Noise sensitivity and collision histories in the one-dimensional balloon process

A paper with proof sketches

Working exposition prepared for Jacob Richey

19 October 2026

Abstract

Grow balloons from independent uniform points of a circle, removing both balloons whenever they meet. With an odd number of starting points, one balloon survives. We describe how resampling the starting points changes that survivor and the collision history. One resampling has a nontrivial limiting probability of changing the survivor, and every growing number of resamplings changes it with probability tending to one. Yet almost all collision pairs survive a sublinear perturbation: the number of changed edges is asymptotic to $(1 + e^{-\gamma})k \log(N/k)$ for every nonzero $k = o(N)$. The proofs follow an insertion through a marked merger tree, construct coupled entrance laws, and sum the effect of a discrepancy over logarithmic scales. We explain these arguments and then indicate how their quantitative forms lead to Gaussian fluctuations, a Brownian sheet, and a Poisson law near complete resampling.

Reading note. This AI-written exposition sketches the working proofs in [W], recorded through Section 7.57. It is not a complete proof paper or an independent verification. The statements below retain that working status; unproved questions are identified explicitly. References to the full arguments are collected at the ends of the relevant sections.

1 The question and the main results

A change to one starting point leaves a long record in the collision history. At any given time, the original and modified systems have at most two discrepant active locations. Over the whole evolution, the change typically passes through order $\log N$ collisions and has a probability bounded away from zero of changing the final survivor. The distinction between an instantaneous discrepancy and its accumulated history is the starting point of this paper.

Let N be odd and let $X_1, \dots, X_N$ be independent uniform points of the circle $\mathbb{R}/\mathbb{Z}$, whose circumference is one. A balloon centered at each point grows at unit speed. When two surviving balloons touch, both disappear. Equivalently, repeatedly match and remove the closest pair of remaining points. Distances are intrinsic circle distances, and ties have probability zero. We measure time by the separation r of a colliding pair; physical time is $r/2$.

Write $C_r(A)$ for the active points of an input A after all collisions of separation at most r, and $\mathcal{M}_r(A)$ for the edges accepted by then. The completed matching is $\mathcal{M}(A) = \mathcal{M}_{1/2}(A)$. Each odd input has $(N-1)/2$ edges and one survivor.

Take an independent uniform sample $Y_1, \dots, Y_N$ and define the nested resampling path

$$A_k = \{Y_1, \dots, Y_k, X_{k+1}, \dots, X_N\}, \qquad 0 \le k \le N.$$

The point in coordinate i keeps label i when its location is replaced. Let $p_N(k)$ be the probability that A_0 and A_k have different survivor labels. Define the spatial history distance by

$$H_N(k;r) = |\mathcal{M}_r(A_0) \triangle \mathcal{M}_r(A_k)|, \qquad H_N(k) = |\mathcal{M}(A_0) \triangle \mathcal{M}(A_k)|.$$

Thus a pair removed from one matching and a new pair added to the other contribute two to $H_N(k)$. Edges here are identified by their physical endpoints. Identifying them by coordinate labels instead gives $H_N^{\mathrm{lab}}(k)$; the main laws below also hold for that convention.

Theorem 1 (Survivor noise)**.** *There is a constant $p_\infty \in (0,1)$ such that $p_N(1) \to p_\infty$. For every fixed k, $p_N(k) \to p(k)$, where*

$$1 - p(k) = \int_0^1 (1-x)^k \, \nu(dx) \tag{1}$$

for a probability measure ν on $[0,1]$ satisfying $\int x\,\nu(dx) = p_\infty$ and $\nu(\{0\}) = 0$. Consequently, for every sequence $k_N \to \infty$ with $k_N \le N$,

$$p_N(k_N) \longrightarrow 1.$$

The fixed-k limits also describe disagreement of survivor positions.

The assertion for growing k_N is stronger than existence of the one-point limit. It rules out any part of the survivor agreement that would persist under arbitrarily many resamplings. Under independent resampling with probability s_N per coordinate, the transition occurs when the expected number Ns_N of replacements is of order one. If $Ns_N \to 0$ there is usually no replacement; if $Ns_N \to \infty$ the survivor labels disagree with probability tending to one. At $s_N = \lambda/N$ the limiting disagreement is $1 - \int e^{-\lambda x}\nu(dx)$.

Theorem 2 (The size of the changed history)**.** *Let γ be Euler's constant. For every nonzero sequence $k = o(N)$, including fixed k,*

$$\frac{H_N(k)}{k\log(N/k)} \longrightarrow 1 + e^{-\gamma} \qquad \text{in } L^1. \tag{2}$$

Moreover, uniformly over $1 \le k \le N$,

$$\mathbb{E}H_N(k) \asymp k\big(1 + \log(N/k)\big), \qquad \sup_{1\le k\le N} \mathbb{E}\left|\frac{H_N(k)}{\mathbb{E}H_N(k)} - 1\right| \longrightarrow 0.$$

The same statements hold for H_N^{lab}.

In particular, when $k \to \infty$ and $k = o(N)$, the eventual survivor changes with high probability while $H_N(k)/N \to 0$. The second theorem quantifies the collision history left by the perturbation. Its logarithm comes from the range of scales between the typical input spacing $1/N$ and the typical spacing $1/k$ between edits.

The construction behind these results is a joint limit of the perturbed processes at every positive separation. It retains which points are shared between copies and where an insertion originally entered. We call this an *entrance law*: a process with finitely many active points at each $r > 0$, obtained by letting the initial number of points tend to infinity. The construction is explained in Section 3; it is the main connection between finite samples and the noise theorem.

There are two relevant bodies of earlier work. The balloon dynamics on infinite spaces is studied by Angel, Ray and Spinka [1]; its matching interpretation belongs to the theory of stable matchings of point processes [2]. In one dimension, removing the closest pair merges three consecutive gaps. This coarsening rule, its cyclic exchangeability argument, and a related recursion for retained mass appear in Bray, Derrida and Godreche [3]. Hierarchical

coalescence and renewal scaling limits are developed by Faggionato, Roberto and Toninelli [4]. We use these connections without identifying the complete dynamics with every model in that literature.

