Mathematical Biology

Haldane's formula in Cannings models: the case of moderately strong selection

Florin Boenkost[1] · **Adrián González Casanova[2]** · **Cornelia Pokalyuk[1]** ·
Anton Wakolbinger[1]

Abstract

For a class of Cannings models we prove Haldane's formula, $\pi(s_N) \sim \frac{2s_N}{\rho^2}$, for the fixation probability of a single beneficial mutant in the limit of large population size N and in the regime of moderately strong selection, i.e. for $s_N \sim N^{-b}$ and $0 < b < 1/2$. Here, s_N is the selective advantage of an individual carrying the beneficial type, and ρ^2 is the (asymptotic) offspring variance. Our assumptions on the reproduction mechanism allow for a coupling of the beneficial allele's frequency process with slightly supercritical Galton–Watson processes in the early phase of fixation.

Keywords Branching process approximation · Cannings model · Directional selection · Probability of fixation

Mathematics Subject Classification Primary 60J10; Secondary 60J80 · 92D15 · 92D25

✉ Florin Boenkost
boenkost@math.uni-frankfurt.de

Adrián González Casanova
adriangcs@matem.unam.mx

Cornelia Pokalyuk
pokalyuk@math.uni-frankfurt.de

Anton Wakolbinger
wakolbinger@math.uni-frankfurt.de

1 Goethe-Universität Frankfurt, FB 12, 60054 Frankfurt, Germany

2 Instituto de Matemáticas, Universidad Nacional Autónoma de México, Área de la Investigación
Científica, Circuito Exterior, Ciudad Universitaria, 04510 Coyoacán, CDMX, Mexico

1 Introduction

Analysing the probability of fixation of a beneficial allele that arises from a single mutant is one of the classical problems in population genetics, see Patwa and Wahl (2008) for a historical overview. A rule of thumb known as Haldane's formula states that the probability of fixation of a single mutant of beneficial type with small selective advantage $s > 0$ and offspring variance ρ^2 in a large population of individuals, whose total number N is constant over the generations, is approximately equal to $2s/\rho^2$. Originally, this was formulated for the (prototypical) model of Wright and Fisher, in which the next generation arises by a multinomial sampling from the previous one (which leads to $\rho^2 = 1 - \frac{1}{N}$ in the neutral case), with the "reproductive weight" of an individual of beneficial type being increased by the (small) factor $1 + s$. A natural generalization of the Wright-Fisher model are the Cannings models; here one assumes *exchangeable* offspring numbers in the neutral case (Cannings 1974; Ewens 2004), and separately within the sets of all individuals of the beneficial and the non-beneficial type in the selective case (Lessard and Ladret 2007).

The reasoning in the pioneering papers by Fisher (1922), Haldane (1927) and Wright (1931) was based on the insight that, as long as the beneficial type is rare, the number of individuals carrying the beneficial type is a slightly supercritical branching process for which the survival probability is

$$\pi(s) \sim \frac{2s}{\rho^2} \qquad \text{as } s \to 0,$$

where $1 + s$ is the offspring expectation and ρ^2 is the offspring variance (see Athreya 1992, Theorem 3). The heuristics then is that the branching process approximation should be valid until the beneficial allele has either died out or has reached a fraction of the population that is substantial enough so that the law of large numbers dictates that this fraction should rise to 1.

Notably, Lessard and Ladret (2007) obtained (for fixed population size N) the result

$$\pi(s) = \frac{1}{N} + \frac{2s}{\rho^2} + o(s) \qquad \text{as } s \to 0, \tag{1}$$

as a special case of their explicit analytic representation of $\pi(s)$ within a quite general class of Cannings models and selection mechanisms.

An interesting parameter regime as $N \to \infty$ is that of *moderate selection*,

$$s_N \sim cN^{-b} \quad \text{with} \quad 0 < b < 1, \quad c > 0, \tag{2}$$

which is between the classical regimes of *weak* and *strong* selection. Is the *Haldane asymptotics*

$$\pi(s_N) \sim \frac{2s_N}{\rho^2} \qquad \text{as } N \to \infty, \tag{3}$$

valid in the regime (2)?

If one could bound in this regime the $o(s)$-term in (1) by $o(N^{-b})$, then (1) would turn into (3). Such an estimate seems, however, hard to achieve in the analytic framework of Lessard and Ladret (2007).

The main result of the present paper is a proof of the Haldane asymptotics using an approximation by Galton–Watson processes in the regime of *moderately strong selection*, which corresponds to (2) for $0 < b < \frac{1}{2}$. Hereby, we assume that the Cannings dynamics admits a *paintbox representation*, whose random weights are exchangeable and *of Dirichlet-type*, and fulfil a certain moment condition, see Sect. 3. Here, the effect of selection is achieved by a decrease of the reproductive weights of the non-beneficial individuals by the factor $1 - s_N$.

An approximation by Galton–Watson processes was used in González Casanova et al. (2017) to prove the asymptotics (3) in the regime of moderately strong selection for a specific Cannings model that arises in the context of experimental evolution, with the next generation being formed by sampling without replacement from a pool of offspring generated by the parents.

In the case $b \geq \frac{1}{2}$ the method developed in the present paper would fail, because then the Galton–Watson approximation would be controllable only up to a time at which the fluctuations of the beneficial allele (that are caused by the resampling) still dominate the trend that is induced by the selective advantage. However, in Boenkost et al. (2021) we proved the Haldane asymptotics (3) for the case of *moderately weak selection*, i.e. under Assumption (2) with $\frac{1}{2} < b < 1$. There a backward point of view turned out to be helpful, which uses a representation of the fixation probability in terms of sampling duality via the *Cannings ancestral selection graph* developed in Boenkost et al. (2021) (see also González Casanova and Spanó 2018).

The results of the present paper together with those of Boenkost et al. (2021) do not cover the boundary case $b = \frac{1}{2}$ between moderately strong and moderately weak selection. We conjecture that the Haldane asymptotics (3) is valid also in this case.

2 A class of Cannings models with selection

This section is a short recap of Boenkost et al. (2021) Section 2; we include it here for self-containedness.

2.1 Paintbox representation in the neutral case

Neutral Cannings models are characterized by the exchangeable distribution of the vector $\nu = (\nu_1, \ldots, \nu_N)$ of offspring sizes; here the ν_i are non-negative integer-valued random variables which sum to N. An important subclass are the *mixed multinomial* Cannings models. Their offspring size vector ν arises in a two-step manner: first, a vector of random weights $\mathcal{W} = (W_1, \ldots, W_N)$ is sampled, which is exchangeable and satisfies $W_1 + \cdots + W_N = 1$ and $W_i \geq 0$, $1 \leq i \leq N$.

In the second step, a die with N possible outcomes $1, \ldots, N$ and outcome probabilities $\mathcal{W} = (W_1, \ldots, W_N)$ is thrown N times, and ν_i counts how often the outcome i occurs. Hence, given the random weights $\mathcal{W}$ the offspring numbers $\nu = (\nu_1, \ldots, \nu_N)$

are Multinomial($N, \mathscr{W}$)-distributed. Following Kingman's terminology, we speak of a *paintbox representation* for v, and call $\mathscr{W}$ the underlying (random) *paintbox*.

This construction is iterated over the generations $g \in \mathbb{Z}$: Let $\mathscr{W}^{(g)} = (W_1^{(g)}, \ldots, W_N^{(g)})$ be independent copies of $\mathscr{W}$, and denote the individuals in generation g by (i, g), $i \in [N]$. Assume that each individual $(j, g+1)$, $j \in [N] := \{1, \ldots, N\}$ in generation $g+1$, chooses its parent in generation g, with conditional distribution

$$\mathbb{P}((i, g) \text{ is the parent of } (j, g+1)|\mathscr{W}^{(g)}) = W_i^{(g)}, \qquad \forall i \in [N].$$

where given $\mathscr{W}^{(g)}$ the choices of the parents for individuals $\{(j, g+1), j \in [N]\}$ are independent and identically distributed. This results in exchangeable offspring vectors $v^{(g)}$ which are independent and identically distributed over the generations g.

For notational simplicity we do not always display dependence of $\mathscr{W}^{(g)}$ on the generation g, and write $\mathscr{W}$ instead. From time to time however we want to emphasise the dependence of $\mathscr{W}$ on N and therefore write $\mathscr{W}^{(N)}$ instead of $\mathscr{W}$.

Some exchangeable offspring vectors do not have a paintbox representation, for example a random permutation of the vector $(2, \ldots, 2, 0, \ldots, 0)$. Prototypical paintboxes are $\mathscr{W} = (\frac{1}{N}, \ldots, \frac{1}{N})$, which leads to the Wright-Fisher model, and the class of Dirichlet$(\alpha, \ldots, \alpha)$-distributed random weights. In particular, the offspring distribution with Dirichlet$(1, \ldots, 1)$-distributed paintbox can be seen as a limiting case of the offspring distribution for the model of experimental evolution considered in Baake et al. (2019) and González Casanova et al. (2017).

2.2 A paintbox representation with selection

Let $\mathscr{W}^{(g)}$, $g \in \mathbb{Z}$, be as in the previous section, and let $s_N \in [0, 1)$. Assume each individual carries one of two types, either the *beneficial type* or the *wildtype*. Depending on the type of individual (i, g) we set

$$\widetilde{W}_i^{(g)} = (1 - s_N)W_i^{(g)}$$

if (i, g) is of wildtype and $\widetilde{W}_i^{(g)} = W_i^{(g)}$ if (i, g) is of beneficial type. The probability that an individual is chosen as parent is now given by

$$\mathbb{P}((i, g) \text{ is parent of } (j, g+1)) = \frac{\widetilde{W}_i^{(g)}}{\sum_{\ell=1}^{N} \widetilde{W}_\ell^{(g)}} \tag{4}$$

for all $i, j \in [N]$. Parents are chosen independently for all $i \in [N]$ and the distribution does not change over the generations. If (i, g) is the parent of $(j, g+1)$ the child $(j, g+1)$ inherits the type of its parent. In particular, this reproduction mechanism leads to offspring numbers that are exchangeable among the beneficial as well as among wildtype individuals.

2.3 The Cannings frequency process

In the previous section we gave a definition for a Cannings model which incorporates selection, by decreasing the random weight of each wildtype individual by the factor $1 - s_N$. This allows to define the Cannings frequency process $\mathcal{X} = (X_g)_{g \geq 0}$ with state space $[N]$ which counts the number of beneficial individuals in each generation g.

Assume there are $1 \leq k \leq N$ beneficial individuals at time g; due to the exchangeability of $\mathcal{W}^{(g)}$ we may assume that the individuals $(1, g), \ldots, (k, g)$ are the beneficial ones. Given $\mathcal{W}^{(g)} = \mathcal{W}$, the probability that individual $(j, g + 1)$ is of beneficial type is then due to (4) equal to

$$\frac{\sum_{i=1}^{k} W_i}{\sum_{i=1}^{k} W_i + (1 - s_N) \sum_{i=k+1}^{N} W_i}, \tag{5}$$

and is the same for all $j \in [N]$. Hence, given $\mathcal{W}^{(g)} = \mathcal{W}$ and given there are k beneficial individuals in generation g, the number of beneficial individuals in generation $g + 1$ has distribution

$$\mathrm{Bin}\left(N, \frac{\sum_{i=1}^{k} W_i}{\sum_{i=1}^{k} W_i + (1 - s_N) \sum_{i=k+1}^{N} W_j}\right); \tag{6}$$

this defines the transition probabilities of the Markov chain $\mathcal{X}$.

3 Main result

Before we state our main result we specify the assumptions on the paintbox and the strength of selection.

Definition 1 (*Dirichlet-type weights*)   We say that a random vector $\mathcal{W}^{(N)}$ with exchangeable components $W_1^{(N)}, \ldots, W_N^{(N)}$ is *of Dirichlet-type* if

$$W_i^{(N)} = \frac{Y_i}{\sum_{\ell=1}^{N} Y_\ell}, \quad i = 1, \ldots, N, \tag{7}$$

where $Y_1, \ldots, Y_N$ are independent copies of a random variable Y with $\mathbb{P}(Y > 0) = 1$.

We assume that

$$\mathbb{E}\left[\exp(hY)\right] < \infty, \tag{8}$$

for some $h > 0$, which implies the finiteness of all moments of Y. The relevance (and possible relaxations) of Condition (8) are discussed further in Remark 2(a), see also the comment in Remark 1(a).

