

Prestige bias in cultural evolutionary dynamics

Saar Egozi¹ and Yoav Ram^{2,3,†}

¹School of Computer Science, Reichman University, Herzliya, Israel

²School of Zoology, Faculty of Life Sciences, Tel Aviv University, Tel Aviv, Israel

³Sagol School of Neuroscience, Tel Aviv University, Tel Aviv, Israel

[†]Corresponding author: yoav@yoavram.com

January 5, 2023

Abstract

A common bias in cultural transmission is to copy whoever appears more successful. This success bias only depends on the performance of the role-model. Because success may be difficult to estimate, it has been suggested that individuals may indirectly estimate success by using prestige as a proxy for success. However, prestige can be defined in different ways: it can reflect performance in another, possibly related, task; but it can also reflect overall influence of the role-model, irrespective of performance. Here, we model prestige bias as a combination of indirect success bias and influence bias, the latter meaning that the choice of a role-model depends on the number of individuals that have already copied from that role-model. We analyze the effect of prestige on cultural evolutionary dynamics using mathematical analysis and stochastic simulations. We find analytic approximations to the stochastic role-model choice process, which facilitate further mathematical analysis and reduce the computational complexity of simulations. We validate these approximations and demonstrate their robustness to model assumptions. We also find approximations to the fixation probability and the fixation time of an invading cultural trait, in both constant and changing environments. These approximations are similar to Kimura's approximations for population-genetic models. These similarities show that indirect success bias effectively plays the role of natural selection, whereas influence bias effectively plays the role of genetic drift. Influence bias also accelerates the evolutionary dynamics, as can be expected in a *rich-getting-richer* process. Our model extends the literature that describes and analyses cultural transmission, especially in human societies where social media is popular and thus influence bias may be expected. Further work is needed to test if this model could explain and predict various phenomena in cultural evolution.

30 Introduction

In cultural transmission, individuals transmit cultural traits (i.e., behaviors, beliefs, norms) to one another, typically by learning and demonstrating (Cavalli-Sforza and Feldman, 1981). Examples for cultural traits in humans are norms, preferences, tools, ideas, and language, potentially transmitted verbally and/or by observations (Creanza et al., 2017). Although cultural transmission is common in humans, it is also observed in other animals, e.g. chimpanzees (Horner et al., 2010; Kendal et al., 2015), dolphins and whales (Whitehead, 2017), elephants (McComb et al., 2001), and even flies (Battesti et al., 2012) and fish (Truskanov et al., 2020).

Cultural transmission may combine vertical transmission, in which parents transmit to their offspring; oblique transmission, in which adults (teachers, leaders, and even strangers) transmit to unrelated offspring; and horizontal transmission, in which peers from the same age cohort transmit to one another (Cavalli-Sforza and Feldman, 1981; Creanza et al., 2017). It has been demonstrated that non-vertical cultural transmission can maintain maladaptive traits, which can be beneficial in changing environments (Ram et al., 2018).

Transmission biases may cause a cultural trait to have a disproportionate probability to be transmitted compared to its frequency in the population. *Success bias* occurs when individuals prefer to copy from role models that demonstrate success in some activity, such as fishing, growing yams, using medicinal plants (Henrich and Broesch, 2011), or hunting (Mesoudi and O'Brien, 2008), and it can increase the probability of learning a successful trait (Borofsky and Feldman, 2022). Indeed, in a tournament between learning strategies conducted by Rendell et al. (2010), most winner strategies included a mixture of success-biased social learning and individual learning, implying that success-biased learning is a good strategy, but that by itself it is not enough to best other strategies, even when success is measured accurately.

Boyd and Richerson (1988, Ch. 5) have suggested that the evaluation of success can be divided to three categories: *direct bias*, *indirect bias* and *frequency-dependent bias*. A direct bias occurs when one phenotype is more attractive than other phenotypes, and is evaluated by *directly* testing the trait. For example, an individual observing a ping-pong match can attempt the observed paddle grips to determine which grip is better. A frequency-dependent bias occurs when the probability to copy a phenotype is higher or lower than the frequency of the phenotype among demonstrators. For example, suppose the common paddle grip is used by 60% of the demonstrators; if the this grip is adopted by 80% of copiers, then transmission is under positive frequency bias, also called *conformity*; if it is adopted by 40% of copiers, then transmission is under negative frequency bias, or *non-conformity* (Molleman et al., 2013). The effects of conformity and non-conformity on cultural evolution have been studied with both models (Denton et al., 2020, 2021, 2022) and experiments (Aljadeff et al., 2020). An indirect bias occurs when an individual uses one phenotype to determine the attractiveness of another phenotype. For example, an observer may copy the paddle grip of the ping-pong player who scored more points in the match, thus indirectly evaluating the grip by the points scored. This, however, may cause mismatches between the copied trait and the rest of the cultural or genetic repertoire of the

