

Critical-window emergence of autocatalytic sets under reaction-channel quotienting

September 14, 2026

Abstract

We determine the limiting probability that a reflexively autocatalytic and food-generated (RAF) set exists in the reversible binary-polymer model in the reaction convention used by the public generator of the finite-core theory of RAF emergence: two split descriptions $u + v \rightleftharpoons uv$ and $v + u \rightleftharpoons vu$ are identified precisely when $uv = vu$. Molecules are the nonempty binary words of length at most n , the food set consists of the six words of lengths one and two, and each molecule catalyses each reaction channel independently with probability p_n , one mark serving both directions. Let $f_n = p_n |J_n|$ be the mean number of channels catalysed by a molecule. For every sequence with $f_n/n \rightarrow \lambda > 0$ the RAF probability converges to

$$\Theta_{\text{qt}}(\lambda) = S_{\text{qt}}(1 - e^{-\lambda}) > 0,$$

where $S_{\text{qt}}(a)$ is the probability that ordinary reversible closure of the food set is unbounded in an infinite field of independent canonical channels of openness a , and the probability of a RAF conditional on a catalysed food gateway converges to $S_{\text{qt}}(1 - e^{-\lambda})/(1 - e^{-34\lambda})$. This resolves the bulk-factor problem left open in the companion paper on exact gateway scaling.

The quotient map preserves closures and RAF witnesses deterministically, but it does not preserve homogeneous Bernoulli ensembles: independent split marks project to heterogeneous OR marks, and a four-molecule example shows that the split and quotient RAF probabilities differ at every parameter. The proof therefore works in the quotient model throughout. A half-openness domination transfers a supplied-seed growth estimate from the split model through fibres of size at most two; a record-word argument proves that unbounded closure generates every fixed word at the original openness; a uniform finite-seed approximation gives continuity of S_{qt} ; an exact law for complete catalytic peeling histories connects static closure to the finite catalytic model; and a bounded-catalyst gateway bound closes the upper estimate. We also prove that Θ_{qt} is nondecreasing and satisfies $S_{\text{sp}}(1 - e^{-\lambda/2}) \leq \Theta_{\text{qt}}(\lambda) \leq \min\{S_{\text{sp}}(1 - e^{-\lambda}), 1 - e^{-30\lambda}\}$, where S_{sp} is the split-position survival function, so that $\Theta_{\text{qt}}(\lambda) \rightarrow 0$ as $\lambda \downarrow 0$ and $\Theta_{\text{qt}}(\lambda) \rightarrow 1$ as $\lambda \rightarrow \infty$. The main theorem and its dependencies are formalised in Lean 4 with no axioms beyond the classical foundation; the few hand-proved comparison results are labelled as such. The results concern structural autocatalysis, not kinetic or thermodynamic viability.

Keywords. autocatalytic sets; RAF theory; binary polymer model; reaction identity; quotient reaction networks; critical window; product measure; formal verification.

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

1 Introduction

Collectively autocatalytic sets were proposed by Kauffman [16, 17], following Eigen’s hypercycle [4], as a mechanism by which a random chemistry could support a self-sustaining metabolism without any single self-replicating molecule. Their combinatorial formalisation is the reflexively

autocatalytic and food-generated set, or RAF, introduced by Steel [23] and developed with Hordijk [9, 14, 8, 24]; see [6, 12, 13] for surveys. A RAF is a nonempty family of reactions that can generate all of its molecules from a prescribed food set and can find a catalyst for each of its reactions inside the resulting closure. RAF theory deliberately separates these structural requirements from any choice of reaction kinetics.

The standard test bed is the binary polymer model [16, 23, 9]: molecules are binary words of length at most n , reactions are ligations and their reverse cleavages, the food set consists of the short words, and each molecule catalyses each reaction independently with a probability tuned so that a molecule catalyses on average f_n reactions. Mossel and Steel [20] showed that f_n growing linearly in n is the relevant scale for RAF occurrence, and simulations [7, 11] display a steep rise of the finite- n RAF probability within a narrow range of f_n/n . A companion manuscript [2] determined the limit on this scale for the textbook *split-position* catalogue, in which a reaction is a product word together with a split position: if $f_n/n \rightarrow \lambda \in (0, \infty)$ then the RAF probability converges to $\Theta_{\text{sp}}(\lambda) = S_{\text{sp}}(1 - e^{-\lambda})$, a continuous nondecreasing function with $0 < \Theta_{\text{sp}}(\lambda) \leq 1 - e^{-36\lambda} < 1$, where S_{sp} is the survival probability of an infinite reversible closure process. Thus the entire linear scale is a nontrivial critical window rather than a step.

Reaction identity

The recent finite-core theory of RAF emergence of Varanasi and Korenaga [28] is accompanied by a public generator [27] whose reaction catalogue differs from the split-position one: it compares displayed reactions and suppresses duplicates. The descriptions $\mathbf{A} + \mathbf{AA} \rightleftharpoons \mathbf{AAA}$ and $\mathbf{AA} + \mathbf{A} \rightleftharpoons \mathbf{AAA}$ are the same channel, while $\mathbf{A} + \mathbf{B} \rightleftharpoons \mathbf{AB}$ and $\mathbf{B} + \mathbf{A} \rightleftharpoons \mathbf{BA}$ remain distinct because their products differ. In general two split descriptions are identified exactly when their factors commute, $uv = vu$. A second companion manuscript [1] worked in this quotient convention: it proved an exact overlap law for the RAF event, showed that for every $n \geq 4$ exactly 34 channels can fire from the six food words, derived the cap $\mathbb{P}(\text{RAF}_n) \leq 1 - (1 - p_n)^{34|X_n|}$, and thereby the limit $1 - e^{-34\lambda}$ for the gateway factor at $f_n = \lambda n$. It factorised the RAF probability as gateway probability times a gateway-conditioned bulk factor B_n , and left the limit of B_n , and the limit of the RAF probability itself, as its first open problem: “it is open whether the two conventions have the same bulk limit.”

Deleting duplicate descriptions changes very little in the reaction count: the number of channels lost is $O(n^2 2^{n/3})$ out of $(n - 2)2^{n+1}$ (Lemma 3.5 and Remark 3.6). It nevertheless changes the random experiment, because in the quotient convention each surviving channel receives a *single* Bernoulli mark per molecule, whereas in the split convention a commuting pair receives two. A four-molecule example (Example 3.2) shows that the two RAF probabilities differ at every parameter $p \in (0, 1)$, even though every closure and every RAF witness corresponds exactly under the quotient map. An asymptotic count ratio therefore cannot identify the two limiting probabilities, and the split-position theorem of [2] does not transfer by relabelling.

Main result

We prove the critical-window law directly in the quotient model. Let X_n be the nonempty binary words of length at most n , J_n the quotient channels, $p_n = f_n/|J_n|$, and $P_n(p)$ the probability of a RAF under independent molecule-channel catalysis with parameter p . Let $S_{\text{qt}}(a)$ be the probability that ordinary reversible closure of the food set is unbounded in an infinite field of independent canonical channels open with probability a (Section 2).

Main Theorem (Theorem 2.2 below). *If $f_n/n \rightarrow \lambda > 0$, then $0 < p_n < 1$ for all large n , and*

$$P_n(p_n) \longrightarrow \Theta_{\text{qt}}(\lambda) := S_{\text{qt}}(1 - e^{-\lambda}) > 0, \quad B_n(p_n) \longrightarrow \frac{S_{\text{qt}}(1 - e^{-\lambda})}{1 - e^{-34\lambda}},$$

where $B_n(p) = P_n(p)/(1 - (1 - p)^{34|X_n|})$ is the gateway-conditioned bulk factor of [1].

The theorem holds for every real sequence with the stated asymptotic intensity, not only $f_n = \lambda n$, and no assumption on a finite initial segment is needed. Section 9 adds that Θ_{qt} is nondecreasing, that

$$S_{\text{sp}}(1 - e^{-\lambda/2}) \leq \Theta_{\text{qt}}(\lambda) \leq \min\{\Theta_{\text{sp}}(\lambda), 1 - e^{-30\lambda}\},$$

and that $\Theta_{\text{qt}}(\lambda) \rightarrow 0$ as $\lambda \downarrow 0$ and $\Theta_{\text{qt}}(\lambda) \rightarrow 1$ as $\lambda \rightarrow \infty$; the sublinear and superlinear regimes are recovered as corollaries. The lower comparison is a Lean-verified consequence of the OR law; the upper comparisons and the endpoint values are short conventional arguments and are labelled as such.

Why the proof cannot be a transfer

The quotient map π_n from split descriptions to channels has fibres of size one or two. It preserves every stage of reversible closure and, for a fixed catalytic configuration, a split RAF exists if and only if a quotient RAF exists for the OR-projected marks (Lemma 3.1). This deterministic equivalence is exact, but the OR projection of independent Bernoulli(p) marks is independent with *heterogeneous* parameters $1 - (1 - p)^{|\pi_n^{-1}(j)|}$, equal to $2p - p^2$ on the double fibres. The quotient model of the generator assigns p on every channel. No homogeneous identity holds, and the two probability laws are different measures on the same channel set. The proof therefore establishes each analytic ingredient of the critical-window argument in the quotient model: the only object imported from the split model is one supplied-seed growth estimate, transferred through a coupling at half openness (Lemma 4.2), and that transfer is an inequality, not an identification.