Remark 1 (a) The biological motivation for considering Dirichlet-type weights comes from seasonal reproductive schemes. At the beginning of a season a set (of size N) of individuals is alive. These individuals and their offspring reproduce and generate a pool of descendants within that season. Only a few individuals from this pool survive till the next season. The number N in the model is assumed to be the total number of individuals that make it to the next season. Dirichlet-type weights arise in the asymptotics of an infinitely large pool of offspring; then sampling with and without replacement coincide. Condition (8), which we will require for the proof of Theorem 1 (see also Remark 2), guarantees that the pool of descendants of a single individual is not too large in comparison to the pool of descendants generated by the other individuals. The simplifying assumption $\mathbb{P}(Y > 0) = 1$ implies that the weight $W_i^{(N)}$ of a parent cannot be equal to zero. Observe, however, that weights of single parents can be arbitrarily small if (e.g.) Y has a density which is continuous and strictly positive in zero.

(b) The case of a deterministic Y corresponds to $W_i^{(N)} \equiv 1/N$, i.e. the classical Wright-Fisher model. If Y has a Gamma(κ)-distribution, then $\mathscr{W}^{(N)}$ is Dirichlet $(\kappa, \dots, \kappa)$-distributed.

Theorem 1 in Huillet and Möhle (2021), gives a classification of a large class of Cannings models with a paintbox of the form (7) with regard to the convergence of their rescaled genealogies.

(c) Let $\nu^{(N)}$ be a sequence of Cannings offspring numbers that are represented by the paintboxes $\mathscr{W}^{(N)}$. It is well known (and easily checked) that

$$\mathrm{Var}\left(\nu_1^{(N)}\right) = N(N-1)\mathbb{E}[(W_1^{(N)})^2]. \qquad (9)$$

If $\mathscr{W}^{(N)}$ is of the form (1) with $\mathbb{E}[Y^2] < \infty$, which is clearly implied by (8), then (see Huillet and Möhle (2021) Theorem 1 (i)) the right hand side of (9) converges to $\frac{\mathbb{E}[Y^2]}{\mathbb{E}[Y]^2}$ as $N \to \infty$.

In view of Remark 1(c) we have for the asymptotic neutral offspring variance

$$\lim_{N \to \infty} \mathrm{Var}\left(\nu_1^{(N)}\right) =: \rho^2 = \frac{\mathbb{E}[Y^2]}{\mathbb{E}[Y]^2}. \qquad (10)$$

Replacing Y by $Y' = \frac{Y}{\mathbb{E}[Y]}$ does not affect (7), hence we can and will assume $\mathbb{E}[Y] = 1$ in the proofs, which simplifies (10) to $\rho^2 = \mathbb{E}[Y^2]$ (and makes (10) consistent with the notation of Huillet and Möhle (2021)). Under the assumption $\mathbb{E}[Y^4] < \infty$, which is implied by (8) as well, the following asymptotics is valid

$$\mathrm{Var}\left(\nu_1^{(N)}\right) = \rho^2 + O(N^{-1}), \text{ as } N \to \infty. \qquad (11)$$

We will discuss the relevance of this asymptotics in Remark 2(b), and prove it in Lemma 2(b).

Turning to the selective advantage, we assume that for a fixed $\eta \in (0, \frac{1}{4})$ the sequence (s_N) obeys

$$N^{-\frac{1}{2}+\eta} \le s_N \le N^{-\eta}, \tag{12}$$

which we call the regime of *moderately strong selection*, thus generalizing the corresponding notion introduced in Sect. 1. (Note that (12) has an analogue in the regime of *moderately weak selection* as discussed in Boenkost et al. (2021)). In order to connect to (2) we define

$$b_N := -\frac{\ln s_N}{\ln N} \tag{13}$$

which is equivalent to $s_N = N^{-b_N}$, with (12) translating to

$$\eta \le b_N \le \frac{1}{2} - \eta.$$

We now state our main result on the asymptotics (as $N \to \infty$) of the fixation probability of the Cannings frequency process $(X_g^{(N)}) = (X_g)$ defined in Sect. 2.3. Note that the Markov chain (X_g) has the two absorbing states 0 and N, with the hitting time of $\{0, N\}$ being a.s. finite for all N.

Theorem 1 (Haldane's formula) *Assume that Conditions* (7), (8) *and* (12) *are fulfilled. Let* $(X_g)_{g \ge 0}$ *be the number of beneficial individuals in generation g, with* $X_0 = 1$. *Let* $\tau = \inf \{g \ge 0 : X_g \in \{0, N\}\}$, *then*

$$\mathbb{P}(X_\tau = N) \sim \frac{2s_N}{\rho^2}, \qquad as \ N \to \infty. \tag{14}$$

We give the proof of Theorem 1 in Sect. 5, after preparing some auxiliary results in Sect. 4. Next we give a strategy of the proof and its main ideas, with an emphasis on the role of Condition (12). In Remark 2 we discuss possible relaxations of Condition (8) and the boundary case $b = \frac{1}{2}$.

The proof of Theorem 1 is divided into three parts, corresponding to three growth phases of $\mathcal{X}$. Concerning the *first phase* we show that the probability to reach the level $N^{b+\delta}$ is $\frac{2s_N}{\rho^2}(1 + o(1))$, for some small $\delta > 0$ and $b := b_N$; this is the content of Proposition 1. The proof is based on stochastic domination from above and below by slightly supercritical Galton–Watson processes $\overline{\mathcal{Z}}$ and $\underline{\mathcal{Z}}$ with respective offspring distributions (29) and (30).

To construct a Galton–Watson stochastic upper bound $\overline{\mathcal{Z}}$ of $\mathcal{X}$ in its initial phase, we recall that the transition probabilities of $\mathcal{X}$ are mixed Binomial specified by (6). Using (7) we approximate (5) from above by

$$\frac{1 + s_N + o(s_N)}{N} \sum_{\ell=1}^{k} Y_\ell. \tag{15}$$

As we will show in Lemma 4, this is possible with probability $1 - O(\exp(-c'N^{1-2\alpha}))$ for some $\alpha < \frac{1}{2}$ and $c' > 0$, and for $k \leq N^{b+\delta}$ with $b + \delta < 1/2$. We will then be able to dominate the mixed Binomial distribution (6) by the mixed Poisson distribution with random parameter (15), again up to an error term of order $o(s_N)$. Noting that (15) is a sum of independent random variables, we arrive at the upper Galton–Watson approximation for a single generation. For any small $\varepsilon > 0$ this can be repeated for $N^{b+\varepsilon}$ generations, which (as an application of Lemma 1 will show) is enough to reach either the level 0 or the level $N^{b+\delta}$ with probability $1 - o(s_N)$.

To obtain a Galton–Watson stochastic lower bound $\underline{\mathcal{Z}}$ of $\mathcal{X}$ in its initial phase, we adapt an approach that was used in González Casanova et al. (2017) in a related situation. As in Sect. 2.1, number the individuals in generation g by (i, g), now with $(1, g), \ldots, (X_g, g)$, being the beneficial individuals, and denote by $\omega_i^{(g)}$ the number of children of the individual (i, g), $1 \leq i \leq X_g$. As will be explained in the proof of Lemma 5, as long as X_g has not reached the level $N^{b+\delta}$, the distribution of $\omega_i^{(g)}$ can be bounded from below by a mixed binomial distribution

$$\mathrm{Bin}\left(N - \lceil N^{b+\delta} \rceil, Y_1 \frac{1 + s_N + o(s_N)}{N}\right)$$

with probability $1 - O(\exp(-N^\varepsilon))$ for some sufficiently small $\varepsilon > 0$, again for $b + \delta < 1/2$. A suitable stopping and truncation at the level $N^{b+\delta}$ will give the Galton–Watson process approximation from below for the first phase.

We will verify in Sect. 4.1 that both slightly supercritical branching processes $\underline{\mathcal{Z}}$ and $\overline{\mathcal{Z}}$ reach the level $N^{b+\delta}$ with probability $\frac{2s_N}{\rho^2}(1 + o(1))$.

As to the *second phase*, we will argue in Sect. 5.2 that, after reaching the level $N^{b+\delta}$ the Cannings frequency process $\mathcal{X}$ will grow to a macroscopic fraction εN with high probability. If the frequency of beneficial individuals is at least $N^{b+\delta}$ (but still below εN), then in a single generation the frequency of beneficial individuals grows in expectation at least by $1 + (1 - \varepsilon)s_N + o(s_N)$. Hence, $cs_N^{-1} \ln N$ generations after $\mathcal{X}$ has reached the level $N^{b+\delta}$, the expected value of the process $\mathcal{X}$ reaches the level $2\varepsilon N$. Similarly one bounds the variance produced in a single generation and derives from this an estimate for the variance accumulated over $cs_N^{-1} \ln N$ generations. This bound being sufficiently small, an application of Chebyshev's inequality yields that (after $cs_N^{-1} \ln N$ generations) $\mathcal{X}$ crosses the level εN with probability tending to 1 after reaching the level $N^{b+\delta}$.

In Sect. 5.3) we deal with the *last phase*, and will show that the fixation probability tends to 1 as $N \to \infty$ if we start with at least εN individuals of beneficial type. Here we use the representation for the fixation probability that is based on a sampling duality between the Cannings frequency process and the Cannings ancestral selection process (CASP) which was provided in Boenkost et al. (2021). For a subregime of moderately weak selection the claim will follow quickly from the representation formula combined with a concentration result for the equilibrium distribution of the CASP that was proved in Boenkost et al. (2021). To complete the proof we will then argue that both the CASP and the representation of the fixation probability depend on the selection parameter in a monotone way.

Remark 2 (a) With some additional work the assumption (8) of the existence of some exponential moment of Y can be relaxed to some weaker moment condition. In order not to overload the present paper, we restrict here to a sketch.

In Lemma 5 we couple the frequency process of the beneficial individuals with Galton–Watson processes for $N^{b+\delta}$ generations. By means of the estimates in Lemma 3 and Lemma 4 we show that these couplings hold for a single generation with probability $1 - O(\exp(-N^{c'}))$ for some appropriate $c' > 0$. Since we need the couplings to hold for $N^{b+\delta}$ generations, it suffices that the couplings hold in a single generation with probability $1 - O(N^{-2(b+\delta)})$ for some $\delta > 0$ (since in this case the probability of the coupling to fail is $o(s_N)$ and therefore can be neglected with regard to (14)). Such probability bounds can also be obtained under weaker assumptions on the distribution of the random variable Y. Assume e.g. that Y has a regularly varying tail, i.e. $\mathbb{P}(Y > x) \sim x^{-\beta} L(x)$ for some $\beta > 0$ and L is a slowly varying function. For the proof of Lemma 5 we need to estimate the probability of the event figuring in Lemma 3 with $b < c \leq 1$ and the probability of the event figuring in Lemma 4 with $b < \alpha < \frac{1}{2}$. To show that these probabilities are of order $O(N^{-2(b+\delta)})$ we only need that $\mathbb{P}\left(\sum_{i=1}^{n} Y_i > x\right) = O(n^{-2(b+\delta)})$ (since the remaining probability in Lemma 3 can be estimated with Hoeffding's inequality, see Hoeffding (1963)) with $n = N$, $x = N^{1-\alpha}$ in Lemma 4 and $n = N^c$, $x = N^c$ in Lemma 3. The asymptotics (3.2) in Mikosch and Nagaev (1998) states that $\mathbb{P}\left(\sum_{i=1}^{n} Y_i > x\right) \sim n x^{-\beta} L(x)$. Consequently, we need to choose $\beta > 0$ such that $N^{1-\beta(1-\alpha)} L(N^{1-\alpha}) = O(N^{-2(b+\delta)})$ as well as $N^{c-\beta c} L(N^c) = O(N^{-2(b+\delta)})$. This works for all choices of $0 < b < \frac{1}{2}$, provided that $\beta \geq 4$.

It would be nice to have a proof of the asymptotics (14) under the assumption that the 4th moment of Y is finite, even without the assumption of a regularly varying tail.

The investigation of the analogue to (14) in the absence of finite second moments, i.e. for Cannings models with heavy-tailed offspring distributions, is the subject of ongoing research, and will be treated in a forthcoming paper.

