 (Interval occupation). *Under the hypotheses of Lemma 5.5 with $T = 200$, for every $1 \leq a \leq b < 200$,*

$$\int_a^b m(t) dt \geq \frac{(b - a)(1.9 - 480\varepsilon) - 54}{640}. \quad (5.7)$$

Proof. Let J and $J + K$ index the holding intervals containing a and b . By Lemma 5.4 the corridor holds up to time b , and by Lemma 5.5 the floor $x \geq (3 \times 10^{18})^{-1}$ holds at every jump time in $[a, b]$ and at a and b , since all these times are at least 1. Apply Lemma 5.6 to the partial interval $[a, t_{J+1})$, to each full holding interval between, and to the partial interval $[t_{J+K}, b]$; the corrected potential is

continuous at jumps, so the increments telescope to $\Phi_{\text{corr}}(b) - \Phi_{\text{corr}}(a)$, while the right-hand sides sum to $(b-a)(1.9 - 480\varepsilon) - 640 \int_a^b m \, dt$ because m is constant on each holding interval. Lemma 5.7 bounds the left-hand side by 54. The checked assembly is `physical_interval_mass_from_noise`. $\square$

This is the whole content of the extension from one window to two: the estimate is stated for an arbitrary actual interval of the same trajectory, with both partial holding intervals retained, and no new initial law is introduced at the middle observation time.

Corollary 5.9 (Export in a window). *Under the hypotheses of Proposition 5.8 and the export reward noise bound with tolerance $\frac{1}{40}$,*

$$\frac{Q(a, b]}{V} \geq \frac{(b-a)(1.9 - 480\varepsilon) - 54}{640} - \frac{1}{20}, \quad (5.8)$$

which for $b - a = 99$ equals $0.1595311015 > 0.1$. The right-hand side exceeds $\frac{1}{10}$ exactly when $b - a \geq 79$; a window of length 78 does not satisfy the sufficient inequality.

Proof. The compensator of Q/V is $\int m \, dt$, so $Q(a, b]/V$ differs from $\int_a^b m \, dt$ by the difference of two prefix noises, each at most $\frac{1}{40}$ (`physical_interval_export_lower`). The numerical values are $(99(1.9 - 480\varepsilon) - 54)/640 = 0.2095311015$ and $150/(1.9 - 480\varepsilon) = 78.95$. $\square$

Applied to $[1, 100]$ and $[100, 199]$ this gives both windows. Lemma 5.5 gives the stock floor at 1, 100 and 199 and Lemma 5.4 the mass bound $\frac{21}{2}V$ there, so endpoint stock holds at all three times; the corridor holds through 199; and the gross-feed compensator is exactly $6t$, so the reward noise bound with tolerance 1 gives $A(199)/V \leq 6 \cdot 199 + 1 = 1195$. The event $F_{n,V}$ therefore holds on the intersection of the six noise bounds.

5.4 One simultaneous probability estimate

Lemma 5.10 (Six-term failure bound). *Let $\omega \in A_n$ with any marks and let V be an integer. For every choice of positive tolerances $d_N, d_M, d_F, d_P, d_E, d_B$ with*

$$\begin{aligned} d_N + \frac{n}{V} &\leq \frac{1}{8}, & d_M + \frac{2}{V} &\leq \frac{1}{80}, & d_F + \frac{2}{V} &\leq 10^{-5}, \\ d_P + \frac{2}{V} &\leq \frac{1}{3 \times 10^{20}}, & d_E + \frac{n}{V} &\leq \frac{1}{40}, & d_B + \frac{n}{V} &\leq 1, \end{aligned} \quad (5.9)$$

the food-only trajectory law satisfies

$$\begin{aligned} \mathbb{P}_\omega(F_{n,V}^c) &\leq 2e^{-E_N(d_N)} + 12e^{-E_C(d_M)} + 4e^{-E_C(d_F)} \\ &\quad + 2e^{-E_C(d_P)} + 2e^{-E_N(d_E)} + 2e^{-E_N(d_B)}, \end{aligned} \quad (5.10)$$

$$E_N(s) = \frac{s^2 V}{4n(96011 \cdot 200 + s)}, \quad E_C(s) = \frac{s^2 V}{4(96000 \cdot 200 + 2s)}. \quad (5.11)$$

Proof. By Section 5.3, $F_{n,V}^c$ is contained in the union of the six noise-bound failures, and the union bound applies; no independence is used. Each failure probability is an exponential martingale tail bound of Freedman type [9] for the compensated jump process stopped at corridor exit: the jump sizes of a normalized coordinate are at most $2/V$ and of the mass and reward counters at most n/V , the predictable quadratic variation on the horizon 200 is bounded on the corridor by the total rate envelope 96000 per unit time for a coordinate and 96011 for the mass and rewards, and the margin conditions (5.9) absorb the jump-size correction so that a noise bound with the tolerance of (5.4)

follows from the centred tail with parameter d . The twelve food terms arise from the composition in Lemma 5.4. The checked statement is `physical_two_window_failure_tail`, whose hypotheses are exactly: empty food rows, the selected incidence with distinct dimer inputs and length-four product, the coefficient bounds (2.5), the food-only initial state, and (5.9). $\square$

Setting all six tolerances equal to d and requiring $V \geq 2n/d$ verifies (5.9) for $n \geq 4$, and every exponent is then at least cV/n because $4(96011 \cdot 200 + d) \leq 8 \times 10^7$ and $4(96000 \cdot 200 + 2d) \leq 8 \times 10^7$; this gives $\mathbb{P}_\omega(F_{n,V}^c) \leq 24e^{-cV/n} = \delta_{n,V}$, which is (4.4) once $V \geq V_*(n) \geq 2n/d$. The choice is convenient rather than optimal; Proposition 4.5(b) exploits the freedom. Under $V \geq V_*(n)$, $cV/n \geq 48$ and $24e^{-48} < \frac{1}{2}$, so the lower bound in (4.5) is positive.

5.5 Source averaging

The bound (4.4) holds for every $\omega \in A_n$ and every mark configuration. Averaging first over the trajectory law, then over the uniform mark law, then over the source restricted to A_n , and using the exact mass (4.1), gives $\mathbb{P}(F_{n,V} \cap A_n) \geq e_a^6 p_{a,n} (1 - \delta_{n,V})$. The exact source mass is computed from the configuration weight without any approximation: the six food rows contribute $\mathbb{P}(D = 0)^6$ and the selected row contributes the probability that a uniformly random D -subset contains r_* , averaged over D , which is $\mathbb{E}D/R_n$ (`source_food_silent_seed_mass`). This proves (4.5).

5.6 The converse and the productive census

Lemma 5.11 (Inherited one-window exclusion [1]). *Let ω be an environment, with arbitrary marks and arbitrary assignments outside the K_* -rectangle, such that at most one incidence lies in the rectangle and no productive incidence is assigned. For every integer $V \geq 2n/d$ with $n/V \leq 1/200000$, from the food-only state,*

$$\mathbb{P}_\omega(F^{\text{old}}) \leq 4e^{-c_{\text{old}}V/n} + 2e^{-E'(1/200)} + 2e^{-E'(1/200000)} \leq 8e^{-c_{\text{old}}V/n}, \quad E'(s) = \frac{s^2V}{4n(96011 \cdot 101 + s)}. \quad (5.12)$$

The first inequality is `physical_output_upper_at_most_one_local`; its mechanism is a conserved product credit that cancels a mixed food–nonfood channel exactly, and a washout penalty that lets the dilution of an outsider catalyst pay for the material it could amplify. The second inequality holds for all n and V because both extra exponents exceed $c_{\text{old}}V/n$ by more than thirty orders of magnitude.

