hexagon:2611.00735v1

MONOTONICITY OF THE SPEED OF BIASED RANDOM WALKS ON GALTON–WATSON TREES

BY

GPT-6 Astra and Heng Ma

ABSTRACT

On a Galton–Watson tree without leaves, the λ-biased random walk (X_n) chooses its next step with weight $\lambda > 0$ for the parent and weight one for each child. When λ strictly less than the mean offspring m, the walk is transient, with a positive speed $v(\lambda) := \lim_{n\to\infty} |X_n|/n$. We study the monotonicity of $v(\cdot)$, and prove that:

- If the GW tree has at least two children, then $v(\cdot)$ is decreasing on $(0, m)$;
- $v(\cdot)$ is decreasing on $(0, 1) \cup (\frac{3+\sqrt{5}}{2}, m)$.

We also sketch a computer-assisted proof of strict decrease on $(0, m)$ using exact checks of about 750,000 rational coefficients. The full proof appears in the companion GitHub version [GM26a].

AI USAGE

The initial proof was found by GPT-6 Astra through extensive interaction with the author in September 2026. This manuscript was substantially generated and revised with the assistance of AI systems. At the time of this version, only part of the arguments and citations have been independently verified in full by the human contributor. **A digested and more streamlined version will replace this preliminary version on both Hexagon and arXiv in the coming months**.

Contents

1. Introduction 2
2. Preliminaries 4
3. Strict decrease on $[0, m)$ when every vertex has at least two children 10
4. Strict decrease for $0 < \lambda < 1$ 13
5. Strict decrease for $(3 + \sqrt{5})/2 \le \lambda < m$ 18
6. Computer-assisted proof for $1 < \lambda < m$ (sketch) 19
References 24

Date: November 20, 2026.

2020 *Mathematics Subject Classification.* 60J10, 60J80, 60K37.

Key words and phrases. Galton–Watson tree, biased random walk, speed monotonicity, escape probability, computer-assisted proof.

Only Heng Ma takes responsibility for this paper.

1. Introduction

Let $\mathbb{T}$ be a Galton–Watson tree with offspring variable ν satisfying

$$\mathbb{P}(\nu = 0) = 0, \qquad 1 < m := \mathbb{E}\nu < \infty. \tag{1.1}$$

At a nonroot vertex with k children, the λ-biased random walk moves to the parent with probability $\lambda/(\lambda + k)$ and to each child with probability $1/(\lambda + k)$. At the root it chooses a child uniformly. Write W_n for the walk started at the root and $|W_n|$ for its distance from the root. In the transient regime $0 \le \lambda < m$, its speed

$$v(\lambda) = \lim_{n\to\infty} \frac{|W_n|}{n}$$

exists almost surely and is a deterministic positive constant; see [LPP97, Aïd14]. Increasing λ strengthens the bias towards the root. Lyons, Pemantle and Peres asked whether v is nonincreasing in λ for every offspring law satisfying (1.1) [LPP97, Question 2.1].

The local drift at a vertex of degree $k + 1$ decreases with λ, but this observation does not settle the question: changing the bias also changes the distribution of the environments visited by the walk. A proof must control both effects. Our analytic arguments make this comparison through the recursive law of the escape probability.

1.1. Analytic results.

For a real-valued function f, denote its upper right Dini derivative by

$$D^+ f(\lambda) = \limsup_{\varepsilon \downarrow 0} \frac{f(\lambda + \varepsilon) - f(\lambda)}{\varepsilon}.$$

The first result covers the entire transient regime when single-child vertices are absent.

Theorem 1.1. *Suppose $\mathbb{P}(\nu \ge 2) = 1$ and $m = \mathbb{E}\nu < \infty$. Then v is strictly decreasing on $[0, m)$. Moreover,*

$$D^+ v(\lambda) < 0, \qquad 1 < \lambda < m.$$

For laws with single-child vertices, we prove two complementary analytic results.

Theorem 1.2. *Suppose (1.1) holds and $\mathbb{P}(\nu = 1) > 0$. Then v is strictly decreasing on $(0, 1)$, and $D^+ v(\lambda) < 0$ for every $0 < \lambda < 1$.*

Theorem 1.3. *Suppose (1.1) holds and $\mathbb{P}(\nu = 1) > 0$. If $m > (3 + \sqrt{5})/2$, then v is strictly decreasing on $[(3 + \sqrt{5})/2, m)$, and*

$$D^+ v(\lambda) < 0, \qquad \frac{3 + \sqrt{5}}{2} \le \lambda < m.$$

All three statements require only a finite first offspring moment. Their proofs in Sections 2–5 are analytic. For laws with single-child vertices, Theorem 6.1 gives the stronger conclusion $v'(\lambda) < 0$ throughout $(1, m)$ by a computer-assisted argument. Section 6 describes the reductions to finite rational inequalities and identifies the exact computations; coefficient data and verification programs are supplied separately. Theorems 1.1–1.3 are proved independently of that extension. Lemma 2.2 proves continuity at $\lambda = 1$. Together with Theorems 1.2 and 6.1, this gives strict decrease throughout $(0, m)$ for laws with single-child vertices as well.

1.2. Relation to earlier work.

Aïdékon's explicit speed formula [Aïd14] provides the starting point for comparisons through the escape law. If the minimum offspring number is $m_1 \ge 2$, Song–Wang–Xiang proved strict decrease on $[0, m_1/(1 + \sqrt{1 - 1/m_1})]$ [SWX16, Theorem 1.2]. This interval contains $\lambda = 1$ in its interior, so it overlaps the interval $(1, m)$ on which we prove strict decrease in Theorem 1.1. Bowditch–Tokushige proved differentiability of speed under an exponential offspring moment assumption [BT20, Theorem 1.1]. The

analytic arguments here establish a strictly negative upper right Dini derivative under a finite first moment.

Remark 1.4 (Concurrent work). Schwarzfischer's recent preprint [Sch26], submitted on November 19, 2026, proves $v'(\lambda) < 0$ on $(0, m)$ under the assumptions $\nu \geq 4$ almost surely and $m < \infty$. Its introduction also states that related techniques can handle $\nu \geq 2$, without presenting that extension. Theorem 1.1 gives a full analytic argument under $\nu \geq 2$, while Theorems 1.2 and 1.3 treat laws with single-child vertices. Our representation of $v'(\lambda)$ is equivalent to that in [Sch26, Proposition 2.9]. Both build on Aïdékon's speed formula [Aïd14] and the linearization of the escape-law recursion. The estimates used here to establish strict negativity of $v'(\lambda)$ differ from those in [Sch26].

The result under $\nu \geq 2$ was obtained in September 2026. An early draft containing this result and its proof is preserved in the repository's initial commit, dated September 25, 2026.[1] Posting to arXiv or Hexagon was postponed while we sought a proof of the full conjecture that would not require computer verification.

1.3. Comparing the speed at two biases. Attach an extra parent to the root of a Galton–Watson tree. For $0 < \lambda < m$, $\lambda \neq 1$, let X, Y be independent copies of the quenched probability that the walk, started at the root, never reaches that parent. The speed representation used throughout the paper is

$$v(\lambda) = \frac{\mathbb{E}\dfrac{XY}{\lambda - 1 + X + Y - XY}}{\mathbb{E}\dfrac{X + Y - XY}{\lambda - 1 + X + Y - XY}}. \tag{1.2}$$

The law of X decreases in stochastic order when λ increases. For $\lambda > 1$, however, the sign of the corresponding speed change is not immediate from (1.2). Fix $v = v(\lambda)$ and subtract v times the denominator expectation from the numerator expectation. To control how this difference changes with the law of X, we subtract a suitable multiple of

$$\mathbb{E}u(X) - \mathbb{E}u\left(\frac{\sum_{i=1}^{\nu} X_i}{\lambda + \sum_{i=1}^{\nu} X_i}\right),$$

where the X_i are independent copies of X, independent of ν. This expression vanishes by the escape recursion. We choose the derivative of u to cancel one term in the change with respect to the law. The remaining change in the law and the change in the bias with the law fixed have the required signs whenever

$$\mathbb{E}X < 2\frac{\mathbb{E}[X/(\lambda - 1 + X)^2]}{\mathbb{E}[1/(\lambda - 1 + X)^2]}. \tag{1.3}$$

Proposition 3.2 makes this implication precise. Weighting the joint law of the children by $(\lambda - 1 + \sum_i X_i)^{-2}$ gives (1.3) for $1 < \lambda < m$ if there are at least two children, and for $(3 + \sqrt{5})/2 \leq \lambda < m$ if there is at least one child.

For $\lambda < 1$, the deterministic bound $X \geq 1 - \lambda$ gives an affine upper bound on the escape probability at a larger bias. The speed quotient preserves stochastic order for laws supported on $[1 - \lambda, 1]$, so the affine bound gives an upper bound on the new speed. We prove that its derivative is negative using the mean resistance $\mathbb{E}(1/X)$ and two applications of Jensen's inequality. Section 4 gives the complete argument.

Section 2.4 proves local Lipschitz continuity of the escape law and a formula for the derivative of its expectations. A positive inverse of the resulting linear operator expresses $v'(\lambda)$ as the negative of a quantity to be estimated, divided by a positive denominator. This formula gives the second proof in Section 3.2 and the starting point for the computer-assisted

[1] See GitHub commit `9374a17`.

estimates in Section 6. Moment inequalities and an identity involving a logarithmic function give lower bounds for that quantity. For each of four bias intervals, estimates on the moments and infinite tails reduce the required strict sign to finitely many rational inequalities. The exact computations verify those inequalities on domains that contain the parameters of every escape law under consideration.

2. Preliminaries

Throughout the paper the offspring variable satisfies $\mathbb{P}(\nu \geq 1) = 1$ and $1 < m = \mathbb{E}\nu < \infty$. The offspring law is fixed when the bias changes. We first express the speed through two independent escape probabilities. For $\lambda > 1$, we establish the kernel identities and linear-response formula used in the two proofs in Section 3 and in the later estimates.