Contributions

- (A) **Quotient critical-window law** (Theorem 2.2). The RAF probability of the quotient model converges at every positive intensity to $S_{\text{qt}}(1 - e^{-\lambda}) > 0$, and the gateway-conditioned bulk factor of [1] converges to $S_{\text{qt}}(1 - e^{-\lambda})/(1 - e^{-34\lambda})$.
- (B) **Deterministic versus distributional quotienting** (Lemma 3.1, Example 3.2). Closures and RAF witnesses correspond exactly across the quotient; independent homogeneous laws do not. The correct projected law is the heterogeneous OR law, and a fibre coupling at reduced openness dominates it by a homogeneous one.
- (C) **Normalisation without enumeration** (Lemma 3.5). An arithmetic uniqueness argument for commuting words bounds the number of deleted descriptions by $n|X_{\lfloor (n-1)/2 \rfloor}|$, so that $n|X_n|/|J_n| \rightarrow 1$ and $p_n|X_n| \rightarrow \lambda$.
- (D) **Quotient analysis at the original parameter** (Sections 5–6). Richness of unbounded closure, a uniform finite-seed approximation, and continuity of S_{qt} on $(0, 1]$ are proved for the quotient field itself, with no auxiliary sprinkling field.
- (E) **Exact peeling-history law and sandwich** (Sections 7–8). A complete-history replacement identity turns adaptive catalytic pruning into a fixed-pool static problem, giving the lower bound; a first-gateway argument with bounded catalysts gives the upper bound.
- (F) **Comparisons and regimes** (Section 9). Monotonicity, two-sided comparison with the split-position profile, the productive-gateway cap $1 - e^{-30\lambda}$, endpoint limits, and the outer scaling regimes.

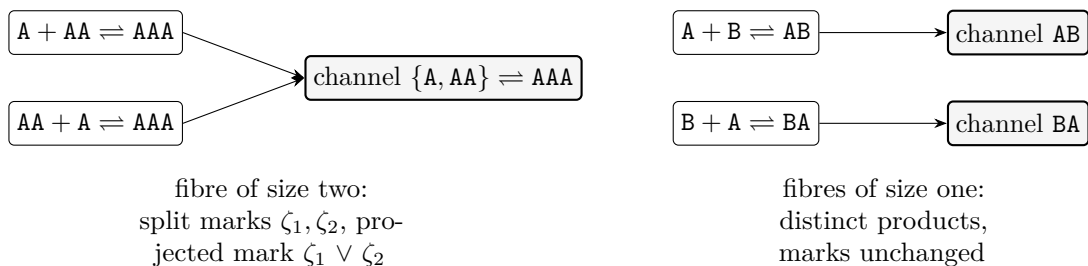


Figure 1: The commuting-factor quotient $\pi_n : R_n \rightarrow J_n$. Every fibre has size one or two. For a fixed catalyst, the projected catalytic mark of a channel is the OR of the split marks in its fibre; under independent Bernoulli(p) split marks a double fibre is open with probability $2p - p^2$, whereas the quotient model assigns probability p to every channel.

(G) **Formal verification** (Appendix A). The main theorem is the Lean 4 declaration `RAFReactionQuotient.reaction_quotient_critical_window_resolution`, checked with warnings promoted to errors against the literal RAF predicate of the generator’s model, with axiom closure `{propext, Classical.choice, Quot.sound}`.

Organisation

Section 2 fixes the model and states the main theorem. Section 3 proves the OR law, the half-openness domination, and the channel count. Section 4 states the supplied-seed estimate and transfers it. Sections 5 and 6 prove richness, finite approximation, and continuity. Section 7 converts static closure into catalytic RAFs through peeling histories, and Section 8 assembles the sandwich and proves Theorem 2.2. Section 9 collects comparisons and regimes. Section 10 explains the consequences for reaction-network representation with a worked example, and Section 11 discusses scope and open problems. Appendix A gives the concordance between the numbered results and the Lean declarations.

2 Model and main theorem

2.1 Molecules, channels, closure

Let $\Sigma = \{A, B\}$ and

$$X_n = \bigcup_{\ell=1}^n \Sigma^\ell, \quad M_n = |X_n| = 2^{n+1} - 2, \quad F = \Sigma \cup \Sigma^2.$$

We work with $n \geq 2$; changing finitely many initial indices affects no limit below. The food set F has six elements, A, B, AA, AB, BA, BB .

A *split description* is an ordered pair of nonempty words (u, v) with $|u| + |v| \leq n$, representing the reversible reaction $u + v \rightleftharpoons uv$. Write R_n for the set of split descriptions; it is in bijection with the set of pairs (product word, split position), so $|R_n| = \sum_{\ell=2}^n (\ell - 1)2^\ell = (n - 2)2^{n+1} + 4$. Two descriptions are *equivalent* if their products are equal and their unordered factor pairs are equal. Write J_n for the quotient and $\pi_n : R_n \rightarrow J_n$ for the quotient map. A nontrivial identification can only exchange the two factors, so it requires $uv = vu$; the canonical representative places the shorter factor first, and equal-length commuting factors are identical (Lemma 3.4), so there is no residual ambiguity. This is exactly the channel set of the generator [27], formalised as `RepositoryChannel n` in [1]. The maps π_n respect the inclusions $R_n \subseteq R_{n+1}$, and $J_\infty = \bigcup_n J_n$ is well defined. Figure 1 illustrates the quotient.

For a channel family $A \subseteq J_n$ and a food cutoff $L \geq 1$, let $\mathcal{C}_{n,L}(A)$ be the smallest subset of X_n containing $X_n \cap X_L$ and closed under both rules

$$u, v \in \mathcal{C}_{n,L}(A) \implies uv \in \mathcal{C}_{n,L}(A), \quad uv \in \mathcal{C}_{n,L}(A) \implies u, v \in \mathcal{C}_{n,L}(A), \quad (1)$$

whenever the channel of (u, v) lies in A . Closure disregards catalysis: it records which molecules can be generated using the specified reactions from the supplied words. It treats availability as a set-valued property and does not count copies, so the multiplicity of a reactant is never a constraint. Write $\mathcal{C}_n = \mathcal{C}_{n,2}$ for closure from the food set.

2.2 Random catalysis and the RAF event

Let $\xi_{j,x}$, $j \in J_n$, $x \in X_n$, be independent Bernoulli variables with parameter $p \in [0, 1]$. The event $\xi_{j,x} = 1$ means that x catalyses channel j in both directions; no separate mark is assigned to cleavage. Food molecules may be catalysts.

Definition 2.1. A nonempty channel set $A \subseteq J_n$ is a *RAF* if every endpoint of every channel in A belongs to $\mathcal{C}_n(A)$, and every $j \in A$ has a catalyst $x \in \mathcal{C}_n(A)$ with $\xi_{j,x} = 1$. Let $P_n(p)$ be the probability that at least one RAF exists.

This is the closure-based RAF definition of [9, 7] for a reversible reaction set and coincides with the predicate **IsRevRAF** of [1]; either side of a reversible channel may initiate its use, after which all its endpoints are available. The definition includes RAFs whose entire closure is food. The expected number of channels catalysed by one molecule is $f = p|J_n|$; following the generator we parametrise by f and put $p_n = f_n/|J_n|$.

2.3 The infinite canonical field and the survival function

Independently of the catalytic model, for $a \in [0, 1]$ give every $j \in J_\infty$ an independent Bernoulli(a) *open* mark, with product law $\mathbb{P}_a$. Let $\mathcal{C}_\infty$ be the ordinary reversible closure of F using the open channels, with no cap on word length. Define

$$S_{\text{qt}}(a) = \mathbb{P}_a\{\mathcal{C}_\infty \text{ contains words of arbitrarily large length}\}. \quad (2)$$

Every membership assertion has a finite derivation, so the survival event is a countable combination of finite-coordinate events and is measurable. The restriction of the infinite field to J_n is the homogeneous product law on J_n ; we write $\mathbb{P}_a$ also for it. The field describes open channels, not catalytic assignments; its role is to name the limit.

2.4 Gateways and the bulk factor

A *food gateway* is a channel with at least one side wholly contained in F . For $n \geq 4$ there are exactly 34 of them (Lemma 8.1). Let G_n be the event that some food gateway is catalysed by some molecule of X_n ; its probability is $1 - (1 - p)^{34M_n}$. Since every RAF contains a catalysed gateway (Section 8), $\{\text{RAF}\} \subseteq G_n$, and the *gateway-conditioned bulk factor* of [1] is

$$B_n(p) = \mathbb{P}_p(\text{RAF} \mid G_n) = \frac{P_n(p)}{1 - (1 - p)^{34M_n}}, \quad 0 < p \leq 1. \quad (3)$$

2.5 Main theorem

Theorem 2.2 (Quotient critical window). *Let (f_n) be a real sequence with $f_n/n \rightarrow \lambda > 0$ and put $p_n = f_n/|J_n|$. Then $0 < p_n < 1$ for all sufficiently large n , and*

$$P_n(p_n) \longrightarrow \Theta_{\text{qt}}(\lambda) := S_{\text{qt}}(1 - e^{-\lambda}) > 0, \quad (4)$$

$$B_n(p_n) \longrightarrow \frac{S_{\text{qt}}(1 - e^{-\lambda})}{1 - e^{-34\lambda}}. \quad (5)$$

Lean: `RAFReactionQuotient.reaction_quotient_critical_window_resolution`.

The Lean statement is the conjunction of four assertions: for all large n the model parameter equals $f_n/|J_n|$ exactly (the totalising clamp used to define a measure for every real f_n is eventually inactive); the RAF probability of the generator’s actual finite catalogue converges to $S_{\text{qt}}(1 - e^{-\lambda})$; that limit is positive; and the actual gateway-conditioned factor converges to the displayed quotient. In particular, the conclusion is not a limit for a newly defined substitute event.

Remark 2.3. The theorem is an exact limit theorem. It provides no finite-size convergence rate, and the supplied-seed cutoff used in the proof is far too large to give one (Remark 4.3). Figure 2 shows how slowly the finite-size channel openness approaches its limit, purely as a consequence of the normalisation.

3 Quotient geometry and normalisation

3.1 Closure projection and the OR law

Lemma 3.1 (Closure and OR projection). *Every fibre of π_n has size one or two. For an active split family $A' \subseteq R_n$, every stage of ordinary reversible closure of A' coincides with the corresponding stage for $\pi_n(A')$. For split catalytic marks $\zeta_{r,x}$ define the projected marks*

$$\bar{\zeta}_{j,x} = \bigvee_{r \in \pi_n^{-1}(j)} \zeta_{r,x}.$$

Then a split RAF exists for ζ if and only if a quotient RAF exists for $\bar{\zeta}$. If the split marks are independent Bernoulli(p), the projected marks are independent with parameters

$$q_j = 1 - (1 - p)^{|\pi_n^{-1}(j)|} \in \{p, 2p - p^2\}. \quad (6)$$

Lean: `canonical_fibre_le_two`, `split_closure_eq_quotient`, `split_raf_iff_quotient`, `fieldOR_independent`, `fieldOR_closed_probability`, `catalytic_OR_distribution`.