individual (Kolodny et al., 2022). Furthermore, Boyd and Richerson (1988, Ch. 8) have suggested that maladaptive traits may spread widely in a population if indirect biases are strong enough by a runaway process caused by a cultural equivalent of sexual selection (Andersson, 1994). For example, Cohen et al. (2021) have shown that helping behaviors can evolve due to horizontal transmission bias even without any benefit to the recipient, or when the benefit is much larger than the cost.

Henrich and Broesch (2011) have studied such indirect success biases, which they call cross-domain success bias or *prestige bias* (e.g., great fishermen may be chosen as role-models for growing yams). They have suggested that such biases, over generations, can lead to cultural adaptations, and that although prestige can lead to maladaptive traits spreading in the population, it can also accelerate the spread of adaptive traits. Boyd and Richerson (1988, Ch. 8) and Fogarty et al. (2017) have suggested that prestige biases are probably more common in humans than success biases, and prestige is often mentioned in the cultural evolution literature, however there are few models of prestige bias.

A broader definition of prestige is the “widespread respect and admiration felt for someone or something on the basis of a perception of their achievements or quality” (New Oxford American Dictionary). Indeed, Chudek et al. (2012) have defined prestige bias as “a tendency to learn from individuals to whom others have preferentially attended, learned or deferred”, and demonstrated its occurrence in 3-4 year old children. To distinguish this form of indirect bias from other definitions of prestige, we call it *influence bias*. This is an important distinction, as influence is a context-based bias, rather than a content-based bias: it does not depend on the phenotype itself but rather on the number of copiers that already copied each role-model, which may be easier and more accurate to estimate compared to success. Influence bias differs from frequency biases such as conformity, which depend on the frequency of a trait in the population or in a sample of role-models, rather than the social dynamics of copying.

In contemporary human society, social media makes it especially easy to estimate the social and cultural influence individuals have over others, and therefore can have an effect on decision making. Online social networks such as *Facebook* and *Instagram* are known to affect the influence of individuals (Anagnostopoulos et al., 2008; Diga and Kelleher, 2009; Peng et al., 2018), and specific marketing practices were invented to capitalize on this effect (Lee et al., 2012).

In the following, we develop a stochastic model of cultural transmission with prestige bias. In our model, prestige combines both cross-domain indirect success bias and influence bias. We develop analytic approximations for this model and analyze its dynamics. We find approximations for the probability and time for fixation of a ‘successful’ phenotype (i.e., that is favored by success bias). Comparing these approximations to Kimura’s approximations for the fixation of a favorable allele (Kimura, 1962; Kimura and Ohta, 1969), we demonstrate that success and influence bias play the role of natural selection and genetic drift, respectively.

Models

104 We begin with a continuous trait model with indirect bias suggested by Boyd and Richerson (1988),
develop an extension with influence bias, and then develop a model with a dichotomous trait. We
106 implemented our stochastic models and approximations, performed statistical analyses, and produced
figures using Python (Van Rossum et al., 2007) with NumPy (Van Der Walt et al., 2011) and Matplotlib
108 (Hunter, 2007). Source code is available at <https://github.com/yoavram-lab/PrestigeBias>.

Continuous trait

110 We follow the model of Boyd and Richerson (1988), assuming only oblique transmission of the trait
and omitting the indirect trait in order to reduce model complexity. We consider a population of N
112 individuals, described by a single trait with a continuous value. Each generation, N naive individuals,
or copiers, choose an individual from the previous generation, or role-models, from which they will
114 copy their trait. Similar to a Wright-Fisher model, we assume non-overlapping generations such that
the entire population is replaced in each generation. The population at time t can be described by
116 $\mathbf{A}(t) = (A_1(t), \dots, A_N(t))$ where $A_i(t)$ is trait value of individual i at time t . We assume the initial
population is drawn from a standard normal distribution, $\mathbf{A}(0) \sim N(0, 1)$. Cultural transmission is
118 modeled by a function F such that

