Supplementary material

In this supplementary material, we provide additional technical details regarding the objective function (1), as well as proofs for the propositions presented in the main article.

Technical details for the objective function (1)

This appendix provides technical details for the objective function that plays a central role in our analysis—Equation (1). While these results have already been formally established in prior work [16, 17, 21, 18], we reproduce them here for convenience and completeness.

The following proposition shows that, the exponential growth rate of the population size of Andorians, $T^{-1}\log n_T$, converges in probability to the log-geometric-average growth rate given by Equation (1).

Proposition A.1. *Under the setup in "The formal model" section, we have*

$$\frac{1}{T}\log n_T \xrightarrow{\text{p}} \mu(p) = \mathbb{E}\left[\log\left(p x_{\text{dis}} + (1-p)x_{\text{nodis}}\right)\right]$$

as T and n_T increase without bound, where "$\xrightarrow{\text{p}}$" represents convergence in probability.

Proposition A.1 illustrates that, $\mu(p)$ given by Equation (1) is the long-run limit of the growth rate of the population size for Andorians with a probability of discrimination p.

The following proposition further demonstrates that, Andorians with a probability of discrimination p that maximize Equation (1) will, over time, dominate the population.

Proposition A.2. *Under the setup in "The formal model" section, let $p, p' \in [0,1]$ such that $\mu(p) < \mu(p')$. Furthermore, let n_T^p and $n_T^{p'}$ denote the number of Andorians in generation T with probabilities of discrimination p and p', respectively. Then, we have*

$$\left(\frac{n_T^p}{n_T^{p'}}\right)^{1/T} \xrightarrow{\text{p}} \exp\left(\mu(p) - \mu(p')\right) < 1$$

as T and n_T increase without bound, where "$\xrightarrow{\text{p}}$" represents convergence in probability. Furthermore, this implies that

$$\frac{n_T^p}{n_T^{p'}} \xrightarrow{\text{p}} 0$$

at an exponential rate.

Proposition A.2 illustrates that, maximizing Equation (1) leads to a "winner-take-all" outcome, as individuals who do not maximize Equation (1) will be rapidly overrun by those who do. This result provides the theoretical foundation for our adoption of the log-geometric-average growth rate as the objective function in our analysis.

Proofs

In this appendix, we provide proofs for the propositions stated in the previous appendix and in the main article.

Proof of Proposition A.1. The number of offspring in generation t is the sum of all offspring produced by individuals in the previous generation:

$$n_t = \sum_{i=1}^{n_{t-1}} x_{i,t}^p = \left(\sum_{i=1}^{n_{t-1}} I_{i,t}^p \right) x_{\text{dis},t} + \left(\sum_{i=1}^{n_{t-1}} \left(1 - I_{i,t}^p\right) \right) x_{\text{nodis},t},$$

where $I_{i,t}^p$ follow a Bernoulli distribution with parameter p and are independent over time t and across individuals i. Using the Law of Large Numbers, we have

$$\frac{\sum_{i=1}^{n_{t-1}} I_{i,t}^p}{n_{t-1}} \xrightarrow{\text{p}} p, \quad \frac{\sum_{i=1}^{n_{t-1}} \left(1 - I_{i,t}^p\right)}{n_{t-1}} \xrightarrow{\text{p}} 1 - p.$$

Under our setup in "The formal model" section, both $x_{\text{dis},t}$ and $x_{\text{nodis},t}$ are bounded. Therefore,

$$\begin{aligned} &\frac{n_t}{n_{t-1}} - (p x_{\text{dis},t} + (1-p) x_{\text{nodis},t}) \\ &\qquad = \left(\frac{\sum_{i=1}^{n_{t-1}} I_{i,t}^p}{n_{t-1}} - p \right) x_{\text{dis},t} + \left(\frac{\sum_{i=1}^{n_{t-1}} \left(1 - I_{i,t}^p\right)}{n_{t-1}} - (1-p) \right) x_{\text{nodis},t} \xrightarrow{\text{p}} 0. \end{aligned}$$