The perturbation question also has a close mean-field counterpart. Ahlberg, Deijfen and Sfragara [5] prove robustness of most matching edges and sensitivity of late edges under resampling of independent edge costs. Here the randomness lies in point locations, so the edge distances are strongly dependent. The results concern the survivor's resampling scale, coupled spatial histories, and the sharp number of changed edges. A full novelty review remains separate from this exposition.

2 Following a single discrepancy

Run two finite inputs using the same increasing list of pair distances. An edge is accepted in a copy precisely when both its endpoints are still active there. Call a point active in only one copy a discrepancy. An edge accepted in both copies removes two common points. An edge accepted in just one copy must contain a discrepancy. If its other endpoint is common, the discrepancy moves to that endpoint and changes which copy contains it. If both endpoints are discrepancies, they disappear together. Therefore

$$|C_s(A) \triangle C_s(B)| \le |C_r(A) \triangle C_r(B)| \le |A \triangle B|, \qquad s \ge r. \tag{3}$$

This argument works in any finite metric space. For one insertion there is exactly one discrepancy at every cutoff. The changed matching edges trace its successive transfers. For several edits these paths may meet and close, but discrepancies never branch.

The useful feature in one dimension is that the path can be encoded without following the inserted process separately. Consider the next background collision, between u and v, with separation s, and let a, c be the adjacent gaps. In local coordinates put $u = 0$, $v = s$ and

$$\alpha = \min(s, a/2), \qquad \beta = \min(s, c/2).$$

An insertion at $z \in (-\alpha, s/2)$ kills u before the background collision, leaving v as the discrepancy. The symmetric event leaves u. Accordingly define

$$T(z) = \begin{cases} s, & -\alpha < z < s/2, \\ 0, & s/2 < z < s + \beta, \\ z, & \text{otherwise.} \end{cases} \tag{4}$$

If $S_B(z)$ denotes the survivor position after inserting z into an even background B, the exact recursion is

$$S_B(z) = S_{B \setminus \{u,v\}}(T(z)). \tag{5}$$

The map records the discrepancy after this background pair has gone. It does not assert that the original insertion physically moves.

Compose the maps in collision order and push Lebesgue measure through them. At an intermediate cutoff, each active gap carries a measure of total mass equal to its length. Its atoms record origins whose discrepancies have already been transferred to old collision vertices. After the even background is exhausted, this gives

$$(S_B)_{\#}\,\mathrm{Leb} = \mathbf{1}_{E_B}(z)\,dz + \sum_{v \in B} \theta_v(B)\delta_v,$$

where $E_B = \{z : S_B(z) = z\}$ and $\theta_v(B) = \mathrm{Leb}\{z : S_B(z) = v\}$. Thus θ_v is the basin mass of the background label v. Let $\Theta_N = \mathrm{Leb}(E_B)$, with $N = |B| + 1$. Two independent insertions into

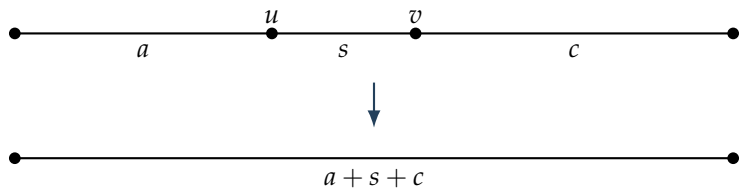


Figure 1: A collision merges three gaps. Reversing the merger gives a tree; recording the collapse map at each vertex tells us how insertion origins pass through that tree. Lengths alone do not contain this information.

the same background agree in a background label with conditional probability $\sum_v \theta_v^2$; they agree in the insertion label with probability Θ_N^2. Consequently

$$p_N(1) = 1 - \mathbb{E}\sum_v \theta_v^2 - \mathbb{E}\Theta_N^2. \tag{6}$$

Exchangeability gives $\mathbb{E}\Theta_N = 1/N$, so the last term vanishes.

This reduction identifies the main analytic difficulty. We need to pass the basin masses to a limit while the number of initial points diverges. Weak convergence of unmarked active sets does not determine them. Even weak convergence of measures alone does not preserve the sum of squared atomic masses. The next construction keeps enough of the collision history to make this passage valid.

Full argument: [W, Sections 2, 7.1, and 7.9].

3 An entrance law that remembers origins

3.1 Renewal gaps and their merger trees

At separation r, every active gap has length at least r. On the line, the entrance law is a stationary renewal process whose gap density is

$$w_r(x) = r^{-1} f(x/r), \qquad \int_1^\infty e^{-sy} f(y)\, dy = \tanh\left(\frac{E_1(s)}{2}\right), \qquad E_1(s) = \int_s^\infty \frac{e^{-t}}{t}\, dt. \tag{7}$$

The triple-merger recursion gives an equation for the Laplace transform $F(s) = \int e^{-sy} f(y)\, dy$:

$$sF'(s) = -\tfrac{1}{2} e^{-s}(1 - F(s)^2), \qquad F(\infty) = 0.$$

Integrating the equation for $\operatorname{arctanh} F$ gives (7). Expanding at zero gives mean gap $2e^\gamma r$, so the intensity is c/r, where

$$c = (2e^\gamma)^{-1}. \tag{8}$$

On a circle the gaps obey a cyclic renewal law with their total length fixed at one and their count restricted to the initial parity. They are not independent circle gaps. Conditional on the gap geometry, however, the past trees inside different gaps are independent.

To describe one tree, protect the endpoints of an interval of length x and suppose its interior has cleared by separation $t \le x$. The last interior collision divides the interval into lengths (a, s, b), where $b = x - a - s$ and s is the separation of that collision. The reverse split has density

$$\mathcal{K}_{t,x}(a, s) = \frac{w_s(a) w_s(s) w_s(b)}{w_t(x)}, \qquad 0 < s < \min(t, x/3), \quad s < a < x - 2s. \tag{9}$$

After the split, the three daughter trees are independent, with inherited cutoff s. The factors in the numerator describe their independent clearances; the denominator normalizes the conditioning on the parent interval. Iterating gives the full merger tree, and attaching (4) at each merger gives its origin measure.