Lemma 5.12 (Productive census). *For every $n \geq 4$ there are exactly 224 productive incidences: 32 food-to-nonfood split identities, each with six food catalysts and its own product as the seventh, hence 192 food-catalyzed and 32 product-self-catalyzed labels. The event T_n is exactly the union over these labels of the events $\{r \in C_z\}$, and*

$$\mathbb{P}_{a,n}(T_n) \leq 224 p_{a,n}. \quad (5.13)$$

Proof. A productive identity has food inputs and nonfood product, so its product has length three or four: $2 \cdot 4 = 8$ monomer–dimer ordered pairs, 8 dimer–monomer pairs, and $4 \cdot 4 = 16$ dimer–dimer pairs, in total 32; the ordered convention counts (0, 01, 001) and (00, 1, 001) separately. For each identity the seven admissible catalysts are distinct choices, so there are $32 \cdot 7 = 224$ labels, all with catalyst length at most 4 and product length at most 4; the same short words occur in every catalogue with $n \geq 4$, which proves the count for all n . The union bound (5.13) uses only the

common marginal $p_{a,n}$ and no independence. The cap-four count, the equivalence of the predicate with the union event, and an injection of the labels of every larger catalogue into those at cap four are checked in `ProductiveCensus.lean`; the reverse inclusion is the sentence above. $\square$

Proof of (4.6) and (4.7). By (3.3), $\mathbb{P}(F_{n,V} \cap T_n^c) \leq \mathbb{P}(F^{\text{old}} \cap T_n^c)$. Split the environments in T_n^c according to whether $L_n \geq 2$. The first part has source probability $\mathbb{P}_{a,n}(L_n \geq 2)$; on the second, Lemma 5.11 applies, since $V \geq V_*(n)$ implies both of its scale hypotheses. This gives (4.7), and $S_{a,n,V} \leq \mathbb{P}(T_n) + \mathbb{P}(F_{n,V} \cap T_n^c)$ with Lemma 5.12 gives (4.6). The formal root proves the same inequalities with the containing rectangle of all catalysts of length at most 4 and all identities of product length at most 6, whose size is $30 \cdot 516 = 15480$. $\square$

5.7 The accounts

The statements about net synthesis and the recovered fraction are Lemma 3.2; the pathwise identity (3.4) is `actual_net_synthesis_identity` and the scalar inequality (3.5) is `mission_material_recovery`. This completes the proof of Theorem 4.1.

6 Limits and consequences

6.1 Proof of Theorem 4.2

The inherited exact-source limits for the capped, shifted sampler at $a_n = 2 - 2/n$ are

$$e_n \rightarrow \frac{6}{\pi^2}, \quad X_n p_n \rightarrow \frac{9}{\pi^2}, \quad \frac{\bar{r}_n}{p_n} \rightarrow 0, \quad \delta_{n,V_n} \rightarrow 0 \quad (6.1)$$

(`collective_empty_row_limit`, `sourceExactCriticalFirstMoment_scaled_tendsto`, `mission_residual_relative_tendsto_zero`, `two_window_noise_tendsto_zero`). The third uses the excess-count estimate of [1] for two incidences in a fixed rectangle, which retains the dependence within a row, together with $e^{-c_{\text{old}} V_n/n} \leq e^{-3n} = o(p_n)$. Dividing (4.5) and (4.6) by p_n proves (4.8). Next,

$$\mathbb{P}(A_n \mid F_{n,V_n}) = \frac{\mathbb{P}(F_{n,V_n} \cap A_n)}{S_n} \geq \frac{e_n^6(1 - \delta_{n,V_n})}{224 + \bar{r}_n/p_n}, \quad (6.2)$$

which gives the first part of (4.9); dividing (4.7) by the positive lower bound in (4.5) gives the second. Finally $X_n/2^n \rightarrow 2$ gives (4.10). The formal root proves the eventual bounds with the coefficients $\gamma/2$ and 15481 and the vanishing conditional probability of no productive singleton; the full constants are the conventional sharpening just given. $\square$

6.2 Screening

For N independent complete experiments, each with fresh chemistry, marks, preparation and clock, the probability of at least one success is $1 - (1 - S_n)^N$, and the least N reaching confidence α is

$$N_\alpha = \left\lceil \frac{\log(1 - \alpha)}{\log(1 - S_n)} \right\rceil = \Theta(2^n), \quad (6.3)$$

by $-\log(1 - s) \sim s$ and (4.10). A finite lower bound on S_n gives a sufficient trial count and a finite upper bound below one a necessary one. Repeating N runs in one sampled environment is a different experiment: its success probability is the mixture $1 - \mathbb{E}_\mu[(1 - q(\omega))^N]$, and replacing it by $1 - (1 - \mathbb{E}_\mu q)^N$ is unjustified, because q is far from constant across environments.

6.3 Reliable environments

If $0 < \eta < 1$ and $\delta_{n,V} \leq \eta$, then

$$e_a^6 p_{a,n} \leq \mu_{a,n} \{q \geq 1 - \eta\} \leq \frac{S_{a,n,V}}{1 - \eta}. \quad (6.4)$$

The lower inclusion is (4.4); the upper bound integrates $q \geq (1 - \eta)\mathbf{1}_{\{q \geq 1 - \eta\}}$. Along the critical sequence the reliable population is therefore $\Theta(p_n)$ with liminf prefactor at least γ and limsup at most $224/(1 - \eta)$. The full law of the reliable population is not identified.

6.4 Transfers that preserve the actual event

If a source or path property B_n satisfies $\mathbb{P}(F^{\text{old}} \cap B_n^c) = o(p_n)$, then by (3.3) and (4.8)

$$\mathbb{P}(B_n^c \mid F_{n,V_n}) \leq \frac{\mathbb{P}(F^{\text{old}} \cap B_n^c)}{S_n} \rightarrow 0. \quad (6.5)$$

Four instances follow.

Unique productive incidence. Every productive label lies in the K_* -rectangle, so two productive incidences force $L_n \geq 2$, an event of source probability $o(p_n)$. Hence, conditional on success, exactly one productive incidence is present with probability tending to one. This says nothing about the relative frequencies of the 224 labels.

Catalytic deletion. Delete every catalytic channel while keeping the basal coefficients, preparation, feed and washout. The companion theorem [2] bounds the disabled success probability of F^{old} by $2e^{-c_{\text{old}}V/n}$ in every environment; by (3.3) the same bounds $F_{n,V}$, and after averaging its ratio to S_n tends to zero. Positive basal chemistry remains, so the disabled probability is not declared zero.

Nesting of the two events. The census converse also bounds $\mathbb{P}(F^{\text{old}}) \leq 224p_n + \bar{r}_n$, so

$$\liminf_n \mathbb{P}(F_{n,V_n} \mid F^{\text{old}}) \geq \frac{\gamma}{224}. \quad (6.6)$$

The stronger task keeps the rarity order of the weaker one; this does not say that most first-window successes repeat.

Conditioning on structure. Every productive label is a singleton RAF in the reversible convention, so $T_n \subseteq G_n$, the event that some RAF exists. The companion structural theorem gives $\mathbb{P}(G_n) \rightarrow \theta > 0$ as a genuine limit (`source_RAF_probability_tendsto`, positive by `sourceCriticalSurvival_pos`). Since $F_{n,V_n} \cap G_n^c \subseteq F_{n,V_n} \cap T_n^c$, (4.7) gives $\mathbb{P}(F_{n,V_n} \cap G_n) = S_n - o(p_n)$, hence $\mathbb{P}(F_{n,V_n} \mid G_n) = \Theta(p_n)$ and $\mathbb{P}(F_{n,V_n} \mid G_n)/S_n \rightarrow \theta^{-1}$. No independent-incidence structural theorem is used.

7 The necessary discrete startup scale

The concentration-level argument of Section 5 says nothing about initiation at small counts. The following bound applies to every food-silent environment and every mark configuration, without a limit in n or V .