$$A_i(t + 1) = F_i(\mathbf{A}(t)) . \quad (1)$$

120 **Success bias.** Boyd and Richerson (1988, Ch. 8, p. 247-249) describe a blended transmission algo-
rithm by defining F as a weighted average of the traits of all role-models,

$$122 \quad F_i(\mathbf{A}) = \sum_{j=1}^N G_{i,j} \cdot A_{i,j} , \quad (2)$$

where $G_{i,j}$ is the success bias of role-model j in the eyes of copier i ,

$$124 \quad G_{i,j} = \frac{\beta(A_{i,j})}{\sum_{k=1}^N \beta(A_{i,k})} , \quad (3)$$

$A_{i,j}$ is the absolute trait value copier i estimates for role-model j with some error $e_i \sim N(0, \eta^2)$,

$$126 \quad A_{i,j} = A_j + e_i, \quad (4)$$

and $\beta(\cdot)$ is the bias function that quantifies the success bias of a role-model,

$$128 \quad \beta(A_{i,j}) = b \cdot \exp \left(- \frac{(A_{i,j} - \hat{A})^2}{2J} \right) , \quad (5)$$

with \hat{A} as the arbitrary optimal trait value, and J and b as parameters that control the bias strength. Therefore, $G_{i,j}$ is a relative success score that copier i assigns to role-model j , equivalent to relative fitness in evolutionary-genetic transmission models.

Boyd and Richerson (1988) note that the deterministic blended transmission algorithm they use has alternatives. We thus develop a similar stochastic model with transmission from a single random role-model. Instead of eq. (2), we define the transmission function F as a random variable with its distribution given by

$$\Pr(F_i(\mathbf{A}) = A_j) = G_{i,j}, \quad (6)$$

such that $G_{i,j}$ gives the probability of copier i to choose to copy the trait of role-model j .

Influence bias. Here we introduce a new element to the model. We assume that in each generation, copiers choose their role-models one by one. We denote $K_{i,j}$ as the number of copiers that chose role-model j after copier i chose a role-model. Thus, i out of N copiers already chose a role-model, $\sum_{j=1}^N K_{i,j} = i$, and there are $N - i$ copiers that have yet to choose a role-model. The stochastic process of role-model choice,

$$\{\mathbf{K}_i = (K_{i,1}, \dots, K_{i,N})\}_{i=1}^N, \quad (7)$$

is described by the recurrence equation

$$K_{i,j} = K_{i-1,j} + S_{i,j}, \quad i, j = 1, 2, \dots, N, \quad (8)$$

where $S_{i,j} = 1$ if the i -th copier chose role-model j and 0 otherwise, and the initial state is $K_{0,j} = 0$.

Following eq. (6), the probability that the i -th copier chose role-model j is given by the prestige of role-model j in the eyes of copier i ,

$$\Pr(S_{i,j} = 1 \mid S_{1,j}, S_{2,j}, \dots, S_{i-1,j}) = G_{i,j}. \quad (9)$$

The prestige $G_{i,j}$ of role-model j in the eyes of copier i is determined by success—the estimated biased trait value $\beta(A_{i,j})$ —and influence—the number of copiers that chose role-model j before copier i , $K_{i-1,j}$, replacing eq. (3) of Boyd and Richerson (1988) with

$$G_{i,j} = \frac{\alpha_{ij} \cdot \beta(A_{i,j}) + (1 - \alpha_{ij}) \cdot K_{i-1,j}}{W_i}. \quad (10)$$

Here, the bias weight α_{ij} is a characteristic of the interaction of role-model j with copier i that determines the relative significance of direct success vs. influence in the role-model's overall prestige in the eyes of the copier.

We assume every individual may evaluate the importance of success and influence differently. Additionally, we assume every role-model displays his influence and success individually. For example, a person with more followers but lack of actual skill may emphasize the number of his followers rather

than his actual skill. We therefore defined $\alpha_{i,j}$ to be the relation between every copier's intrinsic
 162 significance of estimation, and every role-model's personal emphasis.

The trait of role-model j estimated by copier i , $A_{i,j}$, remains as in eq. (4), and W_i is a normalizing
 164 factor that sums the numerator over all role-models ($1 \leq j \leq N$) to ensure $\sum_{j=1}^N G_{i,j} = 1$.