Furthermore, under our setup, we have

$$L \leq \frac{n_t}{n_{t-1}} \leq U, \quad L \leq p x_{\text{dis},t} + (1-p) x_{\text{nodis},t} \leq U, \tag{A.1}$$

where

$$L = \min\left\{\lambda_A^{\text{low}}, \lambda_T^{\text{low}}\right\} > 0, \quad U = \max\left\{\lambda_A^{\text{high}}, \lambda_T^{\text{high}}\right\} > 0.$$

Thus, by Lagrange's mean value theorem,

$$\log\left(\frac{n_t}{n_{t-1}}\right) - \log\left(p x_{\text{dis},t} + (1-p) x_{\text{nodis},t}\right) = \left[\frac{n_t}{n_{t-1}} - (p x_{\text{dis},t} + (1-p) x_{\text{nodis},t})\right] \cdot \frac{1}{\xi} \xrightarrow{\text{p}} 0$$

with a random variable $\xi \in [L, U]$. Averaging over $t = 1, 2, \ldots, T$, we obtain

$$\frac{1}{T}\left[\sum_{t=1}^{T}\log\left(\frac{n_t}{n_{t-1}}\right) - \sum_{t=1}^{T}\log\left(px_{\text{dis},t} + (1-p)x_{\text{nodis},t}\right)\right]$$
$$= \frac{1}{T}\log n_T - \frac{\sum_{t=1}^{T}\log\left(px_{\text{dis},t} + (1-p)x_{\text{nodis},t}\right)}{T} \xrightarrow{\text{p}} 0.$$

Again, by the Law of Large Numbers,

$$\frac{\sum_{t=1}^{T}\log\left(px_{\text{dis},t} + (1-p)x_{\text{nodis},t}\right)}{T} \xrightarrow{\text{p}} \mu(p) = \mathbb{E}\left[\log\left(px_{\text{dis}} + (1-p)x_{\text{nodis}}\right)\right].$$

Therefore,

$$\frac{1}{T}\log n_T \xrightarrow{\text{p}} \mu(p) = \mathbb{E}\left[\log\left(px_{\text{dis}} + (1-p)x_{\text{nodis}}\right)\right],$$

which completes the proof. □

Proof of Proposition A.2. Using Proposition A.1, we have

$$\frac{1}{T}\log n_T^{p} \xrightarrow{\text{p}} \mu(p), \quad \frac{1}{T}\log n_T^{p'} \xrightarrow{\text{p}} \mu(p').$$

Therefore,

$$\frac{1}{T}\log n_T^{p} - \frac{1}{T}\log n_T^{p'} \xrightarrow{\text{p}} \mu(p) - \mu(p').$$

Taking exponentials on both sides yields the result. □

Proof of Proposition 1. We maximize Equation (1) over p. Direct calculation shows that

$$\frac{\partial \mu}{\partial p}(p) = \mathbb{E}\left[\frac{x_{\text{dis}} - x_{\text{nodis}}}{px_{\text{dis}} + (1-p)x_{\text{nodis}}}\right] \tag{A.2}$$

and

$$\frac{\partial^2 \mu}{\partial p^2}(p) = -\mathbb{E}\left[\frac{(x_{\text{dis}} - x_{\text{nodis}})^2}{(px_{\text{dis}} + (1-p)x_{\text{nodis}})^2}\right] < 0.$$

This implies that Equation (1) is concave with respect to p. From Equation (A.2), we have

$$\frac{\partial \mu}{\partial p}(0) = \mathbb{E}\left[\frac{x_{\text{dis}}}{x_{\text{nodis}}}\right] - 1, \qquad \frac{\partial \mu}{\partial p}(1) = 1 - \mathbb{E}\left[\frac{x_{\text{nodis}}}{x_{\text{dis}}}\right].$$