Proof of Theorem 4.3. Before τ every nonfood molecule has zero count and every food molecule has an empty row, so all catalytic propensities vanish and a first nonfood molecule can only be made by a basal ligation of two foods. Put $x = \sum_{|z|=1} x_z$ and $y = \sum_{|z|=2} x_z$. The ordered food pairs whose ligation is nonfood are monomer–dimer, dimer–monomer and dimer–dimer, with ordinary

reactant-product sum $2xy + y^2$; falling factorials are bounded by powers, including for identical dimers, and each basal coefficient is at most 4ε . On the corridor $x + 2y \leq 11$,

$$\lambda_{\text{birth}}(N) \leq 4\varepsilon V(2xy + y^2) \leq 4\varepsilon V(22y - 3y^2) \leq \frac{484}{3}\varepsilon V = \lambda V, \quad (7.1)$$

the maximum being at $y = \frac{11}{3}$, $x = \frac{11}{3}$ (`food_pair_birth_bound`, `startup_intensity_bound`).

Stop the process at $\tau \wedge \sigma$. Before either stopping event only finitely many population states are reachable. Send a corridor exit to a safe absorbing state and a first birth to a distinct absorbing failure state; a birth preserves total monomer mass, so it cannot coincide with a corridor exit. Uniformize the resulting finite chain at a rate q at least its maximal total rate and at least λV : the chain then moves at the ticks of a Poisson process of rate q , and in every nonfailure state the one-tick failure probability is at most $\lambda V/q$. By induction over ticks, conditional survival through k ticks is at least $(1 - \lambda V/q)^k$ whatever the intervening states (`startup_survival_lower`). Averaging over the independent Poisson tick count,

$$\sum_{k \geq 0} e^{-qt} \frac{(qt)^k}{k!} \left(1 - \frac{\lambda V}{q}\right)^k = e^{-\lambda V t}. \quad (7.2)$$

This is a domination argument; the state-dependent birth count itself is not Poisson. The stopped chain agrees with the original path until the first stopping event, which proves (4.11). The event $F_{n,V}$ requires a length-four molecule at time 1 and no corridor exit through 199, so it is contained in $\{\tau \leq 1, \tau < \sigma\}$, and $\mathbb{P}_\omega(F_{n,V}) \leq 1 - e^{-\lambda V}$ follows. Finally $1 - \eta \leq 1 - e^{-\lambda V}$ is equivalent to $\lambda V \geq \log(1/\eta)$ (`startup_necessary_scale`), and $\log(100)/\lambda = 14\,272\,221.65$. $\square$

At $V = 100$ the food-silent mission probability is at most 3.2266×10^{-5} , at $V = 200$ at most 6.4531×10^{-5} , and a birth before time 199 and before corridor exit has probability at most 0.0064 at $V = 100$. These are upper bounds and necessary conditions, not sufficient startup estimates: the first nonfood molecule need not be the selected product, a seed can wash out, and growth to operating stock is a further event. The restriction applies to food-silent environments, the coefficients (2.5), the food-only preparation and the startup deadline 1; a food-catalyzed productive incidence can initiate immediately, and changing ε , the deadline or the preparation changes the question. Figure 2 shows the ceiling.

8 Robustness in the Zipf exponent

The finite Theorem 4.1 holds for every $a > 1$. This section computes how its constants scale and proves Theorem 4.4; the first-moment limit along the families of part (e) is formally verified for every β (`windowZipfMean_source_window_normalized`), and everything else here is a conventional deduction.

8.1 Exact moments and fixed-exponent asymptotics

Write $T_R = \sum_{k \geq R} k^{-a}$ and $R = R_n$. From (2.3),

$$\mathbb{E}D = \frac{\sum_{k=1}^{R-1} k^{1-a} + RT_R}{\zeta(a)} - 1, \quad \mathbb{E}[D(D-1)] = \frac{\sum_{k=1}^{R-1} (k-1)(k-2)k^{-a} + (R-1)(R-2)T_R}{\zeta(a)}. \quad (8.1)$$

The first form follows from $\mathbb{E}D = \zeta(a)^{-1}[\sum_{k < R} (k-1)k^{-a} + (R-1)T_R]$ by writing $(k-1)k^{-a} = k^{1-a} - k^{-a}$ and $\sum_{k < R} k^{-a} = \zeta(a) - T_R$. Both are evaluated without enumeration through Hurwitz zeta values.

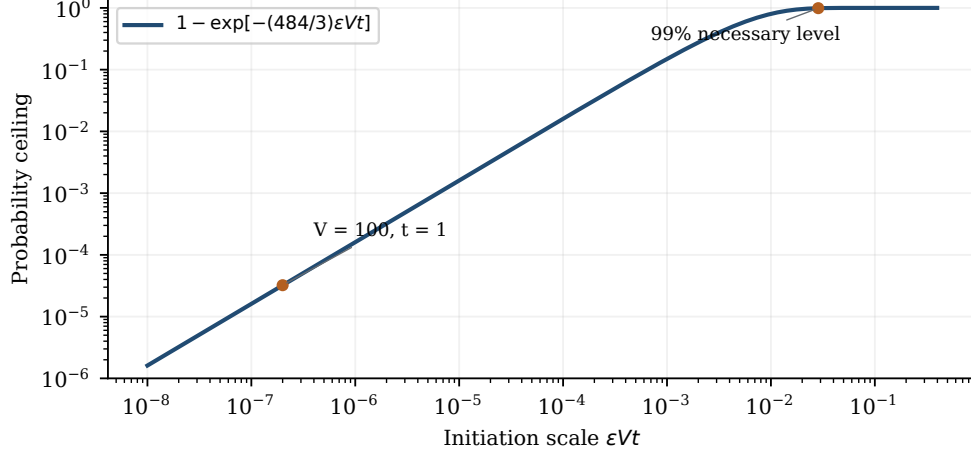


Figure 2: The first-birth ceiling $1 - \exp[-(484/3)\varepsilon Vt]$ as a function of εVt . For the mission $t = 1$, and the bound applies to every food-silent environment and every mark configuration. The marked 99% level is a necessary initiation scale, not a sufficient operating design.

Exponent	$p_{a,n}$	$q_{2,a,n}/p_{a,n}$	singleton dominance, limiting prefactors
$1 < a < 2$	$\Theta(n^{1-a}2^{-(a-1)n})$	$\rightarrow 2(2-a)/(3-a) > 0$	not by the present converse
$a = 2$	$\Theta(2^{-n})$, $2^n p \rightarrow 3 \log 2 / \pi^2$	$\sim 2 / \log R \rightarrow 0$	yes, $[\zeta(2)^{-6}, 224]$
$a_n = 2 - \beta/n$	$\Theta(2^{-n})$, $2^n p \rightarrow C(\beta)$	$\sim 2\beta 2^\beta / [n(2^\beta - 1)] \rightarrow 0$	yes, $[(6/\pi^2)^6, 224]$
$a > 2$	$\Theta(n^{-1}2^{-n})$	$O(R^{2-a} \log R) \rightarrow 0$	yes, $[\zeta(a)^{-6}, 224]$

Table 1: Orders of the incidence probability and of the same-row pair ratio, and the status of singleton dominance, for the volume sequence $V_n \geq V_*(n)$ (Theorem 4.4). In every row the success probability is $\Theta(p_{a,n})$.

Proof of Theorem 4.4(a). For $0 < s < 1$, $\sum_{k < R} k^{-s} = R^{1-s}/(1-s) + O(1)$; for $s = 1$ the sum is $\log R + O(1)$; for $s > 1$ it converges to $\zeta(s)$; and $T_R = R^{1-a}/(a-1) + O(R^{-a})$. For $1 < a < 2$ the two growing terms of $\mathbb{E}D$ are $R^{2-a}/(2-a)$ and $R^{2-a}/(a-1)$, whose sum is $R^{2-a}/[(a-1)(2-a)]$. For $a = 2$ the finite sum is harmonic and $RT_R \rightarrow 1$. For $a > 2$ dominated convergence gives $\mathbb{E}D \rightarrow \zeta(a-1)/\zeta(a) - 1$. Dividing by $R \sim 2n2^n$ gives (4.12); at $a = 2$, $\log R \sim n \log 2$ and $2^n p_{2,n} \rightarrow \log 2 / (2\zeta(2)) = 3 \log 2 / \pi^2$. $\square$

Figure 3 shows the exact values of $2^n p_{a,n}$ for several exponents; Table 1 summarizes the orders.