(b) Relation (11) will be used in the proof of Lemma 6. Moreover, this relation is also instrumental in the companion paper Boenkost et al. (2021) (on the regime of moderately weak selection). The special case $n = 3$ in Lemma 2 a) shows that the assumption $\mathbb{E}[Y^3] < \infty$ implies

$$\mathbb{E}[(W_1^{(N)})^3] = O(N^{-3}).$$

This gives a rate of decay $O(N^{-2})$ for the triple coalescence probability (and is the moment condition (3.6) in Boenkost et al. (2021)).

Condition (8) (on the existence of an exponential moment of Y) guarantees the Haldane asymptotics (14) for Cannings models with weights of Dirichlet type also in the whole regime of moderately weak selection $N^{-1+\eta} \leq s_N \leq N^{-\frac{1}{2}-\eta}$ without any further assumption. In particular the assumption on the finiteness of a negative moment of Y in Boenkost et al. (2021), Lemma 3.7 b), is unnecessary. Indeed, in the proof of Lemma 2(a) we show that $\mathbb{E}[(W_1^{(N)})^n] \leq \left(\frac{2}{N}\right)^n (\mathbb{E}[Y^n] + o(1))$. As shown in the proof of Lemma 3.7(b) in Boenkost et al. (2021) Condition (8)

guarantees that for a sequence (h_N) with $h_N \to \infty$ and $h_N \in O(\log N)$ for all $n \leq 2h_N$ we can estimate $\mathbb{E}[Y^n]$ from above by $C\left(\frac{2h_N}{c}\right)^n$ for appropriate constants $C, c > 0$. Consequently, for N sufficiently large we have $\mathbb{E}[(W_1^{(N)})^n] \leq \left(\frac{Kh_N}{N}\right)^n$ for some appropriate constant $K > 0$, that is Condition (3.8) in Boenkost et al. (2021) is fulfilled.

(c) It seems a mathematically intriguing question whether in the regime of moderate selection all Cannings models which admit a paintbox representation with Dirichlet-type weights and are in Kingman domain of attraction, also follow the Haldane asymptotics (14).

An example of a sequence of Cannings models (with weights *not* of Dirichlet-type) which fulfil Möhle's condition but do not follow the Haldane asymptotics, is the following. In each generation a randomly chosen individual gets weight $N^{-\gamma}$, $0 < \gamma < \frac{1}{2}$ and all the other individuals have a weight of $\frac{1-N^{-\gamma}}{N-1}$. Then we have $\mathbb{E}\left[W_1^2\right] \sim N^{-1-2\gamma}$ and $\mathbb{E}\left[W_1^3\right] = o(\mathbb{E}\left[W_1^2\right])$, therefore by Möhle's criterion the genealogy lies in the attraction of Kingman's coalescent. However, the Haldane asymptotics would predict that the survival probability is of order $s_N/(N^2 N^{-1-2\gamma}) \sim N^{-1-b+2\gamma}$, which for $\gamma < b/2$ is $\ll N^{-1}$. Since the fixation probability of a beneficial allele cannot be smaller than the fixation probability under neutrality (which is $\frac{1}{N}$), (14) must be violated in this example.

(d) The present work together with the approach in Boenkost et al. (2021) does not cover the boundary case $b = \frac{1}{2}$. A quick argument why our arguments cannot be extended simply to the boundary case is the following. We show that once the beneficial type exceeds (in the order of magnitude) the frequency $s_N^{-1} = N^b$ it goes to fixation with high probability. In the regime $b < \frac{1}{2}$ we use couplings with Galton–Watson processes to show that this threshold is reached with probability $\frac{2s_N}{\rho^2}(1 + o(1))$. However, these couplings are not guaranteed as soon as collisions occur, i.e. when beneficial individuals are replacing beneficial individuals. By the well known "birthday problem" collisions are common as soon as $N^{\frac{1}{2}}$ individuals are of the beneficial type. Therefore we require $N^b \ll N^{\frac{1}{2}}$, i.e. $b < \frac{1}{2}$.

In the light of the results of the present paper and of Boenkost et al. (2021), there is little reason to doubt that the assertion of Theorem 1 should fail in the boundary case $b = 1/2$. However, the question remains open (and intriguing) whether then the backward or the forward approach (or a combination of both) is the appropriate tool for the proof.

4 Auxiliary results

4.1 Slightly supercritical Galton–Watson processes

Throughout this subsection, $(s_N)_{N\in\mathbb{N}}$ is a sequence of positive numbers converging to 0, σ^2 is a fixed positive number, and $Z^{(N)} = (Z_n^{(N)})_{n\geq 0}$, $N = 1, 2, \ldots$ are Galton–

Watson processes with offspring expectation

$$\mathbb{E}_1[Z_1^{(N)}] = 1 + s_N + o(s_N), \tag{16}$$

offspring variance $\sigma^2 + o(1)$ and uniformly bounded third moments $\mathbb{E}_1[(Z_1^{(N)})^3]$. Unless stated otherwise we assume that $Z_0^{(N)} = 1$. We write

$$\phi_N := \mathbb{P}\left(\lim_{n\to\infty} Z_n^{(N)} = \infty\right) = 1 - \mathbb{P}\left(Z_n^{(N)} = 0 \text{ for some } n > 1\right) \tag{17}$$

for the survival probability of $(Z^{(N)})$ and observe

$$\phi_N \sim \frac{2s_N}{\sigma^2}. \tag{18}$$

The derivation and discussion of the asymptotics (18) has a venerable history, a few key references being Haldane (1927), Kolmogorov (1938), Eshel (1981), Hoppe (1992), Athreya (1992) Theorem 3, Haccou et al. (2005) Theorem 5.5.

Lemma B.3 in González Casanova et al. (2017) gives a statement on the asymptotic probability that $Z^{(N)}$ either quickly dies out or reaches a certain (moderately) large threshold. The following lemma improves on this in a twofold way. It dispenses with the assumption $s_N \sim cN^{-b}$ for a fixed $b \in (0, 1)$ and more substantially, it gives a *quantitative* estimate for the probability that, given non-extinction, the (moderately) large threshold is reached quickly.

Lemma 1 *Fix $\delta > 0$, and let $T^{(N)} := \inf\{n \geq 0 : Z_n^{(N)} \notin \{1, 2, \ldots, \lceil(\frac{1}{s_N})^{1+\delta}\rceil\}$. Then, for all $\varepsilon > 0$*

$$\mathbb{P}_1\left(T^{(N)} > (1/s_N)^{(1+\varepsilon)}\right) = O\left(\exp\left(-cs_N^{-\varepsilon/2}\right)\right), \tag{19}$$

with $c = -\log\left(\frac{7}{8}\right)$.

Proof Observe that

$$\mathbb{P}_1\left(T^{(N)} > (1/s_N)^{1+\varepsilon}\right) \leq \mathbb{P}_1\left(T^{(N)} > (1/s_N)^{1+\varepsilon}\,\Big|\,Z^N \text{ survives}\right)$$
$$+ \mathbb{P}_1\left(T^{(N)} > (1/s_N)^{1+\varepsilon}\,\Big|\,Z^N \text{ dies out}\right). \tag{20}$$

In Part 1 of the proof we will estimate the first probability on the r.h.s. of (20); this will give the above-mentioned improvement of Lemma B.3 in González Casanova et al. (2017). Part 2 of the proof deals with the second probability on the r.h.s. of (20).

Part 1. Like in the proof of Lemma B.3 in González Casanova et al. (2017) we obtain an upper bound on the time at which the process $Z^{(N)}$ reaches the level $(1/s_N)^{1+\delta}$ given survival, by considering the process $Z^\star = (Z_n^\star)_{n\geq 0}$ consisting of the immortal lines of $Z^{(N)}$ conditioned to non-extinction. (For simplicity of notation we drop a

superscript N in $Z^\star$.) Let ϕ_N denote the survival probability of $Z^{(N)}$ as in (17). The offspring distribution of $Z^\star$ arises from that of $Z^{(N)}$ as

$$\mathbb{P}_1(Z_1^\star = k) = \frac{1}{\phi_N} \mathbb{E}_1\left[\binom{Z_1^{(N)}}{k} \phi_N^k (1 - \phi_N)^{Z_1^{(N)} - k} \right], \quad k \geq 1. \tag{21}$$

(see Lyons and Peres 2017 Proposition 5.28). In particular one has

$$\mathbb{E}_1[Z_1^\star] = \frac{1}{\phi_N} \mathbb{E}_1[Z_1^{(N)} \phi_N] = \mathbb{E}_1[Z_1^{(N)}].$$

Denote, as usual, for a random variable X and an event A by $\mathbb{E}[X; A] := \mathbb{E}[X \mathbb{1}_A]$. Furthermore,

$$\mathbb{E}_1[Z_1^\star; Z_1^\star \geq 3] \leq \frac{1}{\phi_N} \mathbb{E}_1\left[\binom{Z_1^{(N)}}{3} \phi_N^3 \right] = O(\phi_N^2)$$

because of the assumed uniform boundedness of the third moments of $Z_1^{(N)}$. These two relations together with (16), (18) and the fact that $\mathbb{E}_1[Z_1^\star; Z_1^\star = 1] \leq 1$ immediately give a lower bound for $\mathbb{E}_1[Z_1^\star; Z_1^\star = 2] \leq 1$, implying that for any $\beta \in (0, 1)$

$$\mathbb{P}_1(Z_1^\star \geq 2) \geq \beta s_N, \qquad \mathbb{P}_1(Z_1^\star = 1) \leq 1 - \beta s_N. \tag{22}$$

Hence the process $Z^{(N)}$, when conditioned on survival, is bounded from below by the counting process $Z^\star$ of immortal lines, which in turn is bounded from below by the process $\widetilde{Z} = (\widetilde{Z}_n)_{n \geq 0}$ with offspring distribution

$$\nu = (1 - \beta s_N)\delta_1 + \beta s_N \delta_2.$$

So far we closely followed the proof in González Casanova et al. (2017), but now we deviate from that proof to obtain the rate of convergence claimed in (19).

An upper bound for the time $\widetilde{T} := \inf\{n \geq 0 : \widetilde{Z}_n \geq (1/s_N)^{1+\delta}\}$ also gives an upper bound for the time $T^{(N)}$. The idea is now to divide an initial piece of $k \leq (1/s_N)^{(1+\varepsilon)}$ generations into $\lfloor (1/s_N)^{\varepsilon/2} \rfloor$ parts, each of $n_0 \leq (1/s_N)^{(1+\varepsilon/2)}$ generations. Because of the immortality of $\widetilde{Z}$ and the independence between these parts we obtain immediately that

$$\mathbb{P}(\widetilde{T} \geq (1/s_N)^{(1+\varepsilon)}) \leq \mathbb{P}_1(\widetilde{Z}_j \leq (1/s_N)^{1+\delta} \text{ for } j = 1, \ldots, k)$$
$$\leq \left(\mathbb{P}_1(\widetilde{Z}_{n_0} \leq (1/s_N)^{1+\delta}) \right)^{\lfloor (1/s_N)^{\varepsilon/2} \rfloor}$$

We then bound $\mathbb{P}_1(\widetilde{Z}_{n_0} > (1/s_N)^{1+\delta})$ from below by an application of the Paley-Zygmund inequality in its form

$$\mathbb{P}\left(X \geq \frac{\mathbb{E}[X]}{2}\right) \geq \frac{1}{4} \frac{(\mathbb{E}[X])^2}{\mathbb{E}[X^2]}, \tag{23}$$

where X is a non-negative random variable (with finite second moment). For a super-critical Galton–Watson process with offspring expectation m and offspring variance σ^2 the n-th generation offspring expectation and n-th generation offspring variance σ_n^2 are given by m^n and $\sigma^2 m^n (m^n - 1)/(m^2 - m)$ (see Athreya and Ney (1972), p.4). Hence, we obtain

$$\mathbb{E}_1[\tilde{Z}_n] = (1 + \beta s_N)^n,$$
$$\mathrm{Var}_1[\tilde{Z}_n] = \frac{\beta s_N (1 - \beta s_N)(1 + \beta s_N)^n ((1 + \beta s_N)^n - 1)}{(1 + \beta s_N)^2 - (1 + \beta s_N)}. \tag{24}$$