2.1. The escape law and the speed formula. Attach an artificial parent to the root of a descendant tree T, and let $\beta_\lambda(T)$ be the quenched probability, starting from that root, of never visiting its parent. For $0 < \lambda < m$, transience gives $0 < \beta_\lambda(T) < 1$ almost surely. Write $\pi_\lambda = \mathrm{Law}(\beta_\lambda)$ and

$$F_\lambda(s) = \frac{s}{\lambda + s}, \qquad X \overset{\mathrm{d}}{=} F_\lambda(S), \qquad S = \sum_{i=1}^{\nu} X_i, \tag{2.1}$$

where the X_i have law π_λ, independently of one another and of ν. To obtain the recursion, condition on the child subtrees. A step to child i leads to escape before returning to the present vertex with probability X_i; after a return the same experiment restarts. The first-step equation is

$$X = \frac{1}{\lambda + \nu} \sum_{i=1}^{\nu} \{X_i + (1 - X_i)X\},$$

which rearranges to (2.1). At finite depth, terminal escape value one is the boundary condition for reaching that generation before the artificial parent. Those events decrease with the terminal depth, and their probabilities tend to the probability of never hitting the parent. Since F_λ increases in its argument and decreases in λ, induction on the finite recursion followed by this limit gives the same-tree order

$$\beta_\alpha(T) \geq \beta_\gamma(T), \qquad 0 < \alpha < \gamma < m. \tag{2.2}$$

Set

$$t(x, y) = x + y - xy, \qquad A_\lambda(x, y) = \lambda - 1 + t(x, y). \tag{2.3}$$

For independent X, Y with common probability law ρ on $[0, 1]$, define

$$N_\lambda(\rho) = \mathbb{E}_\rho \frac{XY}{A_\lambda(X, Y)}, \qquad Z_\lambda(\rho) = \mathbb{E}_\rho \frac{t(X, Y)}{A_\lambda(X, Y)}, \qquad \mathcal{V}_\lambda(\rho) = \frac{N_\lambda(\rho)}{Z_\lambda(\rho)} \tag{2.4}$$

whenever the integrals are finite and the denominator is positive. If $\lambda > 1$, then $\lambda - 1 \leq A_\lambda \leq \lambda$ on $[0, 1]^2$. If $0 < \lambda < 1$, the finite recursion gives $X \geq 1 - \lambda$ under π_λ, and

$$A_\lambda(x, y) \geq \lambda(1 - \lambda) > 0 \quad \text{for } x, y \in [1 - \lambda, 1]. \tag{2.5}$$

Proposition 2.1 (Symmetric speed formula). *For $0 < \lambda < m$, $\lambda \neq 1$, the speed satisfies*

$$v(\lambda) = \mathcal{V}_\lambda(\pi_\lambda). \tag{2.6}$$

Proof. We use Aïdékon's speed formula [Aïd14, Theorem 1.1], in the leafless form stated in [SWX16, (2.1)]:

$$v(\lambda) = \frac{\mathbb{E}[(\nu - \lambda)X_0/(\lambda - 1 + \sum_{i=0}^{\nu} X_i)]}{\mathbb{E}[(\nu + \lambda)X_0/(\lambda - 1 + \sum_{i=0}^{\nu} X_i)]}. \tag{2.7}$$

The X_i are independent with law π_λ, independently of ν. The speed is unchanged by attaching an artificial parent at the root; see [SWX16, Lemma 2.1]. Conditional exchangeability yields

$$\mathbb{E}\frac{\nu X_0}{\lambda - 1 + X_0 + S} = \mathbb{E}\frac{S}{\lambda - 1 + X_0 + S}.$$

Put $X = X_0$ and $Y = F_\lambda(S)$. Then X, Y are independent with law π_λ and $S = \lambda Y/(1-Y)$. Multiplication of the denominator by $1-Y$ gives $A_\lambda(X,Y)$. The numerator of (2.7) becomes

$$\lambda\mathbb{E}\frac{Y - X + XY}{A_\lambda(X,Y)} = \lambda N_\lambda(\pi_\lambda),$$

because A_λ is symmetric and X, Y have the same law. The denominator of (2.7) becomes $\lambda Z_\lambda(\pi_\lambda)$. The bounds preceding the proposition and $\mathbb{E}\nu < \infty$ justify these operations. ∎

The following lemma will join the strict comparisons on either side of bias one.

Lemma 2.2 (Continuity at bias one). *Suppose that* $\mathbb{P}(\nu \geq 1) = 1$ *and* $1 < m = \mathbb{E}\nu < \infty$. *Then*

$$\lim_{\lambda \to 1} v(\lambda) = v(1) = \mathbb{E}\left[\frac{\nu - 1}{\nu + 1}\right].$$

Proof. For each bias λ near one, let $X_0, X_1, \ldots$ be independent with law π_λ, independently of ν. Couple these variables for different biases by using the same independent trees, with $X_0 = \beta_\lambda(T)$.

Aïdékon's formula (2.7) holds under the stated assumptions for all λ in a neighborhood of one, including $\lambda = 1$ [Aïd14, Theorem 1.1]. Conditional on ν, the common denominator inside the expectations is symmetric in $X_0, \ldots, X_\nu$. Averaging the distinguished variable X_0 over these $\nu + 1$ variables therefore gives

$$v(\lambda) = \frac{\mathbb{E}\left[\dfrac{\nu - \lambda}{\nu + 1}\dfrac{\sum_{i=0}^{\nu} X_i}{\lambda - 1 + \sum_{i=0}^{\nu} X_i}\right]}{\mathbb{E}\left[\dfrac{\nu + \lambda}{\nu + 1}\dfrac{\sum_{i=0}^{\nu} X_i}{\lambda - 1 + \sum_{i=0}^{\nu} X_i}\right]}. \tag{2.8}$$

Fix $\Lambda \in (1, m)$. By (2.2) and transience at bias Λ, almost surely,

$$\sum_{i=0}^{\nu} X_i \geq X_0 = \beta_\lambda(T) \geq \beta_\Lambda(T) > 0, \qquad 0 < \lambda < \Lambda.$$

Consequently, whenever $|\lambda - 1| < \beta_\Lambda(T)/2$,

$$\left|\frac{\sum_{i=0}^{\nu} X_i}{\lambda - 1 + \sum_{i=0}^{\nu} X_i} - 1\right| \leq \frac{2|\lambda - 1|}{\beta_\Lambda(T)} \longrightarrow 0.$$

We next bound the integrands in (2.8) uniformly near one. For $\lambda \geq 1$,

$$0 < \frac{\sum_{i=0}^{\nu} X_i}{\lambda - 1 + \sum_{i=0}^{\nu} X_i} \leq 1.$$

For $0 < \lambda < 1$, the escape recursion with terminal value one gives $X_i \geq 1 - \lambda$. Indeed, if every child value is at least $1 - \lambda$, then their sum S is at least $1 - \lambda$, and

$$\frac{S}{\lambda + S} \geq \frac{1 - \lambda}{\lambda + 1 - \lambda} = 1 - \lambda.$$

Induction followed by the limit in depth proves the bound. Since $\nu \geq 1$, it follows that

$$0 < \frac{\sum_{i=0}^{\nu} X_i}{\lambda - 1 + \sum_{i=0}^{\nu} X_i} = 1 + \frac{1 - \lambda}{\sum_{i=0}^{\nu} X_i - (1 - \lambda)} \leq 1 + \frac{1}{\nu} \leq 2.$$

Also, for $0 < \lambda < \Lambda$,

$$\frac{|\nu - \lambda|}{\nu + 1} \leq 1 + \Lambda, \qquad \frac{\nu + \lambda}{\nu + 1} \leq 1 + \Lambda.$$

Thus both integrands in (2.8) are bounded in absolute value by $2(1+\Lambda)$. Dominated convergence shows that the numerator tends to $\mathbb{E}[(\nu-1)/(\nu+1)]$ and the denominator tends to one. Evaluating (2.8) at $\lambda=1$ gives

$$\lim_{\lambda\to 1} v(\lambda) = \mathbb{E}\left[\frac{\nu-1}{\nu+1}\right] = v(1). \qquad \blacksquare$$

2.2. Continuity of the escape law and speed for $1<\lambda<m$. The comparison of nearby biases requires continuity of π_λ. We prove it from the recursion using only $\mathbb{E}\nu<\infty$.

Lemma 2.3. *The escape law π_λ is continuous in the 1-Wasserstein metric on $(1,m)$. Consequently, v is continuous on $(1,m)$.*

Proof. By (2.2), the one-sided limits of the escape probabilities exist. We will identify their laws by proving uniqueness of the nonzero fixed law of (2.1). To distinguish two ordered fixed laws, we choose a test function whose expected change under one recursion step is strictly monotone along their linear interpolation if the laws differ. That change vanishes at both fixed laws, forcing the laws to agree.

Fix $1<\lambda<m$ and write $c=\lambda-1$. For a nonzero fixed law ρ, define

$$r_\rho(x) = \mathbb{E}_\rho \frac{Y}{A_\lambda(x,Y)^2}, \qquad u_\rho(x) = \int_0^x r_\rho(s)\,\mathrm{d}s.$$

To differentiate expectations after one step of the recursion, we use the following operator. If the children have law σ, let

$$(\mathcal{L}_{\lambda,\sigma} r)(x) = \mathbb{E}_{\nu,\sigma}\left[\frac{\nu\lambda}{(\lambda+x+\sum_{i=2}^{\nu} X_i)^2} r\left(F_\lambda\left(x+\sum_{i=2}^{\nu} X_i\right)\right)\right]. \tag{2.9}$$

The denominator identity

$$A_\lambda(F_\lambda(s),y) = \frac{\lambda(c+s+y)}{\lambda+s} \tag{2.10}$$

gives

$$\frac{\lambda}{(\lambda+s)^2} r_\rho(F_\lambda(s)) = \frac{1}{\lambda}\mathbb{E}_\rho \frac{Y}{(c+s+Y)^2}.$$

The right-hand side decreases in s. If σ stochastically dominates ρ, take independent copies of an ordered coupling, with the σ variable above the ρ variable. This gives $\mathcal{L}_{\lambda,\sigma} r_\rho \le \mathcal{L}_{\lambda,\rho} r_\rho$. At $\sigma=\rho$, the variable Y and the remaining children form ν independent variables with law ρ. Conditional exchangeability and then the fixed-point equation give

$$(\mathcal{L}_{\lambda,\rho} r_\rho)(x) = \frac{1}{\lambda}\mathbb{E}_\rho \frac{\nu Y}{(c+x+\sum_{i=2}^{\nu} X_i + Y)^2} = \frac{1}{\lambda}\mathbb{E}_\rho \frac{S}{(c+x+S)^2} = \mathbb{E}_\rho \frac{Y(1-Y)}{A_\lambda(x,Y)^2}.$$

Subtracting this identity from the definition of r_ρ proves

$$r_\rho(x) - (\mathcal{L}_{\lambda,\rho} r_\rho)(x) = \mathbb{E}_\rho \frac{Y^2}{A_\lambda(x,Y)^2} \ge \frac{\mathbb{E}_\rho Y^2}{\lambda^2} > 0. \tag{2.11}$$

To compare two ordered fixed laws, couple $X_0 \le X_1$ with laws ρ,σ and interpolate $X_s=(1-s)X_0+sX_1$. For the recursion residual

$$\mathfrak{R}_{u_\rho}(\lambda,\sigma) = \mathbb{E}_\sigma u_\rho(X) - \mathbb{E}_{\nu,\sigma} u_\rho\left(F_\lambda\left(\sum_{i=1}^{\nu} X_i\right)\right),$$

the chain rule along independent copies of this coupling yields

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}s}\mathfrak{R}_{u_\rho}(\lambda,\mathrm{Law}(X_s)) &= \mathbb{E}\left[(X_1-X_0)\{r_\rho - \mathcal{L}_{\lambda,\mathrm{Law}(X_s)} r_\rho\}(X_s)\right] \\ &\ge \frac{\mathbb{E}_\rho Y^2}{\lambda^2}\mathbb{E}(X_1-X_0).\end{aligned}$$

For a fixed number of children, this is the ordinary chain rule: the derivative of the sum is the sum of the coupled increments, and their contributions agree by exchangeability. The recursion term is bounded by $\nu\|r_\rho\|_\infty/\lambda$, since each increment is at most one; the direct term is bounded by $\|r_\rho\|_\infty$. These bounds justify differentiation under the offspring expectation. The residual vanishes at both $s = 0$ and $s = 1$, because both endpoint laws satisfy the recursion. Integration in s therefore forces $\mathbb{E}(X_1 - X_0) = 0$ and hence $\rho = \sigma$. Iteration from terminal value one dominates every fixed law: couple its terminal ones above independent samples of that fixed law and iterate. Its limit has law π_λ, so the preceding ordered comparison proves uniqueness of the nonzero fixed law.