There are two distinct convergence steps here. At any fixed cutoff and fixed tree depth, one first proves convergence of the finite geometry in total variation. The finite densities can be written as coefficients of products of Poisson interval weights. Expanding those weights isolates the terms consistent with the total count and parity; the remaining terms vanish. Nonnegativity and convergence of total mass then upgrade density convergence to total variation. This is a finite-depth calculation. Taking infinite depth requires a separate estimate.

The estimate uses a positive representation of f as a geometrically weighted convolution sum of another probability density. Exponential tilting turns it into a renewal density and gives constants $b_-, b_+, \eta > 0$ such that

$$b_- e^{-\eta y} \le f(y) \le b_+ e^{-\eta y}, \qquad y \ge 1. \tag{10}$$

The lower bound matters as much as the upper bound. In (9), it gives a uniformly positive probability of a balanced split, even after conditioning on the parent state.

3.2 Why the unseen tree becomes irrelevant

Suppose we want the mass $\mu([0,z])$ in a shielded interval. Reveal the last split (a,s,b) and express this quantity in terms of the daughter cumulative masses. Outside the middle daughter, only one daughter remains unresolved. Inside it, the two collapse cells must be read simultaneously. With F_1, F_2, F_3 the daughter cumulative masses and $\alpha = \min(s, a/2)$, $\beta = \min(s, b/2)$, for example,

$$\mu([0,z]) = F_1(a-\alpha) + s - F_2(s/2) + F_3(\beta), \qquad a < z < a+s.$$

Thus a cumulative query follows one or three daughter queries, each with coefficient of absolute value one. A query at an endpoint is already known from the total mass.

Give each unresolved terminal interval a budget equal to its length. For any two possible completions below depth d, the difference in the queried masses is at most the total budget Z_d. This is a deterministic bound once the first d levels are known. The balanced split event supplied by (10) has positive probability uniformly in the state and the threshold. On that event the query follows only one side daughter of length at most $3x/4$; otherwise it retains at most the parent length x. Hence, for some $\rho < 1$,

$$\mathbb{E} Z_d(t,x,z) \le x\rho^d. \tag{11}$$

Repeated queries need not be independent. We add their error budgets. The same estimate applied on both sides of a threshold, together with continuous finite-tree split laws, excludes mass at the readout boundaries.

At a fixed positive cutoff there are only finitely many active gaps and finitely many future collision maps. The final basin masses of the active labels can therefore be read using finitely many such queries. Couple the finite and limiting depth-d geometries, apply the same deterministic budgets to their deeper completions, and then let $d \to \infty$. This gives convergence of the relevant mass vector, and hence of its squared norm. Finally send the cutoff to zero. An early-label estimate shows that labels which have disappeared before that cutoff make a vanishing contribution to the agreement probability. The order of limits is essential: sample size, then tree depth, then the positive cutoff.

For coupling insertions, we need the image of the *same origin* in the two trees, rather than just the pushed measure. Follow a uniform origin backwards through its unique daughter

interval. Once it enters a middle daughter, a collapse assigns it to a collision vertex. A killed ancestry estimate shows that the mass of origins which avoid all middle daughters decays with depth. Combining this with the threshold-query bound determines which vertex is chosen. The resulting maps $\Phi^{(d)}$ satisfy, for some $\sigma < 1$,

$$\mathbb{E}\,\mathrm{Leb}\{z : \Phi^{(d)}(z) \neq \Phi(z)\} \leq Cx(x/t)\sigma^d. \tag{12}$$

Equality here means the same vertex. Almost every origin eventually has an exactly determined image, and the limiting origin measure is purely atomic on the countable set of collision vertices.

3.3 Several insertions and shared identities

Represent a comparison with k resamplings as two sets of k fresh probes inserted into a common background of size $N - k$. More generally, take q probes and any fixed family of subsets of them. Stop the background at a small cutoff ε. If all probes lie in distinct, nonadjacent active gaps, their responses superpose exactly at that cutoff, simultaneously for every subset. Their images either add a point or remove the nearest active endpoint, according to the one-insertion rule.

If the active gap lengths are x_i, a union bound controls failure by the empty-background probability and a constant times $q^2 \sum_i x_i^2$. The renewal estimates give

$$\limsup_{N\to\infty} \mathbb{P}(\text{failure at } \varepsilon) \leq Cq^2\varepsilon \tag{13}$$

for fixed q. On the good event, the convergent marked gap trees determine a unique joint law. Push this law to any later positive cutoff. Every subsequential joint limit is within the error in (13) of that same construction. Sending $\varepsilon \downarrow 0$ proves uniqueness.

The marks specify the set of copies containing each background point. The separated construction also rules out accidental coincidence of distinct limiting identities. Survivor agreement is therefore a valid readout of the joint limit. Convergence of the unmarked marginals alone would not establish this.

Random deletions have the same limiting response as fresh insertions. An insertion at the location of an existing point immediately cancels that point and realizes its deletion. The empirical measure of the original sample and Lebesgue measure have asymptotically equal masses on all intervals. At any fixed tree depth this compares their origin budgets, including intervals chosen from the sample itself. The depth error then vanishes by (12). This is the reason a resampling can eventually be represented by two insertions in the entrance process.

Full argument: [W, Sections 7.4–7.6, 7.10, and 7.13–7.14].

4 Why every growing perturbation changes the survivor

4.1 A nontrivial limit and its spectrum

The construction in Section 3 and the basin identity (6) prove existence of p_∞. It is useful to see separately why that limit is strictly between zero and one. Let L be the shorter separation of the last pair of an even background. Its two arcs clear internally without touching their endpoints. Conditional on the last endpoints and the allocations to the arcs, the two interior configurations are independent and reflection-symmetric. A uniform insertion in the shorter arc leaves one of the two endpoints as survivor, with an averaged fair choice between them. An insertion in the other arc can agree with at most one of those choices. Thus

$$p_N(1) \geq \mathbb{E}[L(1 - L)].$$

On $1/4 \le r \le 1/2$ the last-pair density for an even background of size M is explicitly

$$\frac{1-(1-2r)^M}{2r(1-r)}.$$

Integrating just this part gives $\liminf p_N(1) \ge 1/8$. Conversely, on $L > 1/3$ every insertion leaves one of the two last endpoints. Conditional disagreement is then at most $1/2$, yielding $\limsup p_N(1) \le 1 - (\log 2)/4 < 1$. Sharper tree integrals and certified numerical evaluation improve these elementary bounds to

$$0.2016982228 \le p_\infty \le 0.8112015673. \tag{14}$$

The proposed value near 0.658 is still uncertified.