8.2 The rarity order for every exponent

Proof of Theorem 4.4(b). Let $H_K = (2^{K+1} - 2)(K2^{K+3} + 4)$ with $K = K_*$: the number of catalyst-identity pairs in the full K_* -rectangle, an enormous but fixed constant. Since L_n is a sum of at most H_K indicators of common marginal $p_{a,n}$, Markov's inequality gives $\mathbb{P}(L_n \geq 2) \leq \frac{1}{2} \mathbb{E}L_n \leq \frac{1}{2} H_K p_{a,n}$, without any independence. Theorem 4.1 then gives, for $V_n \geq V_*(n)$,

$$\frac{1}{2} c_a^6 p_{a,n} \leq S_{a,n} \leq (224 + \frac{1}{2} H_K) p_{a,n} + 8e^{-c_{\text{old}} V_n / n}, \quad (8.2)$$

using $\delta_{n,V_n} \leq \frac{1}{2}$. Under the envelope, $c_{\text{old}} V_n / n \geq 3n$, so the last term is at most $8e^{-3n}$, which is $o(p_{a,n})$ in every case of (4.12) because $(a-1) \log 2 < 3$. Hence $S_{a,n} = \Theta(p_{a,n})$, with $\liminf S_{a,n}/p_{a,n} \geq \zeta(a)^{-6}$ since $\delta_{n,V_n} \rightarrow 0$. The reliable-population statement is (6.4), valid once $\delta_{n,V_n} \leq \eta$. $\square$

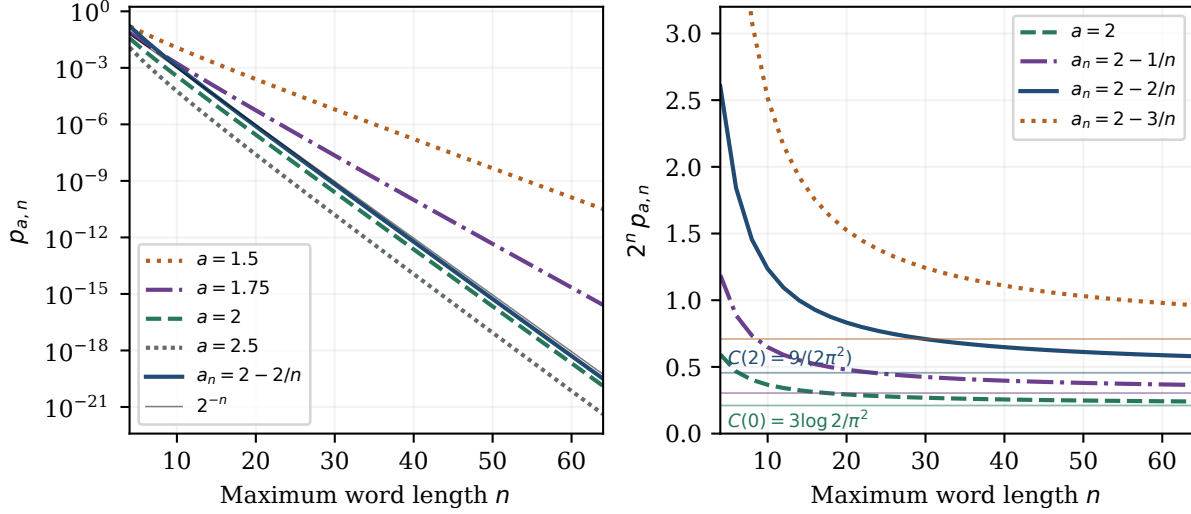


Figure 3: Exact incidence probabilities from (8.1). Left: $p_{a,n}$ for fixed exponents and for $a_n = 2 - 2/n$ against 2^{-n} ; the rarity exponent is $(a - 1) \log 2$ below $a = 2$ and $\log 2$ at and above it. Right: $2^n p_{a,n}$ for $a = 2$ and for the families $a_n = 2 - \beta/n$, $\beta = 1, 2, 3$, with the limits $C(\beta)$ of (4.13) as horizontal lines; the approach is of relative order $1/n$. The curves are source moments, not success probabilities.

The implied upper constant in (8.2) is useless as a finite number; the content of the statement is that the source-to-operation comparison, success of order the incidence probability, survives for every exponent while the exponent itself changes in the computable way (4.12).

8.3 Second moments: dominance for $a \geq 2$ and the boundary

Proof of Theorem 4.4(c) and (d). Two distinct incidences in the rectangle either share a catalyst, with joint probability $q_{2,a,n}$, or lie in different rows, with joint probability $p_{a,n}^2$; a union bound over the fixed rectangle gives

$$\mathbb{P}(L_n \geq 2) \leq C_{\text{same}} q_{2,a,n} + C_{\text{diff}} p_{a,n}^2 \quad (8.3)$$

with constants depending only on the rectangle. From (8.1), expanding $(k - 1)(k - 2) = k^2 - 3k + 2$: for $1 < a < 2$ the growing terms of $\mathbb{E}[D(D - 1)]$ are $R^{3-a}/(3 - a)$ from the sum and $R^{3-a}/(a - 1)$ from the atom, so

$$q_{2,a,n} \sim \frac{2R^{1-a}}{(a - 1)(3 - a)\zeta(a)}, \quad \frac{q_{2,a,n}}{p_{a,n}} \rightarrow \frac{2(2 - a)}{3 - a};$$

for $a = 2$ both terms are $\sim R$, so $q_{2,2,n} \sim 2/(\zeta(2)R)$ and $q_2/p \sim 2/\log R \rightarrow 0$; for $a > 2$ the numerator grows at most like R^{3-a} , $\log R$ or $O(1)$, so $q_2/p = O(R^{2-a} \log R) \rightarrow 0$. In every case $p_{a,n} \rightarrow 0$. Hence for $a \geq 2$ the right-hand side of (8.3) is $o(p_{a,n})$, so $\bar{r}_{a,n,V_n}/p_{a,n} \rightarrow 0$, and the arguments of Section 6 apply verbatim with $e_a = \zeta(a)^{-1}$ in place of e_n : this is (c). For $1 < a < 2$, $\mathbb{P}(L_n \geq 2) \geq q_{2,a,n}$ because two fixed productive labels with a common catalyst lie in the rectangle, and $q_{2,a,n} = \Theta(p_{a,n})$ by the limit just computed: this is (d). $\square$

Part (d) is a substantive boundary rather than a technical inconvenience. For a heavy-tailed row, conditioning on one required incidence leaves a nonvanishing probability of a second prescribed incidence in the same row, so a successful incidence typically arrives with a substantial local catalytic

Term of (5.10)	tolerance	sufficient V at $\eta = 0.01$	deterministic margin
$2e^{-E_N(d_N)}$ (nonfood mass)	0.99/8	$3.556 \times 10^{10} n$	$800n$
$12e^{-E_C(d_M)}$ (six foods, mass)	0.99/80	4.454×10^{12}	16000
$4e^{-E_C(d_F)}$ (selected foods)	0.99×10^{-5}	6.099×10^{18}	2×10^7
$2e^{-E_C(d_P)}$ (selected product)	$0.99/(3 \times 10^{20})$	5.000×10^{49}	6×10^{22}
$2e^{-E_N(d_E)}$ (export)	0.99/40	$8.890 \times 10^{11} n$	$4000n$
$2e^{-E_N(d_B)}$ (gross feed)	0.99	$5.556 \times 10^8 n$	$100n$

Table 2: Unequal-tolerance evaluation of the six-term bound at 99% witness reliability. The product term dominates by thirty orders of magnitude; the n -dependent terms would take over only beyond $n \approx 5.6 \times 10^{37}$.

background, and a converse must weigh that background dynamically rather than exclude it by counting. This does not prove that singleton dominance fails after conditioning on successful dynamics for $1 < a < 2$; it identifies the additional estimate that would be needed.