At any $\lambda \in (1, m)$, each one-sided same-tree limit of β_α as $\alpha \to \lambda$ satisfies (2.1). This follows by bounded convergence, since the number of children is finite almost surely. Choose $\Lambda \in (\lambda, m)$. The one-sided limits dominate $\beta_\Lambda > 0$, and hence their laws are nonzero. Uniqueness identifies both with π_λ. The order (2.2) then gives

$$W_1(\pi_\alpha, \pi_\lambda) = \mathbb{E}|\beta_\alpha - \beta_\lambda| \longrightarrow 0.$$

The kernels in (2.4) are bounded and continuous for biases near $\lambda > 1$. Since $Z_\lambda(\pi_\lambda) > 0$, the speed formula gives continuity of v. ∎

2.3. Kernel identities and the weighted mean. Fix $1 < \lambda < m$, put $c = \lambda - 1$, and let X, Y be independent with law π_λ. In the rest of this subsection, $v = v(\lambda)$, $A = A_\lambda$, $N = N_\lambda(\pi_\lambda)$ and $Z = Z_\lambda(\pi_\lambda)$ are fixed. Define the mean with weight $(c + X)^{-2}$ by

$$k = \frac{\mathbb{E}[X/(c+X)^2]}{\mathbb{E}[1/(c+X)^2]}. \tag{2.12}$$

Define

$$h_v(y) = (1+v)y - v, \qquad (\mathcal{J}f)(x) = \mathbb{E}\frac{f(Y)}{A(x,Y)^2}. \tag{2.13}$$

The derivative of the integrand of $N - vZ$ in its first coordinate is

$$\partial_x \frac{xy - vt(x,y)}{A(x,y)} = \frac{y^2 + ch_v(y)}{A(x,y)^2}. \tag{2.14}$$

With $\mathcal{L} = \mathcal{L}_{\lambda,\pi_\lambda}$ from (2.9), set

$$r(x) = \mathbb{E}\frac{Y}{A(x,Y)^2}, \qquad u(x) = \int_0^x r(s)\,\mathrm{d}s. \tag{2.15}$$

Identity (2.10), exchangeability of the children, and (2.1) yield

$$\begin{aligned}(\mathcal{L}r)(x) &= \frac{1}{\lambda}\mathbb{E}\frac{\nu Y}{(c + x + \sum_{i=2}^{\nu} X_i + Y)^2} = \frac{1}{\lambda}\mathbb{E}\frac{S}{(c+x+S)^2} \\ &= \mathbb{E}\frac{Y(1-Y)}{A(x,Y)^2}.\end{aligned} \tag{2.16}$$

Thus $(I - \mathcal{L})r = \mathcal{J}(y^2)$. This identity will both cancel a term in the direct comparison and give a positive inverse of $I - \mathcal{L}$.

The explicit bias dependence of the speed formula and of the recursion will be expressed through

$$b(x) = \frac{x(1-x)}{\lambda}, \qquad C = \mathbb{E}\frac{v\,t(X,Y) - XY}{A(X,Y)^2}, \qquad K = 2\mathbb{E}[b(X)r(X)] - C. \tag{2.17}$$

By symmetry,

$$\lambda K = \mathbb{E}\frac{XY(2 - X - Y) + \lambda\{XY - vt(X,Y)\}}{A(X,Y)^2}.$$

Use $\lambda - A(x,y) = (1-x)(1-y)$ and $N - vZ = 0$ to remove the term with a single power of A in the denominator. The remaining numerator satisfies

$$\begin{aligned}(1-x)(1-y)\{xy - vt(x,y)\} + xy(2-x-y)\\ = x(1-x)h_v(y) + y(1-y)h_v(x) + (1+v)xy(1-x)(1-y).\end{aligned}$$

After division by $A(x,y)^2$ and averaging, this gives

$$\lambda K = 2\mathbb{E}[X(1-X)(\mathcal{J}h_v)(X)] + (1+v)\mathbb{E}\frac{XY(1-X)(1-Y)}{A(X,Y)^2}. \tag{2.18}$$

Lemma 2.4 (Positivity from the weighted mean). *If $\mathbb{E}X < 2k$, then*

$$\inf_{0\le x\le 1}(\mathcal{J}h_v)(x) > 0, \qquad K > 0.$$

Proof. We first show that $\mathcal{J}h_v$ is uniformly positive. Put $W = \mathbb{E}[X/A(X,Y)]$. Symmetry gives $N + Z = 2W$. For fixed x, the function $y \mapsto A(x,y)^{-1}$ is decreasing; its covariance with Y is nonpositive. Hence $N \le (\mathbb{E}X)W$ and

$$\frac{v}{1+v} = \frac{N}{2W} \le \frac{\mathbb{E}X}{2} < k. \tag{2.19}$$

Relative to the law weighted by $(c+y)^{-2}$, the law weighted by $A(x,y)^{-2}$ has density proportional to $\{(c+y)/A(x,y)\}^2$. This density increases in y, because

$$\frac{\partial}{\partial y}\frac{c+y}{A(x,y)} = \frac{\lambda x}{A(x,y)^2} \ge 0.$$

The mean of Y under this weight is therefore at least k. Since $A \le \lambda$,

$$(\mathcal{J}h_v)(x) \ge \frac{(1+v)k - v}{\lambda^2} > 0, \qquad 0 \le x \le 1. \tag{2.20}$$

By (2.18), this positivity and $0 < X, Y < 1$ give $K > 0$. ∎

2.4. Linear response of the escape law. Fix $1 < \lambda < m$. Throughout this subsection, $c = \lambda - 1$, X, Y are independent with law π_λ, and $v = v(\lambda)$. We retain A, $\mathcal{J}$, $\mathcal{L}$, r, b and C from Section 2.3, evaluated at this bias and law. To differentiate the speed formula we use

$$\Psi = \mathcal{J}\big(y^2 + ch_v(y)\big), \qquad \mathcal{R} = (I - \mathcal{L})^{-1}. \tag{2.21}$$

We first construct the inverse $\mathcal{R}$, then use it to differentiate expectations under π_λ using only the finite mean of ν.

Lemma 2.5 (Positive inverse). *The operator $\mathcal{L}$ is bounded and positive on $C([0,1])$. Its Neumann series $\mathcal{R} = \sum_{j\ge 0}\mathcal{L}^j$ converges in operator norm and defines a positive inverse of $I - \mathcal{L}$. In a neighborhood of any fixed $\lambda \in (1,m)$, the tails of the series admit a common geometric bound in an equivalent norm.*

Proof. For $f \in C([0,1])$, the integrand defining $\mathcal{L}f$ is bounded in absolute value by $\nu\|f\|_\infty/\lambda$. Thus finite mean gives boundedness, and dominated convergence gives continuity in x. Positivity follows directly from the formula for $\mathcal{L}$. Identity (2.16) gives

$$(I - \mathcal{L})r = \mathcal{J}(y^2), \qquad r(x) \ge \frac{\mathbb{E}Y}{\lambda^2} > 0, \qquad \mathcal{J}(y^2)(x) \ge \frac{\mathbb{E}Y^2}{\lambda^2} > 0.$$

The norm $\|f\|_r = \sup_{0\le x\le 1}|f(x)|/r(x)$ is equivalent to the uniform norm. Since $\mathcal{L}$ is positive,

$$|\mathcal{L}f| \le \|f\|_r \mathcal{L}r, \qquad \|\mathcal{L}\|_r \le 1 - \min_{0\le x\le 1}\frac{\mathcal{J}(y^2)(x)}{r(x)} < 1.$$

The geometric series therefore converges and gives the asserted positive inverse.

For the local statement, keep this base function r fixed. If $\alpha \to \lambda$ and $\rho \to \pi_\lambda$ in W_1, then

$$\|\mathcal{L}_{\alpha,\rho} f - \mathcal{L} f\|_\infty \longrightarrow 0 \quad \text{for each fixed } f \in C([0,1]). \tag{2.22}$$

Indeed, for $\nu \le M$ there are only finitely many continuous kernels on compact cubes, and their integrals converge uniformly in x. For biases $\alpha \ge a > 1$, the sum of the two discarded tails is at most $2\|f\|_\infty \mathbb{E}[\nu; \nu > M]/a$. First choose M and then take the limit. Apply (2.22) with $f = r$. The strictly positive minimum of $r - \mathcal{L} r$ persists for nearby operators $\mathcal{L}_{\alpha,\pi_\alpha}$, so their norms in $\|\cdot\|_r$ have a common bound strictly below one. This gives the local geometric bound using convergence on each fixed test function. ∎

Lemma 2.6 (Local Lipschitz continuity). *The map $\lambda \mapsto \pi_\lambda$ is locally Lipschitz in W_1 on $(1, m)$.*

Proof. Continuity was proved in Lemma 2.3 without differentiation. Fix the base parameter, put $u(x) = \int_0^x r(s)\,\mathrm{d}s$, and retain the recursion residual

$$\mathfrak{R}_u(\lambda, \rho) = \mathbb{E}_\rho u(X) - \mathbb{E}_{\nu,\rho} u\left(F_\lambda\left(\sum_{i=1}^{\nu} X_i\right)\right).$$

By (2.22), every law on the ordered interpolation between π_λ and $\pi_{\lambda+\varepsilon}$ satisfies

$$r - \mathcal{L}_{\lambda,\rho} r \ge \tfrac{1}{2} \min_{[0,1]} \mathcal{J}(y^2)$$

when $|\varepsilon|$ is small. Integrating the residual's law derivative along that interpolation, in its increasing direction, gives

$$|\mathfrak{R}_u(\lambda, \pi_{\lambda+\varepsilon})| \ge \tfrac{1}{2} \min_{[0,1]} \mathcal{J}(y^2)\, W_1(\pi_{\lambda+\varepsilon}, \pi_\lambda).$$

The residual at π_λ is zero. The fixed-point equation at $\lambda + \varepsilon$ also gives

$$\mathfrak{R}_u(\lambda, \pi_{\lambda+\varepsilon}) = \mathbb{E}_{\nu,\pi_{\lambda+\varepsilon}}\left[u(F_{\lambda+\varepsilon}(S)) - u(F_\lambda(S))\right].$$

For $\alpha \ge a > 1$ and $s \ge 0$,

$$|\partial_\alpha F_\alpha(s)| = \frac{s}{(\alpha + s)^2} \le \frac{1}{4a}.$$

The residual is consequently at most $|\varepsilon|\|r\|_\infty/(4a)$ in absolute value. Combining the two inequalities proves the bound centered at the fixed parameter. To obtain a common Lipschitz constant on a neighborhood, keep the same base function r and use $\mathfrak{R}_u(\alpha, \pi_\beta)$ for any two nearby biases α, β. The continuity in (2.22) preserves the same positive lower bound for $r - \mathcal{L}_{\alpha,\rho} r$ for all their interpolating laws. The residual at π_α is zero, and its parameter bound is at most $|\beta - \alpha|\|r\|_\infty/(4a)$. Thus the same constant bounds $W_1(\pi_\alpha, \pi_\beta)/|\beta - \alpha|$ throughout a sufficiently small neighborhood. Neither estimate requires a second offspring moment. ∎