We choose the smallest n_0 such that

$$\mathbb{E}_1[\widetilde{Z}_{n_0}] \geq 2(1/s_N)^{1+\delta}. \tag{25}$$

Observe that $n_0 \sim \frac{1}{\beta s_N} \log(2(\frac{1}{s_N})^{1+\delta})$ which ensures that $(1/s_N)^{\varepsilon/2} n_0 \leq (1/s_N)^{1+\varepsilon}$ for N large enough. We now estimate $\mathbb{E}_1[(\widetilde{Z}_{n_0})^2]$ using (24) as follows

$$\mathbb{E}\left[\widetilde{Z}_{n_0}^2\right] = \frac{\beta s_N (1 - \beta s_N)(1 + \beta s_N)^{n_0}((1 + \beta s_N)^{n_0} - 1)}{(1 + \beta s_N)^2 - (1 + \beta s_N)} + (1 + \beta s_N)^{2n_0}$$
$$\leq \frac{\beta s_N (1 + \beta s_N)^{2n_0}}{\beta s_N + (\beta s_N)^2} + (1 + \beta s_N)^{2n_0} \leq 2(1 + \beta s_N)^{2n_0}.$$

Applying (23) with $X := \widetilde{Z}_{n_0}$ yields

$$\mathbb{P}_1\left(\widetilde{Z}_{n_0} \geq \frac{1}{2}\mathbb{E}\left[\widetilde{Z}_{n_0}\right]\right) \geq \frac{1}{4}\frac{(1 + \beta s_N)^{2n_0}}{2(1 + \beta s_N)^{2n_0}} = \frac{1}{8},$$

which because of (25) implies $\mathbb{P}_1(\widetilde{Z}_{n_0} \leq (1/s_N)^{1+\delta}) \leq \frac{7}{8}$. If after time n_0 the process $\widetilde{Z}_{n_0}$ is still smaller than our desired bound $(1/s_N)^{1+\delta}$, we can iterate this argument $\lfloor (1/s_N)^{\varepsilon/2} \rfloor$ times and arrive at

$$(\mathbb{P}_1(\widetilde{Z}_{n_0} \leq (1/s_N)^{1+\delta})^{\lfloor (1/s_N)^{\varepsilon/2} \rfloor} \leq \left(\frac{7}{8}\right)^{\lfloor (1/s_N)^{\varepsilon/2} \rfloor} = \exp(-c\lfloor (1/s_N)^{\varepsilon/2} \rfloor),$$

with $c = -\log \frac{7}{8}$. This gives the desired bound for the first term in (20).

Part 2. We now turn to the second term on the r.h.s. of (20). Define

$$T_0^{(N)} := \inf\{n \geq 0 : Z_n^{(N)} = 0\}.$$

Obviously $T^{(N)} \le T_0^{(N)}$, and so it suffices to prove

$$\mathbb{P}\left(T_0^{(N)} > (1/s_N)^{1+\varepsilon} | Z^{(N)} \text{ dies out}\right) \le \exp(-\beta s_N^{-\varepsilon}(1 + o(1))). \qquad (26)$$

This proof follows closely that of the second part of Lemma B.3 in González Casanova et al. (2017); we include it here for completeness.

We observe

$$
\begin{aligned}
\mathbb{E}_1\left[Z_1^{(N)} | Z^{(N)} \text{ dies out}\right] &= \frac{1}{1 - \phi_N} \mathbb{E}_1\left[(1 - \phi_N)^{Z_1^{(N)}} Z_1^{(N)}\right] \\
&= \mathbb{E}_1\left[(1 - \phi_N)^{Z_1^{(N)}-1} Z_1^{(N)}\right] = \mathbb{P}_1(Z_1^\star = 1),
\end{aligned}
$$

where the first and the last equality follow from the branching property and from (21), respectively. We have shown in (22) that $\mathbb{P}_1(Z_1^\star = 1) \le 1 - \beta s_N + o(s_N)$ and hence we can conclude

$$
\begin{aligned}
\mathbb{E}\left[Z_{\lfloor (1/s_N)^{1+\varepsilon} \rfloor} | Z^{(N)} \text{ dies out}\right] &\le (1 - \beta s_N + o(s_N))^{\lfloor (1/s_N)^{1+\varepsilon} \rfloor} \\
&\le \exp(-\beta s_N^{-\varepsilon}(1 + o(1))).
\end{aligned}
$$

Finally, an application of Markov's inequality yields (26)

$$
\begin{aligned}
\mathbb{P}(T_0^{(N)} > (1/s_N)^{1+\varepsilon} | Z^{(N)} \text{ dies out}) &\le \mathbb{P}(Z_{\lfloor (1/s_N)^{1+\varepsilon} \rfloor} \ge 1 | Z^{(N)} \text{ dies out}) \\
&\le \exp(-\beta s_N^{-\varepsilon}(1 + o(1))).
\end{aligned}
$$

$\square$

4.2 Estimates on the paintbox

The following lemma provides the asymptotics (11) as well as the moment bounds for the Dirichlet-type weights that were addressed in Remark 2(b).

Lemma 2 (Moments of the weights) *Let $Y, Y_1, \ldots, Y_N$ be iid positive random variables with $\mathbb{E}[Y] = 1$ and $\rho^2 = \mathbb{E}[Y^2]$. We abbreviate $H_N := Y_1 + \cdots + Y_N$.*
(a) Assume $\mathbb{E}[Y^n] < \infty$ for some $n \in \mathbb{N}$. Then

$$\mathbb{E}\left[\left(\frac{Y}{(Y + H_N)}\right)^n\right] = O(N^{-n}).$$

(b) Assume $\mathbb{E}[Y^4] < \infty$. Then

$$\mathbb{E}\left[\left(\frac{Y}{(Y + H_N)}\right)^2\right] = \frac{\rho^2}{N^2} + O(N^{-3}).$$

Proof a) Consider the event $F_N := \{\frac{H_N}{N} \leq \frac{1}{2}\}$. First we note that

$$\mathbb{E}\left[\left(\frac{Y}{(Y + H_N)}\right)^n\right] \leq \mathbb{E}\left[\left(\frac{Y}{Y + H_N}\right)^n \mathbb{1}_{F_N}\right] + \mathbb{E}\left[\left(\frac{Y}{H_N}\right)^n \mathbb{1}_{F_N^c}\right]$$

$$\leq \mathbb{P}(F_N) + \left(\frac{2}{N}\right)^n \mathbb{E}[Y^n].$$

Let K be so large that $\mathbb{E}[Y \wedge K] > \frac{1}{2}$. Hoeffding's inequality applied to the sample mean of i.i.d. copies of the bounded random variable $Y \wedge K$ implies that $\mathbb{P}(F_N)$ decays exponentially fast. Since $\mathbb{E}[Y^n]$ is bounded this yields the claim.
(b) We observe that

$$\mathbb{E}\left[\frac{Y^2}{(Y + H_N)^2}\right] = \frac{1}{N^2}\mathbb{E}\left[\frac{Y^2}{\left(\frac{Y}{N} + \frac{H_N}{N}\right)^2}\right].$$

Let $E_N^{(1)} := \{\frac{H_N}{N} > \frac{5}{4}\}$, $E_N^{(2)} := \{\frac{H_N}{N} < \frac{3}{4}\}$ and $E_N^{(3)} := \{Y > \frac{N}{4}\}$. Markov's inequality applied to $\mathbb{P}\left(\left(\sum_{i=1}^{N}(Y_i - 1)\right)^4 \geq \frac{N^4}{4^4}\right)$ together with the assumption $\mathbb{E}[Y^4] < \infty$ implies $\mathbb{P}(E_N^{(1)}) = O(N^{-2})$. Likewise, $\mathbb{P}(E_N^{(3)}) = O(N^{-4})$. Furthermore, let K be so large that $\mathbb{E}[Y \wedge K] > \frac{3}{4}$. Again, Hoeffding's inequality applied to the sample mean of i.i.d.copies of the bounded random variable $Y \wedge K$ together with monotonicity imply that $\mathbb{P}(E_N^{(2)})$ decays exponentially; a fortiori we have $\mathbb{P}(E_N^{(2)}) = O(N^{-3})$.

Let $E_N := E_N^{(1)} \cup E_N^{(2)} \cup E_N^{(3)}$. We have $\mathbb{E}\left[\frac{Y^2}{\left(\frac{H_N}{N} + \frac{Y}{N}\right)^2}\mathbb{1}_{E_N}\right] = O(N^{-1})$ since on $E_N^{(1)}$ we have $\frac{Y^2}{\left(\frac{H_N}{N} + \frac{Y}{N}\right)^2} \leq Y^2$ and on $E_N^{(2)}$ and $E_N^{(3)}$ we have $\frac{Y^2}{\left(\frac{H_N}{N} + \frac{Y}{N}\right)^2} \leq N^2$.

Hence, it remains to show that $\mathbb{E}\left[\frac{Y^2}{\left(\frac{H_N}{N} + \frac{Y}{N}\right)^2}\mathbb{1}_{E_N^c}\right] = \rho^2 + O(N^{-1})$. Define $Z_N := \frac{1}{\sqrt{N}}(H_N - N)$ and observe

$$\mathbb{E}\left[\frac{Y^2}{\left(\frac{H_N}{N} + \frac{Y}{N}\right)^2}\mathbb{1}_{E_N^c}\right] = \mathbb{E}\left[\frac{Y^2}{\left(1 + \frac{Z_N}{\sqrt{N}} + \frac{Y}{N}\right)^2}\mathbb{1}_{E_N^c}\right].$$

Abbreviate $R_N = 2\left(\frac{Z_N}{\sqrt{N}} + \frac{Y}{N}\right) + \left(\frac{Z_N}{\sqrt{N}} + \frac{Y}{N}\right)^2$. On E_N^c we have $-\frac{1}{2} \leq R_N$ and hence

$$1 - R_N \leq \frac{1}{\left(\frac{H_N}{N} + \frac{Y}{N}\right)^2} = \frac{1}{1 + R_N} \leq 1 - R_N + 2R_N^2.$$

Thus

$$\mathbb{E}[Y^2(1 - R_N)\mathbb{1}_{E_N^c}] \leq \mathbb{E}\left[\frac{Y^2}{\left(\frac{H_N - N}{N} + \frac{Y}{N}\right)^2}\mathbb{1}_{E_N^c}\right]$$

$$\leq \mathbb{E}[Y^2(1 - R_N + 2R_N^2)\mathbb{1}_{E_N^c}]. \qquad (27)$$

By Cauchy-Schwarz we have $\mathbb{E}[Y^2\mathbb{1}_{E_N^c}] = \mathbb{E}[Y^2] + O(N^{-1})$. Similarly, $\mathbb{E}[Y^2 Z_N \mathbb{1}_{E_N}] = \mathbb{E}[Y^2 Z_N] + O(N^{-1}) = O(N^{-1})$, since $\mathbb{E}[Y^2 Z_N]$ vanishes due to the independence of Y and Z_N. The remaining terms in (27) are $O(N^{-1})$ as well, which completes the proof of Lemma 2. $\qquad\square$

We now prove a bound on the deviations for the total weight of k individuals.

Lemma 3 (Large deviations bound for a moderate number of random weights) *Let* (Y_i) *and* $(W_i^{(N)})$ *satisfy* (7), (8), $\mathbb{E}[Y] = 1$ *and let* $k = k_N \leq N^c$ *for some* $0 < c \leq 1$. *Then for all* $\varepsilon > 0$ *there exists a positive constant* c_ε *depending only on* ε *and the distribution of* Y *such that*

$$\mathbb{P}\left(\sum_{i=1}^k W_i^{(N)} \geq (1 + \varepsilon)N^{c-1}\right) = O(\exp(-c_\varepsilon N^c)). \qquad (28)$$

Proof This follows by a combination of two Cramér bounds. Indeed, the l.h.s. of (28) is by assumption bounded from above by

$$\mathbb{P}\left(\frac{\sum_{i=1}^{\lceil N^c\rceil} Y_i}{\sum_{j=1}^N Y_j} \geq (1 + \varepsilon)N^{c-1}\right).$$

Abbreviating $E := \{\sum_{j=1}^N Y_j \geq (1 - \varepsilon')N\}$ with ε' such that $(1 + \varepsilon)(1 - \varepsilon') > 1$ we estimate the latter probability from above by

$$\mathbb{P}\left(\sum_{i=1}^{\lceil N^c\rceil} Y_i \geq (1 + \varepsilon)N^{c-1}\sum_{j=1}^N Y_j, E\right) + \mathbb{P}\left(E^c\right)$$

$$\leq \mathbb{P}\left(\sum_{i=1}^{\lceil N^c\rceil} Y_i \geq N^c(1 + \varepsilon)(1 - \varepsilon')\right) + \mathbb{P}\left(E^c\right)$$

$$= O(e^{-N^c I((1+\varepsilon)(1-\varepsilon'))}) + O(e^{-NI(1-\varepsilon')}),$$

denoting by $I(y)$ the rate function of Y. Due to (8) $I(y)$ exists around $\mathbb{E}[Y] = 1$ and is strictly positive for $y \neq 1$ (see Dembo and Zeitouni (1994) Theorem 2.2.3). This yields an upper bound of $O(\exp(-c_\varepsilon N^c))$ with $c_\varepsilon = \min\{I((1 + \varepsilon)(1 - \varepsilon')), I(1 - \varepsilon')\}$. $\quad\square$

The next lemma gives stochastic upper and lower bounds for the sums of the random weights in terms of sums of the independent random variables Y_i.