8.4 Critical families

Proof of Theorem 4.4(e). Along $a_n = 2 - \beta/n$, $\log R/n \rightarrow \log 2$ and $R^{\beta/n} \rightarrow 2^\beta$. In (8.1), $\sum_{k < R} k^{-1+\beta/n} = n(R^{\beta/n} - 1)/\beta + O(1) \sim n(2^\beta - 1)/\beta$ for $\beta \neq 0$ and $\sim n \log 2$ for $\beta = 0$, while $RT_R \rightarrow 2^\beta$ is bounded; so $\mathbb{E}D \sim n(2^\beta - 1)/(\beta\zeta(2))$, and dividing by $R \sim 2n2^n$ gives $C(\beta)$, with $C(0)$ by continuity. This limit is the formally verified statement that $\mathbb{E}D/n \rightarrow (1 - 2^{-b})/(\beta\zeta(2))$ at exponent $2 + b/n$, with $b = -\beta$. For the second moment, $\sum_{k < R} k^{\beta/n} \sim R2^\beta$ and $R^2T_R \sim R2^\beta$, so $\mathbb{E}[D(D-1)] \sim 2 \cdot 2^\beta R/\zeta(2)$ and

$$\frac{q_{2,a_n,n}}{p_{a_n,n}} \sim \frac{2 \cdot 2^\beta / (\zeta(2)R)}{n(2^\beta - 1)/(\beta\zeta(2)R)} = \frac{2\beta 2^\beta}{n(2^\beta - 1)} \rightarrow 0,$$

and $2/(n \log 2)$ at $\beta = 0$. The rest is (8.3) and Section 6, with $e_{a_n} \rightarrow 6/\pi^2$ for every β . The moment estimates are established directly along the sequence from the capped sums; fixed-exponent asymptotics are not invoked as if they were uniform. $\square$

9 Explicit scales and the converse obstruction

9.1 Proof of Proposition 4.5

(a) The uniform bound $\mathbb{P}_\omega(F_{n,V}^c) \leq 24e^{-cV/n}$ was proved in Section 5.4 under $V \geq 2n/d$ alone, so $24e^{-cV/n} \leq \eta$, that is $V \geq (n/c) \log(24/\eta)$, suffices. Numerically $\log(2400)/c = 2.2416 \times 10^{50}$.

(b) Take $(d_N, d_M, d_F, d_P, d_E, d_B) = 0.99 \cdot (\frac{1}{8}, \frac{1}{80}, 10^{-5}, \frac{1}{3 \times 10^{20}}, \frac{1}{40}, 1)$. The margins (5.9) then hold exactly when $V \geq \max\{800n, 16000, 2 \times 10^7, 6 \times 10^{22}, 4000n, 100n\}$. Allocating $\eta/6$ to each term of (5.10), a term $b e^{-E_N(s)}$ is at most $\eta/6$ once $V \geq 4n(96011 \cdot 200 + s)s^{-2} \log(6b/\eta)$, and similarly with $4(96000 \cdot 200 + 2s)$ for an E_C term. At $\eta = 0.01$ the six thresholds are listed in Table 2; rounding upward gives (4.14), and the ratio of 8.966×10^{50} to 5.0002×10^{49} is 17.9. The product term has no factor n because the coordinate exponent E_C has none; replacing all six exponents by the common cV/n introduces that dependence needlessly. Lemma 5.10 admits independent tolerances in exactly this form, so no hidden common-tolerance premise is used. $\square$

9.2 What sets the scale

Three quantities restrict the count scale: the jump correction $n/V \leq d/2$, which needs $V \geq 1.2 \times 10^{21}n$; the exponential suppression needed for a stated reliability, which with the common exponent needs $V \geq (n/c) \log(24/\eta)$ and with the unequal tolerances needs $V \geq 5 \times 10^{49}$; and the absolute product tolerance $\eta_P = (3 \times 10^{18})^{-1}/100$ that enters c and $E_C(d_P)$ squared. The second exceeds the first by 29 orders of magnitude, and the food tolerances are ten orders less demanding, so the scale is set by one number: the absolute noise permitted on a coordinate whose deterministic level at the end of startup is only 7×10^{-19} . The distance from the necessary scale of Theorem 4.3 is therefore not a property of the drift constants or of the census. Closing it requires an estimate whose error is relative to the current stock rather than absolute: a seed-arrival estimate, a finite-count growth comparison from the first molecules to a robust operating region, and only then concentration estimates away from zero. Observable-specific quadratic variation, which omits the reactions that do not move the tracked coordinate, and stopped comparisons retaining food availability and product stock are the natural tools; a scalar stock is not Markovian, and holding food fixed or suppressing basal channels would change the model.

9.3 The finite converse is uninformative at small caps

For $n \leq K_*$ the K_* -rectangle contains the entire catalogue, so L_n is the total number of incidences. With $p_0 = 1/\zeta(a)$ and $p_1 = 2^{-a}/\zeta(a)$, row independence gives the exact value

$$\mathbb{P}(L_n \geq 2) = 1 - p_0^{X_n} - X_n p_1 p_0^{X_n-1}. \quad (9.1)$$

At $n = 4$ and $a = \frac{3}{2}$ its complement is $10^{-11.446}$: the residual $\bar{r}$ is essentially one, and (4.6) reduces to $S \leq 1$. Replacing 15480 by 224 cannot repair this; the remainder counts every potential catalytic coordinate, whereas an unused catalyst with zero population exerts no propensity however many assignments its row contains. An informative finite converse would bound, on the actual path, the integrated adverse drift of the occupied catalysts under the exposed, size-biased row law of Appendix A; no count of potential incidences can supply that estimate. We therefore present the startup ceiling as the informative finite statement and do not present a clipped probability interval as evidence of predictive precision.

10 A matched small-source example

We use the complete catalogue at $n = 4$, all 30 molecules and 68 split identities, with all marks $S = B = H = 1$. The *self* environment has exactly one catalytic incidence, (00, 11, 0011; 0011); the *food* intervention replaces its catalyst by the monomer 0; the *deleted* baseline removes the incidence. Both directions of the channel are changed together, and every basal channel, mark, preparation, feed and washout rule is identical. These are three specified, source-admissible environments and matched interventions. They are not draws from the conditional law of the source, and no frequency over environments is estimated.

The deterministic illustration integrates the complete concentration equations

$$\dot{c} = \mathbf{1}_{\mathcal{F}} - c + \sum_{r=(u,v,uv)} \nu_r \left(b_r + \sum_{z:r \in C_z} k_{rz} c_z \right) (c_u c_v - c_{uv}), \quad (10.1)$$

with the export and both input counters as additional coordinates, using the BDF method at relative tolerance 2×10^{-10} and absolute tolerance 10^{-23} . The concentration limit is the classical one [5, 19];

Environment	$c_{0011}(1)$	$c_{0011}(199)$	$Q(1, 100]/V$	$Q(100, 199]/V$
Self	1.27×10^{-8}	0.2753	102.8	109.0
Food	0.344	0.3441	136.3	136.3
Deleted	1.26×10^{-9}	2.0×10^{-9}	2.209×10^{-5}	2.218×10^{-5}

Table 3: Matched concentration-model illustration at $n = 4$; not finite-count probabilities. All three runs use the full catalogue, the same marks and the same basal chemistry.