Proof. Only exchanging the two factors can give a second representative, so fibres have at most two members. Equivalent descriptions have exactly the same two chemical sides, so each closure stage in (1) is unchanged by projection; induction on stages proves the closure assertion.

Project a split RAF together with its catalytic witnesses: the image family has the same closure and every image channel keeps a witness, so it is a quotient RAF for $\bar{\zeta}$. Conversely, given a quotient RAF A for $\bar{\zeta}$, for each $j \in A$ choose a representative $r \in \pi_n^{-1}(j)$ whose split mark $\zeta_{r,x}$ realises one of the witnesses $x \in \mathcal{C}_n(A)$ of j . The chosen representatives project onto A , so they have the same closure, and each has a catalyst in that closure; they form a split RAF. It is essential to choose the representative using the witness rather than fixing an arbitrary representative in advance.

A projected mark is zero precisely when every mark in its fibre is zero, giving (6). Distinct molecule–channel pairs involve disjoint sets of split coordinates, so the projected marks are independent. $\square$

The OR law is a statement about the split ensemble viewed on the quotient channels. It is not the quotient model: the generator assigns probability p to every channel, including the double fibres. The following example, checked by exact enumeration, shows that the difference is visible in the RAF probability itself.

Example 3.2 (Split and quotient RAF probabilities differ). Take molecules $\{A, AA, AAA, AAAA\}$, food $\{A, AA\}$, and the three split descriptions (A, AA) , (AA, A) , (A, AAA) . The quotient has two

channels, one with a double fibre. With independent Bernoulli(p) marks on the three split descriptions for each of the four molecules, exact enumeration of the 2^{12} configurations gives a split RAF probability of 4077/4096 at $p = 1/2$; the OR-projected law reproduces this value exactly, as Lemma 3.1 predicts. The homogeneous quotient model with parameter $p = 1/2$ on both channels gives $239/256 = 3824/4096$. At $p = 1/3$ the values are 505585/531441 and $5137/6561 = 416097/531441$. The events coincide configuration by configuration; the measures do not.

3.2 Domination at reduced openness

The OR law gives a one-directional comparison between homogeneous fields that is all the proof needs. Let $\mathbb{P}_b^{\text{sp}}$ denote the law of independent Bernoulli(b) marks on R_n and $\mathbb{P}_a^{\text{qt}}$ the law of independent Bernoulli(a) marks on J_n . An event determined by the open channels is *increasing* if it is preserved by opening more channels.

Lemma 3.3 (Half-openness domination). *For $a \in [0, 1]$ and every increasing event E of the quotient field,*

$$\mathbb{P}_{a/2}^{\text{sp}}\{\bar{\omega} \in E\} \leq \mathbb{P}_a^{\text{qt}}(E), \quad (7)$$

where $\bar{\omega}$ is the OR projection of the split field. The same holds with $a/2$ replaced by $1 - \sqrt{1 - a}$.
Lean: product_increasing_mono, OR_half_domination, OR_sharp_domination.

Proof. By Lemma 3.1 the projected field is an independent product with coordinate parameters $q_j \in \{a/2, 1 - (1 - a/2)^2\}$, and $1 - (1 - a/2)^2 = a - a^2/4 \leq a$. Couple the two product measures coordinatewise through independent uniform variables U_j , opening j in the projected field when $U_j < q_j$ and in the quotient field when $U_j < a$. The projected configuration is then pointwise below the quotient configuration, and E is increasing. For the sharper parameter, $1 - (1 - b)^2 = a$ exactly when $b = 1 - \sqrt{1 - a}$, and singletons have parameter $b \leq a$. $\square$

3.3 Counting deleted descriptions

Lemma 3.4 (Commuting factors). *Let $c(w)$ denote the integer whose binary expansion (with $A = 0$, $B = 1$ and leading zeros retained through the length) is w . If $uv = vu$ with $|u| = r$ and $|v| = s$, then $(2^r - 1)c(v) = (2^s - 1)c(u)$. Consequently, for a fixed word u and a fixed length s , there is at most one word v of length s with $uv = vu$, and commuting factors of equal length are equal.*

Lean: commuting_factor_unique, commuting_equal_length.

Proof. Concatenation acts on codes by $c(uv) = 2^s c(u) + c(v)$ and $c(vu) = 2^r c(v) + c(u)$. Equating and rearranging gives the displayed identity. Since $2^r - 1 \neq 0$, the code $c(v)$ is determined by u and s , and a code with a prescribed length determines the word. If $r = s$ the identity gives $c(u) = c(v)$. $\square$

Lemma 3.5 (Deleted descriptions and normalisation). *Let $D_n = |R_n| - |J_n|$. Then*

$$|R_n| = (n - 2)2^{n+1} + 4, \quad D_n \leq n M_{\lfloor (n-1)/2 \rfloor}, \quad \frac{nM_n}{|J_n|} \longrightarrow 1.$$

Lean: deletedKey_injective, deleted_card_bound, split_count_le_quotient_add_loss_bound, quotient_normalization.

Proof. The split count is $\sum_{\ell=2}^n (\ell - 1)2^\ell$, which telescopes to the displayed formula. A deleted description is a noncanonical pair (v, u) with $uv = vu$, $u \neq v$, and by Lemma 3.4 the factors have different lengths; say $|u| = r < |v| = s$. Associate to the deleted description the pair (u, s) . By Lemma 3.4 this association is injective. Because $r < s$ and $r + s \leq n$ we have

$r \leq \lfloor (n-1)/2 \rfloor$, so $u \in X_{\lfloor (n-1)/2 \rfloor}$, while $s \leq n$; this proves the bound. Since $M_k = 2^{k+1} - 2$, the bound gives $D_n/(nM_n) \leq 2^{\lfloor (n-1)/2 \rfloor + 1}/(2^{n+1} - 2) \rightarrow 0$, and the exact split formula gives $|R_n|/(nM_n) \rightarrow 1$. Subtracting and inverting gives the limit. $\square$

Remark 3.6 (Exact count). By the Lyndon–Schützenberger theorem [19, 18], $uv = vu$ with $u \neq v$ if and only if $u = w^a$, $v = w^b$ for a unique primitive word w and $a \neq b$, which gives the exact closed form $D_n = \sum_{d \leq n/3} P_d \lfloor (\lfloor n/d \rfloor - 1)^2/4 \rfloor$ with P_d the number of primitive binary words of length d [1]. The formal development uses only the elementary bound of Lemma 3.5, which suffices for normalisation and needs no theory of primitive words. The closed form agrees with exact enumeration for $n \leq 12$ (Table 1); for instance $D_{12} = 96$ against the bound $12 \cdot M_5 = 744$.

Consequently the parameters in Theorem 2.2 satisfy

$$p_n \longrightarrow 0, \quad p_n M_n \longrightarrow \lambda. \quad (8)$$

Eventual positivity of p_n follows from $f_n/n \rightarrow \lambda > 0$ and eventual $p_n < 1$ from $p_n \rightarrow 0$. In Sections 4–8 we prove the RAF limit for every sequence $p_n \in [0, 1]$ satisfying (8); the general statement is `quotient_critical_limit`, and `quotientParameter_limits` discharges its premises for $p_n = f_n/|J_n|$.

4 Uniform growth from a supplied seed

The one quantitative input imported from the split-position development [2] is a supplied-seed growth estimate. It is an unconditional theorem about ordinary reversible closure in the split field, proved there by a Peierls-type contour argument [21, 5] on binary-word geometry; it makes no reference to RAFs or to any survival profile. We state it with its quantitative form so that its scope is explicit. For a cap N and a food cutoff L , $\mathcal{C}_{N,L}^{\text{sp}}$ denotes reversible closure in X_N of the supplied seed $X_L \cap X_N$ under the open split descriptions.

Lemma 4.1 (Split supplied-seed estimate). *Fix $b > 0$ and an integer $m \geq 1$. There is $L > 0$ such that for every $N \geq 1$, under independent Bernoulli(b) marks on R_N ,*

$$\mathbb{P}_b^{\text{sp}} \{m |\mathcal{C}_{N,L}^{\text{sp}}| < (m-1)M_N\} \leq \frac{1}{m}. \quad (9)$$

Lean: `HordijkSteelThreshold.static_bulk_near_full, static_bulk_deficit_bound`.

The quantitative statement behind Lemma 4.1 is as follows. Put

$$C_* = 2 \cdot 7^{64} \cdot 81, \quad h_k = C_*(1 - b^2)^k, \quad \delta_k = \frac{h_k}{1 - h_k}. \quad (10)$$

For every k with $h_k < 1$, with $L = 10k$, every $N \geq 1$ and every integer $d \geq 1$,

$$\mathbb{P}_b^{\text{sp}} \{|\mathcal{C}_{N,L}^{\text{sp}}| + d \leq M_N\} \leq \delta_k + \frac{M_N \delta_k}{d}. \quad (11)$$

The verified argument constructs a food-contracted word graph whose effective connections are realised by open two-reaction detours through short words. A contour of size r separating a molecule from the seed has kr disjoint detours available, so it fails with probability at most $(1 - b^2)^{kr}$; the per-molecule contour envelope has at most C_*^r members of size r ; summing over $r \geq 1$ bounds each molecule’s probability of being cut off by δ_k . If the root is not cut off, every molecule that is not cut off lies in its effective component, which lies in ordinary reversible closure. A union bound for the root and Markov’s inequality for the number of cut-off molecules give (11). The contour-envelope count and the disjoint-detour construction are theorems of the split development, not consequences of a numerical test; Appendix A lists their modules. Only (11) is used here.