Lemma 2.7 (Derivatives of expectations and of the speed). *For every $f \in C^1([0,1])$,*

$$\frac{\mathrm{d}}{\mathrm{d}\lambda} \mathbb{E} f(\beta_\lambda) = -\mathbb{E}[b(X)(\mathcal{R} f')(X)]. \tag{2.23}$$

The derivative is continuous on $(1, m)$. In particular, $v \in C^1((1, m))$ and

$$v'(\lambda) = -\frac{D_\lambda}{Z}, \qquad D_\lambda = 2\mathbb{E}[b(X)(\mathcal{R}\Psi)(X)] - C, \qquad Z = Z_\lambda(\pi_\lambda) > 0. \tag{2.24}$$

Proof. Couple $X = \beta_\lambda$ and $X_\varepsilon = \beta_{\lambda+\varepsilon}$ on the same tree. For $\varepsilon \ne 0$ small, define a positive finite measure by

$$\eta_\varepsilon(a) = \mathbb{E}\left[\frac{1}{\varepsilon} \int_{X_\varepsilon}^{X} a(s)\,\mathrm{d}s\right], \qquad a \in C([0,1]).$$

Positivity holds for both signs of ε, by the same-tree order. Lemma 2.6 bounds its mass:

$$\eta_\varepsilon(1) = \frac{W_1(\pi_{\lambda+\varepsilon}, \pi_\lambda)}{|\varepsilon|}.$$

Let u be a primitive of a fixed continuous a. Then $(\mathbb{E}u(X_\varepsilon) - \mathbb{E}u(X))/\varepsilon = -\eta_\varepsilon(a)$. Use the fixed-point equations to compute the same difference by first changing the bias at fixed child law $\pi_{\lambda+\varepsilon}$, and then changing the child law at fixed bias λ. The parameter contribution tends to $-\mathbb{E}[b(X)a(X)]$. Its integrand is bounded by $\|a\|_\infty/(4a_0)$ for a positive lower bound a_0 on the nearby biases, because $|\partial_\alpha F_\alpha(s)| \le 1/(4a_0)$.

For the child-law contribution, let $X_t = (1-t)X + tX_\varepsilon$ and $\rho_t = \mathrm{Law}(X_t)$. Differentiating the finite-child recursion and using exchangeability gives the exact quotient

$$-\int_0^1 \mathbb{E}\left[\frac{X - X_\varepsilon}{\varepsilon}(\mathcal{L}_{\lambda,\rho_t}a)(X_t)\right] \mathrm{d}t.$$

For each fixed a, finite-mean truncation as in (2.22) yields

$$\sup_{0\le t\le 1} \|\mathcal{L}_{\lambda,\rho_t}a - \mathcal{L}a\|_\infty \longrightarrow 0.$$

Multiplying this uniform error by the bounded mass $\eta_\varepsilon(1)$ shows that the child-law quotient is $-\eta_\varepsilon(\mathcal{L}a) + o(1)$. Hence

$$\eta_\varepsilon((I - \mathcal{L})a) = \mathbb{E}[ba] + o(1). \tag{2.25}$$

Take $a = \mathcal{R}f'$, which is continuous by Lemma 2.5. Equation (2.25) gives (2.23) for two-sided increments. This passage uses bounded expectations of $(X - X_\varepsilon)/\varepsilon$ and a uniform test-function error; it does not require uniform integrability of the random difference quotients.

For continuity of the derivative, nearby operators converge on each fixed continuous test by (2.22). Each finite Neumann sum therefore converges strongly, by induction on its length and uniform boundedness of the operators. The common geometric tail bound in Lemma 2.5 then proves strong convergence of the resolvents. Together with W_1 continuity of π_λ and continuity of b, this proves continuity in (2.23).

Finally, freeze $v = v(\lambda)$ in

$$H_\alpha(x,y) = \frac{xy - vt(x,y)}{\alpha - 1 + t(x,y)}.$$

On any compact bias interval in $(1, m)$, this kernel and its first derivatives are bounded and uniformly continuous on $[0,1]^2$. Its averaged coordinate derivative at the base parameter is Ψ, by (2.14); its averaged explicit parameter derivative is C. Applying (2.23) to the two marginals gives

$$\frac{\mathrm{d}}{\mathrm{d}\alpha}\mathbb{E}H_\alpha(\beta_\alpha, \widetilde{\beta}_\alpha)\Big|_{\alpha=\lambda} = C - 2\mathbb{E}[b\mathcal{R}\Psi].$$

Here the two escapes are independent. For precision, the averaged test in one marginal can depend on the increment through the other marginal. Its derivative converges uniformly to the base averaged derivative, by compact uniform continuity and W_1 convergence. The resulting difference-quotient error is bounded by that uniform difference times $\eta_\varepsilon(1)$, so it tends to zero. The same argument applied separately to N_λ and Z_λ first establishes differentiability of their quotient. Since the last display equals $Zv'(\lambda)$, it proves (2.24). Strong resolvent continuity proves that v' is continuous. ∎

3. Strict decrease on $[0, m)$ when every vertex has at least two children

We give two proofs of Theorem 1.1: one compares nearby biases directly, and the other uses the linear-response formula. Both use the inequality $\mathbb{E}X < 2k$ for $1 < \lambda < m$, which we prove first. We weight the joint law of the children by $(\lambda - 1 + \sum_i X_i)^{-2}$. Under this

measure, the mean of the first child is at least k. Conditional exchangeability then turns $\nu \geq 2$ into a lower bound on their sum.

LEMMA 3.1. *Let $\lambda > 1$, $c = \lambda - 1$, and let $X \in (0,1)$ satisfy (2.1) with $\nu \geq 2$ almost surely. For k defined in (2.12), one has $\mathbb{E}X < 2k$.*

Proof. Let M be the probability law obtained by weighting the law of X with $(c+X)^{-2}$, and write

$$k = \mathbb{E}_M X, \qquad j = \mathbb{E}_M X^2, \qquad \ell = \mathbb{E}_M X^3. \tag{3.1}$$

Undoing this change of measure expresses the original mean as

$$\mathbb{E}X = \frac{c^2 k + 2cj + \ell}{c^2 + 2ck + j}. \tag{3.2}$$

We will use $\ell \leq j$ and an upper bound on j to prove $\mathbb{E}X < 2k$. To obtain the bound on j, weight the joint law of $(\nu, X_1, X_2, \ldots)$ by $(c+S)^{-2}$ and denote the resulting probability measure by Q. Under the original law, X_1 is independent of $H = \sum_{i=2}^{\nu} X_i$. Its marginal under Q, relative to M, has density proportional to

$$R(x) = \mathbb{E}\left[\left(\frac{c+x}{c+x+H}\right)^2\right].$$

The function R is increasing. The covariance inequality for two increasing functions therefore gives

$$\mathbb{E}_Q X_1 = \frac{\mathbb{E}_M[XR(X)]}{\mathbb{E}_M R(X)} \geq k.$$

For example, the covariance sign follows by taking independent U, V with law M and writing $2\,\mathrm{Cov}_M(X, R(X)) = \mathbb{E}[(U-V)(R(U)-R(V))] \geq 0$. Conditional exchangeability under the symmetric weight and $\nu \geq 2$ now imply

$$\mathbb{E}_Q S = \mathbb{E}_Q[\nu X_1] \geq 2\mathbb{E}_Q X_1 \geq 2k. \tag{3.3}$$

By (2.1), S has the same law as $\lambda X/(1-X)$ and $c+S$ has the same law as $(c+X)/(1-X)$. Thus

$$\mathbb{E}_Q S = \lambda \frac{k-j}{1-2k+j}.$$

The denominator is $1-2k+j = \mathbb{E}_M(1-X)^2 > 0$. Combining this formula with (3.3) gives

$$j \leq \frac{k(c-1+4k)}{c+1+2k}. \tag{3.4}$$

Since $j \geq k^2$ and

$$\frac{k(c-1+4k)}{c+1+2k} - k^2 = \frac{k(1-k)(c+2k-1)}{c+1+2k},$$

we also have $c+2k-1 \geq 0$.

If $k \geq 1/2$, then $\mathbb{E}X < 1 \leq 2k$. Suppose $k < 1/2$. First use $\ell \leq j$ in (3.2). Since $2k - 2c - 1 < 0$, the upper bound for j in (3.4) then gives a lower bound for $2k - \mathbb{E}X$:

$$\begin{aligned}
&(2k - \mathbb{E}X)(c^2 + 2ck + j) \\
&\quad \geq c^2 k + 4ck^2 + (2k - 2c - 1)\frac{k(c-1+4k)}{c+1+2k} \\
&\quad = \frac{k}{c+1+2k}\Big[(c+2k-1)^3 + 2(c+2k-1)^2 \\
&\qquad\qquad + \{2 + 2k(1-2k)\}(c+2k-1) + 2(1-k)(1-2k)\Big] > 0.
\end{aligned}$$

Each term in brackets is nonnegative, and the last is strictly positive. This proves $\mathbb{E}X < 2k$. ∎

For $1 < \lambda < m$, the assumption $\nu \geq 2$ enters through (3.3). Independence is used under the original law and conditional exchangeability under Q. If one child is allowed, the same calculation gives only $\mathbb{E}_Q S \geq k$. Section 5 uses this weaker inequality for $(3+\sqrt{5})/2 \leq \lambda < m$.

3.1. Proof by direct comparison. We first prove a criterion that also applies when single-child vertices are allowed. Fix $1 < \lambda < m$ and use the notation of Section 2.3, evaluated at this bias and escape law.

Proposition 3.2. *If $\mathbb{E}X < 2k$, then $D^+ v(\lambda) < 0$.*

Proof. We compare a larger bias with λ by changing the escape law and the bias separately. The escape law decreases in stochastic order. We modify $N_\alpha(\rho) - vZ_\alpha(\rho)$ by a term that vanishes when $\rho = \pi_\alpha$, so that the modified expression increases with ρ near π_λ and has a negative derivative in α at (λ, π_λ). Lemma 2.4 gives the two signs needed below.

For a nearby parameter $\alpha > 1$ and a child law ρ on $[0,1]$, keep u and v fixed and set

$$\mathscr{F}(\alpha, \rho) = N_\alpha(\rho) - vZ_\alpha(\rho) - 2\left\{ \mathbb{E}_\rho u(X) - \mathbb{E}_{\nu,\rho} u\left(F_\alpha\left(\sum_{i=1}^{\nu} X_i \right) \right) \right\}. \tag{3.5}$$

The expression in braces vanishes at $\rho = \pi_\alpha$ by the recursion. Consequently,

$$\mathscr{F}(\alpha, \pi_\alpha) = Z_\alpha(\pi_\alpha)\{v(\alpha) - v(\lambda)\}. \tag{3.6}$$

For an ordered coupling $X_0 \leq X_1$, write $X_s = (1-s)X_0 + sX_1$ and $\rho_s = \mathrm{Law}(X_s)$. Differentiating (3.5) gives

$$\frac{\mathrm{d}}{\mathrm{d}s}\mathscr{F}(\alpha, \rho_s) = 2\mathbb{E}[(X_1 - X_0) G_{\alpha,\rho_s}(X_s)], \tag{3.7}$$

$$G_{\alpha,\rho}(x) = \int \frac{y^2 + (\alpha-1)h_v(y)}{A_\alpha(x,y)^2} \rho(\,\mathrm{d}y) - r(x) + (\mathcal{L}_{\alpha,\rho} r)(x).$$

At $(\alpha, \rho) = (\lambda, \pi_\lambda)$, the choice of r gives $G_{\lambda,\pi_\lambda} = c\mathcal{J}h_v$, which is uniformly positive by (2.20). The map $(\alpha, \rho) \mapsto G_{\alpha,\rho}$ is continuous in the uniform norm when ρ varies in W_1. For the pair-kernel term this follows from uniform continuity on the compact parameter and coordinate domain. For $\mathcal{L}_{\alpha,\rho} r$, truncate ν at M. For each of the finitely many possible offspring numbers, couple the child variables so their mean absolute differences tend to zero. Uniform continuity of the finite-child integrand then gives convergence uniformly in $x \in [0,1]$. The tail of each operator, on a compact interval with $\alpha \geq a > 1$, is bounded by

$$\frac{\|r\|_\infty}{a} \mathbb{E}[\nu; \nu > M] \longrightarrow 0.$$

The same bound justifies the differentiation in (3.7). Hence $G_{\alpha,\rho} > 0$ throughout a neighborhood of (λ, π_λ).