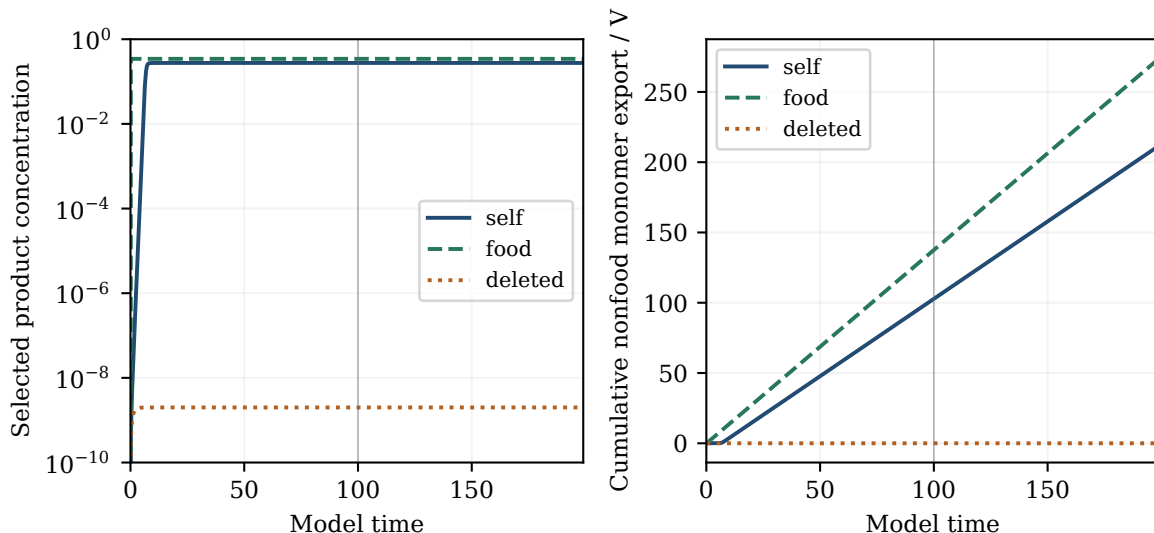


Figure 4: Matched full-catalogue concentration trajectories for the three specified environments. All curves are numerical ODE solutions, not certified count trajectories. The panels separate the selected product from the cumulative collected nonfood mass; the deleted baseline retains positive basal chemistry.

it is used as a picture, and no ODE probability or finite-count error is inferred from it. Table 3 and Figure 4 give the results: the self environment reaches selected product concentration 0.275 and exports 102.8V and 109.0V in the two windows, the food environment starts immediately, and the deleted baseline exports at the basal level. The input ledger is identical in the three cases, 1194 normalized food molecules and 1990 monomer equivalents by time 199 in addition to the initial 10, and the largest deviation of total mass from 10 is 3.7×10^{-13} . Large cumulative exports are consistent with bounded instantaneous mass because feed continues throughout.

The count-process diagnostic runs the exact stochastic simulation algorithm [10] with falling-factorial propensities, all channels, the food feed and washout, from the food-only state to the deadline 1, at $V = 100$ for the three environments and at the predeclared second setting $V = 200$ for the self environment; each row has 200 independently seeded clocks, and corresponding intervention rows share seed labels, which does not by itself provide a pathwise monotone coupling. Each path records the first nonfood molecule, the first selected product, the first corridor exit, and the counts. Table 4 separates these observables, which the mission conflates.

Two remarks fix what the diagnostic can and cannot show. First, a zero-birth row of 200 runs has exact two-sided 95% binomial upper endpoint 0.01828, whereas the analytic ceilings of Theorem 4.3 are 3.2×10^{-5} at $V = 100$ and 6.5×10^{-5} at $V = 200$: observing no initiation is compatible with the bound but cannot resolve its value, and the $V = 200$ row was a prospective prediction that the

Environment	V	Runs	Nonfood by 1	Selected by 1	Births before exit	Exits by 1
Self	100	200	0	0	0	5
Self	200	200	0	0	0	0
Food	100	200	200	200	200	5
Deleted	100	200	0	0	0	5

Table 4: Exact-algorithm count-process startup diagnostics at $n = 4$. In the food row every first nonfood molecule is the selected product and appears before time 0.0154. The five corridor exits at $V = 100$ are feed fluctuations crossing $11V$ within one time unit, a finite-count effect absent from the deterministic model; they are retained, not conditioned away. Each path has between 1100 and 2600 events.

bound survived, not a measurement of it. Second, the mechanistic distinction is visible without statistics: food catalysis can act at the first event, whereas an absent product cannot catalyze its own first appearance, and at $V = 100$ the total basal birth rate is at most $\lambda V = 3.2 \times 10^{-5}$ per unit time, so the food-silent rows had no realistic chance to initiate by the deadline. That is the content of the necessary scale.

11 Dimensions, interpretation, and scope

An illustrative dimensionalization uses a concentration scale c_* in mol/L, a physical volume Ω in litres and a residence time τ_{res} :

$$V = N_A \Omega c_*, \quad t_{\text{phys}} = \tau_{\text{res}} t, \quad k_{\text{phys}} = \frac{\kappa}{\tau_{\text{res}} c_*^{m-1}}, \quad (11.1)$$

for a dimensionless coefficient κ of full molecularity m . Forward ligation and reverse cleavage have different molecularities, and an explicit catalyst adds a reactant factor, so equal dimensionless paired coefficients do not have equal dimensional units. At $c_* = 1 \mu\text{M}$ the necessary 99% startup scale $V = 1.43 \times 10^7$ corresponds to about 23.7 pL and to an initial food population of $6V \approx 8.6 \times 10^7$ molecules; the envelope scale 2.5×10^{61} at $n = 4$ to 4.2×10^{43} L, the direct witness scale 8.97×10^{50} to 1.5×10^{33} L, and the unequal-tolerance scale 5.0×10^{49} to 8.3×10^{31} L. The first number is a necessary condition and not a reactor design; the others demonstrate the conservatism of the concentration-level analysis and not achievable equipment. Taking $\tau_{\text{res}} = 1 \text{ h}$ gives a 199-hour mission at dilution rate 1 h^{-1} ; it does not calibrate the reaction coefficients to any chemistry, and there is no experimentally determined mapping from binary words to molecules in this study.

Table 5 lists the modelling choices. The immediate methodological content is concrete: define the operating task before screening networks, distinguish source occurrence from operating reliability, expose the initiation bottleneck separately from the growth argument, and require compatible material and uncertainty accounts. A physically grounded successor would start from an explicit molecular reaction set with measurable rates and preserve stoichiometry, rate units, catalyst multiplicities, driving species and reservoir conditions; nothing here transfers automatically from binary words to activated oligomers or enzymatic ligation.

12 Discussion

Three requirements, three scales. Availability of a productive organization, creation of its first molecules, and execution of a specified task are different requirements with different quantitative

Feature	Treatment here	Consequence
Discrete counts	explicit falling-factorial jump process	first-molecule barriers and demographic noise are represented
Reversibility	both directions of every split identity	signed synthesis and reverse losses are accounted for
Resources	maintained feed with a finite observed expenditure allowance	finite mission accounting, not an autonomously depleting reservoir
Background chemistry	every nonfood row retained in the witness	a full-catalogue statement within the declared source
Degradation and waste	uniform washout; no damage or inhibition	selective destruction needs new drift estimates
Spatial structure	well mixed; no diffusion or membrane	no transport-limited prediction
Product identity	aggregate nonfood export; existential length-four stock	no target-specific activity or purification
Thermochemistry	paired dimensionless coefficients	material accounting is not a work or dissipation theorem

Table 5: What the model includes and omits.

statements. Availability is a source event of probability $224p_{a,n} + o(p_{a,n})$ for the productive singletons and $\theta > 0$ for RAFs in general. Initiation in a food-silent environment is a first-passage event with hazard at most λV , hence a necessary scale $V \geq \log(1/\eta)/\lambda$. Execution, on one uninterrupted trajectory with stock at three times and export in two windows, has probability of order $p_{a,n}$ at a sufficient scale, and conditioning on it selects a single productive incidence. Treating “a RAF exists” as an operating prediction collapses these distinctions.

What the two-window event adds. The event demands two actual output windows and endpoint stock on the same trajectory while accounting for the gross supply. The proof handles this without a restart theorem: the interval estimate (5.7) is stated for an arbitrary actual interval and telescopes across the middle observation time, and the six noise controls are imposed once on the whole horizon. This is weaker than uniform restart from every state in the stock region and weaker than indefinite operation, and it is stated as such. The material identity shows that the exported mass is net internal synthesis, since the reactor starts with no nonfood inventory.