Derivation of Lemma 4.1 from (11). Since $b > 0$, choose $k \geq 1$ with $\delta_k \leq 1/(m(m+1))$, take $L = 10k$ and $d = \lfloor M_N/m \rfloor + 1$. The event in (9) implies $|\mathcal{C}_{N,L}^{\text{sp}}| + d \leq M_N$, while $M_N/d \leq m$, so (11) gives probability at most $(m+1)\delta_k \leq 1/m$. $\square$

Lemma 4.2 (Quotient supplied-seed estimate). *For every $a_0 > 0$ and integer $m \geq 1$ there is $L > 0$ such that*

$$\mathbb{P}_a\{m|\mathcal{C}_{N,L}| < (m-1)M_N\} \leq \frac{1}{m} \quad \text{for all } a \in [a_0, 1] \text{ and } N \geq 1. \quad (12)$$

Lean: `quotient_supplied_seed_bulk`.

Proof. Apply Lemma 4.1 at $b = a_0/2$ to obtain L . By Lemma 3.1 the closure of the projected split family equals the quotient closure of the projected marks, so the failure event in (12) is the projection of the failure event in (9); it is a decreasing event. Lemma 3.3 at every $a \geq a_0$ (its increasing complement, with $a_0/2 \leq a/2$) gives $\mathbb{P}_a\{\text{failure}\} \leq \mathbb{P}_{a_0/2}^{\text{sp}}\{\text{failure}\} \leq 1/m$. The same L therefore works for all $a \geq a_0$ and all caps. $\square$

This is the only transfer from split to quotient channels in the entire proof. No quotient contour enumeration is needed.

Remark 4.3 (The cutoff is existential, not practical). The constant C_* in (10) makes the seed cutoff enormous. For $a_0 = 0.1$ and $m = 10$ one needs k with $C_*(1-b^2)^k < \eta/(1+\eta)$, $\eta = 1/110$, i.e. $k \geq 53667$ and $L = 536670$; for $a_0 = 0.5$ the values are $k = 2082$, $L = 20820$. These are sufficient theoretical cutoffs, not convergence rates. A finite certified bracket $\mathbb{P}_a(H_{N,L}) - 1/m \leq S_{\text{qt}}(a) \leq e_K(a)$ follows from the results below, but with these L the lower bracket is not computable by enumeration, and we do not replace it by an unsupported numerical curve.

5 Richness at the original openness

We now work in the infinite quotient field of Section 2 at a fixed openness $a > 0$. The argument uses only the coordinates of that field; no auxiliary sprinkling is introduced.

The infinite field is realised on the coordinate set $\{(w, i)\}$ of split descriptions, indexed by product word and split position, and the quotient coordinates are the canonical indices $(w, i) \mapsto (w, i')$ that reverse the split position exactly when the two factors commute and the right factor is shorter. Canonicalisation preserves the product word and hence preserves every finite prefix of the field indexed by product length; its restriction to J_n has the homogeneous product law on J_n .

Lean: `canonicalIndex_prefix`, `infinite_quotient_restriction_map`.

Lemma 5.1 (Survival generates every fixed word). *For $a > 0$ and every nonempty word v ,*

$$\mathbb{P}_a\{\mathcal{C}_\infty \text{ is unbounded and } v \notin \mathcal{C}_\infty\} = 0.$$

Lean: `quotient_own_cap_record`, `quotient_record_support`, `quotient_cylinder_deficit`, `quotient_missing_target_null`, `quotient_target_ae`.

Proof. Expose channels by the length of their product. On unbounded closure there are arbitrarily long words w generated using channels whose products have length at most $|w|$: take a finite derivation reaching beyond a given length and choose a longest word in it; that word is generated within its own cap. For each cutoff B , the first such *record* above B , together with its finite prefix configuration and a fixed choice of a record word, gives a countable partition of the survival event into disjoint finite-prefix record events.

Write $v = v_1 \cdots v_\ell$. From a record word w , append the target letters using the channels of

$$w + v_1 \rightleftharpoons wv_1, \quad wv_1 + v_2 \rightleftharpoons wv_1v_2, \quad \dots, \quad wv_1 \cdots v_{\ell-1} + v_\ell \rightleftharpoons wv,$$

and finally use $w + v \rightleftharpoons wv$ in the cleavage direction. Every letter is food, so if all of these channels are open then v is generated. After canonicalisation the support contains at most $\ell + 1$ distinct coordinates, each with product length strictly greater than $|w|$. Canonicalisation preserves the product, so every required coordinate lies outside the record prefix: the repair marks are *fresh*. Conditional on any record event, the repair succeeds with probability at least $a^{\ell+1}$.

To use freshness without assuming independence between successive records, let D be the event in the lemma and let C be any event determined by a finite product-length prefix. Partition $D \cap C$ according to the first record above that prefix. On each piece the repair coordinates are fresh relative to the piece's prefix and the repair must fail (since $v \notin \mathcal{C}_\infty$), so summing the disjoint pieces gives

$$\mathbb{P}_a(D \cap C) \leq (1 - a^{\ell+1}) \mathbb{P}_a(C). \quad (13)$$

Finite-prefix events generate the product σ -algebra, so D can be approximated in measure by such events C . Passing to the limit in (13) gives $\mathbb{P}_a(D) \leq (1 - a^{\ell+1}) \mathbb{P}_a(D)$, hence $\mathbb{P}_a(D) = 0$. $\square$

Survival therefore implies generation of every fixed finite seed with probability one. This is an almost-sure implication; it does not assert that every surviving configuration generates every word.

6 Finite approximation and continuity

For $L \geq 1$ let H_L be the event that every word of X_L is eventually generated from F in the infinite field, and let $H_{N,L}$ require generation within cap N , so that $H_{N,L} \uparrow H_L$ as $N \rightarrow \infty$. For $K \geq 2$ let E_K be the event that some word of length greater than K is generated. A first escape from X_K uses a ligation of two available words of length at most K , so its product has length at most $2K$; hence E_K is determined by the channels of J_{2K} even though closure is taken in the infinite field. Moreover $E_K \downarrow \{\mathcal{C}_\infty \text{ unbounded}\}$ as $K \rightarrow \infty$.

Write $h_{N,L}(a) = \mathbb{P}_a(H_{N,L})$ and $e_K(a) = \mathbb{P}_a(E_K)$. Each is a finite sum of terms $a^r(1-a)^{|J|-r}$ over configurations of a finite channel set and is therefore a polynomial in a . We have

$$h_{N,L}(a) \uparrow \mathbb{P}_a(H_L) \quad (N \rightarrow \infty), \quad e_K(a) \downarrow S_{\text{qt}}(a) \quad (K \rightarrow \infty), \quad (14)$$

and Lemma 5.1 gives $S_{\text{qt}}(a) \leq \mathbb{P}_a(H_L)$ for every $a > 0$ and L .

Lean: `quotient_finite_seed_tendsto`, `quotient_escape_tendsto`, `quotientSurvival_le_seed`, `continuous_quotient_event_probability`.

Lemma 6.1 (Uniform seed approximation). *For every $a_0 > 0$ and integer $m \geq 2$ there is $L > 0$ such that*

$$S_{\text{qt}}(a) \leq \mathbb{P}_a(H_L) \leq S_{\text{qt}}(a) + \frac{1}{m} \quad (a \in [a_0, 1]). \quad (15)$$

Lean: `quotient_same_field_seed_mass`, `quotient_uniform_seed_approximation`.

Proof. Choose L from Lemma 4.2. In the same field at cap N , if $H_{N,L}$ occurs then the closure of the supplied seed $X_L \cap X_N$ is contained in the closure of F , because every seed word has already been generated from F . Consequently

$$\mathbb{P}_a(H_{N,L}) \leq \mathbb{P}_a\{m |\mathcal{C}_{N,2}| \geq (m-1)M_N\} + \frac{1}{m}, \quad (16)$$

a union bound for two events in the same field; no independence is used. Fix K . For all sufficiently large N , a set of at least $(1 - 1/m)M_N$ words cannot fit inside X_K , so the mass event in (16) implies E_K . Letting $N \rightarrow \infty$ gives $\mathbb{P}_a(H_L) \leq e_K(a) + 1/m$, and letting $K \rightarrow \infty$ gives the upper bound in (15) by (14). The lower bound is Lemma 5.1. $\square$

Proposition 6.2 (Continuity of the survival function). *S_{qt} is continuous at every $a \in (0, 1]$ relative to $[0, 1]$.*

Lean: `quotientSurvival_tendsto`.

Proof. Let $a_i \rightarrow a > 0$ and $\varepsilon > 0$. Choose K with $e_K(a) < S_{\text{qt}}(a) + \varepsilon$. Since e_K is a polynomial and $S_{\text{qt}}(a_i) \leq e_K(a_i)$, $\limsup_i S_{\text{qt}}(a_i) \leq S_{\text{qt}}(a) + \varepsilon$.

For the lower bound take $a_0 = a/2$ and $m \geq 2$ with $1/m < \varepsilon$, and choose L from Lemma 6.1. Choose N with $h_{N,L}(a) > S_{\text{qt}}(a) - \varepsilon$, using (14) and richness. Eventually $a_i \geq a_0$, so

$$S_{\text{qt}}(a_i) \geq \mathbb{P}_{a_i}(H_L) - \frac{1}{m} \geq h_{N,L}(a_i) - \frac{1}{m},$$

and continuity of the polynomial $h_{N,L}$ gives $\liminf_i S_{\text{qt}}(a_i) \geq S_{\text{qt}}(a) - 2\varepsilon$. Let $\varepsilon \downarrow 0$. No uniform finite witness cap is asserted or needed. $\square$

7 From static closure to catalytic RAFs

We return to the finite catalytic model on $J_n \times X_n$. For a realised catalytic matrix ξ and a molecule set $C \subseteq X_n$ define the channels catalysed from C ,

$$A_\xi(C) = \{j \in J_n : \exists x \in C, \xi_{j,x} = 1\}.$$

Starting from all possible catalysts, *peel* by

$$A_0 = A_\xi(X_n), \quad C_t = C_n(A_t), \quad A_{t+1} = A_\xi(C_t). \quad (17)$$

The channel sets decrease and stabilise by time $|J_n|$; write A_* for the terminal set and $C_* = C_n(A_*)$ in this section. This is the standard RAF pruning algorithm of [9].