Therefore, if both $\frac{\partial \mu}{\partial p}(0) > 0$ and $\frac{\partial \mu}{\partial p}(1) > 0$, Equation (1) is increasing with respect to p and the optimal solution is $p^* = 1$; if both $\frac{\partial \mu}{\partial p}(0) < 0$ and $\frac{\partial \mu}{\partial p}(1) < 0$, Equation (1) is decreasing with respect to p and the optimal solution is $p^* = 0$. Finally, if both $\frac{\partial \mu}{\partial p}(0) \geq 0$ and $\frac{\partial \mu}{\partial p}(1) \leq 0$, the optimal p^* is reached when Equation (A.2) is equal to zero. This completes the proof. $\square$

Proof of Proposition 2. By direct calculation, the growth rate in Equation (1) is

$$\begin{aligned}\mu(p) =& \mathbb{E}\left[\log\left(px_{\text{dis}} + (1-p)x_{\text{nodis}}\right)\right]\\ =& \mathbb{E}\left[\log\left[p(\beta_{\text{dis}} - \beta_{\text{nodis}})(\lambda_T - \lambda_A) + \beta_{\text{nodis}}\lambda_T + (1-\beta_{\text{nodis}})\lambda_A\right]\right]\\ =& qr\log\lambda^{\text{low}} + (1-q)(1-r)\log\lambda^{\text{high}}\\ &+ q(1-r)\log\left[(\beta_{\text{dis}} - \beta_{\text{nodis}})(\lambda^{\text{high}} - \lambda^{\text{low}})p + \beta_{\text{nodis}}\lambda^{\text{high}} + (1-\beta_{\text{nodis}})\lambda^{\text{low}}\right]\\ &+ (1-q)r\log\left[(\beta_{\text{dis}} - \beta_{\text{nodis}})(\lambda^{\text{low}} - \lambda^{\text{high}})p + \beta_{\text{nodis}}\lambda^{\text{low}} + (1-\beta_{\text{nodis}})\lambda^{\text{high}}\right].\end{aligned}$$

Thus,

$$\begin{aligned}\frac{\partial\mu}{\partial p}(p) =& q(1-r)\frac{(\beta_{\text{dis}} - \beta_{\text{nodis}})(\lambda^{\text{high}} - \lambda^{\text{low}})}{p(\beta_{\text{dis}} - \beta_{\text{nodis}})(\lambda^{\text{high}} - \lambda^{\text{low}}) + \beta_{\text{nodis}}\lambda^{\text{high}} + (1-\beta_{\text{nodis}})\lambda^{\text{low}}}\\ &+ (1-q)r\frac{(\beta_{\text{dis}} - \beta_{\text{nodis}})(\lambda^{\text{low}} - \lambda^{\text{high}})}{p(\beta_{\text{dis}} - \beta_{\text{nodis}})(\lambda^{\text{low}} - \lambda^{\text{high}}) + \beta_{\text{nodis}}\lambda^{\text{low}} + (1-\beta_{\text{nodis}})\lambda^{\text{high}}}.\end{aligned}$$

According to the proof of Proposition 1, we have $\frac{\partial^2\mu}{\partial p^2}(p) < 0$, implying that $\frac{\partial\mu}{\partial p}(p)$ is decreasing with respect to p. Therefore, if $\frac{\partial\mu}{\partial p}(1) > 0$, we have $\frac{\partial\mu}{\partial p}(p) > 0$ and the optimal solution is $p^* = 1$; if $\frac{\partial\mu}{\partial p}(0) < 0$, we have $\frac{\partial\mu}{\partial p}(p) < 0$ and the optimal solution is $p^* = 0$. One can directly verify that $\frac{\partial\mu}{\partial p}(1) > 0$ is equivalent to $r > r_{\text{upper}}$, and $\frac{\partial\mu}{\partial p}(0) < 0$ is equivalent to $r < r_{\text{lower}}$. Finally, $\frac{\partial\mu}{\partial p}(0) \geq 0$ and $\frac{\partial\mu}{\partial p}(1) \leq 0$ are equivalent to $r_{\text{lower}} \leq r \leq r_{\text{upper}}$, and in this case, the optimal solution is reached when $\frac{\partial\mu}{\partial p}(p^*) = 0$, which is equivalent to $p^*(q,r) = p^{*\text{partial}}(q,r)$. This completes the proof. $\square$