Where the constants come from. The loss coefficients 23 and 25 are the corridor bound $\sum_z x_z \leq 11$ counted in two input positions plus the number of cleavage positions; the drift constant 2 is $4 \cdot \frac{3}{4} - 1$; the endpoint constant 54 is the range of $\log x - 4H$ between the stock floor and the mass cap; and the window length 79 is the least duration at which (5.8) exceeds $\frac{1}{10}$. None of these sets the count scale. The scale is set by the absolute product tolerance $(3 \times 10^{18})^{-1}/100$, entering squared in c ; Proposition 4.5 shows exactly how far the estimate can be improved by choosing tolerances and where a new inequality is required.

Open problems. (1) A sufficient count scale within reach of the necessary one, through a seed-to-established-stock comparison with relative rather than absolute error. (2) A finite converse that weighs occupied catalysts on the actual path under the size-biased row law, replacing the count of potential incidences; this is also what singleton dominance for $1 < a < 2$ would need. (3) The relative operational weights of the 224 productive labels, that is, the limits of the environment-averaged conditional success probabilities identified in [1]. (4) A restart theorem from an arbitrary state

in the stock region, and more than two windows. (5) A thermochemically consistent source in which the exported mass can be related to work and dissipation [23]. None of these alters the finite statements proved here.

Data and code availability. The repository location will be supplied with the author information. The editable source package contains the paper, the bibliography, the figure data and scripts, the exact-arithmetic replay of every printed number, the raw simulation records, and the two strict verification receipts (Appendix B).

A Explicit constants and source formulas

The common-tolerance choice. With $d_N = \dots = d_B = d$ and $V \geq 2n/d$, the margins (5.9) hold for $n \geq 4$ because $n/V \leq d/2$ and $2/V \leq d/2$, and $d+d/2 \leq \frac{1}{8}, \frac{1}{80}, 10^{-5}, \frac{1}{3 \times 10^{20}}, \frac{1}{40}, 1$ respectively; the last inequality is the only one with slack of less than a factor two, since $d = \frac{1}{2} \cdot \frac{1}{3 \times 10^{20}}$. Every exponent is then at least cV/n , and the six coefficients sum to 24. At $V = V_*(n)$, $\delta_{n,V} = 24e^{-48} \approx 3.4 \times 10^{-20}$ for $n = 4$ and decreases with n .

Source moments. The moments (8.1) are evaluated as $\sum_{k=1}^{R-1} k^{-s} = \zeta(s) - \zeta(s, R)$ for $s \neq 1$, with $\zeta(s, R)$ the Hurwitz zeta function, and as the harmonic number for $s = 1$; this is how the values in Figure 3 and the checks of the replay script are computed. At $a_n = 2 - 2/n$ the normalized first moment satisfies $\mathbb{E}D/n \rightarrow (1 - e^{2 \log 2})/[-2\zeta(2)] = 9/\pi^2$, the profile of the formal development at the parameter -2 , and with $X_n \sim 2 \cdot 2^n$, $R_n \sim 2n2^n$ this is (4.10). Numerically $2^n p_n$ equals 2.604, 1.455, 0.928, 0.693 and 0.580 at $n = 4, 8, 16, 32, 64$; the approach to 0.456 is slow, of relative order $1/n$.

The conditional sampler. Conditioning a row on containing one specified incidence changes the degree law to $\mathbb{P}(D = d \mid r \in C_z) = d \mathbb{P}(D = d) / \mathbb{E}D$, after which the remaining $d - 1$ coordinates are uniform. For k required and ℓ forbidden distinct coordinates the likelihood multiplier is $(d)_k (R_n - d)_\ell / (R_n)_{k+\ell}$. At $n = 4$ the cap atom has probability about 0.09 unconditionally and about 0.56 after requiring one incidence. Planting an incidence into an unconditioned row is therefore not an exact conditional sampler, and any finite converse that weighs occupied catalysts must use this law.

Exact values of the finite obstruction. At $n = 4$, $a = \frac{3}{2}$: $X_4 = 30$, $p_0 = 1/\zeta(\frac{3}{2}) = 0.3828$, $p_1 = 2^{-3/2} p_0$, and the complement in (9.1) is $p_0^{30} + 30p_1 p_0^{29} = 10^{-11.446}$.

B Formal scope and reproducibility

The Lean development is a single project pinned to `leanprover/lean4:v4.30.0` with a pinned Mathlib manifest; the reactor namespace `RandomViability` imports the source namespace `PowerLawSmallRAF`, and both use one configuration weight for the capped, shifted Zipf sampler. The core declaration is

```
RandomViability.c6_two_window_resolution
```

in `proofs/RepeatedFunction/C6Resolution.lean`. Its strict receipt records exit code zero, warnings treated as errors, empty compiler output, 425 dependency sources with their SHA-256 hashes,

the pinned toolchain and manifest, and the transitive axioms `propext`, `Classical.choice` and `Quot.sound` only; the source hash of the root is `a18c3766...3060`. The declaration asserts, for every integer volume sequence $V_n \geq V_*(n)$: the finite bracket $0 < \text{lower} \leq S \leq \text{upper}$ for every $n \geq 4$ and $a > 1$, with the lower bound $e_a^6 p_{a,n} (1 - 24e^{-cV/n})$ and the upper bound $15480 p_{a,n} + \bar{r}$; the eventual bounds $\frac{1}{2}(6/\pi^2)^6 p_n \leq S_n \leq 15481 p_n$ at $a_n = 2 - 2/n$; the eventual posterior bound $\frac{1}{2}(6/\pi^2)^6 / 15481$ for the seed class; the vanishing conditional probability of no productive singleton; and, for every environment and mark configuration, the almost-sure signed synthesis $J(199) > V/5$ on the mission. Its event is exactly Definition 3.1 (`twoWindowMission`), including stock at time 1.

A second strict receipt, for `proofs/RepeatedFunction/PublicationBundle.lean`, checks `ProductiveCensus.lean` and `StartupNecessary.lean`: the count 224 at cap four by kernel evaluation, the injection of the labels of every catalogue into cap four, the equivalence of the productive predicate with the union event, the food-pair maximum $121/3$, the rate consequence, the finite-kernel survival induction, the logarithmic necessary-scale inequality and the scalar material-ratio inequality; 428 dependency entries, empty compiler output, the same three axioms. Table 6 maps the statements of this paper to their evidence.

Statement	Evidence
Lemma 5.1 (loss envelopes)	<code>physical_food_drift_bound</code> , <code>physical_product_drift_bound</code> ; derivation in the text
Lemma 5.2	<code>collectivePotential</code> and its derivative lemma; proof in the text
Lemmas 5.4, 5.5	<code>physical_mass_localization</code> , <code>physical_startup_product_from_noise</code>
Lemmas 5.6, 5.7	<code>physical_holding_potential_gain</code> , <code>physical_potential_endpoint_penalty</code>
Proposition 5.8, Corollary 5.9	<code>physical_interval_potential_gain</code> , <code>physical_interval_mass_from_noise</code> , <code>physical_interval_export_lower</code> , <code>interval_export_margin</code>
Lemma 5.10 and (4.4)	<code>physical_two_window_failure_tail</code> , <code>physical_two_window_uniform_tail</code> , <code>concrete_two_window_success_lower</code>
Source averaging (4.5)	<code>source_food_silent_seed_mass</code> , <code>fully_averaged_two_window_lower</code>
Lemma 5.11	<code>physical_output_upper_at_most_one_local</code> (companion [1])
Lemma 5.12, count and predicate (4.6), (4.7) with 15480	<code>productive_labels_four</code> , <code>productive_labels_card_le</code> , <code>productive_union_iff</code> ; union bound (5.13) conventional <code>averaged_mission_finite_upper</code> , <code>two_window_finite_source_resolution</code>
Lemma 3.2	<code>actual_net_synthesis_identity</code> , <code>food_only_mission_net_synthesis</code> , <code>mission_material_recovery</code>
Theorem 4.2	limits (6.1) formal; eventual bounds with $\gamma/2$, 15481 in the root; full constants conventional
Section 6.4	<code>two_local_incidence_mass_div_incidence_tendsto_zero</code> , <code>source_RAF_probability_tendsto</code> ; transfers conventional
Theorem 4.3	<code>food_pair_birth_bound</code> , <code>startup_intensity_bound</code> , <code>startup_survival_lower</code> , <code>startup_necessary_scale</code> ; stopped-generator identification and Poisson mixture conventional
Theorem 4.4	first-moment limit along families <code>windowZipfMean_source_window_normalized</code> ; all else conventional
Proposition 4.5	conventional evaluation of Lemma 5.10
Sections 10, 11	numerical illustration; no proof premise

Table 6: Claims and their evidence. Lean names live in `proofs/RandomViability/`, `proofs/RepeatedFunction/` and `proofs/PowerLawSmallRAF/`.