We next vary α with the law fixed at π_λ. Since $\partial_\alpha F_\alpha(s)|_{\alpha=\lambda} = -b(F_\lambda(s))$, the derivative of $\mathscr{F}(\alpha, \pi_\lambda)$ at $\alpha = \lambda$ equals $C - 2\mathbb{E}[br] = -K < 0$, by Lemma 2.4.

We now combine the two comparisons. Couple $\beta_{\lambda+\varepsilon} \leq \beta_\lambda$ on the same tree. By Lemma 2.3, the entire linear interpolation of these variables has law converging to π_λ in W_1, uniformly along the interpolation. Equation (3.7) therefore gives, for small $\varepsilon > 0$,

$$\begin{aligned} Z_{\lambda+\varepsilon}(\pi_{\lambda+\varepsilon})\{v(\lambda+\varepsilon) - v(\lambda)\} &= \mathscr{F}(\lambda+\varepsilon, \pi_{\lambda+\varepsilon}) \\ &\leq \mathscr{F}(\lambda+\varepsilon, \pi_\lambda) = -K\varepsilon + o(\varepsilon). \end{aligned}$$

Since $Z_{\lambda+\varepsilon}(\pi_{\lambda+\varepsilon}) \to Z > 0$, we obtain $D^+ v(\lambda) \leq -K/Z < 0$. ∎

We will also use the elementary fact that a continuous function with strictly negative upper right Dini derivative at every point of an interval is strictly decreasing. Indeed, if

$f(b) \geq f(a)$ for some $a < b$, a minimum on $[a, b]$ is attained at some $x < b$. All sufficiently small right difference quotients at that x are nonnegative, a contradiction.

First proof of Theorem 1.1. For each $1 < \lambda < m$, Lemma 3.1 and Proposition 3.2 give $D^+v(\lambda) < 0$. Lemma 2.3 and the Dini criterion above imply strict decrease on $(1, m)$.

To include $\lambda \leq 1$, we use a known interval of strict decrease that overlaps $(1, m)$. Let $m_1 = \min\{k : \mathbb{P}(\nu = k) > 0\} \geq 2$. Song–Wang–Xiang [SWX16, Theorem 1.2 and Lemma 2.5] prove strict decrease on

$$\left[0, \frac{m_1}{1 + \sqrt{1 - 1/m_1}}\right]. \tag{3.8}$$

The right endpoint is larger than one and smaller than $m_1 \leq m$. Therefore (3.8) overlaps $(1, m)$, and the two strict comparisons yield strict decrease on $[0, m)$. ∎

3.2. Proof via linear response. The second proof uses the same weighted-mean inequality to determine the sign of the derivative formula from Section 2.4.

Second proof of Theorem 1.1. Fix $1 < \lambda < m$. Lemma 3.1 gives $\mathbb{E}X < 2k$, so Lemma 2.4 gives $\mathcal{J}h_v > 0$ and $K > 0$. Recall $\Psi = \mathcal{J}(y^2 + ch_v)$ and the positive inverse $\mathcal{R} = (I - \mathcal{L})^{-1}$ from (2.21) and Lemma 2.5. Applying $\mathcal{R}$ to $(I - \mathcal{L})r = \mathcal{J}(y^2)$ yields

$$\mathcal{R}\Psi = r + c\mathcal{R}\mathcal{J}h_v. \tag{3.9}$$

Substituting this identity into (2.24) gives

$$\begin{aligned} D_\lambda &= 2\mathbb{E}[b(X)r(X)] - C + 2c\mathbb{E}[b(X)(\mathcal{R}\mathcal{J}h_v)(X)] \\ &= K + 2c\mathbb{E}[b(X)(\mathcal{R}\mathcal{J}h_v)(X)] \geq K > 0. \end{aligned}$$

Here $c > 0$, $b \geq 0$ and positivity of $\mathcal{R}$ give the inequality. Since $Z > 0$, Lemma 2.7 gives $v'(\lambda) = -D_\lambda/Z < 0$. The mean value theorem now gives strict decrease on $(1, m)$. Combining this interval with the overlapping Song–Wang–Xiang interval (3.8) proves strict decrease on $[0, m)$. ∎

The second proof gives $v'(\lambda) < 0$ on $(1, m)$ under the finite first-moment assumption. If the offspring law has an exponential moment, Bowditch–Tokushige [BT20, Theorem 1.1] give differentiability of the speed throughout $(0, m)$ in the leafless case. The bound $D^+v(\lambda) < 0$ for $1 < \lambda < m$, together with [SWX16, Lemma 2.5], then gives $v'(\lambda) < 0$ for every $0 < \lambda < m$. The strict monotonicity in Theorem 1.1 itself uses only the finite first moment.

4. Strict decrease for $0 < \lambda < 1$

We now allow vertices with a single child. For $0 < \lambda < 1$, the escape probability is at least $1 - \lambda$ on every tree. This bound allows us to compare the escape laws at two biases. We combine the comparison with a resistance estimate to show that the decrease in escape probabilities outweighs the explicit dependence on λ in the speed formula.

4.1. An affine bound for the escape probabilities.

Lemma 4.1. *For $0 < \lambda < \alpha < 1$, the same-tree escape probabilities satisfy*

$$\beta_\alpha \leq T_{\lambda,\alpha}(\beta_\lambda), \qquad T_{\lambda,\alpha}(x) = 1 - \frac{\alpha(1+\lambda)}{\lambda(1+\alpha)}(1 - x). \tag{4.1}$$

The affine map $T_{\lambda,\alpha}$ is increasing and sends $[1 - \lambda, 1]$ into $[1 - \alpha, 1]$. Moreover, $\lambda \mapsto \beta_\lambda(T)$ is locally Lipschitz on $(0, 1)$, uniformly in the tree.

Proof. We first bound the derivative of the finite-depth recursion, integrate the bound, and then let the depth tend to infinity. Take terminal value one; the resulting recursion is differentiable for positive bias. Put $d = 1 - \lambda$ and $a_\lambda = 1/[\lambda(1+\lambda)]$. At a vertex with $n \geq 1$ children, write $x = S/(\lambda + S)$, where $S = \sum_{i=1}^{n} x_i$. If every child value is at least d, then $S \geq nd \geq d$ and $S/(\lambda+S) \geq d/(\lambda+d) = d$. Thus induction gives $d \leq x \leq 1$ at every depth. With $q = -\partial_\lambda x$ and $q_i = -\partial_\lambda x_i$,

$$q = \frac{S + \lambda \sum_{i=1}^{n} q_i}{(\lambda+S)^2}, \qquad 0 \leq q \leq \frac{x}{d} \leq \frac{1}{d}. \tag{4.2}$$

For the upper bound, the induction hypothesis $q_i \leq x_i/d$ gives

$$q \leq \frac{S + \lambda S/d}{(\lambda+S)^2} = \frac{S}{d(\lambda+S)^2} \leq \frac{x}{d},$$

because $\lambda + S \geq \lambda + d = 1$. At the terminal generation the derivative is zero, so the induction starts.

We also claim $q \geq a_\lambda(1-x)$. The claim holds at the terminal vertices. If it holds for all children, subtracting the required numerator from the numerator in (4.2) gives

$$\begin{aligned} S + \lambda a_\lambda(n-S) - a_\lambda\lambda(\lambda+S) &= \frac{n - \lambda - dS}{1+\lambda} \\ &\geq \frac{\lambda(n-1)}{1+\lambda} \geq 0, \end{aligned}$$

where $S \leq n$ was used. Integration of $(1-x)' \geq a_\lambda(1-x)$ between λ and α yields (4.1) at finite depth, since

$$\exp\left(\int_\lambda^\alpha \frac{\mathrm{d}s}{s(1+s)}\right) = \frac{\alpha(1+\lambda)}{\lambda(1+\alpha)}.$$

The affine map $T_{\lambda,\alpha}$ is continuous, so letting the depth tend to infinity proves the comparison for β_λ and β_α. On every compact interval $[a,b] \subset (0,1)$, the finite-depth derivative bound is at most $1/(1-b)$. Integrating that bound between any two biases and then passing to the depth limit gives

$$|\beta_{\lambda_1}(T) - \beta_{\lambda_2}(T)| \leq \frac{|\lambda_1 - \lambda_2|}{1-b}, \qquad \lambda_1, \lambda_2 \in [a,b],$$

uniformly in the tree. Finally,

$$T_{\lambda,\alpha}(1-\lambda) = 1 - \frac{\alpha(1+\lambda)}{1+\alpha} \geq 1 - \alpha, \qquad T_{\lambda,\alpha}(1) = 1,$$

which verifies the support assertion. ∎

4.2. Stochastic order and the speed quotient. Fix $0 < \lambda < 1$ and write $d = 1 - \lambda$. For a probability law σ supported on $[d,1]$, put $N = N_\lambda(\sigma)$ and $Z = Z_\lambda(\sigma)$. Its speed quotient $v = N/Z = \mathcal{V}_\lambda(\sigma)$ satisfies

$$\frac{d}{1+\lambda} \leq v \leq 1. \tag{4.3}$$

Indeed $\mathcal{V}_\lambda(\sigma)$ is a weighted average of $xy/t(x,y)$, which lies between $d/(2-d)$ and one on $[d,1]^2$. With v fixed at this quotient, let

$$\begin{aligned} P_v(y) &= y^2 - d\{(1+v)y - v\}, \\ \Psi(x) = \mathbb{E}_\sigma \frac{P_v(Y)}{A_\lambda(x,Y)^2}, &\qquad C = \mathbb{E}_\sigma \frac{v\,t(X,Y) - XY}{A_\lambda(X,Y)^2}. \end{aligned} \tag{4.4}$$

The coordinate derivative of $(xy - vt(x,y))/A_\lambda(x,y)$ is $P_v(y)/A_\lambda(x,y)^2$. Since

$$P_v(y) = y(y-d) + dv(1-y) > 0, \qquad d \leq y \leq 1,$$

increasing the escape variables increases the quotient. To see this, couple $X_0 \le X_1$ with laws supported on $[d,1]$ and put $X_s = (1-s)X_0 + sX_1$. Write $v_s = \mathcal{V}_\lambda(\mathrm{Law}(X_s))$, and let Y_s be an independent copy. Quotient differentiation gives

$$\frac{\mathrm{d}}{\mathrm{d}s}\mathcal{V}_\lambda(\mathrm{Law}(X_s)) = \frac{2}{Z_\lambda(\mathrm{Law}(X_s))}\mathbb{E}\left[(X_1 - X_0)\frac{P_{v_s}(Y_s)}{A_\lambda(X_s,Y_s)^2}\right] \ge 0.$$

For each s, the quotient obeys (4.3), so $P_{v_s} > 0$ by the displayed factorization. The positive denominator bound (2.5) bounds every kernel derivative on this compact square and justifies differentiation under the expectation. Thus $\mathcal{V}_\lambda$ preserves stochastic order among laws supported on $[1-\lambda, 1]$. More generally, if X varies differentiably within this interval with uniformly bounded derivative $\dot{X}$, the same quotient rule gives the derivative $2\mathbb{E}[\dot{X}\Psi(X)]/Z$. Changing λ with the law fixed gives the derivative C/Z.