Proof of Proposition 3. One can directly prove this result by, for example, calculating the partial derivatives of $p^{*\text{partial}}(q,r)$ with respect to r and q, respectively. $\square$

Proof of Proposition 4. Using Bayes' formula, the posterior probabilities are

$$\begin{aligned}\mathbb{P}(r = r_0|\lambda_T^1,\dots,\lambda_T^N) &= \frac{\mathbb{P}(r = r_0)\mathbb{P}(\lambda_T^1,\dots,\lambda_T^N|r = r_0)}{\mathbb{P}(r = r_0)\mathbb{P}(\lambda_T^1,\dots,\lambda_T^N|r = r_0) + \mathbb{P}(r = r_1)\mathbb{P}(\lambda_T^1,\dots,\lambda_T^N|r = r_1)}\\ &= \frac{\pi_0 r_0^m(1-r_0)^{N-m}}{\pi_0 r_0^m(1-r_0)^{N-m} + \pi_1 r_1^m(1-r_1)^{N-m}}\end{aligned}$$

and

$$\mathbb{P}(r = r_1|\lambda_T^1,\dots,\lambda_T^N) = 1 - \mathbb{P}(r = r_0|\lambda_T^1,\dots,\lambda_T^N) = \frac{\pi_1 r_1^m(1-r_1)^{N-m}}{\pi_0 r_0^m(1-r_0)^{N-m} + \pi_1 r_1^m(1-r_1)^{N-m}}.$$

Therefore, the Bayesian estimation of r^* is given by Equation (7). □

Proof of Proposition 5. The conditional expectation can be computed by

$$\begin{aligned}
&\mathbb{E}\left[\log\left(px_{\text{dis}}+(1-p)x_{\text{nodis}}\right)|\lambda_T^1,\lambda_T^2,\ldots,\lambda_T^N\right]\\
=&\mathbb{E}\left[\log\left[p\left(\beta_{\text{dis}}\lambda_T+(1-\beta_{\text{dis}})\lambda_A\right)+(1-p)\left(\beta_{\text{nodis}}\lambda_T+(1-\beta_{\text{nodis}})\lambda_A\right)\right]|\lambda_T^1,\lambda_T^2,\ldots,\lambda_T^N\right]\\
=&q\mathbb{E}[r|\lambda_T^1,\lambda_T^2,\ldots,\lambda_T^N]\\
&\quad\cdot\log\left[p\left(\beta_{\text{dis}}\lambda_T^{\text{low}}+(1-\beta_{\text{dis}})\lambda_A^{\text{low}}\right)+(1-p)\left(\beta_{\text{nodis}}\lambda_T^{\text{low}}+(1-\beta_{\text{nodis}})\lambda_A^{\text{low}}\right)\right]\\
&+q(1-\mathbb{E}[r|\lambda_T^1,\lambda_T^2,\ldots,\lambda_T^N])\\
&\quad\cdot\log\left[p\left(\beta_{\text{dis}}\lambda_T^{\text{high}}+(1-\beta_{\text{dis}})\lambda_A^{\text{low}}\right)+(1-p)\left(\beta_{\text{nodis}}\lambda_T^{\text{high}}+(1-\beta_{\text{nodis}})\lambda_A^{\text{low}}\right)\right]\\
&+(1-q)\mathbb{E}[r|\lambda_T^1,\lambda_T^2,\ldots,\lambda_T^N]\\
&\quad\cdot\log\left[p\left(\beta_{\text{dis}}\lambda_T^{\text{low}}+(1-\beta_{\text{dis}})\lambda_A^{\text{high}}\right)+(1-p)\left(\beta_{\text{nodis}}\lambda_T^{\text{low}}+(1-\beta_{\text{nodis}})\lambda_A^{\text{high}}\right)\right]\\
&+(1-q)(1-\mathbb{E}[r|\lambda_T^1,\lambda_T^2,\ldots,\lambda_T^N])\\
&\quad\cdot\log\left[p\left(\beta_{\text{dis}}\lambda_T^{\text{high}}+(1-\beta_{\text{dis}})\lambda_A^{\text{high}}\right)+(1-p)\left(\beta_{\text{nodis}}\lambda_T^{\text{high}}+(1-\beta_{\text{nodis}})\lambda_A^{\text{high}}\right)\right].
\end{aligned}$$