Dichotomous trait

166 To allow for mathematical analysis of the model, we introduce a simplified version where the trait
 only has two phenotypes: the optimal phenotype \hat{A} and the sub-optimal phenotype A . All role-models
 168 with the same phenotype will contribute to the probability of that phenotype to be transmitted, and
 thus influence is determined by the number of copiers that have already chosen a role-model with
 170 either phenotype. In addition, for simplicity and for easier mathematical analysis, we assume α is
 homogeneous, such that we have exchangeability between role-models. Therefore, the probability of
 172 the i -th copier to copy phenotype A is

$$G_{i,A} = \frac{(N - X)\alpha'\beta(A) + K_{i,A}}{i - 1 + (N - X)\alpha'\beta(A) + \alpha'X}, \quad (11)$$

174 where X is the number of role-models with trait \hat{A} , $K_{i,A}$ is the number of copiers that already chose
 A when copier i chooses a role-model, $\beta(\hat{A}) = 1$, and $\alpha' = \frac{\alpha}{1-\alpha}$ is the odd ratio of the bias weight.
 176 Complementing this, the probability of the i -th copier to copy phenotype \hat{A} is $G_{i,\hat{A}} = 1 - G_{i,A}$. The
 rest of the details follow the continuous trait model.

178 Results

Our model is defined by two nested stochastic processes. Change over multiple generations is described
 180 by the phenotype distribution at each generation, $\{\mathbf{A}(t)\}_t$. The transition from one generation to the
 next is described by the number of copiers each role-model has after i copiers have chosen a role-
 182 model, $\{\mathbf{K}_i\}_{i=1}^N$. We emphasize that the models are nested: $\mathbf{A}(t + 1)$ can be computed from $\mathbf{A}(t)$ by
 evaluating \mathbf{K}_N . However, the former requires iterating over eqs. (8) and (9). Thus, we sought to
 184 find an equivalent stochastic process that has the same joint distribution as \mathbf{K}_N . We found two such
 approximations, summarized here and explained in detail below. In both we assume either α_j is a
 186 trait of role-model j only (contrary to $\alpha_{i,j}$ that included the relation with copier i), or a completely
 homogeneous α .

188 **Generalized binomial approximation.** The number of copiers of a specific role-model at each step,
 $K_{i,j}$, follows the *generalized binomial distribution* (Drezner and Farnum, 1993) and therefore,

190 (i) the expected number of copiers of role-model j equals its prestige in the eyes of the first copier,
 multiplied by the total number of copiers, that is, $E[K_{N,j}] = N \cdot G_{1,j}$ if trait estimation error is uniform
 192 for all copiers ($e = e_i$ for $i = 1, \dots, N$); and (ii) the expected number of copiers of each role-model
 equals its relative biased trait value, similar to the role of relative fitness in population-genetic models,

194 that is, $E[K_{N,j}] = \beta(A_j + e)/\bar{\beta}$ if the bias weight is uniform for all role-models ($\alpha = \alpha_j$ for $j = 1, \dots, N$),
 where $\bar{\beta} = 1/N \sum_{j=1}^N \beta(A_j + e)$ is the population mean estimated trait value.

196 **Dirichlet-Multinomial approximation.** The role-model choice process, $\{\mathbf{K}_i\}_{i=1}^N$, is equivalent to a
Pólya urn model if trait estimation error is uniform for all copiers ($e = e_i$ for all $i = 1, \dots, N$), meaning
 198 there's no meaning for the order of copiers.

Note that here $G_{i,j}$ is only a function of the trait values A_j and the bias weights α_j , as the estimation
 200 error is uniform, meaning all copiers are identical.

Generalized binomial distribution

202 The generalized binomial distribution (GBD) emerges from a series of dependent Bernoulli trials (in
 contrast to the standard binomial distribution in which trials are independent) and is parameterized by
 204 $GBD(n, p, \theta)$ where n is the number of trials, p is the probability of success of the first trial, and θ is
 the correlation between trials ($\theta = 0$ gives the standard binomial distribution).