From now on, X, Y are independent with law π_λ at the fixed parameter λ, $N = N_\lambda(\pi_\lambda)$, $Z = Z_\lambda(\pi_\lambda)$ and $v = v(\lambda)$. All expectations in (4.4) refer to this law. The affine comparison in Lemma 4.1 bounds the speed at a larger bias by $\mathcal{V}_\alpha(\mathrm{Law}(T_{\lambda,\alpha}(X)))$, with equality at $\alpha = \lambda$. Its derivative there is

$$\frac{1}{Z}\left(C - \frac{2}{\lambda(1+\lambda)}\mathbb{E}[(1-X)\Psi(X)]\right).$$

We will prove that this derivative is negative. Identity (4.9) below expresses the negative of its numerator with a positive coefficient of C and a nonpositive coefficient of $\mathbb{E}[1/A_\lambda(X,Y)]$. The next three lemmas provide the corresponding bounds, together with $v\mathbb{E}(1/X) \le 1$.

Lemma 4.2 (Speed and mean resistance). *Write $\rho = \mathbb{E}(1/X)$ and $e = \mathbb{E}(1/\nu)$. Then*

$$v \le 1 - \lambda e, \qquad 1 < \rho \le \frac{1}{1-\lambda e}, \qquad v\rho \le 1. \tag{4.5}$$

Proof. To bound v, we write (2.7) as a weighted average of local drifts. Let $T_j = \sum_{i=1}^{j} X_i$ for independent X_i with law π_λ. Conditional on $\nu = n$, the weight in the denominator is

$$W_{n+1}, \qquad W_j = \frac{j-d}{j}\mathbb{E}\frac{T_j}{T_j - d} = 1 + d\mathbb{E}\frac{1-X_1}{T_j - d}, \qquad j \ge 2.$$

The identity for W_j follows by exchangeability: $\mathbb{E}[X_1/(T_j - d)] = j^{-1}\mathbb{E}[T_j/(T_j - d)]$. Coupling all T_j by one independent sequence shows that W_j decreases with j, since the nonnegative numerator $1 - X_1$ stays fixed while the denominator increases. The degree formula is exactly

$$v = \frac{\mathbb{E}\left[W_{\nu+1}(\nu-\lambda)/(\nu+\lambda)\right]}{\mathbb{E}W_{\nu+1}}.$$

The local drift $(n-\lambda)/(n+\lambda)$ increases with n, whereas W_{n+1} decreases. Their covariance is nonpositive. All terms are integrable: $T_j - d \ge (j-1)d$ and the local drift is bounded. Consequently,

$$v \le \mathbb{E}\frac{\nu-\lambda}{\nu+\lambda} = 1 - 2\lambda\mathbb{E}\frac{1}{\nu+\lambda} \le 1 - \lambda\mathbb{E}\frac{1}{\nu}.$$

For the last inequality, $\lambda < 1 \le \nu$ implies $2/(\nu+\lambda) \ge 1/\nu$.

The resistance $R = 1/X$ is bounded by $1/d$ and satisfies

$$R \overset{\mathrm{d}}{=} 1 + \frac{\lambda}{\sum_{i=1}^{\nu} R_i^{-1}}.$$

The harmonic-mean inequality gives $(\sum_{i=1}^{\nu} R_i^{-1})^{-1} \le \nu^{-2}\sum_{i=1}^{\nu} R_i$. Conditional on $\nu = n$, independence gives the expectation of the right-hand side as ρ/n. Thus $\rho \le 1 + \lambda e\rho$. Since $\lambda e < 1$, this gives $\rho \le (1-\lambda e)^{-1}$. Both v and ρ are positive, and their respective upper bounds give $v\rho \le (1-\lambda e)/(1-\lambda e) = 1$. Strictness of $\rho > 1$ follows from $X < 1$ almost surely. ∎

LEMMA 4.3 (Sign of the explicit parameter term). *The quantity C in (4.4), evaluated under π_λ, is nonnegative.*

Proof. We expand C and the identity $vZ - N = 0$ in the same powers of λ^{-1}. Their coefficients will express C as a covariance of two increasing sequences. Put $a_j = \mathbb{E}(1-X)^j$ and $b_j = a_j - a_{j+1}$ for $j \geq 0$. Since $1 - X \leq \lambda$, the geometric expansion of $A_\lambda^{-1} = [\lambda - (1-X)(1-Y)]^{-1}$ gives absolutely convergent series

$$N = \lambda^{-1} \sum_{j\geq 0} b_j^2 \lambda^{-j},$$
$$Z = \lambda^{-1} \sum_{j\geq 0} b_j(2a_j - b_j)\lambda^{-j},$$
$$C = \lambda^{-2} \sum_{j\geq 0} (j+1)\{2va_jb_j - (1+v)b_j^2\}\lambda^{-j}.$$

For example $a_j, b_j \leq \lambda^j$, so the series for C is dominated by a constant times $\sum(j+1)\lambda^j$. The expansion follows from

$$\frac{1}{A_\lambda(X,Y)} = \frac{1}{\lambda} \sum_{j\geq 0} \left(\frac{(1-X)(1-Y)}{\lambda} \right)^j, \quad b_j = \mathbb{E}[X(1-X)^j].$$

The expansion of A_λ^{-2} has the additional factors $j+1$. Since $0 < X < 1$, all a_j, b_j are positive. Cauchy–Schwarz gives $a_{j+1}^2 \leq a_j a_{j+2}$, so a_{j+1}/a_j increases and $b_j/a_j = 1 - a_{j+1}/a_j$ decreases. Therefore $2v - (1+v)b_j/a_j$ increases with j. Its weighted sum is

$$\sum_{j\geq 0} a_j b_j \lambda^{-j} \left(2v - (1+v)\frac{b_j}{a_j} \right) = \lambda(vZ - N) = 0.$$

Normalize the summable positive weights $a_j b_j \lambda^{-j}$ to a probability distribution on the nonnegative integers. The covariance of the increasing sequences $j+1$ and $2v - (1+v)b_j/a_j$ is nonnegative. The zero weighted mean shows that the series for C equals a positive constant times this covariance. Its first moment is finite by the geometric bound already proved, so $C \geq 0$. ∎

LEMMA 4.4 (Reciprocal kernel estimate). *With $\rho = \mathbb{E}(1/X)$, set $B_\rho = 2\rho - 1 - d\rho^2$. Then $B_\rho > 0$ and*

$$K := \mathbb{E}\frac{1}{A_\lambda(X,Y)} \leq \frac{\rho^2}{B_\rho}. \tag{4.6}$$

Proof. For $R, S \in [1, 1/d]$, the reciprocal kernel is

$$\frac{1}{A_\lambda(1/R, 1/S)} = \frac{RS}{R + S - 1 - dRS}.$$

Its second derivative in R equals

$$-\frac{2S(S-1)(1-dS)}{(R+S-1-dRS)^3} \leq 0,$$

and the same holds in S. All denominators are positive on $[1, 1/d]^2$, since $R + S - 1 - dRS = RSA_\lambda(1/R, 1/S)$. Apply Jensen's inequality first in $R = 1/X$, conditional on $S = 1/Y$, and then in S:

$$K \leq \mathbb{E}\frac{\rho S}{\rho + S - 1 - d\rho S} \leq \frac{\rho^2}{2\rho - 1 - d\rho^2}.$$

Independence makes the conditional mean of R equal to ρ. Since $\rho \in [1, 1/d]$, the same denominator identity gives $B_\rho = \rho^2 A_\lambda(1/\rho, 1/\rho) > 0$. ∎

4.3. A negative upper bound for $D^+v(\lambda)$.

Proof of Theorem 1.2. For $\alpha > \lambda$ close to λ, Lemma 4.1 and stochastic monotonicity of $\mathcal{V}_\alpha$ on $[1-\alpha, 1]$ give

$$v(\alpha) \le \mathcal{V}_\alpha\big(\mathrm{Law}(T_{\lambda,\alpha}(\beta_\lambda))\big). \tag{4.7}$$

Equality holds at $\alpha = \lambda$. We will show that the right-hand side has a strictly negative derivative there.

At $\alpha = \lambda$, the derivative of $T_{\lambda,\alpha}(X)$ with respect to α is $-a_\lambda(1-X)$ by Lemma 4.1, where $a_\lambda = 1/[\lambda(1+\lambda)]$. By the derivative formulas following (4.4), the derivative of the upper bound in (4.7) is $-D_a/Z$, where

$$D_a = 2a_\lambda \mathbb{E}[(1-X)\Psi(X)] - C. \tag{4.8}$$

We prove $D_a > 0$ using the three preceding lemmas. Write $\Delta = (1+\lambda)v - d \ge 0$, where nonnegativity follows from (4.3). The exact identity is

$$\lambda(1+\lambda)D_a = (1+v) - \Delta K + d\lambda C. \tag{4.9}$$

To verify (4.9), expand

$$\begin{aligned}(1-x)P_v(y) + (1-y)P_v(x) - \lambda(1+\lambda)\{vt(x,y) - xy\}\\ = A_\lambda(x,y)\{x+y-(2-d+d^2)v\} - d\lambda(xy - dv).\end{aligned}$$

Divide by $A_\lambda(x,y)^2$ and average. Using $xy - dv = xy - vt(x,y) + vA_\lambda(x,y)$ and $(2-d+d^2) + d\lambda = 2$, we obtain

$$\begin{aligned}\lambda(1+\lambda)D_a &= \mathbb{E}\frac{X+Y-2v}{A_\lambda(X,Y)} + d\lambda C\\ &= (1+v)(1+dK) - 2vK + d\lambda C\\ &= (1+v) - \Delta K + d\lambda C.\end{aligned}$$

The second equality uses $x + y = t(x,y) + xy$, $N = vZ$ and $Z = 1 + dK$. This proves (4.9).