Lemma 7.1 (Terminal test). *Every RAF is contained in every A_t . Conversely, let A_*^u be the set of channels of A_* all of whose endpoints lie in C_* . A RAF exists if and only if $A_*^u \neq \emptyset$. In particular $|C_*| > 6$ is sufficient for a RAF.*

Lean: `quotient_raf_subset_peeling, quotient_terminal_iff, quotient_terminal_hasRAF`.

Proof. If A is a RAF then $A \subseteq A_0$, and if $A \subseteq A_t$ then $C_n(A) \subseteq C_t$, so every catalyst of A lies in C_t and $A \subseteq A_{t+1}$. Conversely, removing from A_* the channels with an endpoint outside C_* does not change the closure, because a channel used in a derivation has all of its endpoints in the final closure; so $C_n(A_*^u) = C_*$. Every channel of A_*^u has all endpoints in this closure and, by terminal stability $A_* = A_\xi(C_*)$, a catalyst in it. Hence a nonempty A_*^u is a RAF. If $|C_*| > 6$ then some nonfood word is generated, so some channel of A_* has all endpoints in C_* and $A_*^u \neq \emptyset$. $\square$

The terminal test retains RAFs whose closure is food; the sufficient condition $|C_*| > 6$ is used only for the lower bound and does not redefine the RAF event.

For a *fixed* catalyst pool C of size k , distinct channels are catalysed independently with probability

$$q(p, k) = 1 - (1 - p)^k. \quad (18)$$

The pools in (17) are adaptive, not fixed. The following identity is what permits the comparison.

Lemma 7.2 (Complete-history replacement). *Fix an ordering of X_n and let I_k be its first k elements. The entire finite sequence of channel sets in (17) has the same joint law if $A_\xi(C)$ is replaced at every step by $A_\xi(I_{|C|})$.*

Lean: `quotient_column_history, quotient_history_map, quotient_pool_map`.

Proof. Fix a proposed decreasing history $A_0 \supseteq A_1 \supseteq \dots \supseteq A_T$. Its prescribed pools are deterministic and nested: $D_0 = X_n$ and $D_t = \mathcal{C}_n(A_{t-1})$ for $t \geq 1$. Put $s = 1 - p$ and $k_t = |D_t|$. For a single channel j , the column $(\xi_{j,x})_{x \in X_n}$ determines the first time d at which j leaves the history, namely the first t with $A_\xi(D_t) \not\ni j$; the probability of each value is

$$\begin{cases} s^{k_0}, & d = 0, \\ s^{k_d} - s^{k_{d-1}}, & 1 \leq d \leq T, \\ 1 - s^{k_T}, & d = T + 1 \text{ (catalysed from every pool)}. \end{cases} \quad (19)$$

Nestedness gives the middle line by subtraction: having no catalyst in D_{d-1} implies having none in D_d . The event that the realised history equals the proposed one is the intersection over channels of the corresponding column constraints; columns are independent, so its probability is the product of the weights (19). Replacing the pools D_t by the prefixes I_{k_t} preserves their cardinalities and their nesting, hence every weight. Histories that are not decreasing have probability zero under both processes. Equality of the probabilities of all finite histories proves the claim. The proof does not condition the original columns on a random pool and then assert independence; it compares the two joint laws directly. $\square$

Lemma 7.3 (Static mass barrier). *If $6 < k \leq M_n$ then*

$$\mathbb{P}_{q(p,k)}\{|\mathcal{C}_{n,2}| \geq k\} \leq P_n(p). \quad (20)$$

Lean: `quotient_static_barrier, quotient_mass_ge_static`.

Proof. Consider the prefix process of Lemma 7.2 and suppose the fixed open family $A_\xi(I_k)$ generates at least k molecules from food. It is contained in $A_0 = A_\xi(X_n)$. Whenever it is contained in the current active family, the current closure has size at least k , so the next prefix pool contains I_k and the family is contained in the next active family. By induction it is contained in every active family, so the terminal closure has at least $k > 6$ elements and Lemma 7.1 yields a RAF. Apply Lemma 7.2 to transfer the probability to the original process, and note that $A_\xi(I_k)$ has the fixed-pool law (18). $\square$

For $m \geq 2$ put $k_{n,m} = (m-1)\lfloor M_n/m \rfloor$. If $q \leq q(p, k_{n,m})$ and $k_{n,m} > 6$, monotonicity of product measures in the parameter gives

$$\mathbb{P}_q\{m|\mathcal{C}_{n,2}| \geq (m-1)M_n\} \leq P_n(p), \quad (21)$$

since the mass threshold implies $|\mathcal{C}_{n,2}| \geq k_{n,m}$ and (20) applies.

Lean: `quotient_near_full_lower`.

8 The probability sandwich

Assume $p_n \in [0, 1]$ and (8), and let $a = 1 - e^{-\lambda} \in (0, 1)$. The elementary binomial limit gives

$$q(p_n, M_n) \longrightarrow a, \quad (22)$$

$$q(p_n, k_{n,m}) \longrightarrow 1 - e^{-(1-1/m)\lambda} \quad (m \geq 2 \text{ fixed}), \quad (23)$$

because $\log(1 - p_n) = -p_n + O(p_n^2)$ and $M_n p_n^2 = (M_n p_n) p_n \rightarrow 0$; the rounding in $k_{n,m}$ changes the exponent by a vanishing amount.

Lean: `generic_pool_limit, generic_near_pool_mass`.

8.1 Lower bound

Fix $q \in (0, a)$ and $m \geq 2$. Richness, finite seed approximation and Lemma 4.2 imply

$$\liminf_{n \rightarrow \infty} \mathbb{P}_q\{m |\mathcal{C}_{n,2}| \geq (m-1)M_n\} \geq S_{\text{qt}}(q) - \frac{1}{m}. \quad (24)$$

Indeed, take the L of Lemma 4.2 for $a_0 = q$, apply the same-field inclusion (16) at cap n , let $n \rightarrow \infty$ using $h_{n,L}(q) \uparrow \mathbb{P}_q(H_L)$, and use $\mathbb{P}_q(H_L) \geq S_{\text{qt}}(q)$.

Choose m so large that $q < 1 - e^{-(1-1/m)\lambda}$. By (23) and $k_{n,m} \rightarrow \infty$, inequality (21) applies for all large n , so $\liminf_n P_n(p_n) \geq S_{\text{qt}}(q) - 1/m$. This holds for all sufficiently large m , hence $\liminf_n P_n(p_n) \geq S_{\text{qt}}(q)$. Finally let $q \uparrow a$ and use Proposition 6.2:

$$\liminf_n P_n(p_n) \geq S_{\text{qt}}(a). \quad (25)$$

8.2 Upper bound

Lemma 8.1 (The 34 gateways). *For $n \geq 4$ there are exactly 34 food gateways: four with a product of length two, fourteen with a product of length three, and sixteen with a product of length four. Every nonempty food-generated channel family contains a gateway.*

Proof. A gateway has a side wholly in F . If the product side is in F the product has length two and both factors are letters: four channels. Otherwise both factors are food words. Two letters and a letter with a dimer give the sixteen ordered length-three descriptions, among which $(A, AA) \sim (AA, A)$ and $(B, BB) \sim (BB, B)$ are the only commuting pairs, leaving fourteen channels. Two dimers give sixteen ordered descriptions; equal-length commuting factors are equal (Lemma 3.4), so no two distinct descriptions merge and all sixteen are channels. A dimer with a trimer is not a gateway. For the second assertion, examine the first channel of the family that is enabled in a derivation of its own closure from food; its enabling side lies in F . If no derivation step is ever needed, all endpoints already lie in F and any member is a gateway. $\square$

The count 34 was established for the generator's catalogue in [1] by kernel evaluation, and the split-position count is 36 [2]; the two missing channels are exactly the two identifications at product length three. A RAF supplies a catalyst for its gateway inside its own closure, so $\{\text{RAF}\} \subseteq G_n$, as used in (3).

Fix $K \geq 2$. The closure of the full-pool active family $A_\xi(X_n)$ either escapes X_K , or it does not; in the second case every RAF, being contained in A_0 by Lemma 7.1, has closure inside X_K , so its gateway has a catalyst of length at most K . Therefore

$$P_n(p_n) \leq e_K(q(p_n, M_n)) + 34 M_K p_n. \quad (26)$$

The first term uses the fixed-pool law (18) of $A_\xi(X_n)$ and the fact that a first escape from X_K is determined by channels of product length at most $2K$; the second is a union bound over at most 34 gateways and at most M_K catalysts of bounded length. No independence is asserted after conditioning on failure to escape.

Lean: `quotient_bounded_seed_le, quotient_raf_upper`.

For fixed K , continuity of the polynomial e_K , the limit (22) and $p_n \rightarrow 0$ give $\limsup_n P_n(p_n) \leq e_K(a)$. Letting $K \rightarrow \infty$ and using (14),

$$\limsup_n P_n(p_n) \leq S_{\text{qt}}(a). \quad (27)$$

Together with (25) this proves the convergence in (4) for every sequence satisfying (8).

Lean: `quotient_critical_limit`.

8.3 Positivity, the actual catalogue, and the conditional limit

The comparison of Lemma 3.3 passes from the increasing finite events E_K to their decreasing limit, so

$$S_{\text{qt}}(a) \geq S_{\text{sp}}(a/2) > 0 \quad (a > 0), \quad (28)$$

where S_{sp} is the split-position survival function of [2]. Positivity of S_{sp} at every positive openness is a theorem of that development; it can also be read off Lemma 4.1: at a fixed positive openness, $m = 2$ and each finite escape cutoff, a finite supplied seed survives with probability at least $1/2$, forcing the finitely many channels that generate that seed from F to be open has positive probability, conditioning finitely many independent coordinates to be open can only increase supplied-seed survival, and on that event the food closure contains the seed closure.