206

Result 1 (Generalized binomial approximation). *The number of copiers of role-model j after i copiers
 208 have chosen a role-model follows the GBD, $K_{i,j} \sim GBD(i, \alpha_i \cdot \beta(A_j + e))$ if $e_i = e$ for all role-models
 $i = 1, \dots, N$.*

210 *Proof.* We denote $Q_j(k, i) = P(K_{i,j} = k \mid K_{i-1,j})$ as the probability that exactly k out of i copiers
 choose role-model j given $K_{i-1,j}$ our of $i - 1$ copiers chose role-model j . Using conditional probability
 212 and eq. (8),

$$Q_j(k, i) = P_j(S_{i,j} = 1 \mid k - 1, i - 1) \cdot Q_j(k - 1, i - 1) + P_j(S_{i,j} = 0 \mid k, i - 1) \cdot Q_j(k, i - 1), \quad (12)$$

214 where $S_{i,j} = 1$ when the i -th copier chooses role-model j . Equation (12) is equivalent to eq. (2.1) by
 Drezner and Farnum (1993), which completes the proof.

216 This result gives the following two corollaries on the expected number of followers of a given role-
 model by the end of the role-model choice process. The proof of Corollary 2 is in Appendix A.

218

Corollary 1. *The expected number of copiers of role-model j after all copiers have chosen a role-
 220 model is $E[K_{N,j}] = N \cdot G_{1,j}$, where $G_{1,j}$ is the probability of the first copier to copy role-model
 j .*

222 *Proof.* Drezner and Farnum (1993, eq. (2.3)) show that $E[k] = N \cdot Q_j(1, 1)$ (modified from their
 notation). $Q_j(1, 1)$ is the initial probability to choose role-model j , before any role-model choices are
 224 made. $Q_j(1, 1) = G_{1,j}$ by definition, which completes the proof.

Corollary 2. $E[K_{N,j}] = \alpha_j \cdot \beta(A_j + e)/\overline{\alpha \cdot \beta(A + e)}$, where the averaging in the denominator is over
 226 the role-models index, j .

The special case where the bias weight is uniform for all role-models ($\alpha = \alpha_j$ for $j = 1, \dots, N$) is
 228 interesting, because we can evaluate the expected number of copiers using a linear equation

$$E[K_{Nj}] = N \cdot \frac{\alpha \cdot \beta(A_j + e)}{\sum_{m=1}^N \alpha \cdot \beta(A_m + e)} = \beta(A_j + e) \left/ \overline{\beta(A + e)} \right., \quad (13)$$

230 where the only variable is $A_j + e$, because $\overline{\beta(A + e)}$ is the mean of the distribution of the trait values, modified by some constant parameters of β . We can then denote $L = 1/\overline{\beta(A + e)}$ and write

$$232 \quad E[K_{Nj}] = L \cdot \beta(A_j + e). \quad (14)$$

Numerical validation. To validate our results we ran 1,000 simulations of the full model, and
 234 compared the results with Corollary 2. We compare the distribution of number of copiers by plotting the histograms of both our simulations results and the expected values based on Corollary 2. We used
 236 the average of the 1,000 simulations to eliminate drift errors, as can be seen in Figure 1.

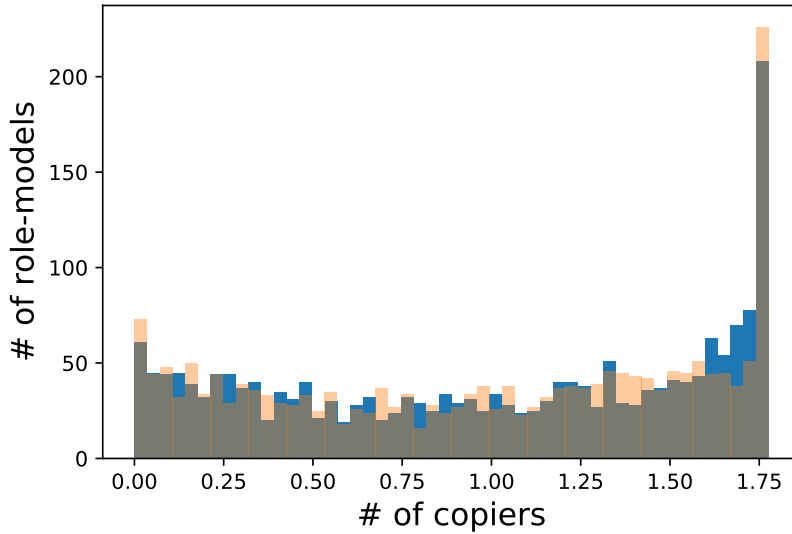


Figure 1: Numerical validation of the GB approximation. The approximation (orange) fits simulation results (blue) well when using 1,000 simulations for both models. Here, population size, $N = 2,000$; bias weight, $\alpha = 0.1$; idea phenotype value, $\hat{A} = 1$; role-model traits $\mathbf{A} \sim N(0, 1)$; success bias value, $\beta(A) = 0.956$.