For the full resampling profile, encode the survivor label as a unit vector $F_N \in \mathbb{R}^N$. Decompose this function of the independent coordinates into orthogonal components indexed by subsets of coordinates. Let $w_{N,j}$ be the total squared norm of components using exactly j coordinates. Their nonnegativity gives

$$1 - p_N(k) = \sum_{j=0}^{N} w_{N,j} \frac{\binom{N-k}{j}}{\binom{N}{j}}, \qquad \sum_j w_{N,j} = 1. \tag{15}$$

Indeed a component contributes to agreement exactly when none of its coordinates is refreshed. Averaging the refreshed set of size k gives the displayed avoidance probability. Coordinate exchangeability makes this equivalent to replacing the first k coordinates.

Place mass $w_{N,j}$ at j/N to obtain ν_N. For fixed k, replacing the avoidance factor by $(1-j/N)^k$ costs at most $\binom{k}{2}/N$. The coupled limit gives convergence of $p_N(k)$ for every fixed k, and these moments determine a unique measure ν on $[0,1]$. This proves (1). Formula (15) also makes $p_N(k)$ increasing in k.

What it does not prove is $\nu(\{0\}) = 0$. Although $w_{N,0} = 1/N$, mass at orders $j = o(N)$ could accumulate at zero. Excluding this possibility is the main additional step in the noise argument.

4.2 Removing an invariant atom

The right state for this step is the full entrance history $\omega = (C_\omega(r) : r > 0)$, recorded at rational cutoffs. It contains the merger trees and hence all the origin maps. Let Π_{e} and Π_{o} be its even and odd laws. Inserting at origin z gives a measurable transformation $\mathcal{T}_z$ which exchanges these parities. The same-origin construction and deletion comparison give

$$\mathcal{T}_z^2 = I, \qquad \mathcal{T}_z \mathcal{T}_w = \mathcal{T}_w \mathcal{T}_z, \tag{16}$$

almost everywhere, and each $\mathcal{T}_z$ preserves the equal mixture $\Pi = (\Pi_{\mathrm{e}} + \Pi_{\mathrm{o}})/2$ for almost every z.

These identities come from the joint finite constructions. Reversing an iid add-one pair is a uniform deletion from its larger member. Keeping the tagged origin in this reversal identifies the inverse with the same $\mathcal{T}_z$. Inserting two independent probes in either order identifies the two compositions. The empirical-origin comparison from Section 3 is needed here; convergence of unmarked configurations would not justify either identity.

Let $Kf(\omega) = \int_{\mathbb{R}} / \mathbb{Z} f(\mathcal{T}_z \omega)\, dz$. This is a self-adjoint Markov operator between the parity sectors. On odd histories, one refresh has kernel $Q = K^2$, a positive self-adjoint contraction. Let $\mathcal{I}$ be its invariant sigma-field, and let Λ_ω be the conditional law of the survivor location $U(\omega)$ given $\mathcal{I}$. The spectral interpretation of long refresh times gives

$$\nu(\{0\}) = \mathbb{E} \sum_x \Lambda_\omega(\{x\})^2. \tag{17}$$

One way to obtain this identity is to partition the circle into dyadic intervals. For each interval, the long-time limit of its survivor indicator correlation is the squared norm of its invariant projection. Refining the partition leaves exactly the diagonal atomic mass.

We can now target a hypothetical atom. Suppose X is an $\mathcal{I}$-measurable location. Its invariant representative extends to even histories so that $X(\mathcal{T}_z\omega) = X(\omega)$ almost everywhere. To see this extension, apply K to $e^{2\pi i X}$ and use equality in the L^2 contraction: the average still has modulus one, so the values being averaged agree.

Starting from an odd stationary history, choose the insertion origin z uniformly in a small interval of radius h about $X(\omega)$. This adaptive choice still gives an exactly stationary even history. Indeed, writing a_h for the uniform density on that interval centered at zero, measure preservation and invariance of X imply

$$\begin{aligned}\int\int a_h(z - X(\omega))\, f(\mathcal{T}_z\omega)\, dz\, \Pi_{\mathrm{o}}(d\omega) \\ = \int\int a_h(z - X(\omega'))f(\omega')\, dz\, \Pi_{\mathrm{e}}(d\omega') = \Pi_{\mathrm{e}}(f).\end{aligned}$$

For each fixed h, the sampling density is bounded, so the almost-everywhere identities remain applicable.

There is one further geometric input: insertions sufficiently close to a protected endpoint remove it at a fixed cutoff with probability tending to one. Its proof follows the endpoint threshold backwards through the leftmost or rightmost tree branch. A uniformly positive chance of settling the query, together with control of the inherited cutoff, supplies a whole neighborhood of successful origins. Almost-everywhere convergence of origin maps by itself would not supply this neighborhood.

Thus, after the adaptive insertion, X is absent at any fixed cutoff r with probability tending to one as $h \downarrow 0$. Make one more, independent uniform insertion. This restores the odd stationary law and preserves X. For X to be the resulting survivor, it must either have survived the adaptive removal or be restored from the insertion mass carried by vertices already absent at cutoff r. If B_r is that early basin mass in the intermediate even history,

$$\mathbb{P}(U = X) \le \mathbb{P}\{X \text{ remains active at } r \text{ after targeting}\} + \mathbb{E}_{\Pi_{\mathrm{e}}} B_r. \tag{18}$$

The first term vanishes as $h \downarrow 0$. The second vanishes as $r \downarrow 0$: the atomic origin measure has total mass one, and each of its collision vertices has a positive lifetime.

We have proved $\mathbb{P}(U = X) = 0$ for every invariant selected location. If Λ_ω had an atom, measurably selecting a largest one would contradict this conclusion. Hence Λ_ω is diffuse, and (17) gives $\nu(\{0\}) = 0$. This proof does not require ergodicity of the full refresh chain.