Lemma 4 (Bounds for the random weights) *Assume that Conditions* (7) *and* (8) *are fulfilled and* $\mathbb{E}[Y] = 1$. *Let* $0 < \alpha < \frac{1}{2}$, *then for* $k = k_N \leq N$

$$\mathbb{P}\left(\frac{1 - N^{-\alpha}}{N} \sum_{i=1}^{k} Y_i \leq \sum_{i=1}^{k} W_i^{(N)} \leq \frac{1 + N^{-\alpha}}{N} \sum_{i=1}^{k} Y_i\right)$$
$$\geq 1 - \exp\left(-c' N^{1-2\alpha}\right) (1 + o(1)),$$

for some $c' > 0$.

Proof It suffices to show

$$\mathbb{P}\left(\left|\frac{N}{\sum_{j=1}^{N} Y_j} - 1\right| \geq N^{-\alpha}\right) = O(\exp(-N^{1-2\alpha})).$$

For $0 < c < 1$ we have

$$\mathbb{P}\left(\sum_{i=1}^{N} Y_i < cN\right) = O(\exp(-N I(c))),$$

where $I(y)$ is the rate function of Y. Condition (8) ensures that $I(c) > 0$ for $\mathbb{E}[Y_1] = 1 \neq c$, see Dembo and Zeitouni (1994) Theorem 2.2.3.

For any $a, a' \geq 1$ one has $\left|\frac{1}{a} - \frac{1}{a'}\right| \leq |a - a'|$. This yields

$$\mathbb{P}\left(\left|\frac{N}{\sum_{j=1}^{N} Y_j} - 1\right| \geq N^{-\alpha}\right) = \mathbb{P}\left(\frac{1}{c}\left|\frac{Nc}{\sum_{j=1}^{N} Y_j} - c\right| \geq N^{-\alpha}\right)$$
$$\leq \mathbb{P}\left(\frac{1}{c^2}\left|\frac{\sum_{j=1}^{N} Y_j}{N} - 1\right| \geq N^{-\alpha}\right) + O(e^{-N I(c)})$$
$$= \mathbb{P}\left(\frac{1}{\sqrt{N}}\left|\sum_{i=1}^{N}(Y_i - 1)\right| \geq c^2 N^{\frac{1}{2} - \alpha}\right) + O(e^{-N I(c)}).$$

Using Cramér (1938) Theorem 1, the probability on the r.h.s. can, with a suitable $\widetilde{c} > 0$, be estimated from above by

$$\exp\left(\widetilde{c} N^{1-3\alpha}\right) \exp\left(-\frac{c^4}{2} N^{1-2\alpha}\right) (1 + O(N^{-\alpha} \log N))$$
$$= \exp\left(-\frac{c^4}{2} N^{1-2\alpha}\right) (1 + o(1)),$$

which gives the desired result. □

5 Proof of the main result

Recall from (13) that we denote the order of the selection strength by $b_N = -\frac{\log s_N}{\log N}$. To simplify notation we will drop the subscript and simply write $b := b_N$. As mentioned already in the sketch of the proof of Theorem 1 we assume without loss of generality that $\mathbb{E}[Y] = 1$.

The proof of the Theorem is divided into three parts, which correspond to three phases of growth for the Cannings frequency process $\mathcal{X}$. The initial phase is decisive: due to Proposition 1, the probability that $\mathcal{X}$ reaches the level $N^{b+\delta}$ for some sufficiently small δ is given by the r.h.s. of (14). Lemmas 6 and 7 then guarantee that, once having reached the level $N^{b+\delta}$, the process $\mathcal{X}$ reaches N with high probability. The proof of the Theorem is then a simple combination of these three results and the strong Markov property. Indeed, with τ_1, τ_2, τ_3 as in Proposition 1, Lemmas 6 and 7, and with $\delta, \delta', \varepsilon$ fulfilling the requirements specified there, the fixation probability in the l.h.s. of (14) can be rewritten as

$$
\begin{aligned}
\mathbb{P}(X_\tau = N) &= \mathbb{P}(X_{\tau_3} = N | X_{\tau_2} \geq \varepsilon N) \mathbb{P}(X_{\tau_2} \geq \varepsilon N | X_{\tau_1} \geq N^{b+\delta}) \mathbb{P}_1(X_{\tau_1} \geq N^{b+\delta}) \\
&= (1 - o(1))(1 - O(N^{-\delta'})) \frac{2 s_N}{\rho^2}(1 + o(1)) \\
&\sim \frac{2 s_N}{\rho^2}.
\end{aligned}
$$

5.1 First phase: from 1 to $N^{b+\delta}$

In this section we show that as long as $X_g \leq N^{b+\delta}$ the process $\mathcal{X}$ can be upper and lower bounded (with sufficiently high probability) by two slightly supercritical branching processes $\underline{Z} = (\underline{Z}_g)_{g \geq 0}$ and $\overline{Z} = (\overline{Z}_g)_{g \geq 0}$. To construct the upper bound $\overline{Z}$ we take the highest per capita selective advantage, which occurs when only a single individual is beneficial. Using Lemmas 3 and 4, we will approximate the thus arising mixed binomial distribution by a mixed Poisson distribution, which leads for $\overline{Z}$ to the offspring distribution

$$
\text{Pois}\left(Y_1(1 + s_N + o(s_N))\right), \tag{29}
$$

where Y_1 is the random variable figuring in (7). To arrive at the lower bounding Galton–Watson process $\underline{Z}$ we note that the per capita selective advantage is bounded from below by the one when $\lceil N^{b+\delta} \rceil$ beneficial individuals are present in the parent generation, as long as the process $\mathcal{X}$ has not reached the level $N^{b+\delta}$. Again using Lemmas 3 and 4 we will show that the offspring distribution of $\underline{Z}$ can be chosen as the mixed binomial distribution

$$
\text{Bin}\left(N - \lceil N^{b+\delta} \rceil, \frac{Y_1}{N}(1 + s_N + o(s_N))\right). \tag{30}
$$

Lemma 5 (Coupling with Galton–Watson processes) *Let δ and α be such that $0 < \delta < \eta$ and $\frac{1}{2} - \eta < \alpha < \frac{1}{2}$, and put*

$$\tau_1 = \inf\{g \geq 0 : X_g = 0 \text{ or } X_g \geq N^{b+\delta}\}.$$

Then $\mathcal{X}$ can be defined on one and the same probability space together with two branching process $\underline{\mathcal{Z}}$ and $\overline{\mathcal{Z}}$ with offspring distributions (30) and (29), respectively, such that for $j = 1, 2, \ldots$

$$\mathbb{P}(\underline{Z}_{j\wedge\tau_1} \wedge \lceil N^{b+\delta}\rceil \leq X_{j\wedge\tau_1} \wedge \lceil N^{b+\delta}\rceil \leq \overline{Z}_{j\wedge\tau_1} \mid \underline{Z}_{j-1\wedge\tau_1} \leq X_{j-1\wedge\tau_1} \leq \overline{Z}_{j-1\wedge\tau_1})$$
$$\geq 1 - e^{-c'N^{1-2\alpha}}(1 + o(1)), \tag{31}$$

with c' as in Lemma 4.

Applying the latter estimate g times consecutively yields immediately the following corollary:

Corollary 1 *Let $\delta, \alpha, \tau_1, \underline{\mathcal{Z}}$ and $\overline{\mathcal{Z}}$ be as in Lemma 5. If $X_0 \leq N^{b+\delta}$, then for all $g \in \mathbb{N}_0$*

$$\mathbb{P}(\underline{Z}_{g\wedge\tau_1} \wedge \lceil N^{b+\delta}\rceil \leq X_{g\wedge\tau_1} \wedge \lceil N^{b+\delta}\rceil \leq \overline{Z}_{g\wedge\tau_1} \mid \underline{Z}_0 \leq X_0 \leq \overline{Z}_0)$$
$$\geq \left(1 - O(\exp(-c'N^{1-2\alpha}))\right)^g. \tag{32}$$

Proof *of Lemma 5.* We proceed inductively, assuming that for $g = 1, 2, \ldots$ we have constructed $\mathcal{X}, \overline{\mathcal{Z}}$ and $\underline{\mathcal{Z}}$ up to generation $g-1$ such that (31) holds for $j = 1, \ldots, g-1$. Together with X_g we will construct $\overline{Z}_g$ and $\underline{Z}_g$, and check the asserted probability bound for the coupling.

Given $\{X_{g-1} = k\}$ and the weights (W_i) in generation $g-1$, the number of beneficial individuals X_g in generation g has the binomial distribution (6). Aiming first at the construction of the upper bound $\overline{\mathcal{Z}}$, we relate (6) to (29) in terms of stochastic order. For $p, p' \geq 0$, a $\mathrm{Bin}(N, p)$-distributed random variable B is stochastically dominated by a $\mathrm{Pois}(Np')$-distributed random variable P if

$$e^{-p'} \leq (1 - p), \tag{33}$$

see (1.21) in Klenke and Mattner (2010). Indeed, in this case the probability of the outcome zero is not larger for a $\mathrm{Pois}(p')$-distributed random variable P_1 than for a Bernoulli(p)-distributed random variable B_1, which yields $B_1 \preceq P_1$, where $\preceq$ denotes the usual stochastic ordering of the random variables. Consequently

$$B \stackrel{d}{=} \sum_{i=1}^{N} B_i \preceq \sum_{i=1}^{N} P_i \stackrel{d}{=} P.$$

with B_i and P_i being independent copies of B_1 and P_1, respectively. In particular, for $p \geq 0$ and $p' = p(1 + N^{b+2\delta-1})$ we have

$$e^{-p'} \leq 1 - p' + (p')^2 = 1 - p(1 + N^{b+2\delta-1}) + p^2(1 + N^{b+2\delta-1})^2.$$

Hence Condition (33) holds if

$$p(1 + N^{b+2\delta-1})^2 < N^{b+2\delta-1}. \tag{34}$$

Given (W_i), the success probability of the binomial distribution (6) is bounded from above via

$$p := \left(\sum_{i=1}^{k} W_i \right) / (1 - s_N).$$

Thus by Lemma 3, (34) is fulfilled with probability $1 - O(\exp(-c_\varepsilon N^{b+\delta}))$ with c_ε as in Lemma 3. In this sense the number of beneficial offspring is dominated by a $\text{Pois}\left(N \frac{\sum_{i=1}^{k} W_i}{(1-s_N)}(1 + N^{b+2\delta-1}) \right)$-distributed random variable with high probability. Applying Lemma 4 yields that with probability $1 - \exp(-c'N^{1-2\alpha})(1 + o(1))$ the following chain of inequalities is valid:

$$N \frac{\sum_{i=1}^{X_{g-1}} W_i}{(1 - s_N)}(1 + N^{b+2\delta-1}) \leq \frac{\sum_{i=1}^{X_{g-1}} Y_i}{(1 - s_N)}(1 + N^{b+2\delta-1})(1 + N^{-\alpha})$$

$$= \sum_{i=1}^{X_{g-1}} Y_i(1 + s_N + o(s_N)) \leq \sum_{i=1}^{\overline{Z}_{g-1}} Y_i(1 + s_N + o(s_N)).$$

In this way $\mathcal{X}$ can be coupled with a branching process $\overline{\mathcal{Z}}$ with a mixed Poisson offspring distribution of the form (29).