Conventional proofs. The union bound that replaces 15480 by 224, the limiting constants of Theorem 4.2, the screening formula, the reliability inequality, the four transfers, the stopped-generator identification and Poisson mixture in Theorem 4.3, Theorem 4.4 except its first-moment limit, and Proposition 4.5 are proved in the text and are not additional terminal Lean exports. The finite-kernel induction alone does not certify that a particular physical generator meets its hypotheses; that identification is the explicit argument of Section 7.

Reproduction. The exact-arithmetic script `check_paper.py` distributed with the source replays every printed number: the catalogue counts and the census, the noise constants and margins, the drift and endpoint algebra, the window margins and the threshold 79, the material ratio, the hazard and the necessary scale, the source moments and every asymptotic formula of Theorem 4.4 against exact Hurwitz-zeta evaluations, the unequal-tolerance thresholds, the finite obstruction, the dimensional conversions, and the numerical tables from the retained raw records; it also checks the two receipts for verification status, dependency counts and axioms. The figure script regenerates the four figures from the retained data. The simulation script, its full-catalogue definitions, and the raw per-path records are included for provenance; the production run used seeds 91000–91199, a single numerical worker, and a checked event cap that no path reached. Formal verification uses the repository bridge, which pins the Mathlib working directory, treats warnings as errors and writes the receipts; no dependency update is performed.

References

- [1] Anonymous. Diffuse RAFs and productive operation: asymptotic singleton dominance in a random polymer model. Companion manuscript, 2026; arXiv preprint in preparation, 2026.
- [2] Anonymous. Structural autocatalysis can be common while productive operation is rare: an exact separation theorem in a capped-Zipf binary polymer model. Companion manuscript, 2026; arXiv preprint in preparation, 2026.
- [3] David F. Anderson and Thomas G. Kurtz. Continuous time Markov chain models for chemical reaction networks. In Heinz Koepl, Gianluca Setti, Mario di Bernardo, and Douglas Densmore, editors, *Design and Analysis of Biomolecular Circuits*, pages 3–42. Springer, New York, 2011. doi: 10.1007/978-1-4419-6766-4_1.
- [4] Alex Blokhuis, David Lacoste, and Philippe Nghe. Universal motifs and the diversity of autocatalytic systems. *Proceedings of the National Academy of Sciences*, 117(41):25230–25236, 2020. doi: 10.1073/pnas.2013527117.
- [5] R. W. R. Darling and J. R. Norris. Differential equation approximations for Markov chains. *Probability Surveys*, 5:37–79, 2008. doi: 10.1214/07-PS121.
- [6] 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. doi: 10.1007/978-3-030-79876-5_37.
- [7] Stewart N. Ethier and Thomas G. Kurtz. *Markov Processes: Characterization and Convergence*. Wiley, New York, 1986. doi: 10.1002/9780470316658.

- [8] J. Doyne Farmer, Stuart A. Kauffman, and Norman H. Packard. Autocatalytic replication of polymers. *Physica D: Nonlinear Phenomena*, 22(1–3):50–67, 1986. doi: 10.1016/0167-2789(86)90233-2.
- [9] David A. Freedman. On tail probabilities for martingales. *The Annals of Probability*, 3(1): 100–118, 1975. doi: 10.1214/aop/1176996452.
- [10] Daniel T. Gillespie. Exact stochastic simulation of coupled chemical reactions. *The Journal of Physical Chemistry*, 81(25):2340–2361, 1977. doi: 10.1021/j100540a008.
- [11] Varun Giri and Sanjay Jain. The origin of large molecules in primordial autocatalytic reaction networks. *PLoS ONE*, 7(1):e29546, 2012. doi: 10.1371/journal.pone.0029546.
- [12] Wim Hordijk and Mike Steel. Detecting autocatalytic, self-sustaining sets in chemical reaction systems. *Journal of Theoretical Biology*, 227(4):451–461, 2004. doi: 10.1016/j.jtbi.2003.11.020.
- [13] Wim Hordijk and Mike Steel. Autocatalytic sets extended: dynamics, inhibition, and a generalization. *Journal of Systems Chemistry*, 3:5, 2012. doi: 10.1186/1759-2208-3-5.
- [14] Wim Hordijk and Mike Steel. Autocatalytic sets in polymer networks with variable catalysis distributions. *Journal of Mathematical Chemistry*, 54(10):1997–2021, 2016. doi: 10.1007/s10910-016-0666-z.
- [15] Wim Hordijk and Mike Steel. Chasing the tail: the emergence of autocatalytic networks. *Biosystems*, 152:1–10, 2017. doi: 10.1016/j.biosystems.2016.12.002.
- [16] 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. doi: 10.3390/ijms12053085.
- [17] Wim Hordijk, Leonard Hasenclever, Jie Gao, Dilyana Mincheva, and Jotun Hein. An investigation into irreducible autocatalytic sets and power law distributed catalysis. *Natural Computing*, 13(3):287–296, 2014. doi: 10.1007/s11047-014-9429-6.
- [18] Stuart A. Kauffman. Autocatalytic sets of proteins. *Journal of Theoretical Biology*, 119(1): 1–24, 1986. doi: 10.1016/S0022-5193(86)80047-9.
- [19] Thomas G. Kurtz. The relationship between stochastic and deterministic models for chemical reactions. *The Journal of Chemical Physics*, 57(7):2976–2978, 1972. doi: 10.1063/1.1678692.
- [20] Elchanan Mossel and Mike Steel. Random biochemical networks: the probability of self-sustaining autocatalysis. *Journal of Theoretical Biology*, 233(3):327–336, 2005. doi: 10.1016/j.jtbi.2004.10.011.
- [21] Praneet Nandan, Philippe Nghe, and Jérémie Unterberger. Autocatalytic cores in the diluted regime: classification and properties. *Journal of Mathematical Biology*, 92:36, 2026. doi: 10.1007/s00285-026-02357-7.
- [22] Zhen Peng, Jeff Linderorth, and David A. Baum. The hierarchical organization of autocatalytic reaction networks and its relevance to the origin of life. *PLoS Computational Biology*, 18(9): e1010498, 2022. doi: 10.1371/journal.pcbi.1010498.

- [23] Riccardo Rao and Massimiliano Esposito. Nonequilibrium thermodynamics of chemical reaction networks: wisdom from stochastic thermodynamics. *Physical Review X*, 6(4):041064, 2016. doi: 10.1103/PhysRevX.6.041064.
- [24] Alessandro Ravoni. Impact of composition on the dynamics of autocatalytic sets. *Biosystems*, 198:104250, 2020. doi: 10.1016/j.biosystems.2020.104250.
- [25] Mike Steel. The emergence of a self-catalysing structure in abstract origin-of-life studies. *Applied Mathematics Letters*, 13(3):91–95, 2000. doi: 10.1016/S0893-9659(99)00191-3.
- [26] 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. doi: 10.1098/rsif.2018.0808.
- [27] 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. doi: 10.1145/3372885.3373824.
- [28] Jérémie Unterberger and Philippe Nghe. Stoichiometric and dynamical autocatalysis for diluted chemical reaction networks. *Journal of Mathematical Biology*, 85:26, 2022. doi: 10.1007/s00285-022-01798-0.