Finally, monotonicity transfers the conclusion to any growing finite resampling count. For fixed k, eventually $k_N \ge k$, so

$$\liminf_N p_N(k_N) \ge p(k).$$

Now send $k \to \infty$ in (1). The same adaptive argument applied to the conditional active counting measure shows that two copies have no common active location at any fixed positive cutoff after a growing number of edits. This establishes disjointness, not independence of the configurations.

Full argument: [W, Sections 3, 7.6, 7.8–7.10, and 7.13–7.15].

5 Where the logarithm and its coefficient come from

5.1 The mass captured by one collision

For a background edge e, let W_e be the Lebesgue measure of insertion origins that destroy that edge. Conditional on the background, the expected number of its edges destroyed by one uniform insertion is $\sum_e W_e$. The discrepancy path alternates between lost background edges and new edges. Apart from an endpoint correction, its total length is twice the number of lost background edges.

There is a useful locality bound for W_e. The two neighboring gaps and the collision gap merge into a parent interval I_e of length L_e. An origin outside that interval cannot destroy e, so $W_e \le L_e$. At active count m, cyclic gap exchangeability gives

$$\mathbb{E}\sum_i x_i^2 \le C/m, \qquad \mathbb{E}L_e^2 \le C/m^2. \tag{19}$$

For example, conditional on the unordered gaps and their minimum s, the two neighbors of s are a uniform ordered pair of the other gaps. Expanding the increase of the gap-square sum under their merger gives an expected increment at most C/m^2. Summing over successive counts proves the first estimate, and another expansion gives the second.

For two independent insertions, an edge destroyed by both has conditional probability W_e^2. By (19), $\mathbb{E}\sum_e W_e^2$ is bounded uniformly in the initial size. The overlap of the two discrepancy paths therefore contributes only a bounded error to the expected resampling distance:

$$\mathbb{E}H_N(1) = 4\mathbb{E}\sum_e W_e + O(1). \tag{20}$$

The factor four is two paths, each counting both lost and new edges.

To evaluate the sum, let $s_{M,m}$ be the next collision separation in a background of initial size M, just before its active count falls from m to $m-2$. Each gap of length x_i carries its pushed origin measure μ_i. Define its two corner regions to have widths $\min(s_{M,m}, x_i/2)$, and let $A_{M,m}$ be their total mass over all gaps. The central collision gap contributes all its mass $s_{M,m}$. The two neighboring gaps contribute their corner masses. Exchangeability of the *decorated* gaps therefore gives the exact expectation identity

$$\mathbb{E}W_{M,m} = \mathbb{E}s_{M,m} + \frac{\mathbb{E}A_{M,m} - \mathbb{E}s_{M,m}}{m-1}. \tag{21}$$

The decoration is essential: the expected origin measure in a gap is not simply Lebesgue measure.

The two inputs to (21) are

$$ms_{M,m} \longrightarrow c, \qquad A_{M,m} \longrightarrow \tfrac{1}{2} \quad \text{in } L^1, \qquad m \to \infty, \quad m/M \to 0, \tag{22}$$

where c is defined in (8). The first is inversion of the renewal count law. To understand the second, consider a deterministic cutoff r instead. An insertion changes the active count by -1 when its image lies in a corner and by $+1$ otherwise. Hence, with $K_M(r) = |C_r(B_M)|$,

$$\mathbb{E}A_M(r) = \frac{1 - \{\mathbb{E}K_{M+1}(r) - \mathbb{E}K_M(r)\}}{2}. \tag{23}$$

At a mesoscopic cutoff, the leading count is c/r for either initial parity, and the exact renewal formulas make the difference on the right tend to zero. Independent marked gaps give a renewal-reward law and therefore concentration of the corner mass around $1/2$.

Transferring this statement uniformly to fixed initial count takes some care. For a Poisson background the marked first and second moments give the law throughout $r \to 0$, $Mr \to \infty$.

Coupling Poisson and fixed counts changes its corner mass in mean by at most $C\sqrt{M}\,r$, which covers small cutoffs. Finite-tree coefficient estimates cover the complementary larger cutoffs. A separate bracket around the random collision separation then proves (22). Substituting a random time directly into (23) would not be justified.

Now sum (21) over $j > m$ of the appropriate parity. The active count decreases by two, so the harmonic sum is $\frac{1}{2}\log(M/m) + O(1)$. Equations (22) give

$$\mathbb{E}\sum_{j>m} W_{M,j} \sim a_{\mathrm{H}} \log(M/m), \qquad a_{\mathrm{H}} = \frac{c + 1/2}{2} = \frac{1 + e^{-\gamma}}{4}. \tag{24}$$

A central range of counts gives the asymptotic; deterministic bounds $js_{M,j} \le 1$ and $A_{M,j} \le 1$ control its omitted ends. Equation (20) now explains the coefficient in Theorem 2.

5.2 Why the logarithm stops at the edit scale

For many edits, use a background B of size M and q independent probes, divided into two sets P_1, P_2. Resampling corresponds to $M = N - k$ and $q = 2k$. Stop the background when its active count reaches $m > q$, and call that separation τ. Let $G_B(z) = |\mathcal{M}_\tau(B) \triangle \mathcal{M}_\tau(B + z)|$ be one probe's early cost. The useful comparison is

$$\mathbb{E}\left| H(B + P_1, B + P_2) - \sum_{i=1}^{q} G_B(Z_i) \right| \le C\left(m + q + \frac{q^2}{m - q} \right). \tag{25}$$

Here $H(A, A') = |\mathcal{M}(A) \triangle \mathcal{M}(A')|$, and Z_i lists all the probes.

The terms have a concrete meaning. After τ, discrepancy contraction leaves only $O(m + q)$ possible further edges. Before τ, telescope the effects of the probes. A failure of additivity has a mixed two-probe witness: in one of the four configurations with neither, either, or both probes present, a parent interval contains both sources. Summing squared parent masses over collisions above count $m - q$ costs $C/(m - q)$. There are order q^2 possible pairs. The argument uses the appropriate positive configuration for each witness; it does not place every witness in the original background. That distinction allows boundary cascades.