The lower bound also uses a comparison with a Galton–Watson process, now with a mixed binomially distributed offspring distribution:

Number the individuals in generation $g - 1$ by $(i, g - 1)$, with $(1, g - 1), \ldots,$ $(X_{g-1}, g - 1)$, being the beneficial individuals. Given $\mathcal{W}$, we use a sequence of coin tossings to determine which of the individuals from generation g are the children of $(i, g - 1)$. The first N tosses determine which individuals are the children of $(1, g - 1)$. Denoting the number of these children by ω_1^{g-1}, the next $N - \omega_1^{g-1}$ tosses (with an updated success probability) determine which individuals are the children of $(2, g - 1)$, etc. Observe that as long as $X_{g-1} \leq N^{b+\delta}$, and given $\mathcal{W}$ and $\sum_{\ell=1}^{i-1} \omega_\ell^{(g-1)} =: h$, then $\omega_i^{(g-1)}$ for $i \leq X_{g-1}$ has distribution

$$\text{Bin}\left(N - h, \frac{W_i}{\sum_{\ell=i}^{X_{g-1}} W_\ell + (1 - s_N) \sum_{\ell=X_{g-1}+1}^{N} W_\ell} \right). \tag{35}$$

Note that the success probability in (35) can be estimated from below by

$$\frac{W_i}{\sum_{\ell=1}^{X_{g-1}} W_\ell + (1 - s_N) \sum_{\ell=X_{g-1}+1}^{N} W_\ell} = \frac{W_i}{1 - s_N + s_N \sum_{\ell=1}^{X_{g-1}} W_\ell}.$$

As long as $X_{g-1} \leq \lceil N^{b+\delta} \rceil$, Lemma 3 ensures that for $\varepsilon > 0$

$$\frac{W_i}{1 - s_N + s_N \sum_{j=1}^{X_{g-1}} W_j} \geq \frac{W_i}{1 - s_N + (1 + \varepsilon) N^{\delta-1}} \tag{36}$$

with probability $1 - O(\exp(-c_\varepsilon N^{b+\delta}))$. Lemma 4, in turn, yields that the r.h.s. of (36) is bounded from below by

$$\frac{Y_i}{N(1 - s_N + (1 + \varepsilon) N^{\delta-1})} (1 + N^{-\alpha}) = \frac{Y_i}{N} (1 + s_N + o(s_N))$$

with probability at least $1 - \exp(-c' N^{1-2\alpha})(1 + o(1))$.

Thus, if $\omega_1^{(g-1)} + \cdots + \omega_{i-1}^{(g-1)} = h \leq \lceil N^{b+\delta} \rceil$, then the distribution of $\omega_i^{(g-1)}$ specified in (35) is bounded from below by

$$\mathrm{Bin}\left(N - \lceil N^{b+\delta} \rceil, \frac{W_i}{1 - s_N + (1 + \varepsilon) N^{\delta-1}}\right)$$

with probability $1 - O(\exp(-c_\varepsilon N^{b+\delta}))$.

If $\omega_1^{(g-1)} + \cdots + \omega_{i-1}^{(g-1)} = h > \lceil N^{b+\delta} \rceil$, then we have $\underline{Z}_{g \wedge \tau_1} \wedge \lceil N^{b+\delta} \rceil \leq X_{g \wedge \tau_1} \wedge \lceil N^{b+\delta} \rceil$. Consequently $\mathcal{X}$ can be coupled with a Galton–Watson process $\underline{Z}$ with offspring distribution of the form (30) such that also the lower estimate in (32) is fulfilled. This completes the proof of Lemma 5.   □

We are now ready to prove that $\mathcal{X}$ reaches the level $N^{b+\delta}$ with probability $\frac{2s_N}{\rho^2}(1 + o(1))$.

Proposition 1 (Probability to reach the critical level) *Assume Conditions* (7), (8) *and* (12) *are fulfilled and define* $\tau_1 = \inf\{g \geq 0 : X_g \geq N^{b+\delta} \text{ or } X_g = 0\}$ *with* $0 < \delta < \eta$, *then*

$$\mathbb{P}(X_{\tau_1} \geq N^{b+\delta}) = \frac{2s_N}{\rho^2}(1 + o(1)). \tag{37}$$

Proof We use the couplings of $\mathcal{X}$ with the slightly supercritical branching processes $\underline{Z}$ and $\overline{Z}$ from Corollary 1 and show that both processes reach the level $N^{b+\delta}$ with probability $\frac{2s_N}{\rho^2}(1 + o(1))$. Let $\delta' > 0$ and E be the event that the stochastic ordering between $\underline{Z}, \mathcal{X}$ and $\overline{Z}$ holds until generation $n_0 = \lceil N^{b+\delta'} \rceil$, that is

$$E = \{\underline{Z}_0 \wedge \lceil N^{b+\delta} \rceil \leq X_0 \wedge \lceil N^{b+\delta} \rceil \leq \overline{Z}_0, \ldots, \underline{Z}_{n_0} \wedge \lceil N^{b+\delta} \rceil \leq X_{n_0} \wedge \lceil N^{b+\delta} \rceil \leq \overline{Z}_{n_0}\}.$$

We show below that the stopping time τ_1 fulfils

$$\mathbb{P}(\tau_1 \geq \lceil N^{b+\delta'} \rceil) = o(s_N). \tag{38}$$

For some g that is polynomially bounded in N, the r.h.s. of (32) is bounded from above by $1 - o(s_N)$. Thus, combining Corollary 1 and (38) we deduce

$$\mathbb{P}(E, \tau_1 \leq \lceil N^{b+\delta'} \rceil) = 1 - o(s_N).$$

We are now going to bound (37) from above by estimating the corresponding probability for $\overline{\mathcal{Z}}$ and the stopping time $\overline{\tau}_1 = \inf\{g \geq 0 : \overline{Z}_g \geq N^{b+\delta} \text{ or } \overline{Z}_g = 0\}$. More precisely,

$$
\begin{aligned}
\mathbb{P}(X_{\tau_1} \geq N^{b+\delta}) &= \mathbb{P}(X_{\tau_1} \geq N^{b+\delta}, \tau_1 \leq \lceil N^{b+\delta'} \rceil, E) + o(s_N) \\
&\leq \mathbb{P}(\overline{Z}_{\overline{\tau}_1} \geq N^{b+\delta}, \tau_1 \leq \lceil N^{b+\delta'} \rceil, E) + o(s_N) \\
&\leq \mathbb{P}(\overline{Z}_{\overline{\tau}_1} \geq N^{b+\delta}) + o(s_N).
\end{aligned}
$$

To obtain an upper bound for the probability of $\overline{\mathcal{Z}}$ to reach the level $N^{b+\delta}$ it suffices to estimate the survival probability of $\overline{\mathcal{Z}}$. For notational simplicity let us write $\{\overline{\mathcal{Z}} \text{ survives}\}$ for the event $\{\forall g \geq 0 : \overline{Z}_g > 0\}$ and similarly $\{\overline{\mathcal{Z}} \text{ dies out}\}$ for the event $\{\exists g \geq 0 : \overline{Z}_g = 0\}$. We have

$$
\begin{aligned}
&\mathbb{P}_1\left(\overline{Z}_{\overline{\tau}_1} \geq N^{b+\delta}\right) \\
&\quad \leq \mathbb{P}_1\left(\overline{Z}_{\overline{\tau}_1} \geq N^{b+\delta} | \overline{\mathcal{Z}} \text{ survives}\right) \mathbb{P}_1(\overline{\mathcal{Z}} \text{ survives}) + \mathbb{P}_1\left(\overline{Z}_{\overline{\tau}_1} \geq N^{b+\delta} | \overline{\mathcal{Z}} \text{ dies out}\right) \\
&\quad \leq \mathbb{P}_1(\overline{\mathcal{Z}} \text{ survives}) + \mathbb{P}\left(\text{all } \lceil N^{b+\delta} \rceil \text{ individuals die out}\right) \\
&\quad = \mathbb{P}_1(\overline{\mathcal{Z}} \text{ survives}) + (1 - \mathbb{P}_1(\overline{\mathcal{Z}} \text{ survives}))^{\lceil N^{b+\delta} \rceil}.
\end{aligned}
$$

The survival probability of $\overline{\mathcal{Z}}$ will now be estimated by means of (18). To this purpose we calculate the expectation and variance of the offspring distribution (29). The expectation is $1 + s_N + o(s_N)$ and the variance is given by

$$
\begin{aligned}
&\text{Var}\left(\text{Pois}\left(Y_1(1 + s_N + o(s_N))\right)\right) \\
&\quad = \text{Var}\left(\mathbb{E}\left[\text{Pois}\left(Y_1(1 + s_N + o(s_N))\right) \Big| Y_1\right]\right) \\
&\qquad + \mathbb{E}\left[\text{Var}\left(\text{Pois}\left(Y_1(1 + s_N + o(s_N))\right) \Big| Y_1\right)\right] \\
&\quad = \text{Var}\left(Y_1(1 + s_N + o(s_N))\right) + \mathbb{E}[Y_1(1 + s_N + o(s_N))] \\
&\quad = (1 + s_N + o(s_N))^2 \text{Var}(Y_1) + 1 + s_N + o(s_N) = \rho^2(1 + o(1)).
\end{aligned}
$$

Equation (18) yields that the survival probability of the process $\overline{\mathcal{Z}}$ is given by $\frac{2s_N}{\rho^2}$ $(1 + o(1))$. The lower bound in (37) follows by similar arguments by considering the process $\underline{\mathcal{Z}}$ instead.

It remains to show (38). Define $\overline{\tau}^{(0)}$ and $\underline{\tau}^{(u)}$ as the stopping times that the process $\overline{\mathcal{Z}}$ reaches 0 and the process $\underline{\mathcal{Z}}$ reaches the upper bound $N^{b+\delta}$, respectively, i.e.

$$\overline{\tau}^{(0)} = \inf\{g \geq 0 : \overline{Z}_g = 0\}, \qquad \underline{\tau}^{(u)} = \inf\{g \geq 0 : \underline{Z}_g \geq N^{b+\delta}\},$$

with the convention that the infimum over an empty set is infinity. Then

$$\begin{aligned}
\mathbb{P}(\tau_1 \geq \lceil N^{b+\delta'}\rceil) &\leq \mathbb{P}(\tau_1 \geq \lceil N^{b+\delta'}\rceil, E) + \mathbb{P}(E^c) \\
&\leq \mathbb{P}(\underline{\tau}^{(u)} \geq \lceil N^{b+\delta'}\rceil, \overline{\tau}^{(0)} \geq \lceil N^{b+\delta'}\rceil, E) + \mathbb{P}(E^c) \\
&= \mathbb{P}(\underline{\tau}^{(u)} \geq \lceil N^{b+\delta'}\rceil, \overline{\tau}^{(0)} \geq \lceil N^{b+\delta'}\rceil, E, \underline{\mathcal{Z}} \text{ dies out }) \\
&\quad + \mathbb{P}(\underline{\tau}^{(u)} \geq \lceil N^{b+\delta'}\rceil, \overline{\tau}^{(0)} \geq \lceil N^{b+\delta'}\rceil, E, \underline{\mathcal{Z}} \text{ survives }) + \mathbb{P}(E^c) \\
&= \mathbb{P}(\underline{\tau}^{(u)} \geq \lceil N^{b+\delta'}\rceil, \overline{\tau}^{(0)} \geq \lceil N^{b+\delta'}\rceil, E, \underline{\mathcal{Z}} \text{ dies out }) \\
&\quad + O(e^{-c'N^{\frac{\delta'}{2}}}) + O(N^{b+\delta'}e^{-\frac{1}{2}N^{1-2\alpha}}),
\end{aligned} \tag{39}$$

by an application of Lemma 1 and Corollary 1 and $\alpha < \frac{1}{2}$ as defined there. To keep the notation simple we denote by e_N terms of the order $\exp(-N^c)$ for some $c > 0$. Proceeding with (39) we obtain

$$\begin{aligned}
(39) &= \mathbb{P}(\underline{\tau}^{(u)} \geq \lceil N^{b+\delta'}\rceil, \overline{\tau}^{(0)} \geq \lceil N^{b+\delta'}\rceil, E, \underline{\mathcal{Z}} \text{ dies out }, \overline{\mathcal{Z}} \text{ survives}) \\
&\quad + \mathbb{P}(\underline{\tau}^{(u)} \geq \lceil N^{b+\delta'}\rceil, \overline{\tau}^{(0)} \geq \lceil N^{b+\delta'}\rceil, E, \underline{\mathcal{Z}} \text{ dies out }, \overline{\mathcal{Z}} \text{ dies out}) + e_N \\
&\leq \mathbb{P}(\underline{\mathcal{Z}} \text{ dies out }, \overline{\mathcal{Z}} \text{ survives}, E) + e_N,
\end{aligned} \tag{40}$$

again by an application of Lemma 1. Note that

$$\begin{aligned}
\mathbb{P}(\underline{\mathcal{Z}} \text{ dies out }, \overline{\mathcal{Z}} \text{ survives}, E) &+ \mathbb{P}(\underline{\mathcal{Z}} \text{ survives }, \overline{\mathcal{Z}} \text{ survives}, E) \\
&= \mathbb{P}(\overline{\mathcal{Z}} \text{ survives}, E) = \frac{2s_N}{\rho^2}(1 + o(1)).
\end{aligned}$$