Lemmas 4.3 and 4.4 imply

$$\lambda(1+\lambda)D_a \ge 1 + v - \Delta\frac{\rho^2}{B_\rho}.$$

The right-hand side decreases as a function of v, since

$$B_\rho - (1+\lambda)\rho^2 = -(\rho-1)^2 - \rho^2 < 0.$$

Using $v \le 1/\rho$ from Lemma 4.2 gives

$$D_a \ge \frac{1 - 1/\rho}{\lambda(1+\lambda)B_\rho} > 0. \tag{4.10}$$

It remains to justify differentiating the right-hand side of (4.7). On a sufficiently small right interval $[\lambda, \lambda+\delta] \subset (0,1)$, the variables $T_{\lambda,\alpha}(\beta_\lambda)$ lie in $[1-\alpha, 1]$, so (2.5) bounds the denominators away from zero. Both $T_{\lambda,\alpha}(\beta_\lambda)$ and its derivative in α are uniformly bounded there. Differentiation under the expectation is therefore legitimate, and the derivative of the upper bound at $\alpha = \lambda$ is $-D_a/Z$, by (4.4) and (4.8). Subtracting the common value at $\alpha = \lambda$ in (4.7), dividing by $\alpha - \lambda$, and taking the upper limit gives $D^+v(\lambda) \le -D_a/Z < 0$. The local Lipschitz bound in Lemma 4.1 is uniform over trees. Together with the same denominator bound and the speed formula, it gives continuity of v on $(0,1)$. The Dini criterion in Section 3.1 completes the proof. ∎

5. Strict decrease for $(3+\sqrt{5})/2 \le \lambda < m$

When $\mathbb{P}(\nu = 1) > 0$, the change of measure from Section 3 gives $\mathbb{E}_Q S \ge k$ in place of $\mathbb{E}_Q S \ge 2k$. We show that this still implies $\mathbb{E}X < 2k$ for $\lambda \ge (3+\sqrt{5})/2$, so Proposition 3.2 applies.

LEMMA 5.1. *Let $\lambda \ge (3+\sqrt{5})/2$, $c = \lambda - 1$, and let $X \in (0,1)$ satisfy (2.1) with $\nu \ge 1$. Then the weighted mean (2.12) satisfies $\mathbb{E}X < 2k$.*

Proof. Use the probability measure M obtained by weighting the law of X with $(c+X)^{-2}$, and let $k = \mathbb{E}_M X$, $j = \mathbb{E}_M X^2$, $\ell = \mathbb{E}_M X^3$. As in Lemma 3.1, weight the joint law of $\nu, X_1, X_2, \dots$ by $(c+S)^{-2}$ to obtain Q. These normalizing constants are finite and positive because $c > 0$ and $X, S > 0$.

The comparison of marginal densities in Lemma 3.1 still applies when $\nu = 1$: the remainder $H = \sum_{i=2}^{\nu} X_i$, which is independent of X_1 under the original law, is then zero, and $((c+x)/(c+x+H))^2$ is the constant one. For all $H \ge 0$ this ratio is nondecreasing in x. Thus $\mathbb{E}_Q X_1 \ge k$. Conditional exchangeability gives $\mathbb{E}_Q S = \mathbb{E}_Q[\nu X_1] \ge \mathbb{E}_Q X_1 \ge k$. The recursion transforms this expectation into

$$\mathbb{E}_Q S = \frac{\mathbb{E}[\lambda X(1-X)/(c+X)^2]}{\mathbb{E}[(1-X)^2/(c+X)^2]} = \lambda \frac{k-j}{1-2k+j}.$$

Here $1 - 2k + j = \mathbb{E}_M(1-X)^2 > 0$. Rearranging $\mathbb{E}_Q S \ge k$, with $\lambda = c+1$, gives

$$\lambda \frac{k-j}{1-2k+j} \ge k, \qquad j \le \frac{k(c+2k)}{c+1+k}. \tag{5.1}$$

Undoing the change of measure gives exactly (3.2), with positive denominator $c^2 + 2ck + j = \mathbb{E}_M(c+X)^2$. If $k \ge 1/2$, then $\mathbb{E}X < 1 \le 2k$. Suppose $k < 1/2$. Since $0 < X < 1$, we have $\ell \le j$, and therefore

$$(2k - \mathbb{E}X)(c^2 + 2ck + j) = c^2 k + 4ck^2 + (2k - 2c)j - \ell \ge c^2 k + 4ck^2 + (2k - 2c - 1)j.$$

The coefficient $2k - 2c - 1$ is negative. Substituting the upper bound for j from (5.1) consequently gives the lower bound

$$\begin{aligned}
&(2k - \mathbb{E}X)(c^2 + 2ck + j) \\
&\quad \ge c^2 k + 4ck^2 + (2k - 2c - 1)\frac{k(c+2k)}{c+1+k} \\
&\quad = \frac{k}{c+1+k}\left[c(c^2 - c - 1) + (5c^2 + 2c - 2)k + 4(c+1)k^2\right] > 0.
\end{aligned} \tag{5.2}$$

Indeed $c \ge (1+\sqrt{5})/2$, so $c^2 - c - 1 \ge 0$ and $5c^2 + 2c - 2 > 0$. Since $k > 0$, strictness holds even at the threshold. ∎

Proof of Theorem 1.3. For each $(3+\sqrt{5})/2 \le \lambda < m$, Lemma 5.1 verifies the hypothesis of Proposition 3.2. Thus $D^+ v(\lambda) < 0$, including at the left endpoint. Continuity and the Dini criterion in Section 3.1 give strict decrease on $[(3+\sqrt{5})/2, m)$. ∎

For $1 < \lambda < (3+\sqrt{5})/2$, the constant term in the brackets in (5.2) is negative, so this estimate no longer proves $\mathbb{E}X < 2k$. The computer-assisted argument in Section 6 treats the remaining interval.

6. Computer-assisted proof for $1 < \lambda < m$ (sketch)

We outline the proof for laws with single-child vertices. All bias intervals in this section are restricted to $\lambda < m$. Section 6 of the companion GitHub version [GM26a] gives the complete proof. The coefficient tables and verification programs are in [GM26b].

THEOREM 6.1 (Computer-assisted monotonicity). *Fix an offspring law μ with $\mu(0) = 0$, $\mu(1) > 0$ and $1 < m = \mathbb{E}_\mu \nu < \infty$. With the root and transition conventions of Section 1,*

$$v_\mu \in C^1((1, m)), \qquad v'_\mu(\lambda) < 0 \quad (1 < \lambda < m). \tag{6.1}$$

6.1. THE SPEED DERIVATIVE. Fix $1 < \lambda < m$, put $c = \lambda - 1$, and let X, Y be independent with law π_λ. We retain the notation of Sections 2.3–2.4, with $v = v(\lambda)$ and $Z = Z_\lambda(\pi_\lambda)$. Lemma 2.7 proves $v \in C^1((1, m))$ and gives

$$v'(\lambda) = -\frac{D_\lambda}{Z}, \qquad D_\lambda = 2\mathbb{E}[b(X)(\mathcal{R}\Psi)(X)] - C, \qquad Z > 0, \tag{6.2}$$

where $\Psi = \mathcal{J}(y^2 + ch_v)$ and $\mathcal{R} = (I - \mathcal{L})^{-1}$ is positive. It remains to prove $D_\lambda > 0$ for laws allowing single-child vertices.

6.2. BOUNDS FROM THE ESCAPE RECURSION. To estimate D_λ, we use the restrictions that the recursion places on moments of the escape probability. Put

$$U = \frac{X}{c + X}, \quad u = \mathbb{E}U, \quad h = \frac{cu}{1 - u}, \quad d = 1 - h,$$
$$a_n = \mathbb{E}U^n, \quad \beta_n = \frac{h}{\lambda^n - d}, \quad \rho_n = \frac{a_n}{\beta_n}.$$

Here $0 < U < 1/\lambda$ and $0 < h < 1$. The numbers β_n form a comparison sequence. Applying the recursion and a covariance inequality to one child and the sum of its siblings gives

$$\rho_0 = \rho_1 = 1, \qquad \rho_n > \rho_{n+1} > 0 \ (n \geq 1), \qquad \rho_n \longrightarrow 0. \tag{6.3}$$

Expanding the pair denominator gives absolutely convergent sums

$$T = \sum_{n \geq 1} \lambda^{n-1} a_n^2, \qquad Q = \sum_{n \geq 1} \lambda^{n-1} a_n a_{n-1}, \qquad v = \frac{cT}{2Q - (\lambda + 1)T}.$$

Comparison with the sequence β_n yields $v < h$ and $\inf_{[0,1]} \Psi > 0$; see [GM26a, Lemmas 6.2–6.3].

We next bound the positive measure $\eta(f) = \mathbb{E}[b(X)(\mathcal{R}f)(X)]$ without truncating the series for $\mathcal{R}$. The required comparison follows from an identity for the function

$$g(x) = \frac{x(c + x)\log(1 + c/x)}{\lambda c \log \lambda} - \frac{x}{c} \quad (x > 0), \qquad g(0) = 0.$$

The function g is strictly concave on $[0, \infty)$ and nonnegative on $[0, 1]$. For independent child escapes and $S = \sum_{i=1}^{\nu} X_i$, define

$$\delta_g = F'_\lambda(S) \left(\sum_{i=1}^{\nu} g(X_i) - g(S) \right).$$

Concavity and $g(0) = 0$ give $\delta_g \geq 0$, with strict inequality on $\{\nu \geq 2\}$. Moreover, δ_g is bounded by a constant times ν. The identity satisfied by g under the recursion yields, for $f \in C([0, 1])$,

$$\eta(f) = \mathbb{E}[g(X)f(X)] + \mathbb{E}[\delta_g(\mathcal{R}f)(F_\lambda(S))]. \tag{6.4}$$

For $f \geq 0$ this implies $\eta \geq g\pi_\lambda$; the Neumann series also gives $\eta \geq b\pi_\lambda$, hence $\eta \geq \max(g, b)\pi_\lambda$.

To use this measure bound, subtract a multiple of $\mathcal{J}(y^2)$ from Ψ so that the remainder is nonnegative:

$$\alpha = \min\left\{1, \frac{\Psi(0)}{\mathcal{J}(y^2)(0)}\right\}, \qquad H_* = \Psi - \alpha\mathcal{J}(y^2) \geq 0, \qquad F_* = 2\alpha\mathbb{E}[br] - C + 2\mathbb{E}[gH_*].$$

Since $(I - \mathcal{L})r = \mathcal{J}(y^2)$, we have $\mathcal{R}\Psi = \alpha r + \mathcal{R}H_*$. Substitution in (6.4) gives

$$D_\lambda = F_* + 2\mathbb{E}[\delta_g(\mathcal{R}H_*)(F_\lambda(S))] \geq F_*. \tag{6.5}$$

The case $\alpha = 1$ already gives $D_\lambda > 0$. Otherwise $0 < \alpha < 1$, $\alpha > \lambda^2 u$, $u < \lambda^{-2}$ and $h < 1/(\lambda + 1)$. These inequalities specify the parameter range left to consider. Their proofs are in [GM26a, Lemmas 6.4–6.5].

6.3. Moment estimates for $1 < \lambda \le 707/500$. We need a bound on the differences between successive moment ratios in addition to (6.3). Extending a_s, β_s to real $s \ge 1$ and putting $\rho(s) = a_s/\beta_s$, the escape recursion gives a probability measure σ on $(0,1)$ such that

$$\rho(s) = \int_0^1 t^{s-1}\{s-(s-1)t\}\,\sigma(\mathrm{d}t),$$
$$\Delta_s := \rho(s) - \rho(s+1) = s\int_0^1 t^{s-1}(1-t)^2\,\sigma(\mathrm{d}t),$$
$$-\rho'(s) > \left(1 - \frac{1}{2s}\right)\Delta_s.$$

This representation, proved in [GM26a, Lemma 6.6], uses the same measure for every order, so the differences Δ_s are constrained together. Expanding the kernels containing g produces the terms $\mathbb{E}[U^s \log(1/U)]$. The bound on $-\rho'(s)$ therefore gives a strict lower bound for their contribution in terms of a_s and Δ_s.