Although basic, Figure 1 shows good fit of the GB approximation. This validation is initial, and the
 238 more extensive validations we do on the Dirichlet-Multinomial approximation, because it is what we will use in our analysts.

240 Dirichlet-Multinomial distribution approximation

Pólya urn model. This stochastic process consists of N draws from an urn with an initial amount of
 242 colored balls of M colors. When a ball is drawn, it is then placed back in the urn together with an additional new ball of the same color. Let $\mathbf{U}_i = \{u_{i,1}, u_{i,2}, \dots, u_{i,M}\}$ where $u_{i,j}$ is the number of balls of

244 the j -th color in the urn after i draws. Let $S_{i,j} = 1$ when drawing a j -colored ball on the i -th draw, and 0 otherwise. The probability that $S_{i,j} = 1$ given \mathbf{U}_{i-1} is

$$246 \quad P(S_{i,j} = 1 \mid \mathbf{U}_{i-1}) = \frac{u_{i-1,j}}{\sum_{m=1}^M u_{i-1,m}} = \frac{o_j + w_{i-1,j}}{\sum_{m=1}^M o_m + w_{i-1,m}} = \frac{o_j + w_{i-1,j}}{i-1 + \sum_{m=1}^M o_m}, \quad (15)$$

where o_j is the initial number of balls of the color j in the urn, and $w_{i,j}$ is the cumulative number of
248 new balls that were added to the urn after i draws of the color j .

250 **Result 2** (Pólya urn model). *The role-model choice process, $\{\mathbf{K}_i\}_{i=1}^N$, is equivalent to a Pólya urn model if both trait estimation error and bias weight are uniform in the population, $e = e_j$ and $\alpha = \alpha_j$ for all $j = 1, \dots, N$.*

Proof. Denote $\alpha' = \frac{\alpha}{1-\alpha}$ as the bias weight ratio, and $A'_j = A_j + e$. From eq. (10) and because
254 $\sum_{j=1}^N K_{i,j} = i$, we have

$$G_{i,j} = \frac{\alpha' \beta(A'_j) + K_{i-1,j}}{\sum_{m=1}^N \alpha' \beta(A'_m) + K_{i-1,m}} = \frac{\alpha' \beta(A'_j) + K_{i-1,j}}{i-1 + \sum_{m=1}^N \alpha' \beta(A'_m)}. \quad (16)$$

256 Substituting $M = N$, $o_j = \alpha' \beta(A'_j)$, and $w_{i,j} = K_{i,j}$ in eq. (15) gives eq. (16), thus completing the proof.

258 Frigyik et al. (2010, section 2) prove that the proportion of different colored balls in a *Pólya urn model* converges to the Dirichlet distribution as the number of draws approaches infinity, based on
260 the *Martingale Convergence Theorem* (Durrett, 1999). We therefore have an approximation for the relative prestige each role-model has when evaluated by copiers. Thus, choosing the role-models for
262 all copiers is equivalent to drawing from a Multinomial distribution where the parameters are the modified weights from a Dirichlet distribution and we have the following corollary.

264 **Corollary 3.** *The number of copiers of each role-model follows a Dirichlet-Multinomial distribution, $\mathbf{K}_i \sim \text{DM}(N, \mathbf{G}_1)$, under the conditions of Result 2.*

Numerical validation. To validate our analytical result (Corollary 3) and test its sensitivity to the
268 assumptions ($e_i = e$ and $\alpha_i = \alpha$ for $i = 1, \dots, N$) we compare it to results of stochastic simulations of the full model. First, we computed an observed distribution of the number of copiers from the average
270 empirical distribution of multiple simulations. We then compared this observed distribution with the expected theoretical DM distribution as can be seen in Figure 2 (a). The difference in distributions
272 was not perceived when plotting both distributions on the same figure, so we used the difference instead. The maximum difference is 0.5 role-models, which indicate a very good fit. In addition, we
274 tested the likelihood of the observed data to be drawn from the DM distribution, against a shuffle of the parameters vector of the DM distribution itself, as seen in Figure 2 (b). We see that the negative

log likelihood of the observed data is much higher than any other shuffled version of the parameters vector, supporting our approximation more.