Lean: `split_half_survival_le_quotient`, `quotientSurvival_pos`, `quotientTransitionLimit_positive`.

The probability $P_n(p)$ above is defined through a channel-indexed Boolean matrix. The generator, and the formal development of [1], describe a configuration by its family of catalyst subsets $(A_j)_{j \in J_n}$ and assign a configuration with z present coordinates the weight $p^z(1 - p)^{M_n|J_n|-z}$. The two descriptions are related by an explicit bijection under which the RAF predicates agree, and the weights of corresponding singletons coincide; summing them identifies the two finite probabilities.

Lean: `quotient_probability_eq_catalogue`.

Finally, the gateway coordinates are 34 disjoint columns of M_n independent marks, so

$$\mathbb{P}_{p_n}(G_n) = 1 - (1 - p_n)^{34M_n} \longrightarrow 1 - e^{-34\lambda} > 0$$

by (8). Dividing the proved limit (4) by this positive limit gives (5). Lemma 3.5 discharges (8) for $p_n = f_n/|J_n|$, which completes the proof of Theorem 2.2.

Lean: `repository_gateway_critical_window`, `repository_bulk_factor_critical_window`.

9 Comparisons, caps, and regimes

This section records what the proof gives beyond the limit itself. Results marked “Lean” are compiled declarations; those marked “conventional” are short hand proofs, and we keep the two carefully apart.

9.1 Monotonicity

Proposition 9.1. *S_{qt} is nondecreasing on $[0, 1]$, and $\Theta_{\text{qt}}(\lambda) = S_{\text{qt}}(1 - e^{-\lambda})$ is nondecreasing on $(0, \infty)$.*

Lean: `quotientSurvival_mono`, `quotientTransitionLimit_mono`.

Proof. Coordinatewise uniform coupling makes each finite escape event E_K increasing in a ; pass to the decreasing limit in K . Compose with the increasing map $\lambda \mapsto 1 - e^{-\lambda}$. $\square$

9.2 Two-sided comparison with the split-position model

Proposition 9.2 (Lower comparison). *For $a \in [0, 1]$, $S_{\text{sp}}(1 - \sqrt{1 - a}) \leq S_{\text{qt}}(a)$. Consequently $\Theta_{\text{sp}}(\lambda/2) = S_{\text{sp}}(1 - e^{-\lambda/2}) \leq \Theta_{\text{qt}}(\lambda)$.*

Lean: `OR_sharp_domination`, `split_sharp_survival_le_quotient`.

Proof. At split openness $b = 1 - \sqrt{1 - a}$ a double fibre projects to OR probability $1 - (1 - b)^2 = a$ and a singleton to $b \leq a$, so Lemma 3.3 in its sharp form dominates the projected finite escape events by the quotient ones; let $K \rightarrow \infty$. Since $1 - \sqrt{e^{-\lambda}} = 1 - e^{-\lambda/2}$, the second statement follows. $\square$

Proposition 9.3 (Upper comparison). *For $a \in [0, 1]$, $S_{\text{qt}}(a) \leq S_{\text{sp}}(a)$, hence $\Theta_{\text{qt}}(\lambda) \leq \Theta_{\text{sp}}(\lambda)$. Conventional proof; not formalised.*

Proof. Couple the quotient field with a split field of the same openness by reading, for each channel, the split mark of its canonical representative. Representatives are distinct, so these are independent Bernoulli(a) coordinates with the quotient law. Every open quotient channel then has an open split representative with the same two chemical sides, so every quotient derivation is pointwise a split derivation and $\mathcal{C}_{\infty}^{\text{qt}} \subseteq \mathcal{C}_{\infty}^{\text{sp}}$ configuration by configuration. Survival is preserved. $\square$

The two comparisons say that the quotient profile is sandwiched between the split profile at half intensity and at full intensity. Whether $\Theta_{\text{qt}} < \Theta_{\text{sp}}$ strictly is discussed in Section 11. Neither comparison orders the gateway-normalised factors, whose denominators $1 - e^{-34\lambda}$ and $1 - e^{-36\lambda}$ differ.

9.3 Gateway caps and endpoints

Proposition 9.4 (Caps). *For every $\lambda > 0$,*

$$\Theta_{\text{qt}}(\lambda) \leq 1 - e^{-34\lambda} \quad \text{and} \quad \Theta_{\text{qt}}(\lambda) \leq 1 - e^{-30\lambda}.$$

The first bound combines two compiled statements, $P_n(p) \leq 1 - (1 - p)^{34M_n}$ from [1] (`repositoryRAFBernoulliProbability_le_gateway`) and the gateway limit of Theorem 2.2. The second follows from the conventional infinite-survival bound $S_{\text{qt}}(a) \leq 1 - (1 - a)^{30}$. Conventional proof; not formalised.

Proof. The first bound is immediate from the two cited limits. For the second, unbounded closure requires the generation of a nonfood word. Cleavage of a food word yields food words, so the first nonfood word ever produced is the product of a ligation whose two factors are food: a gateway with nonfood product. By Lemma 8.1 there are exactly $14 + 16 = 30$ such *productive* gateways. If all thirty are closed, the closure remains equal to F . Their marks are independent, so $S_{\text{qt}}(a) \leq 1 - (1 - a)^{30}$; substitute $a = 1 - e^{-\lambda}$. $\square$

The finite RAF event retains food-only RAFs, so the 30-gateway bound is a statement about infinite survival and about the limit, not about $P_n(p)$ at finite n , for which the 34-gateway cap of [1] is the correct statement. For the conditional limit, Proposition 9.4 gives

$$\lim_n B_n(p_n) \leq \frac{1 - e^{-30\lambda}}{1 - e^{-34\lambda}}, \quad \limsup_{\lambda \downarrow 0} \lim_n B_n(p_n) \leq \frac{30}{34}.$$

No equality and no small- λ expansion is asserted.

Proposition 9.5 (Endpoints). *$S_{\text{qt}}(0) = 0$ and $S_{\text{qt}}(1) = 1$. Consequently $\Theta_{\text{qt}}(\lambda) \rightarrow 0$ as $\lambda \downarrow 0$ and $\Theta_{\text{qt}}(\lambda) \rightarrow 1$ as $\lambda \rightarrow \infty$. Conventional proof; not formalised.*

Proof. At $a = 0$ no channel is open and $\mathcal{C}_{\infty} = F$. At $a = 1$ every channel is open; by induction on length, a word of length $\ell + 1$ is the ligation of its first letter with its length- ℓ suffix, both available, so every word is generated and closure is unbounded. The limit at $\lambda \downarrow 0$ follows from the cap $1 - e^{-34\lambda}$; the limit at $\lambda \rightarrow \infty$ follows from $S_{\text{qt}}(1) = 1$ and continuity of S_{qt} at $a = 1$ (Proposition 6.2, Lean), since $1 - e^{-\lambda} \rightarrow 1$. $\square$

9.4 Outer scaling regimes

For the two regimes outside the linear window the parameter $f_n/|J_n|$ need not lie in $[0, 1]$ without an explicit assumption; $f_n/n \rightarrow \infty$ does not by itself give $f_n \leq |J_n|$, and $f_n/n \rightarrow 0$ does not give $f_n \geq 0$. We therefore assume eventual validity.

Corollary 9.6 (Sublinear and superlinear catalysis). *Let (f_n) satisfy $0 \leq f_n \leq |J_n|$ for all n .*

(i) *If $f_n/n \rightarrow 0$ then $P_n(p_n) \rightarrow 0$.*

(ii) *If $f_n/n \rightarrow \infty$ then $P_n(p_n) \rightarrow 1$.*

Part (i) is a theorem of [1]; part (ii) is a conventional proof.

Lean: `sublinear_catalysis_forces_actual_repository_raf_vanishing` for (i).

Proof. (i) is the gateway union bound $P_n(p_n) \leq 34M_n p_n$ together with Lemma 3.5. For (ii) fix $c > 0$. Eventually $f_n \geq cn$, so $p_n \geq cn/|J_n| =: p'_n$, and $P_n(\cdot)$ is nondecreasing in the parameter by coordinatewise coupling. Theorem 2.2 applied to $f'_n = cn$ gives $\liminf_n P_n(p_n) \geq \Theta_{\text{qt}}(c)$. Let $c \rightarrow \infty$ and apply Proposition 9.5. No assumption $p_n \rightarrow 0$ is used in (ii). $\square$

9.5 Independent heterogeneous catalysis

The general limit theorem behind Theorem 2.2 takes an arbitrary parameter sequence with $p_n \rightarrow 0$ and $p_n M_n \rightarrow \lambda$ (`quotient_critical_limit`). Combined with coupling it gives a robustness statement for heterogeneous but independent catalysis.

Corollary 9.7 (Envelope bounds). *Keep the chemical model and let the coordinates $\xi_{j,x}$ be independent with $\mathbb{P}(\xi_{j,x} = 1) = p_{n,j,x}$, where $\ell_n \leq p_{n,j,x} \leq u_n$ for all j, x with $\ell_n M_n \rightarrow \lambda_-$ and $u_n M_n \rightarrow \lambda_+$ in $(0, \infty)$. Then*

$$S_{\text{qt}}(1 - e^{-\lambda_-}) \leq \liminf_n \mathbb{P}(\text{RAF}_n) \leq \limsup_n \mathbb{P}(\text{RAF}_n) \leq S_{\text{qt}}(1 - e^{-\lambda_+}).$$

Conventional proof; not formalised.

Proof. Couple the three independent product fields through the same uniforms; RAF existence is increasing in the configuration, so $P_n(\ell_n) \leq \mathbb{P}(\text{RAF}_n) \leq P_n(u_n)$. Both envelopes tend to zero with $\ell_n M_n, u_n M_n$ convergent, so the general limit theorem applies to each. $\square$

The corollary covers uniformly small perturbations of the homogeneous law. It does not cover correlated catalytic marks, nor heterogeneity in the channel set itself; coordinate independence is essential to the coupling.