This is exactly the (unconditional) expected growth rate defined by Equation (1) with r replaced by $\mathbb{E}[r|\lambda_T^1,\lambda_T^2,\ldots,\lambda_T^N]$. Therefore, the optimal solution is $p^*\left(q,\hat{r}_N(\lambda_T^1,\lambda_T^2,\ldots,\lambda_T^N)\right)$. □

Proof of Proposition 6. This result holds because $\hat{p}_N^*(\lambda_T^1,\lambda_T^2,\ldots,\lambda_T^N)$ given by Equation (8) depends on $\hat{r}_N(\lambda_T^1,\lambda_T^2,\ldots,\lambda_T^N)$ given by Equation (7), which further relies on m, the number of Tellarians observed with adverse events. Because $\lambda_T^1,\lambda_T^2,\ldots,\lambda_T^N$ are IID, m follows a binomial distribution with parameters N and r^*. This proves the result. □

Proof of Proposition 7. We first prove Equation (12) for the case of $r^*<\tilde{r}$. For $\hat{r}_N$ given by Equation (7), we have

$$\hat{r}_N=\frac{\pi_0 r_0^m(1-r_0)^{N-m}r_0+\pi_1 r_1^m(1-r_1)^{N-m}r_1}{\pi_0 r_0^m(1-r_0)^{N-m}+\pi_1 r_1^m(1-r_1)^{N-m}}=\frac{r_0+\frac{\pi_1}{\pi_0}\left(\frac{r_1}{r_0}\right)^m\left(\frac{1-r_1}{1-r_0}\right)^{N-m}r_1}{1+\frac{\pi_1}{\pi_0}\left(\frac{r_1}{r_0}\right)^m\left(\frac{1-r_1}{1-r_0}\right)^{N-m}}.\tag{A.3}$$

Because m follows a binomial distribution with parameters N and r^*, by the strong law of large numbers, we have

$$\frac{1}{N}\log\left[\left(\frac{r_1}{r_0}\right)^m\left(\frac{1-r_1}{1-r_0}\right)^{N-m}\right] = \frac{m}{N}\log\frac{r_1}{r_0} + \frac{N-m}{N}\log\frac{1-r_1}{1-r_0}$$
$$\xrightarrow{a.s.} r^*\log\frac{r_1}{r_0} + (1-r^*)\log\frac{1-r_1}{1-r_0}.$$

When $r^* < \tilde{r}$, the limit above is smaller than zero. Hence, as N increases without bound,

$$\left(\frac{r_1}{r_0}\right)^m\left(\frac{1-r_1}{1-r_0}\right)^{N-m} \xrightarrow{a.s.} 0,$$

and therefore, $\hat{r}_N \xrightarrow{a.s.} r_0$. The proof for the case of $r^* > \tilde{r}$ is similar, which we omit.

Next we prove Equation (13). By the central limit theorem,

$$\sqrt{N}\cdot\frac{\frac{m}{N} - r^*}{\sqrt{r^*(1-r^*)}} \xrightarrow{d} \mathcal{N}(0,1).$$