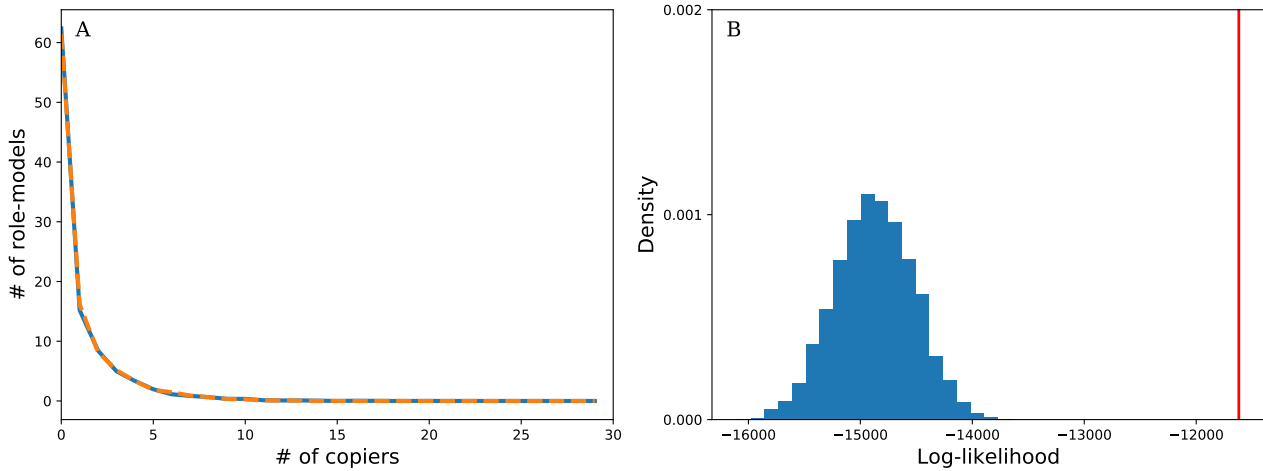


Figure 2: Numerical validation of the DM approximation. We performed computational "experiments" of the role-model choice process and compared them to the DM distribution. **(A)** The difference between the DM distribution (orange) and the empirical distribution of the experiments (blue) is very small. **(B)** The log-likelihood of the DM distribution for results of the experiments (red vertical line) is much higher compared to the log-likelihood of permutations of experiments (blue histogram). Here, population size, $N = 100$; number of experiments, $m = 100$; phenotype values, $\hat{A} = 1$, $A \sim N(0, 1)$; success-bias weight, $\alpha = 0.5$. No estimation error or bias is applied, and traits are estimated and copied perfectly.

Next, we examined the fixation probability and fixation time of a favored phenotype \hat{A} when invading a population of phenotype A and compared results from the full model and the DM approximation. Thus, we assume the population has a single individual with phenotype \hat{A} and $N - 1$ individuals with phenotype A . We find that the number of simulations needed to sufficiently approximate our model with the DM approximation is roughly 1,000 (Figure 3). We examined the robustness of the DM approximation to relaxing the approximation assumptions. First, we relaxed our assumption of no estimation error e . Estimation error in the original model was drawn from a normal distribution, and added to the trait value before evaluation of the bias ($A_{ij} = A_j + e_i$). When estimation error is applied, we sample J_i for each copier i from a normal distribution with varying scale (variance). Even when the standard deviation is 0.1, the fixation probability and time is similar (Figure 4). We also relaxed our assumption of a uniform bias weight α (i.e., $\alpha_i = \alpha$). We allowed α to vary in the population, drawing α_j for each role-model j from a normal distribution such that $\alpha_j \sim N(0.5, q)$ where $q \in [10^{-7}, 10^{-1}]$. We found again that results of the DM approximation are similar to those from stochastic simulations of the full model (Figure 5).

Fixation probability and time

After finding that the DM distribution is a good approximation of the (within-generation) role-model choice process, we turn our attention to the (between-generation) evolutionary dynamics. We focus on the fixation probability and conditional fixation time (conditioned on the population reaching fixation) of a favored phenotype, using a diffusion-equation approximation approach, similar to analyses of