Choose m of the correct parity with $m \asymp 4q$. Then the error in (25) is $O(q)$. Conditional on B,

$$\mathbb{E}[G_B(Z) \mid B] = 2\sum_{j>m} W_{M,j} + O(1).$$

Combining with (24) gives

$$\mathbb{E}H(B + P_1, B + P_2) \sim 2q a_{\mathrm{H}} \log(M/q).$$

For $q = 2k$ and $M = N - k$, this is $(1 + e^{-\gamma})k\log(N/k)$. The range below the edit spacing produces the logarithm; the interacting late part costs only order k in mean.

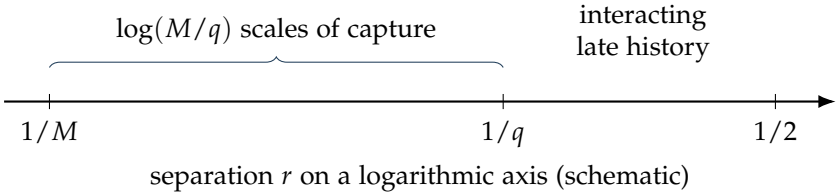


Figure 2: With q sparse probes, most of the history cost accumulates between the input scale and the edit scale. The later part has mean order q and centered fluctuations of order $\sqrt{q}$.

Full argument: [W, Sections 7.11–7.12 and 7.19–7.22].

6 Why the sharp mean is also a typical value

The fixed-edit and growing-edit laws use different averaging mechanisms. When there are many probes, average over origins in one background. When there are only finitely many, average along logarithmic scales. Keeping the two arguments separate avoids an unjustified appeal to independence between scales or between perturbed systems.

For growing $q = o(M)$, the captures $G_B(Z_i)$ in (25) are conditionally iid given B. A single insertion has a second moment bounded by $C(1 + \log M)^2$, so

$$\mathbb{E}\left|\sum_{i=1}^{q} G_B(Z_i) - q\mathbb{E}[G_B(Z) \mid B]\right| \le C\sqrt{q}(1 + \log M). \tag{26}$$

This is negligible compared with $q \log(M/q)$ for every growing $q = o(M)$. For $q \le \sqrt{M}$ use $\log(M/q) \ge \frac{1}{2}\log M$; for $q > \sqrt{M}$ the factor $q^{-1/2}$ controls the remaining logarithm.

The background conditional mean must also concentrate. Reveal the cyclic order of decorated gaps only as mergers require it: at each step expose the two neighbors of the minimum and contract that three-gap block. The unexposed cyclic order remains uniform, even given the entire revealed history. In this filtration,

$$W_{M,j} - \left(s_{M,j} + \frac{A_{M,j} - s_{M,j}}{j-1}\right)$$

is a martingale difference. The sum above active count m has variance at most $\sum_{j>m} \mathbb{E}L_{M,j}^2 \le C/m$. The predictable part is controlled in L^1 by (22) and the harmonic-sum argument. Consequently

$$\frac{\sum_{j>m} W_{M,j}}{\log(M/m)} \longrightarrow a_{\mathrm{H}} \quad \text{in } L^1.$$

Together with (26) and the $O(q)$ comparison error, this proves the growing-edit part of Theorem 2.

For fixed q, take the stationary marked line entrance process and insert one origin at zero. Let $J(a, b)$ count background edges with separations in $(a, b]$ destroyed by this insertion. Scaling the entrance law makes the costs in successive unit logarithmic bands stationary, with

$$\mathbb{E}J(a, b) = a_{\mathrm{H}} \log(b/a). \tag{27}$$

The marked renewal process is translation-ergodic. This also forces ergodicity of its dilation by e: conjugating a fixed translation by repeated dilations makes that translation arbitrarily small. An L^2 function invariant under dilation is therefore invariant under every translation, and hence constant. The ergodic theorem gives

$$\frac{J(e^{-T}, 1)}{T} \longrightarrow a_{\mathrm{H}} \quad \text{in } L^1.$$

This concerns spatial dilation of the line process, not refresh ergodicity on the circle.

There is still a finite-sample transfer to make. Keep a central range of separations $M^{-1+\epsilon} \le r \le M^{-\epsilon}$, cap each unit-band reward, and group a fixed number of bands into a block. Marked local convergence compares each finite block with its line counterpart, uniformly over the central range. The triangle inequality then compares the average of all blocks without requiring their independence. First let $M \to \infty$, then increase the block size using line ergodicity. Finally remove the cap and let $\epsilon \downarrow 0$.

The last removal uses positivity: the capped central capture is bounded by the full capture, whose mean is already known sharply. Thus the discarded tails cannot hide a macroscopic

part of the normalized cost. A fixed number of probes has only a bounded expected interaction error, so the same law holds for every fixed q. Combining fixed and growing subsequences proves (2) for all nonzero $k = o(N)$. For label histories, the spatial and label distances differ by at most $2k$, which is negligible on this scale. At positive edit density, independent permutations of the refreshed labels give the stronger uniform mean bound $\mathbb{E}(H_N - H_N^{\mathrm{lab}}) \le 4$.

Full argument: [W, Sections 7.24–7.25].

7 From the law of large numbers to fluctuation laws

The preceding argument explains the main theorem package. Its quantitative extensions describe the rest of the resampling path. We give shorter sketches here, emphasizing the additional estimates needed to pass from means to centered fluctuations.

7.1 Sparse edits: captures and their remainder

For the single line insertion, logarithmic capture bands satisfy a functional central limit theorem. Their covariance decays sufficiently fast, and

$$\frac{\operatorname{Var} J(e^{-T}, 1)}{T} \longrightarrow v_{\mathrm{H}} > 0.$$

The mechanism combines the origin-map approximation with marked renewal blocks separated by long gaps. Dilation turns spatial separation estimates into logarithmic-time decorrelation. Moment factorization controls higher cumulants and yields Brownian fluctuations. A separate argument rules out zero variance; v_{H} has not been explicitly evaluated.

For q probes, twice the sum of their captures suggests variance $4qv_{\mathrm{H}} \log(M/q)$. Establishing that suggestion requires much more than the $O(q)$ mean error in (25). Such an error can exceed the proposed standard deviation. The actual interaction remainder must be centered and shown to have variance $O(q)$; the late history needs the same bound.