Thus, using the fact that

$$\frac{1}{N}\log\left[\left(\frac{r_1}{r_0}\right)^m\left(\frac{1-r_1}{1-r_0}\right)^{N-m}\right] = \frac{m}{N}\log\frac{r_1}{r_0} + \frac{N-m}{N}\log\frac{1-r_1}{1-r_0}$$
$$= \left(\log\frac{r_1}{r_0} + \log\frac{1-r_0}{1-r_1}\right)\frac{m}{N} + \log\frac{1-r_1}{1-r_0},$$

we have

$$\sqrt{N}\cdot\frac{\frac{1}{N}\log\left[\left(\frac{r_1}{r_0}\right)^m\left(\frac{1-r_1}{1-r_0}\right)^{N-m}\right] - \left[r^*\left(\log\frac{r_1}{r_0} + \log\frac{1-r_0}{1-r_1}\right) + \log\frac{1-r_1}{1-r_0}\right]}{\sqrt{r^*(1-r^*)\left(\log\frac{r_1}{r_0} + \log\frac{1-r_0}{1-r_1}\right)^2}} \xrightarrow{d} \mathcal{N}(0,1). \quad \text{(A.4)}$$

Furthermore, Equation (A.3) implies that

$$\left(\frac{r_1}{r_0}\right)^m\left(\frac{1-r_1}{1-r_0}\right)^{N-m} = \frac{\pi_0}{\pi_1}\cdot\frac{\hat{r}_N - r_1}{r_0 - \hat{r}_N},$$

and combining this with Equation (A.4) leads to Equation (13).

Finally, for any $x \in (r_0, r_1)$,

$$\mathbb{P}(\hat{r}_N \le x) = \mathbb{P}\left(\frac{\log\frac{\hat{r}_N - r_0}{r_1 - \hat{r}_N} + \log\frac{\pi_0}{\pi_1} - N\left[r^* \log\frac{r_1}{r_0} - (1-r^*)\log\frac{1-r_0}{1-r_1}\right]}{\sqrt{Nr^*(1-r^*)\left(\log\frac{r_1}{r_0} + \log\frac{1-r_0}{1-r_1}\right)^2}} \right.$$
$$\left. \le \frac{\log\frac{x-r_0}{r_1-x} + \log\frac{\pi_0}{\pi_1} - N\left[r^* \log\frac{r_1}{r_0} - (1-r^*)\log\frac{1-r_0}{1-r_1}\right]}{\sqrt{Nr^*(1-r^*)\left(\log\frac{r_1}{r_0} + \log\frac{1-r_0}{1-r_1}\right)^2}} \right)$$
$$\approx \Phi\left(\frac{\log\frac{x-r_0}{r_1-x} + \log\frac{\pi_0}{\pi_1} - N\left[r^* \log\frac{r_1}{r_0} - (1-r^*)\log\frac{1-r_0}{1-r_1}\right]}{\sqrt{Nr^*(1-r^*)\left(\log\frac{r_1}{r_0} + \log\frac{1-r_0}{1-r_1}\right)^2}} \right),$$

where Φ is the distribution function of $\mathcal{N}(0,1)$. Taking the derivative with respect to x leads to Equation (14). □

Proof of Proposition 8. According to Equation (8), we have $\hat{p}^*_N = p^*(q, \hat{r}_N)$. Thus, due to the fact that $\hat{r}_N \overset{a.s.}{\to} \hat{r}_\infty$ as N increases without bound and $p^*(\cdot,\cdot)$ is a continuous function, Equation (15) holds.

If $r_{\text{lower}} < \hat{r}_\infty < r_{\text{upper}}$, by Proposition 2, for sufficiently large N, we have

$$p^*_N = \frac{[\beta_{\text{nodis}}\lambda^{\text{high}} + (1-\beta_{\text{nodis}})\lambda^{\text{low}}](1-q)\hat{r}_N - [\beta_{\text{nodis}}\lambda^{\text{low}} + (1-\beta_{\text{nodis}})\lambda^{\text{high}}]q(1-\hat{r}_N)}{(\beta_{\text{nodis}} - \beta_{\text{dis}})(\lambda^{\text{high}} - \lambda^{\text{low}})(q + \hat{r}_N - 2q\hat{r}_N)},$$

which implies that

$$\hat{r}_N = \frac{A + B\hat{p}^*_N}{C\hat{p}^*_N + D}.$$

Combining this with Equation (13) leads to Equation (16). Finally, the proof of Equation (17) is similar to that for Equation (14), and therefore we omit the proof. □