10 Interpretation for reaction-network analysis

10.1 Descriptions, channels, and probability laws

The distinction between a reaction description and a reaction channel matters already in the smallest network. Consider one catalyst x and the two descriptions of $A + AA \rightleftharpoons AAA$. If a database lists both, and a random-catalysis model assigns independent marks with probability p to each listed description, then the chemically correct merged channel is catalysed by x with probability $2p - p^2$: $3/4$ at $p = 1/2$. If instead one first merges the descriptions and then assigns probability p to the merged channel, the value is $1/2$. These are different probabilistic models on the same chemical channel, and Example 3.2 shows that the difference propagates to the RAF probability. If, on the other hand, the two records are duplicate observations of the *same* catalytic event, their marks coincide and the merged probability remains p ; the

Table 1: Exact channel census and finite-size openness. $M_n = |X_n|$, $|R_n|$ split descriptions, $|J_n|$ quotient channels; $q_n(\lambda) = 1 - (1 - \lambda n/|J_n|)^{M_n}$ is the full-pool channel openness at $f_n = \lambda n$, whose limits are $1 - e^{-0.1} = 0.095163$ and $1 - e^{-1} = 0.632121$. Counts for $n \leq 12$ are by enumeration; larger n use the closed form. Values are exact formula evaluations, not RAF probabilities.

n	M_n	$ R_n $	$ J_n $	$nM_n/ J_n $	$q_n(0.1)$	$q_n(1)$
4	30	68	64	1.87500	0.171459	0.855743
8	510	3076	3048	1.33858	0.125301	0.738244
12	8190	81924	81828	1.20106	0.113174	0.699150
16	131070	1835012	1834798	1.14297	0.108007	0.681131
20	2097150	37748740	37748322	1.11112	0.105162	0.670811
24	33554430	738197508	738196506	1.09091	0.103351	0.664089

OR formula with independent inputs must not be applied to duplicated observations without examining their dependence.

For a network analyst the immediate lesson is a preprocessing discipline: specify chemical equivalence of reaction descriptions first; determine which catalytic records represent independent events; and only then assign a probability law on the merged channels. In a biochemical database this requires preserving species identity, stoichiometry, direction, and compartment wherever these distinguish processes; AB and BA must remain distinct products. The binary-word quotient is a precise model of one such operation. It is not a universal equivalence rule for biochemical databases, and Theorem 2.2 is a theorem about one model, not a transfer principle for every convention.

10.2 A reproducible normalisation example

Table 1 and Figure 2 illustrate the normalisation used by the theorem, using exact counts only. The channel count $|J_n|$ is obtained for $n \leq 12$ by enumerating words and canonicalising commuting swapped factors, and for larger n by the closed form of Remark 3.6; the two agree wherever both are available. The displayed openness $q_n = 1 - (1 - p_n)^{M_n}$ is the exact full-pool channel probability (18) evaluated numerically; it is *not* an estimate of the RAF probability, and no sampling is involved. Its limit is $1 - e^{-\lambda}$, the argument of S_{qt} in Theorem 2.2. The finite-size correction is material: at $n = 12$ and $\lambda = 1$ the openness is 0.699 against the limit 0.632, because $nM_n/|J_n|$ approaches 1 only like $1 + 2/(n - 2)$. Any comparison between finite- n simulations and the limiting profile must account for this normalisation before invoking S_{qt} .

10.3 What the theorem does and does not say about chemistry

Cooperative RNA catalytic networks [26] are an experimental motivation for studying the organisation of catalysis, and RAF theory has been used to interpret them [15]. Theorem 2.2 does not calibrate such an experiment. The model has a fixed six-word food set, binary sequences, reversible set-valued availability, and independent catalytic marks; it omits concentrations, rates, inhibition, degradation, and any energy budget. Structural RAF existence therefore does not imply positive net flux, kinetic persistence, or thermodynamic feasibility; the extensions to dynamics and inhibition [10] and to partitioned networks [22] make these distinctions explicit. The intensity λ is dimensionless and carries no implied rate or concentration unit. Applying the result to a particular chemical system would require a separate model-to-data calibration, which we do not attempt.

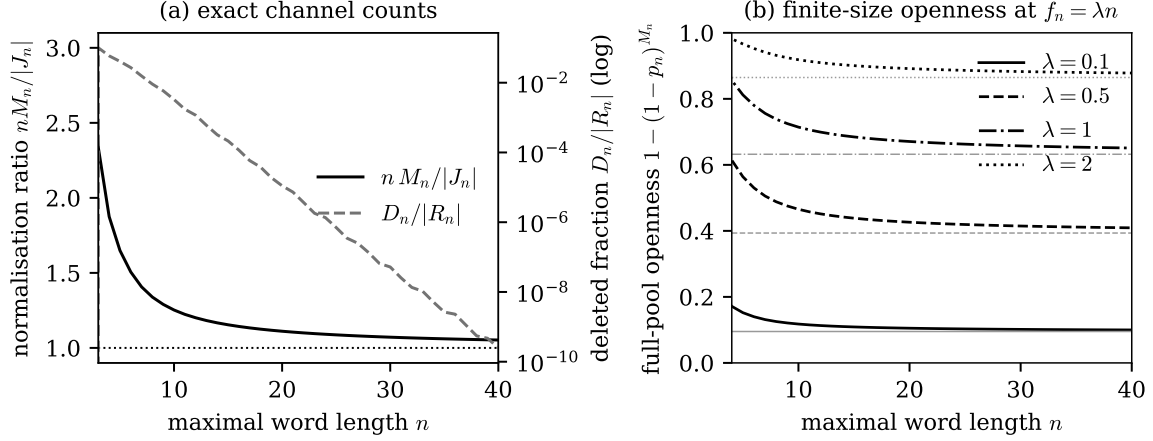


Figure 2: (a) The normalisation ratio $nM_n/|J_n|$ (solid, left axis) and the fraction $D_n/|R_n|$ of split descriptions deleted by the quotient (dashed, right axis, logarithmic), from the exact counts. (b) The full-pool channel openness $1 - (1 - p_n)^{M_n}$ at $f_n = \lambda n$ for four intensities, with the limits $1 - e^{-\lambda}$ as thin horizontal lines. Both panels are exact evaluations; neither is a RAF probability.

11 Discussion

What was resolved

The companion paper [1] isolated all unresolved linear-scale behaviour of the generator’s model in the bulk factor B_n and asked whether the split-position and quotient conventions share a bulk limit. Theorem 2.2 determines $\lim B_n = S_{\text{qt}}(1 - e^{-\lambda})/(1 - e^{-34\lambda})$, and Propositions 9.2 and 9.3 locate Θ_{qt} between $\Theta_{\text{sp}}(\lambda/2)$ and $\Theta_{\text{sp}}(\lambda)$. The two profiles therefore have the same qualitative shape, a continuous nondecreasing curve from 0 to 1 across the whole linear scale, and the finite step conjecture of [11, Eq. (21)] fails in the quotient convention for the same reason as in the split convention: no finite λ separates limit zero from limit one.

Why a direct proof was necessary

Every deterministic object transfers across the quotient: closures, RAF witnesses, caps, and record structure. What does not transfer is the measure. The failed shortcut, which we record because it is tempting, is to argue that the deleted fraction $D_n/|R_n| = O(n 2^{-2n/3})$ is negligible and hence that the two homogeneous RAF probabilities have the same limit. The OR law shows the flaw: the split ensemble seen on the quotient channels is a *different* product measure, with parameter $2p - p^2$ on every commuting pair, and Example 3.2 shows that the difference reaches the RAF event. A negligible count ratio does not control a difference of measures whose supports are the same channel set. The proof therefore re-establishes richness, uniform approximation, continuity, and the sandwich in the quotient model, importing only one inequality from the split model.

Open problems

- (1) *Strictness.* Is $\Theta_{\text{qt}}(\lambda) < \Theta_{\text{sp}}(\lambda)$ for every $\lambda > 0$? A conventional route couples the quotient field (canonical marks only) with the full split field, forces the noncanonical description $\text{AA} + \text{A} \rightleftharpoons \text{AAA}$ open and the other productive split gateways closed, so that the quotient is trapped at food while the split field escapes through AAA ; a finite

compatible repair support then generates a supplied seed from AAA, and Lemma 4.1 gives positive split survival on this event. The exact geometry of the trap has been checked by enumeration (a repair support of 246 split descriptions at seed cutoff 6), but the formal docking of the coupling, the trap and the forcing step is incomplete, and we make no strictness claim. Strictness of the total survival probabilities would not by itself order the gateway-normalised factors.

- (2) *Rates.* Theorem 2.2 has no finite-size convergence rate, and the cutoffs of Remark 4.3 show that the present supplied-seed constant cannot supply a useful one. A contour estimate with a realistic constant, or a different amplification mechanism, would make finite certified brackets computable.
- (3) *The value of S_{qt} .* Nothing here identifies $S_{\text{qt}}(a)$ beyond the bounds $S_{\text{sp}}(1 - \sqrt{1 - a}) \leq S_{\text{qt}}(a) \leq \min\{S_{\text{sp}}(a), 1 - (1 - a)^{30}\}$ and the endpoints. Rigorous numerical bounds for S_{qt} or S_{sp} at a given openness are open.
- (4) *Formalisation gaps.* The upper comparison, the productive-gateway cap, the endpoint values, and the superlinear regime are conventional proofs; each is short and should be formalised.
- (5) *Beyond independence.* Corollary 9.7 handles independent heterogeneous marks. Correlated catalysis, such as the specificity-structured catalysis of [15] or power-law catalysis distributions [11], is outside the product-measure method.

Data and code availability

The Lean sources are in the public repository <https://github.com/mmislán/Erdos> under `proofs/RAFReactionQuotient/` (this paper), `proofs/HordijkSteelThreshold/` (split-position development) and `proofs/OverlapCorrectedRAF/` (quotient model and gateway). The exact census, the fixture enumeration of Example 3.2, the verification receipts, and the script producing Table 1 and Figure 2 are under `problem_workspaces/RAF_critical_window_reaction_quotient/` and `key_results/RAFs/`.