In order to show that (40) is $o(s_N)$ it suffices to prove that

$$\mathbb{P}(\underline{\mathcal{Z}} \text{ survives }, \overline{\mathcal{Z}} \text{ survives}, E) = \frac{2s_N}{\rho^2}(1 + o(1)).$$

Considering again the event $\{\underline{\tau}^{(u)} \leq \lceil N^{b+\delta'}\rceil\}$ and applying (18) one obtains

$$\begin{aligned}
\mathbb{P}(\underline{\mathcal{Z}} \text{ survives }, \overline{\mathcal{Z}} \text{ survives}, E) &= \mathbb{P}(\underline{\mathcal{Z}} \text{ survives }, \overline{\mathcal{Z}} \text{ survives}, E, \underline{\tau}^{(u)} \leq \lceil N^{b+\delta'}\rceil) + e_N \\
&= \mathbb{P}(\underline{\mathcal{Z}} \text{ survives }, E, \underline{\tau}^{(u)} \leq \lceil N^{b+\delta'}\rceil) + e_N,
\end{aligned}$$

since the events E and $\{\underline{\tau}^{(u)} \leq \lceil N^{b+\delta'}\rceil)\}$ imply that $\overline{Z}_g \geq N^{b+\delta}$ for some $g \leq N^{b+\delta'}$ and the probability for $\overline{\mathcal{Z}}$ to die out after reaching $N^{b+\delta}$ is $(1 - \frac{2s_N}{\rho^2}(1+o(1)))^{N^{b+\delta}} = e_N$.

One more application of (18) yields

$$\mathbb{P}(\underline{\mathcal{Z}} \text{ survives }, E, \underline{\tau}^{(u)} \leq \lceil N^{b+\delta'} \rceil) = \frac{2s_N}{\rho^2}(1 + o(1)),$$

which finishes the proof. □

5.2 Second phase: from $N^{b+\delta}$ to εN

In this section we show that $\mathcal{X}$, once having reached the level $N^{b+\delta}$, will reach the level εN with probability tending to 1 as $N \to \infty$.

Lemma 6 (From $N^{b+\delta}$ to εN with high probability) *Assume $X_0 \geq N^{b+\delta}$ with $0 < \delta < \eta$, let $0 < \varepsilon < \frac{\delta}{2-2\eta-\delta}$ and define the stopping time*

$$\tau_2 = \inf\{g \geq 0 : X_g \notin \{1, 2, \ldots, \lfloor \varepsilon N \rfloor\}\}.$$

Then there exists some $\delta' > 0$ such that

$$\mathbb{P}\left(X_{\tau_2} \geq \varepsilon N\right) = 1 - O(N^{-\delta'}).$$

Proof By monotonicity it is enough to prove the claim for $X_0 = \lceil N^{b+\delta} \rceil$. By definition we have

$$\mathscr{L}(X_{g+1}|X_g) = \text{Bin}\left(N, \frac{\sum_{i=1}^{X_g} W_i}{\sum_{i=1}^{X_g} W_i + (1 - s_N)\sum_{i=X_g+1}^{N} W_i}\right). \qquad (41)$$

Next we lower-bound $\mathcal{X}$ by the process $\widetilde{\mathcal{X}} = (\widetilde{X}_g)_{g \geq 0}$, $\widetilde{X}_0 = X_0$, with conditional distribution

$$\mathscr{L}(\widetilde{X}_{g+1}|\widetilde{X}_g) = \text{Bin}\left(N, \frac{\sum_{i=1}^{\widetilde{X}_g} W_i}{1 - s_N \sum_{i=\varepsilon N+1}^{N} W_i}\right) \qquad (42)$$

as long as $\widetilde{X}_g \leq \varepsilon N$. If $\widetilde{X}_g > \varepsilon N$ we assume that $\widetilde{X}_{g+1}$ is distributed as a slightly supercritical branching process with $\text{Pois}(Y_1 q_N)$ distributed offspring, where

$$q_N = N\mathbb{E}\left[\frac{W_1}{1 - s_N \sum_{i=\varepsilon N+1}^{N} W_i}\right]. \qquad (43)$$

We will see that by this definition in each generation the expectation of $\widetilde{\mathcal{X}}$ increases by the factor q_N, see (45) and (46). The generation-wise increase of the variance conditioned on the current state can be estimated from above by a factor $\rho^2(1 + o(1))$, see (45) and (47), leading to an iterative estimate on the variance of the form (48).

As long as $X_g \geq \widetilde{X}_g$, the success probability in the mixed Binomial distribution on the r.h.s. of (41) dominates the corresponding one on the r.h.s. of (42). Consequently, starting $\widetilde{X}$ and X both in $\lfloor N^{b+\delta} \rfloor$ we can couple them, such that $\widetilde{X}_g \leq X_g$ as long as $\widetilde{X}$ did not cross the level εN. In particular, we have for the stopping time $\widetilde{\tau} = \inf\{g \geq 0 : \widetilde{X}_g \notin \{1, 2, \dots, \lfloor \varepsilon N \rfloor\}\}$

$$\mathbb{P}(X_\tau \geq \varepsilon N) \geq \mathbb{P}(\widetilde{X}_{\widetilde{\tau}} \geq \varepsilon N). \tag{44}$$

To show $\mathbb{P}(\widetilde{X}_{\widetilde{\tau}} \geq \varepsilon N) = 1 - O(N^{-\delta'})$ we will estimate the first and second moment of $\widetilde{X}_{g_0}$ for a suitably chosen $g_0 \in \mathbb{N}$ and then use Chebyshev's inequality to show that $\widetilde{X}_{g_0}$ is above εN with sufficiently high probability. For this purpose we consider $m(x)$ and $v(x)$, the one-step conditional expectation and variance of $\widetilde{X}$ at $x \in \mathbb{N}$, that is

$$m(x) = \mathbb{E}\left[\widetilde{X}_1 | \widetilde{X}_0 = x\right], \qquad v(x) = \text{Var}\left(\widetilde{X}_1 | \widetilde{X}_0 = x\right).$$

From the definition of $\widetilde{\mathcal{X}}$ as a branching process above εN we have for $x > \varepsilon N$

$$m(x) = q_N x, \qquad v(x) = \rho^2 x(1 + o(1)). \tag{45}$$

Next we show that $m(x)$ and $v(x)$ fulfil relations similar to (45) also for $x \leq \varepsilon N$, which will allow to estimate the expectation and the variance of $\widetilde{X}_{g_0}$.

For $x \leq \varepsilon N$ we have due to (42)

$$\begin{aligned}
m(x) &= N\mathbb{E}\left[\frac{\sum_{i=1}^x W_i}{1 - s_N \sum_{i=\varepsilon N+1}^N W_i}\right] \\
&= N\mathbb{E}\left[\sum_{i=1}^x W_i \left(1 + s_N \sum_{i=\varepsilon N+1}^N W_i + O(s_N^2)\right)\right] \\
&= x(1 + s_N(1 - \varepsilon)N^2 \mathbb{E}\left[W_1 W_{\varepsilon N+1}\right] + O(s_N^2)) \\
&= x\left(1 + s_N(1 - \varepsilon)(1 + O(N^{-1})) + O(s_N^2)\right) \\
&= x\left(1 + (1 - \varepsilon)s_N + O(s_N^2)\right).
\end{aligned}$$

In the penultimate equality we used $\mathbb{E}[W_1 W_2] = \frac{1}{N^2} + O(N^{-3})$ which results from the fact

$$1 = \mathbb{E}\left[\left(\sum_{i=1}^N W_i\right)^2\right] = N\mathbb{E}\left[W_1^2\right] + N(N-1)\mathbb{E}[W_1 W_2]$$

and (11). Consequently, we have for all $x \in \mathbb{N}$, recalling (43),

$$m(x) = x q_N = x\left(1 + (1 - \varepsilon)s_N + O(s_N^2)\right). \tag{46}$$

Next we analyse $v(x)$, again for $x \leq \varepsilon N$. In view of (42), a decomposition of the variance gives

$$
v(x) = \mathrm{Var}\left(N \frac{\sum_{i=1}^{x} W_i}{1 - s_N \sum_{i=\varepsilon N+1}^{N} W_i} \right)
$$

$$
+ \mathbb{E}\left[N \frac{\sum_{i=1}^{x} W_i}{1 - s_N \sum_{i=\varepsilon N+1}^{N} W_i} \left(1 - \frac{\sum_{i=1}^{x} W_i}{1 - s_N \sum_{i=\varepsilon N+1}^{N} W_i} \right) \right]
$$

$$
\leq \mathbb{E}\left[\left(N \frac{\sum_{i=1}^{x} W_i}{1 - s_N \sum_{i=\varepsilon N+1}^{N} W_i} \right)^2 \right] - \mathbb{E}\left[N \frac{\sum_{i=1}^{x} W_i}{1 - s_N \sum_{i=\varepsilon N+1}^{N} W_i} \right]^2
$$

$$
+ \mathbb{E}\left[N \frac{\sum_{i=1}^{x} W_i}{1 - s_N} \right]
$$

$$
\leq \mathbb{E}\left[\left(N \frac{\sum_{i=1}^{x} W_i}{1 - s_N} \right)^2 \right] - \mathbb{E}\left[N \sum_{i=1}^{x} W_i \right]^2 + \mathbb{E}\left[N \frac{\sum_{i=1}^{x} W_i}{1 - s_N} \right].
$$

Because of the negative correlation of the W_i, the sum of the first and the second term is not larger than $x N^2 \mathrm{Var}\,(W_1)$, which because of (11) is $\leq x(\rho^2 - 1) + O(N^{-1})$. Since the third term is $x(1 + O(s_N))$, we have for all $x \leq \varepsilon N$

$$
v(x) \leq \rho^2 x (1 + o(1)). \tag{47}
$$

Combining (46) and (47) allows us to estimate the variance $\mathrm{Var}\,(\widetilde{X}_g)$ for $g \in \mathbb{N}$, again by decomposing the variance:

$$
\begin{aligned}
\mathrm{Var}\,(\widetilde{X}_g) &= \mathrm{Var}\,\left(\mathbb{E}\left[\widetilde{X}_g | \widetilde{X}_{g-1} \right] \right) + \mathbb{E}\left[\mathrm{Var}\,(\widetilde{X}_g | \widetilde{X}_{g-1}) \right] \\
&= \mathrm{Var}\,\left(m(\widetilde{X}_{g-1}) \right) + \mathbb{E}\left[v(\widetilde{X}_{g-1}) \right] \\
&\leq q_N^2 \mathrm{Var}\,(\widetilde{X}_{g-1}) + \rho^2 \mathbb{E}\left[\widetilde{X}_{g-1} \right] (1 + o(1)) \\
&= q_N^2 \mathrm{Var}\,(\widetilde{X}_{g-1}) + \rho^2 q_N^{g-1} \widetilde{X}_0 (1 + o(1)).
\end{aligned} \tag{48}
$$

Iterating this argument yields

$$
\begin{aligned}
\mathrm{Var}\,(\widetilde{X}_g) &= \rho^2 \widetilde{X}_0 q_N^{g-1} \sum_{j=0}^{g-1} q_N^j (1 + o(1)) \\
&= \rho^2 \widetilde{X}_0 q_N^{g-1} \frac{q_N^g - 1}{q_N - 1} (1 + o(1)).
\end{aligned}
$$

Choose the minimal $g_0 \in \mathbb{N}$ such that $2\varepsilon N \leq \mathbb{E}\left[\widetilde{X}_{g_0}\right] = q_N^{g_0} X_0$, which yields recalling the initial condition $X_0 = \lceil N^{b+\delta} \rceil$

$$g_0 = \left\lceil \frac{\log(2\varepsilon N X_0^{-1})}{\log q_N} \right\rceil = \left\lceil \frac{\log(2\varepsilon N^{1-b-\delta})}{(1-\varepsilon)s_N + O(s_N^2)} \right\rceil.$$

Applying Chebyshev's inequality with $\widetilde{X}_0 = X_0$, we obtain

$$\mathbb{P}\left(|\widetilde{X}_{g_0} - \mathbb{E}\left[\widetilde{X}_{g_0}\right]| \geq \varepsilon N\right) \leq \frac{\rho^2 \widetilde{X}_0 q_N^{g_0-1} \frac{q_N^{g_0}-1}{q_N-1}(1+o(1))}{\varepsilon^2 N^2}$$

$$\leq \frac{\rho^2 N^{b+\delta} q_N^{2g_0} \frac{N^b}{(1-\varepsilon)}(1+o(1))}{\varepsilon^2 N^2}$$

$$= \frac{\rho^2}{\varepsilon^2(1-\varepsilon)} N^{2b+\delta-2}(1 + (1-\varepsilon)s_N + O(s_N^2))^{2g_0}(1+o(1))$$

$$\leq c_{\rho,\varepsilon} N^{2b+\delta-2} \exp(2g_0 s_N(1 + O(s_N)))(1+o(1))$$

$$\leq c_{\rho,\varepsilon} N^{2b+\delta-2} N^{\frac{2}{1-\varepsilon}(1-b-\delta)}(1+o(1))$$

$$= O(N^{-\delta'}),$$

for some small $\delta' > 0$ due to the assumptions on ε.