Proof of Proposition 9. Because $p^*(q, r^*) = \text{argmax}_p\, \mu(p)$, we have $\mu\left(\hat{p}^*_N(\lambda^1_T, \lambda^2_T, \ldots, \lambda^N_T)\right) \le \mu\left(p^*(q, r^*)\right)$. Hence, by taking expectations of both sides, $\mathbb{E}\left[\mu\left(\hat{p}^*_N(\lambda^1_T, \lambda^2_T, \ldots, \lambda^N_T)\right)\right] \le \mu\left(p^*(q, r^*)\right)$. In addition, $\mu\left(\hat{p}^*_\infty(\lambda^1_T, \lambda^2_T, \ldots)\right) = \mu\left(p^*(q, r^*)\right)$ holds because of Propositions 7 and 8. □

Proof of Proposition 10. Without loss of generality, let us assume $r_m > r^*$. After incorporating mutations, the new unconditional probability of adverse events for Tellarians is

$r^{**} = p_m r_m + (1 - p_m) r^* > r^*$. Define

$$\begin{aligned}\mu(p, q, r^{**}) =& q r^{**} \log \lambda^{\text{low}} + (1-q)(1-r^{**}) \log \lambda^{\text{high}} \\ &+ q(1-r^{**}) \log \left[(\beta_{\text{dis}} - \beta_{\text{nodis}})(\lambda^{\text{high}} - \lambda^{\text{low}})p + \beta_{\text{nodis}}\lambda^{\text{high}} + (1-\beta_{\text{nodis}})\lambda^{\text{low}}\right] \\ &+ (1-q) r^{**} \log \left[(\beta_{\text{dis}} - \beta_{\text{nodis}})(\lambda^{\text{low}} - \lambda^{\text{high}})p + \beta_{\text{nodis}}\lambda^{\text{low}} + (1-\beta_{\text{nodis}})\lambda^{\text{high}}\right],\end{aligned}$$

which is the expected growth rate under Assumption 1, as given in the proof of Proposition 2. The key to this proof is considering the following function:

$$f(\varrho) = \mu(p^*(q, \varrho), q, r^{**}), \quad \varrho \in (0, 1),$$

where $p^*(\cdot, \cdot)$ is given by Equation (4). By the definition of $p^*(\cdot, \cdot)$ and the strict concavity of $\mu(p, q, r^{**})$ with respect to p (see the proof of Proposition 2), $f(\varrho)$ reaches its highest value if and only if $p^*(q, \varrho) = p^*(q, r^{**})$. Our goal is to prove $\mathbb{E}[f(\hat{r}_N)] > f(\hat{r}_\infty)$ for some N.

Now we prove the result under the following claim:

Claim. We can always find parameters such that

$$r_{\text{lower}} < r^{**} < r_{\text{upper}}, \tag{A.5}$$

and

$$r^* = r_{\text{lower}}, \tag{A.6}$$

where r_{lower} and r_{upper} are defined in Proposition 2.

Under this claim, $p^*(q, \varrho)$ is strictly increasing with respect to ϱ (Proposition 3). Therefore, $f(\varrho)$ is strictly increasing when $\varrho \in (r_{\text{lower}}, r^{**})$ and strictly decreasing when $\varrho \in (r^{**}, r_{\text{lower}})$. In addition, for $\varrho \leq r_{\text{lower}}$, we have $f(\varrho) = f(r_{\text{lower}})$, and for $\varrho \geq r_{\text{upper}}$, we have $f(\varrho) = f(r_{\text{upper}})$.

Case 1. If $f(r_{\text{lower}}) \geq f(r_{\text{upper}})$, let $r_0 = r^* = r_{\text{lower}}$ and $r_1 \in (r^{**}, r_{\text{upper}}]$ such that $f(r_1) = f(r_0)$. By Proposition 4, for any N and any $\pi_0, \pi_1 \in (0, 1)$, we have $\hat{r}_N \in (r_0, r_1)$. This implies that $f(\hat{r}_N) > f(r_1) = f(r_0)$. Therefore, due to the fact that $\hat{r}_\infty \in \{r_0, r_1\}$ (Proposition 7), we have $f(\hat{r}_N) > f(\hat{r}_\infty)$, which further implies $\mathbb{E}[f(\hat{r}_N)] > f(\hat{r}_\infty)$.