For the early remainder, an empty background cutoff gap separates the histories of all probe subsets. Consecutive occupied gaps form clusters. Their rewards include cascades across adjacent gaps, which can be arbitrarily long in a deterministic configuration. A fine-scale interaction is rare, however, and its probability supplies a summable gain in each logarithmic band. This produces uniform block moments and the $O(q)$ variance bound. A bound using only pairs of probes in the same original parent interval is false; the cluster decomposition is what accommodates changing boundaries.

The fixed-count input cannot be obtained by conditioning a whole Poisson configuration and accepting the resulting density loss. Instead, expose complete original records only inside probe-occupied gaps. The unused gaps retain random original counts. Those hidden counts form a reservoir whose local central limit estimate absorbs the conditioning on the prescribed total. Separate moment bounds transfer centers and variances; total variation convergence alone would not transfer them.

Late edges use marginal parent locality. With high probability, nearby fully shielded intervals determine a cell's history against arbitrary changes farther away. Their failure probability decreases exponentially with the search radius. Summing localized covariances over dyadic separation bands gives centered late variance $O(q)$. This remains valid when the two copies have few common active points.

For k resamplings, the complete sparse result is consequently

$$\frac{H_N(k) - \mathbb{E}H_N(k)}{\sqrt{8kv_{\mathrm{H}} \log(N/k)}} \Longrightarrow \mathcal{N}(0,1), \qquad 1 \le k = o(N), \tag{28}$$

with the corresponding variance asymptotic. Bounded counts use a joint comparison with finitely many independent line captures; growing counts use the centered remainder and reservoir estimates.

Uniform versions resolve two parameters at once. For $k \to \infty$, $k = o(N)$, put $M = N - k$, $q = 2k$, $L = \log(M/q)$ and

$$r_N(t) = \begin{cases} M^{-1}(M/q)^t, & 0 \le t \le 1, \\ q^{-1}(q/2)^{t-1}, & 1 \le t \le 2. \end{cases}$$

After centering by its exact mean and dividing by $\sqrt{8kv_{\mathrm{H}}L}$, the field $H_N(\lfloor ku \rfloor; r_N(t))$ converges, with bilinear interpolation, to

$$W(u, t \wedge 1), \qquad \mathrm{Cov}(W(u,s), W(v,t)) = \min(u,v)\min(s,t). \tag{29}$$

The interpolation uses edit mesh $1/k$ and time mesh $1/\lceil L \rceil$; the raw field is uniformly close to it. Shared edit prefixes explain $\min(u,v)$; shared logarithmic bands explain $\min(s,t)$. Uniform centered control of the late history explains why the time coordinate stops at one. Tightness requires maximal bounds over both parameters; finite collections of capture limits alone do not give the sheet theorem.

Full argument: [W, Sections 7.26, 7.37–7.42, and 7.48–7.56].

7.2 Positive density: a profile and a conditioned Gaussian process

At edit fraction θ, take independent Poisson inputs on the line with common intensity $1 - \theta$ and private intensities θ, θ. Define $\mathfrak{h}(\theta)$ as the expected changed-edge density, assigning each edge to its left endpoint. At a fixed separation cutoff b, gaps longer than b in the union input separate iid blocks. A renewal-reward law gives the truncated density. The unprocessed edge density tends to zero as $b \to \infty$, yielding

$$H_N(k)/N \longrightarrow \mathfrak{h}(\theta) \quad \text{in } L^2 \quad \text{when } k/N \to \theta.$$

The function is continuous, increasing and concave, with $\mathfrak{h}(0) = 0$, $\mathfrak{h}(1) = 1$, and

$$\mathfrak{h}(\theta) \sim (1 + e^{-\gamma})\theta \log(1/\theta) \quad (\theta \downarrow 0). \tag{30}$$

This dense law, combined with the sparse laws, gives the uniform relative concentration stated in Theorem 2.

For fluctuations, the Poisson block covariance must be corrected because the finite model has prescribed counts of each input type. Let R be a block's vector of rewards, C its vector of color counts, and L_b its length. Center each by its mean rate times L_b and write the covariance per unit length as

$$\begin{pmatrix} V_b & B_b \\ B_b^{\mathsf{T}} & D \end{pmatrix}.$$

Here D is the diagonal matrix of color intensities, with zero-intensity colors omitted. Conditioning the counts gives the reward covariance

$$\Gamma_b = V_b - B_b D^{-1} B_b^{\mathsf{T}}. \tag{31}$$

The subtraction removes fluctuations caused merely by random input counts. Local count reservoirs justify this conditioning, and late locality removes b.

For the whole nested resampling path, colors specify intervals of old and fresh priorities. Rare changes of priority give increment moment bounds at fixed b. A further uniform late bound is

$$\left\| \max_{0\le j\le N} \sup_{r\ge b/N} |H_N(j;r) - H_N(j;b/N) - \mathbb{E}\{H_N(j;r) - H_N(j;b/N)\}| \right\|_p \le C_p\sqrt{N/b}\log(2+b),$$

where the separation supremum ends at $1/2$. This makes the cutoff removable simultaneously for all edit fractions. Together with a uniform mean bias $|\mathbb{E}H_N(j) - N\mathfrak{h}(j/N)| \le C\log^2(N+2)$, it gives, after linear interpolation,

$$\left(\frac{H_N(\lfloor N\theta\rfloor) - N\mathfrak{h}(\theta)}{\sqrt{N}}\right)_{0\le\theta\le1} \Longrightarrow (G(\theta))_{0\le\theta\le1}. \tag{32}$$

The limit is continuous and Gaussian, with covariance $\Gamma = \lim_{b\to\infty}\Gamma_b$. It has $G(0) = G(1) = 0$ and $\operatorname{Var} G(\theta) = v_{\mathrm{D}}(\theta) > 0$ for $0 < \theta < 1$. Its covariance is not explicitly identified; these facts do not make it a Brownian bridge.