Since $\mathbb{E}\left[\widetilde{X}_{g_0}\right] \geq 2\varepsilon N$, this implies

$$\mathbb{P}\left(\widetilde{X}_{\widetilde{\tau}} \geq \varepsilon N\right) \geq \mathbb{P}(\widetilde{X}_{g_0} \geq \varepsilon N) \geq 1 - O(N^{-\delta'})$$

and due to (44) this finishes the proof. □

5.3 Third phase: from εN to N

Lemma 7 below concerns the last step of the proof, showing that once the process $\mathcal{X}$ has reached the level $\lfloor \varepsilon N \rfloor$, it goes to fixation with high probability. Our proof relies on a representation of the fixation probability of $\mathcal{X}$ in terms of (a functional of) the equilibrium state $A_{\text{eq}} := A_{\text{eq}}^{(N)}$ of the counting process $\mathcal{A} := \mathcal{A}^{(N)} = (A_m)_{m \geq 0}$ of the *potential ancestors* in the time discrete Cannings ancestral selection graph as provided by Boenkost et al. (2021). The process $\mathcal{A}^{(N)}$ is called *Cannings ancestral selection process (CASP)* in Boenkost et al. (2021); for fixed N, it is a recurrent, $[N]$-valued Markov chain whose transition probabilities are specified in Boenkost et al. (2021) Section 2.3.

Theorem 3.1 and Formula (3.2) (see also Corollary 3.3) in Boenkost et al. (2021) provide the following sampling duality representation of the fixation probability of $\mathcal{X}$

when started with k individuals:

$$\mathbb{P}_k(\mathcal{X} \text{ eventually hits } N) = 1 - \mathbb{E}\left[\frac{(N-k)(N-k-1)\cdots(N-k-A_{\text{eq}}+1)}{N(N-1)\cdots(N-A_{\text{eq}}+1)}\right]. \tag{49}$$

Intuitively, this says that $\mathcal{X}$ goes extinct if and only if a random sample of (random) size A_{eq}, drawn without replacement from the population of size N, avoids the k beneficial individuals.

Formula (49) implies

$$\mathbb{P}_{\lceil \varepsilon N \rceil}(\mathcal{X} \text{ eventually hits } N) \geq 1 - \mathbb{E}\left[(1-\varepsilon)^{A_{\text{eq}}^{(N)}}\right]. \tag{50}$$

The representation of the transition probabilities of $\mathcal{A}$ in Boenkost et al. (2021) Section 2.3 in terms of two half steps yields that for fixed N CASPs with different selection parameters can be coupled in such a way that $A_{\text{eq}}^{(N)}$ is increasing in s_N. Take a sequence $(\tilde{s}_N)$ satisfying $\tilde{s}_N \leq s_N$ and Condition (1.2) in Boenkost et al. (2021), i.e.

$$N^{-1+\eta} \leq \tilde{s}_N \leq N^{-2/3+\eta}.$$

Let $\tilde{A}_{\text{eq}}^{(N)}$ be the equilibrium state belonging to $\tilde{s}_N$ (and to the same Dirichlet-type paintbox as that of $\mathcal{X}$). The central limit result Boenkost et al. (2021), Corollary 6.10, implies that $\tilde{A}_{\text{eq}}^{(N)} \to \infty$ in probability as $N \to \infty$. Because of the just mentioned monotonicity in the selection coefficient, the same convergence holds true for the sequence $\left(A_{\text{eq}}^{(N)}\right)$. The following lemma is thus immediate from (50) and dominated convergence:

Lemma 7 (From εN to N with high probability) *Let $\mathcal{X}$ be a Cannings frequency process with $X_0 = k \geq \varepsilon N$ for some $0 < \varepsilon < 1/2$. Assume that Conditions (7), (8) and (12) are fulfilled. Define $\tau_3 := \inf\{g \geq 0 : X_g \in \{0, N\}\}$. Then*

$$\mathbb{P}_k(X_{\tau_3} = N) = 1 - o(1).$$

6 Discussion

The analysis of fixation probabilities of slightly beneficial mutants is at the heart of population genetics; some seminal and more modern references are given in the Introduction. Our main result concerns Haldane's asymptotics (3) for the fixation probability in a regime of moderate selection.

Our framework is that of Cannings models with selection (as reviewed in Sect. 2), where the corresponding neutral genealogies are assumed to belong to the domain of attraction of Kingman's coalescent. This class of models is motivated by seasonal reproduction cycles in which within each season a large number of offspring is generated but only a comparatively small number (concentrated around a carrying capacity

N) of randomly sampled offspring survive to the next season. In this setting it is reasonable to approximate sampling without replacement by sampling with replacement. Thus, under the assumption of neutrality, the probability that the j-th offspring that survives till the next generation is a child of parent i is approximately given by the random weight

$$W_i = \frac{Y_i}{\sum_{\ell=1}^{N} Y_\ell},$$

where $Y_1, \ldots, Y_N$ are the sizes of (potential one-generation) offspring of parents $1, \ldots, N$. These sizes are assumed to be independent and identically distributed in the present paper, leading to the concept of weights *of Dirichlet type*. The subsequent generation then arises by a multinomial sampling with random weights, and to add selection the weights of wildtype parents are decreased by the factor $(1 - s_N)$. For a closely related model with a specific distribution of Y_i (and sampling without replacement) in the context of Lenski's long-term evolution experiment see González Casanova et al. (2017) and Baake et al. (2019).

We prove Haldane's asymptotics in the case of *moderately strong selection*, see Theorem 1, in which the selection strength s_N obeys

$$N^{-\frac{1}{2}+\eta} \leq s_N \leq N^{-\eta}$$

for some $\eta > 0$ and a large population size N. In the companion paper Boenkost et al. (2021) the range of *moderately weak selection* was considered, i.e. in the case

$$N^{-1+\eta} \leq s_N \leq N^{-\frac{1}{2}-\eta}$$

for some $\eta > 0$. Since $s_N \gg N^{-1}$ in the regime of moderate selection, selection acts in this case on a faster timescale than genetic drift.

In Boenkost et al. (2021) an ancestral selection graph for the just described class of Cannings models with selection was defined, and it was shown that the fixation probability π_N is equal to the expeced value $\mathbb{E}\left[\frac{A_{eq}^{(N)}}{N}\right]$, where $A_{eq}^{(N)}$ is the number of lines of the ancestral selection graph in its equilibrium. While we could analyse directly the asymptotics of that quantity in the regime of moderately weak selection, we were facing too large fluctuations of A_{eq}^{N} in the regime of moderately strong selection in order to be successful with this approch. Conversely, it turned out that the classical idea of branching process approximation is suitable precisely in that latter regime.

For highly skewed offspring distributions an asymptotics for the fixation probability arises which is different from (3). In cases where the neutral genealogy is attracted by a Beta$(2 - \alpha, \alpha)$-coalescent, Okada and Hallatschek (2021) argue that the fixation probability is proportionally to $s_N^{\frac{1}{\alpha-1}}$, if $1 \gg s_N \gg N^{-(\alpha-1)}$. Thus the probability of fixation is substantially smaller than in Haldane's asymptotics, which is reasonable since the offspring variance is diverging as $N \to \infty$. Notably, since the evolutionary timescale of Cannings models in the domain of attraction of Beta-coalescents is of

the order $N^{\alpha-1}$, the case $1 \gg s_N \gg N^{-(\alpha-1)}$ corresponds to the regime of moderate selection; note also that the case of coalescents being in the domain of attraction of a Kingman coalescent corresponds formally to $\alpha = 2$.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

References

Athreya K (1992) Rates of decay for the survival probability of a mutant gene. J Math Biol 30(6):577–581

Athreya K, Ney P (1972) Branching processes. Grundlehren Math. Wiss., Band, vol 196. Springer, New York

Baake E, González Casanova A, Probst S, Wakolbinger A (2019) Modelling and simulating Lenski's long-term evolution experiment. Theor Popul Biol 127:58–74

Boenkost F, González Casanova A, Pokalyuk C, Wakolbinger A (2021) Haldane's formula in Cannings models: the case of moderately weak selection. Electron J Probab 26(4):1–36

Cannings C (1974) The latent roots of certain Markov chains arising in genetics: a new approach, I. Haploid models. Adv Appl Probab 6(2):260–290

Cramér H (1938) Sur un nouveau théoreme-limite de la théorie des probabilités. Actual Sci Ind 736:5–23. English translation: Cramér H, Touchette H (2018) On a new limit theorem in probability theory. Preprint: arXiv:1802.05988v3

Dembo A, Zeitouni O (1994) Large deviations techniques and applications, 2nd edn. Springer, New York

Eshel I (1981) On the survival probability of a slightly advantageous mutant gene with a general distribution of progeny size—a branching process model. J Math Biol 12(3):355–362

Ewens W (2004) Mathematical population genetics: I. Theoretical introduction, 2nd edn. Springer, New York

Fisher R (1922) On the dominance ratio. Proc R Soc Edinb 42:321–341

González Casanova A, Spanó D (2018) Duality and fixation in Ξ-Wright–Fisher processes with frequency-dependent selection. Ann Appl Probab 28:250–284

González Casanova A, Kurt N, Wakolbinger A, Yuan L (2017) An individual-based model for the Lenski experiment, and the deceleration of the relative fitness. Stoch Process Appl 126:2211–2252

Haccou P, Jagers P, Vatutin VA (2005) Branching processes: variation, growth, and extinction of populations, 5th edn. Cambridge University Press, Cambridge

Haldane J (1927) A mathematical theory of natural and artificial selection 23:607–615

Hoeffding W (1963) Probability inequalities for sums of bounded random variables. J Am Stat Assoc 58:13–30

Hoppe FM (1992) Asymptotic rates of growth of the extinction probability of a mutant gene. J Math Biol 30(6):547–566

Huillet T, Möhle M (2021) Asymptotic genealogies for a class of generalized Wright–Fisher models, preprint: arXiv:2106.10939v1

Klenke A, Mattner L (2010) Stochastic ordering of classical discrete distributions. Adv Appl Probab 42(2):392–410

Kolmogorov A (1938) Zur Lösung einer biologischen Aufgabe. Comm Math Mech Chebyshev Univ Tomsk 2(1):1–12. English translation: Kolmogorov AN (1992) Selected works of A. N. Kolmogorov. Vol II. Probability theory and mathematical statistics, Chapter 25. Kluwer, Dordrecht

Lessard S, Ladret V (2007) The probability of fixation of a single mutant in an exchangeable selection model. J Math Biol 54(5):721–744

Lyons R, Peres Y (2017) Probability on trees and networks, vol 42. Cambridge University Press, Cambridge

Mikosch T, Nagaev AV (1998) Large deviations of heavy-tailed sums with applications in insurance. Extremes 1(1):81–110

Okada T, Hallatschek O (2021) Dynamic sampling bias and overdispersion induced by skewed offspring distributions. Genetics, iyab135

Patwa Z, Wahl L (2008) The fixation probability of beneficial mutations. J R Soc Interface 5(28):1279–1289

Wright S (1931) Evolution in Mendelian populations. Genetics 16(2):97–159