When using an upper bound on v, we keep the law π_λ fixed. This distinction matters because the identity

$$C = \mathbb{E}\left[b(X)\mathcal{J}\big((1-y)\{2v-(1+v)y\}\big)(X)\right]$$

uses $N - vZ = 0$, which need not hold after replacing v by an upper bound. Lemmas 6.7–6.8 of [GM26a] keep the resulting remainder and justify the substitution. They give an absolutely convergent series in the moments and their differences that is strictly smaller than F_*. It therefore suffices to prove that this series is nonnegative.

For $1 < \lambda < 2/\sqrt{3}$, the series estimate uses positive square terms to control the negative terms containing the differences Δ_s. The comparison sequence β_s bounds the entire tail. The proof splits the parameters $q = h/c$ and h into six ranges:

$$q \le 1/15; \quad 1/15 < q \le 3; \quad 3 < q \le 10; \quad q \ge 21;$$
$$10 < q < 21,\ h \ge 1/4; \qquad 10 < q < 21,\ h < 1/4.$$

In the last range, an analytic estimate treats all terms with index at least 25. The remaining indices require 48 polynomials, one for each index $1, \dots, 24$ at two parameter endpoints. The other ranges use rational tail bounds, finite coefficient tests and, where needed, exact Sturm sequences. The estimates give a strictly positive lower bound in each range, and hence $D_\lambda > 0$; see [GM26a, Lemmas 6.9–6.14 and Proposition 6.15].

For $2/\sqrt{3} \le \lambda \le 707/500$, a bound on the comparison series gives an upper bound on v that retains the first two moments; see [GM26a, Lemma 6.16]. To use it in the series estimate, the proof also needs the recursion inequality

$$\frac{a_3}{a_2} < \frac{1}{\lambda + 1 - \lambda a_2/u};$$

see [GM26a, Lemma 6.17]. The proof estimates the whole series before discarding a nonnegative tail. For $u \le 1/2$, the retained terms have indices $1, \dots, 4$ and are estimated together using the moment constraints. For $u > 1/2$, the first two terms suffice. In this case $\alpha < 1$ is impossible for $\lambda \ge 11/8$, while for $\lambda < 11/8$ a curvature argument reduces the bound to two positive endpoint values. Exact polynomial checks on rational rectangles give $D_\lambda > 0$ at both endpoints and throughout the interval [GM26a, Section 6.5, especially Lemmas 6.20, 6.27 and 6.30].

6.4. Kernel and integral bounds for $707/500 \le \lambda < (3+\sqrt{5})/2$. We use the formula for D_λ in (6.2) and the measure bound from (6.4). Define

$$P_c(y) = (1-y)\{2v - (1+v)y\}.$$

The identity $C = \mathbb{E}[b\mathcal{J}P_c]$, positivity of Ψ, and $\eta \ge \max(g,b)\pi_\lambda$ imply

$$D_\lambda \ge \mathbb{E}\left[2\max\{g(X),b(X)\}\Psi(X) - b(X)(\mathcal{J}P_c)(X)\right]. \tag{6.6}$$

Here $g(x) \ge b(x)$ for $x \le 1/(\lambda+1)$ and $b(x) \ge g(x)$ for $x \ge 1/(\lambda+1)$. On $707/500 \le \lambda \le 39/20$, the full proof establishes

$$\begin{aligned} 2\frac{g(x)}{b(x)}\Psi(x) - (\mathcal{J}P_c)(x) > 0, &\qquad 0 < x \le 1/(\lambda+1),\\ 2\Psi(x) - (\mathcal{J}P_c)(x) > 0, &\qquad 1/(\lambda+1) \le x < 1. \end{aligned}$$

Multiplying by $b(x) > 0$ and using (6.6) gives $D_\lambda > 0$. The products inside the expectation in (6.6) are bounded. The kernel estimates use the first four moments and a geometric bound on all remaining terms. Concavity reduces one moment variable to its two endpoints; two integrations of rational functions then extend the endpoint bounds to the required parameter range. Exact coefficient signs establish the required boundary values and derivative signs. These steps are developed in [GM26a, Section 6.6, Lemmas 6.31–6.35].

For $39/20 \le \lambda < (3+\sqrt{5})/2$, a different representation of D_λ keeps the dependence on the root offspring number ν. With $S = \sum_{i=1}^{\nu} X_i$, put

$$I = \mathbb{E}\frac{2S^2 + 2\{c - (\lambda+1)v\}S + (1+v)\sum_{i=1}^{\nu} X_i^2 + 2v(\nu - \lambda^2)}{(c+S)^2}. \tag{6.7}$$

The moment bounds give $c - (\lambda+1)v \ge 0$. A Laplace representation of the reciprocal square expresses D_λ using two factors: $\eta(e^{-t\cdot})$ and the expectation of the numerator in (6.7) under the weight $e^{-tS}/\mathbb{E}e^{-tS}$. Both factors decrease with $t > 0$. The covariance inequality, applied with respect to the finite measure $te^{-ct}\mathbb{E}[e^{-tS}]\,\mathrm{d}t$, gives

$$D_\lambda \ge \frac{I}{\lambda}\frac{\displaystyle\int_0^\infty te^{-ct}\mathbb{E}[e^{-tS}]\eta(e^{-t\cdot})\,\mathrm{d}t}{\displaystyle\int_0^\infty te^{-ct}\mathbb{E}[e^{-tS}]\,\mathrm{d}t}.$$

The quotient is finite and strictly positive. It therefore remains to prove $I > 0$.

The recursion and two comparisons of affine functions at their endpoints reduce $I > 0$ to the 63 coefficient inequalities in [GM26a, Lemma 6.36]. The tail estimates and the reduction to these inequalities are proved analytically. The denominator in (6.7) is positive because $c > 0$; moreover $S^2/(c+S)^2 \le 1$, $\sum_i X_i^2 \le S$, and $\mathbb{E}\nu < \infty$ justify the signed integral under the stated first-moment assumption. This argument includes $\lambda = 39/20$.

For $(3+\sqrt{5})/2 \le \lambda < m$, Theorem 1.3 gives $D^+v(\lambda) < 0$. Differentiability from (6.2) identifies this with $v'(\lambda)$, so no computer-assisted estimate is needed there.

6.5. Exact arithmetic and completion of the proof. Each remaining inequality concerns an explicit rational function on a specified domain. The full version proves positivity of the denominators before multiplying by them. An affine change of the bounded variables to unit intervals then gives a polynomial with rational coefficients. For example, in two variables write

$$P(x,y)=\sum_{i=0}^{M}\sum_{j=0}^{N} b_{ij}\binom{M}{i}x^i(1-x)^{M-i}\binom{N}{j}y^j(1-y)^{N-j}.$$

If all $b_{ij}>0$, then $P>0$ on the closed unit square. If some coefficients vanish, the proof checks that a positive basis term remains at every point arising from the escape law where strictness is required. For an unbounded nonnegative variable, the proof uses nonnegative coefficients in the ordinary power basis. The exceptional univariate polynomials are handled by exact Sturm sequences with positive endpoint values and no interior zero. Rational subdivisions are recorded explicitly. The algebraic endpoint tests at $2/\sqrt{3}$ reduce to exact rational comparisons by checking signs and comparing squares. None of these arguments infers a sign from a numerical mesh.

The table lists the finite inequalities used in each bias interval.

Bias interval	Proof and finite inequalities in the full version
$(1, 2/\sqrt{3})$	Section 6.4; Lemmas 6.11–6.14; Proposition 6.15.
$[2/\sqrt{3}, 707/500]$	Section 6.5; Lemmas 6.18, 6.25, 6.27 and 6.30.
$[707/500, 39/20]$	Section 6.6; Lemmas 6.31–6.35.
$[39/20, (3+\sqrt{5})/2)$	Section 6.6; Lemma 6.36.

The references in this table are to the November 20, 2026 GitHub version [GM26a]. Its supplement supplies the input polynomials, exact coefficient arrays, complete exceptional Sturm sequences and programs that reconstruct the expressions. The supplementary guide maps each finite lemma to its data and executable check. Its verification uses integer and rational arithmetic. The full manuscript also proves that these domains contain the parameters of the escape law and that the infinite tails satisfy the stated bounds.

To finish Theorem 6.1, fix $1<\lambda<m$. Equation (6.2) reduces the conclusion to $D_\lambda>0$. The case $\alpha=1$ following (6.5) gives this sign directly. If $\alpha<1$, the four intervals in the table together with $(3+\sqrt{5})/2\le\lambda<m$ in Theorem 1.3 cover $(1,m)$, including their shared endpoints. Hence $v'(\lambda)<0$ everywhere on $(1,m)$. The mean value theorem on each compact subinterval gives strict decrease on $(1,m)$. Lemma 2.2 gives continuity at $\lambda=1$. Together with Theorem 1.2, these conclusions yield strict decrease throughout $(0,m)$.

References

[Aïd14] Élie Aïdékon. Speed of the biased random walk on a Galton–Watson tree. *Probability Theory and Related Fields*, 159:597–617, 2014. Theorem numbering follows arXiv:1111.4313v4. 2, 3, 4, 5

[BT20] Adam Bowditch and Yuki Tokushige. Differentiability of the speed of biased random walks on Galton–Watson trees. *ALEA. Latin American Journal of Probability and Mathematical Statistics*, 17:609–642, 2020. arXiv:1906.07913v2. 2, 13

[GM26a] GPT-6 Astra and Heng Ma. Monotonicity of the speed of biased random walks on Galton–Watson trees: full (GitHub) version. `https://github.com/hengmamath/monotone-speed-biased-RW/blob/main/paper/github.pdf`, 2026. Companion full manuscript, November 20, 2026, Section 6. Complete proofs and accompanying exact-computation supplement. 1, 19, 20, 21, 22, 23

[GM26b] GPT-6 Astra and Heng Ma. Supplementary exact computations and source records for speed monotonicity. `https://github.com/hengmamath/monotone-speed-biased-RW/tree/main/paper/supplement`, 2026. Exact coefficient data, executable verification programs and source records accompanying this article. See the supplementary README for the correspondence with the finite arithmetic lemmas. 19

[LPP97] Russell Lyons, Robin Pemantle, and Yuval Peres. Unsolved problems concerning random walks on trees. In *Classical and Modern Branching Processes*, volume 84 of *IMA Volumes in Mathematics and its Applications*, pages 223–238. Springer, New York, 1997. Corrected author version, Question 2.1. 2

[Sch26] Felix Schwarzfischer. A proof of the Lyons–Pemantle–Peres monotonicity conjecture for high offspring laws. arXiv:2611.25882v1, 2026. Submitted November 19, 2026; Theorem 1.1 and the discussion following it. 3

[SWX16] He Song, Longmin Wang, and Kainan Xiang. Monotonicity of speed for biased random walk on Galton–Watson tree. arXiv:1610.08151v1, 2016. Theorem 1.2, Lemma 2.1, equation (2.1), and Lemma 2.5. 2, 4, 5, 13