The mean comparison needs sharper forms near the boundaries. They give the common terminal normalization

$$\frac{H_N(k) - N\mathfrak{h}(k/N)}{\sqrt{Nv_{\mathrm{D}}(k/N)}} \Longrightarrow \mathcal{N}(0,1) \quad \text{whenever } 1 \le k \le N, \quad (N-k)^2/N \to \infty. \tag{33}$$

Here $\operatorname{Var} H_N(k) \sim Nv_{\mathrm{D}}(k/N)$, and $v_{\mathrm{D}}(\theta) \sim 8v_{\mathrm{H}}\theta\log(1/\theta)$ at zero. Replacing the center by only the leading expression in (30) is not justified at the fluctuation scale.

Full argument: [W, Sections 7.23, 7.43, and 7.46–7.57].

7.3 Near complete resampling: rare shared edges

Let $M = N - k$ be the number of unchanged coordinates, now small relative to N. If J_N is the number of common spatial edges, then exactly

$$H_N(k) = N - 1 - 2J_N. \tag{34}$$

A common edge needs two unchanged endpoints. Among pairs at distance of order $1/N$, the number of opportunities is therefore of order M^2/N. Conditional on a candidate pair, the two surrounding private inputs become asymptotically independent.

Let $a(t)$ be the probability that two forced points at distance t are matched in a unit-rate Poisson input on the line, and set

$$\kappa = \int_0^\infty a(t)^2\,dt.$$

For bounded rescaled lengths, disjoint-pair factorial moments give the Poisson law; locality supplies the two acceptance factors $a(t)$. A shared pair of distinct matching edges cannot reuse an endpoint, which controls the remaining factorial-moment terms. Tail estimates then remove the length cutoff. Thus

$$M^2/N \to \lambda < \infty \quad \Longrightarrow \quad J_N \Longrightarrow \mathrm{Pois}(\kappa\lambda). \tag{35}$$

There is also a marked point-process version, with intensity $\lambda\,dx\,a(t)^2\,dt$ for edge location x and rescaled length t. The coefficient has the certified enclosure

$$0.194237478218 \le \kappa \le 0.194237478220.$$

For label edges there is an additional asymptotically independent $\mathrm{Pois}(1/2)$ coincidence count, so the bounded endpoint law depends on which edge convention is used.

When $M = o(N)$ and $M^2/N \to \infty$, the shared-edge count is instead asymptotically Gaussian, with mean and variance asymptotic to $\kappa M^2/N$. This matches (33) because

$$v_{\mathrm{D}}(\theta) \sim 4\kappa(1-\theta)^2 \quad (\theta \uparrow 1).$$

The same constant appears in the separate integer matching spectrum

$$1 - \mathfrak{h}(\theta) = \sum_{\ell \geq 2} \pi_\ell (1-\theta)^\ell, \qquad \pi_2 = 2\kappa.$$

The weights satisfy $\pi_\ell \geq 0$ and $\sum_{\ell \geq 2} \pi_\ell = 1$. This spectrum records orders of matching indicators. It is different from the rescaled survivor spectrum ν in (1).

The endpoint in (35) is complete *resampling*, not the last physical collision. For the latter, if R_N is the last collision separation of an odd sample, the working result is

$$R_N \Longrightarrow \frac{1}{1 + Y_1 + Y_2},$$

where Y_1, Y_2 are iid with density f from (7). Its physical time is half this quantity. This constrained renewal law supplies no assertion of Poisson terminal clumps. The Poisson mechanism above concerns rare pairs shared by almost independently resampled inputs.

Full argument: [W, Sections 7.7, 7.23, and 7.44–7.47].

8 What the arguments leave open

The survivor theorem is qualitative at its sharp scale: it identifies when agreement vanishes, but does not identify ν explicitly. The working arguments give logarithmic upper bounds on agreement and power bounds for finite-sample approximation, with uncalibrated constants. A narrow certified interval for p_∞ remains a separate numerical and analytic problem. The historical estimate near 0.658 and the original small-sample calculations still lack their complete original reproduction package.

The history laws reveal more explicit structure. Their leading sparse coefficient comes from the renewal density and the corner-mass identity; their complete-refresh coefficient comes from the squared forced-pair acceptance probability. Between these endpoints, the profile $\mathfrak{h}$, the variance rate v_{H}, and the covariance of G remain largely implicit. Proposed Brownian tangents of G at its two boundaries are not established results of this paper. Nor has the full circle refresh chain been proved ergodic.

The proof sketches also identify where a detailed verification should concentrate. The marked construction must preserve identities and origins, the adaptive insertion must preserve the stationary law, and the corner law must be uniform through the required finite-sample regimes. For fluctuations, the centered remainder estimates and the exact-count conditioning are additional substantive inputs. Finite checks of collapse maps, parent witnesses, and label coincidences support those constructions, but do not prove their asymptotic bounds.

The full record includes spatial active-count fields and fluctuations of weighted parent histories. They concern related observables and help develop the renewal machinery; they are not needed to state the two principal theorems here. The common thread of this exposition is the origin of a perturbation: how its discrepancy moves, how the merger tree records that movement, and how its accumulated captures determine both the final survivor response and the changed collision history.

References

[W] *The balloon process on the circle*. AI-written working manuscript prepared with Jacob Richey, mathematical record through 16 October 2026, Sections 1–10 and 7.1–7.57. Canonical proof record, with source `balloon-process.tex`. The present article is an exposition of that record, not an independent source of verification.

[1] O. Angel, G. Ray and Y. Spinka. *A tale of two balloons*. arXiv:2103.06847, 2021.

[2] A. E. Holroyd, R. Pemantle, Y. Peres and O. Schramm. *Poisson matching*. arXiv:0712.1867, 2007, revised 2008.

[3] A. J. Bray, B. Derrida and C. Godreche. *Non-trivial algebraic decay in a soluble model of coarsening*. Europhys. Lett. **27** (1994), 175–180. Authors' scan.

[4] A. Faggionato, C. Roberto and C. Toninelli. *Universality for one-dimensional hierarchical coalescence processes with double and triple merges*. Ann. Appl. Probab. **24** (2014), 476–525. arXiv:1112.4630.

[5] D. Ahlberg, M. Deijfen and M. Sfragara. *Mean field stable matchings*. arXiv:2406.04911, 2024.