Case 2. If $f(r_{\text{lower}}) < f(r_{\text{upper}})$, let $r_0 = r^* = r_{\text{lower}}$ and $r_1 = r_{\text{upper}}$. Consider the case of $N = 1$. We have

$$\hat{r}_1 = \begin{cases} \hat{r}_1^- := \frac{\pi_0 r_0^2 + \pi_1 r_1^2}{\pi_0 r_0 + \pi_1 r_1}, & \lambda_T^1 = \lambda_T^{\text{low}}, \\ \hat{r}_1^+ := \frac{\pi_0 (1-r_0) r_0 + \pi_1 (1-r_1) r_1}{\pi_0 (1-r_0) + \pi_1 (1-r_1)}, & \lambda_T^1 = \lambda_T^{\text{high}}. \end{cases}$$

It is easy to verify that $r_0 < \hat{r}_1^- < \hat{r}_1^+ < r_1$. We select π_0 and π_1 such that $\hat{r}_1^- > r'$, where $r' \in (r_{\text{lower}}, r^{**})$ satisfies $f(r') = f(r_{\text{upper}}) = f(r_1)$. Hence, we have both $f(\hat{r}_1^-) > f(r_1) > f(r_0)$ and $f(\hat{r}_1^+) > f(r_1) > f(r_0)$. Therefore, due to the fact that $\hat{r}_\infty \in \{r_0, r_1\}$ (Proposition 7), we have $f(\hat{r}_1) > f(\hat{r}_\infty)$, which further implies $\mathbb{E}[f(\hat{r}_1)] > f(\hat{r}_\infty)$.

We finally prove the claim. By direct calculation and the fact of $r^* < r^{**}$, Equations (A.5) and (A.6) are equivalent to

$$\beta_{\text{dis}} < \frac{[(1-r^{**})\lambda^{\text{high}} + r^{**}\lambda^{\text{low}}]q - r^{**}\lambda^{\text{low}}}{(\lambda^{\text{high}} - \lambda^{\text{low}})(r^{**} + q - 2r^{**}q)} \tag{A.7}$$

and

$$\beta_{\text{nodis}} = \frac{[(1-r^{*})\lambda^{\text{high}} + r^{*}\lambda^{\text{low}}]q - r^{*}\lambda^{\text{low}}}{(\lambda^{\text{high}} - \lambda^{\text{low}})(r^{*} + q - 2r^{*}q)}. \tag{A.8}$$

Furthermore, it is easy to verify that both right-hand sides of Equations (A.7) and (A.8) being smaller than 1 is equivalent to

$$q < \frac{\lambda^{\text{high}} r^{**}}{\lambda^{\text{low}}(1-r^{**}) + \lambda^{\text{high}} r^{**}}, \qquad q < \frac{\lambda^{\text{high}} r^{*}}{\lambda^{\text{low}}(1-r^{*}) + \lambda^{\text{high}} r^{*}}. \tag{A.9}$$

Therefore, to satisfy Equations (A.5) and (A.6), first, we set $\lambda^{\text{high}} = 1$. Second, we select q satisfying

$$0 < q < \min\left\{\frac{\lambda^{\text{high}} r^{**}}{1 - r^{**} + \lambda^{\text{high}} r^{**}}, \frac{\lambda^{\text{high}} r^{*}}{1 - r^{*} + \lambda^{\text{high}} r^{*}}\right\},$$

which further satisfies Equation (A.9). Third, we select $\lambda^{\text{low}} \in (0, 1)$ small enough such that the right hand sides of Equations (A.7) and (A.8) are both greater than 0. Finally, we select $\beta_{\text{dis}}, \beta_{\text{nodis}} \in (0, 1)$ satisfying Equations (A.7) and (A.8). This completes the proof. □