A Formal verification

A.1 The root declaration

The main theorem is the Lean 4 [3] declaration

`RAFReactionQuotient.reaction_quotient_critical_window_resolution`

in `proofs/RAFReactionQuotient/Resolution.lean`. Its universal parameters are $f : \mathbb{N} \rightarrow \mathbb{R}$ and $\lambda : \mathbb{R}$; its only hypotheses are $0 < \lambda$ and $f(n)/n \rightarrow \lambda$. Its conclusion is the conjunction of four statements.

- (i) For all large n the model parameter equals $f(n)/|J_n|$ exactly.
Lean: `quotientParameter_eventually_exact`.
- (ii) The generator’s RAF probability converges to the survival probability of the infinite canonical field at openness $1 - e^{-\lambda}$.
Lean: `repositoryRAFBernoulliProbability, quotientTransitionLimit, repository_critical_window`.
- (iii) That limit is positive.
Lean: `quotientTransitionLimit_positive`.

- (iv) The gateway-conditioned factor of [1] converges to the quotient of the limit by $1 - e^{-34\lambda}$.

Lean: repositoryGatewayConditionedBulkProbability, repository_bulk_factor_critical_window.

The finite RAF predicate, the channel type `RepositoryChannel`, the Bernoulli catalogue, and the gateway factor are imported unchanged from the development of [1]; the supplied-seed estimate is imported unchanged from the development of [2].

A.2 Toolchain and audit

Verification uses Lean v4.30.0 and the frozen Mathlib [25] manifest checked into the repository (SHA-256 a8df0b60...). The strict receipt records successful compilation of the root file with warnings promoted to errors, empty compiler output, and the authenticated declaration interface; the root source hash is ca3264cf... and the receipt hash 0f091921.... The axiom closure of the root is {propext, Classical.choice, Quot.sound}; there are no sorrys, no problem-specific axioms, and no native unchecked computation. An independent freshness audit checks the source and compiled-artefact hashes of the 197 recorded dependency modules and confirms that all 32 local modules are covered by the root’s dependency closure or by a successful supplementary receipt. The audit checks saved evidence and is not a substitute for the kernel check; formal verification establishes the encoded statement, and the model definitions of Section 2 are what make that statement scientifically interpretable.

The supplementary module `PublicationConsequences.lean` compiles Propositions 9.1 and 9.2 (quotientSurvival_mono, quotientTransitionLimit_mono, OR_sharp_domination, split_sharp_survival_le_quotient) with its own successful receipt. Propositions 9.3, 9.4 (second bound) and 9.5, Corollary 9.6(ii), and Corollary 9.7 are conventional and are labelled as such in the text.

A.3 Concordance

Table 2 maps the results of the paper to their source modules. Modules in the main block are in `proofs/RAFReactionQuotient/`, and the model, catalogue and gateway are in `proofs/OverlapCorrectedRAF/`. The split supplied-seed estimate is in the directory `proofs/HordijkSteelThreshold/`, in the modules `StaticBulk` and `StaticClosureMass`, whose geometric inputs are the modules

Lean: WordMoleculeContourCount, WordContourProbability, WordContourMarginals, WordEffectiveClosure.

A.4 Reproduction

From the repository root, with the pinned Lean working directory enforced by the wrapper (line breaks are for display only):

```
.venv/Scripts/python.exe scripts/verify_proof.py
proofs/RAFReactionQuotient/Resolution.lean -timeout 900
-declaration RAFReactionQuotient.reaction_quotient_critical_window_resolution
-output problem_workspaces/RAF_critical_window_reaction_quotient/
verification/Resolution.verify.json

.venv/Scripts/python.exe
problem_workspaces/RAF_critical_window_reaction_quotient/audit_resolution.py
```

The first command recompiles the root and its dependency closure and writes a receipt; the second replays the freshness audit. The table and figure of Section 10 are produced by `make_figures.py` in the manuscript source directory, which also re-derives the census of the workspace pilot and asserts agreement between enumeration and closed form. No dependency update is needed to reproduce the proof.

Table 2: Formal locations of the proof steps. Module names are files under the directories named in the text; declaration names are given beneath each result in the body of the paper.

Result	Modules
Lemma 3.1 (closure and OR projection)	ClosureProjection, QuotientMap, SplitAdapter, Fibres, Surjectivity, CapCompatibility, ORLaw, ORDistribution, CatalyticOR
Lemma 3.3 (domination)	ProductDomination, PublicationConsequences
Lemmas 3.4–3.5 (count, normalisation)	CommutingUniqueness, CountLoss, Counts, SourceParameters
Lemma 4.1 (split supplied seed)	HordijkSteelThreshold/StaticBulk, StaticClosureMass
Lemma 4.2 (quotient supplied seed)	SeedBulk
Infinite law, records, Lemma 5.1 (14), Lemma 6.1, Prop. 6.2	InfiniteSource, InfiniteDock, Richness FiniteApproximation, EventContinuity, SeedApproximation, SurvivalContinuity
Lemmas 7.1–7.3 (peeling)	History, Terminal, PoolLaw, SourceBounds
Section 8 (sandwich, positivity, catalogue)	PoolLimits, CriticalLimit, SurvivalPositive, CatalogueBridge, Resolution
Propositions 9.1, 9.2	PublicationConsequences

References

- [1] [Author name to be inserted]. Exact overlap partitions and gateway scaling in Kauffman RAF networks. Companion manuscript, 2026. Defines the commuting-factor quotient of reaction channels, proves the exact 34-channel food gateway and the bound $\Pr(\text{RAF}_n) \leq 1 - (1 - p_n)^{34|X_n|}$, and leaves the gateway-conditioned bulk factor open, 2026.
- [2] [Author name to be inserted]. A nontrivial critical window for RAF emergence in the binary polymer model. Companion manuscript, 2026. Split-position reaction catalogue; proves that the RAF probability converges to $S(1 - e^{-\lambda})$ when $f_n/n \rightarrow \lambda$, with $0 < S(1 - e^{-\lambda}) \leq 1 - e^{-36\lambda}$, 2026.
- [3] 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.
- [4] Manfred Eigen. Selforganization of matter and the evolution of biological macromolecules. *Naturwissenschaften*, 58(10):465–523, 1971.
- [5] Geoffrey Grimmett. *Percolation*, volume 321 of *Grundlehren der mathematischen Wissenschaften*. Springer, Berlin, 2nd edition, 1999.
- [6] Wim Hordijk, Jotun Hein, and Mike Steel. Autocatalytic sets and the origin of life. *Entropy*, 12(7):1733–1742, 2010.
- [7] 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.
- [8] Wim Hordijk, Joshua I. Smith, and Mike Steel. Algorithms for detecting and analysing autocatalytic sets. *Algorithms for Molecular Biology*, 10:15, 2015.
- [9] Wim Hordijk and Mike Steel. Detecting autocatalytic, self-sustaining sets in chemical reaction systems. *Journal of Theoretical Biology*, 227(4):451–461, 2004.
- [10] Wim Hordijk and Mike Steel. Autocatalytic sets extended: Dynamics, inhibition, and a generalization. *Journal of Systems Chemistry*, 3:5, 2012.
- [11] Wim Hordijk and Mike Steel. Autocatalytic sets in polymer networks with variable catalysis distributions. *Journal of Mathematical Chemistry*, 54(10):1997–2021, 2016.

- [12] Wim Hordijk and Mike Steel. Chasing the tail: The emergence of autocatalytic networks. *Biosystems*, 152:1–10, 2017.
- [13] Wim Hordijk and Mike Steel. Autocatalytic networks at the basis of life’s origin and organization. *Life*, 8(4):62, 2018.
- [14] Wim Hordijk, Mike Steel, and Stuart Kauffman. The structure of autocatalytic sets: Evolvability, enablement, and emergence. *Acta Biotheoretica*, 60(4):379–392, 2012.
- [15] Wim Hordijk, Peter R. Wills, and Mike Steel. Autocatalytic sets and biological specificity. *Bulletin of Mathematical Biology*, 76(1):201–224, 2014.
- [16] Stuart A. Kauffman. Autocatalytic sets of proteins. *Journal of Theoretical Biology*, 119(1):1–24, 1986.
- [17] Stuart A. Kauffman. *The Origins of Order: Self-Organization and Selection in Evolution*. Oxford University Press, New York, 1993.
- [18] M. Lothaire. *Combinatorics on Words*. Cambridge Mathematical Library. Cambridge University Press, Cambridge, 1997.
- [19] Roger C. Lyndon and Marcel-Paul Schützenberger. The equation $a^m = b^n c^p$ in a free group. *Michigan Mathematical Journal*, 9(4):289–298, 1962.
- [20] Elchanan Mossel and Mike Steel. Random biochemical networks: the probability of self-sustaining autocatalysis. *Journal of Theoretical Biology*, 233(3):327–336, 2005.
- [21] Rudolf Peierls. On Ising’s model of ferromagnetism. *Mathematical Proceedings of the Cambridge Philosophical Society*, 32(3):477–481, 1936.
- [22] Joshua I. Smith, Mike Steel, and Wim Hordijk. Autocatalytic sets in a partitioned biochemical network. *Journal of Systems Chemistry*, 5:2, 2014.
- [23] Mike Steel. The emergence of a self-catalysing structure in abstract origin-of-life studies. *Applied Mathematics Letters*, 13(3):91–95, 2000.
- [24] 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.
- [25] 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.
- [26] Niles Vaidya, Michael L. Manapat, Irene A. Chen, Ramon Xulvi-Brunet, Eric J. Hayden, and Niles Lehman. Spontaneous network formation among cooperative RNA replicators. *Nature*, 491(7422):72–77, 2012.
- [27] Vijay Varanasi. Emergence-of-Autocatalysis-in-Prebiotic-Reaction-Networks: public source and data repository. <https://github.com/VJ-Varanasi/Emergence-of-Autocatalysis-in-Prebiotic-Reaction-Networks>, 2026. Reaction generator and catalysis sampler, as retrieved 1 September 2026.
- [28] Vijay Varanasi and Jun Korenaga. Emergence of autocatalysis in prebiotic reaction networks. *Physical Review E*, 113:044304, 2026.