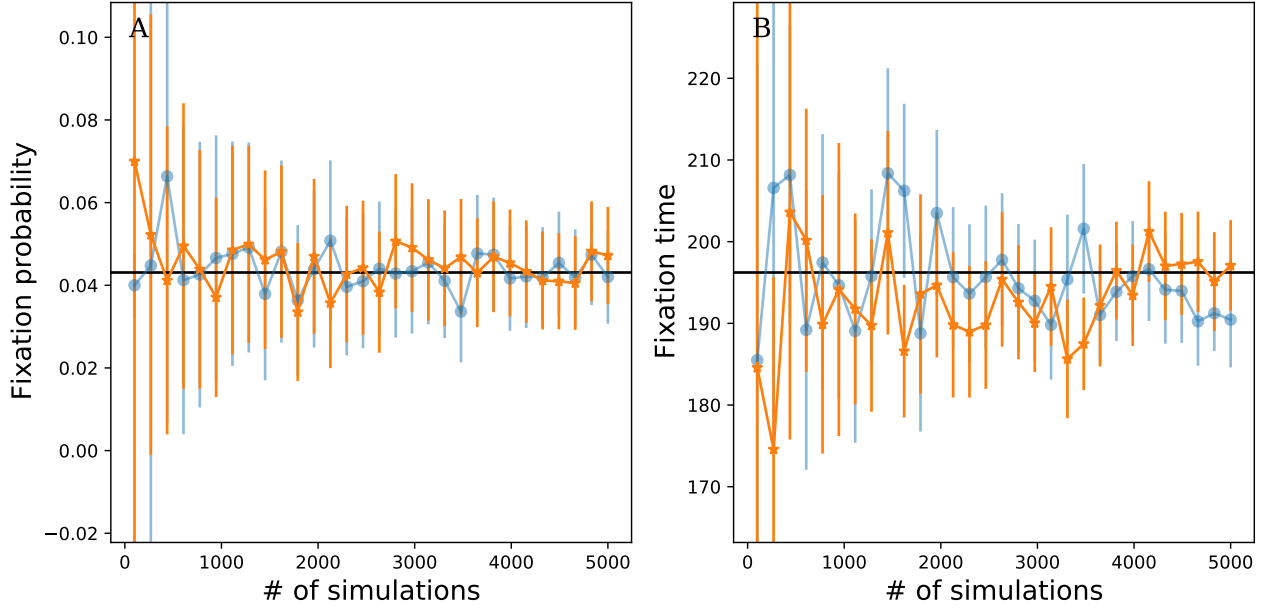


Figure 3: DM Approximation precision as function of number of simulations. Our DM approximation (orange) agrees with stochastic simulation results (blue) when using 1,000 or more simulations. Both fluctuate around the analytic fixation probability approximation (black; eq. (19)). Markers are averages across simulations, error bars are 95% confidence intervals. Here, population size, $N = 1000$; success-bias weight, $\alpha = 0.5$; phenotype values, $\hat{A} = 1$, $A = 0.7$; success-bias value, $\beta(A) = 0.956$.

population-genetic models (Kimura, 1962; Kimura and Ohta, 1969; Otto and Whitlock, 2006). We are
 298 mainly interested in the effect of the bias weight, α , which determines the relative effect of success and
 influence on prestige bias, given by eq. (10). For simplicity, we do not include role-model estimation
 300 error in this analysis, i.e. $e_i = 0$ for every copier i . As shown above, transmission in our model is
 approximately DM distributed (Corollary 3 and eq. (16)).

302 We start by finding the expectation and variance of the change in frequency from one generation to the
 next, which are the drift and diffusion terms of the diffusion equation. The proof is in Appendix B.

304 **Result 3** (Drift and diffusion terms in a constant environment). *Let x and x' be the frequency of type \hat{A}
 306 in a population with N individuals in the current and next generation, and β is the success coefficient
 of phenotype A , $\beta = \beta(A) < \beta(\hat{A}) = 1$. Then,*

$$308 \quad E[x' - x] \approx x(1 - x)(1 - \beta), \quad V(x' - x) \approx x(1 - x) \left(\frac{1}{\alpha N + (1 - \alpha)} \right). \quad (17)$$

This analysis gives a surprising result relating the parameters α and β to parameters of the classical
 310 Wright-Fisher model from population genetics: the selection coefficient s , a measure of the effect of
 natural selection on the change in frequency of genotypes, and the effective population size, N_e , a
 312 measure of the effect of random genetic drift on the change in frequency of genotypes. In a diffusion-
 equation approximation of the classical Wright-Fisher model, the expectation and variance of the
 314 change in frequency are $E[x' - x] = x + x(1 - x)s + o(s)$ and $V[x' - x] = x(1 - x)/N_e$ (Kimura, 1962,

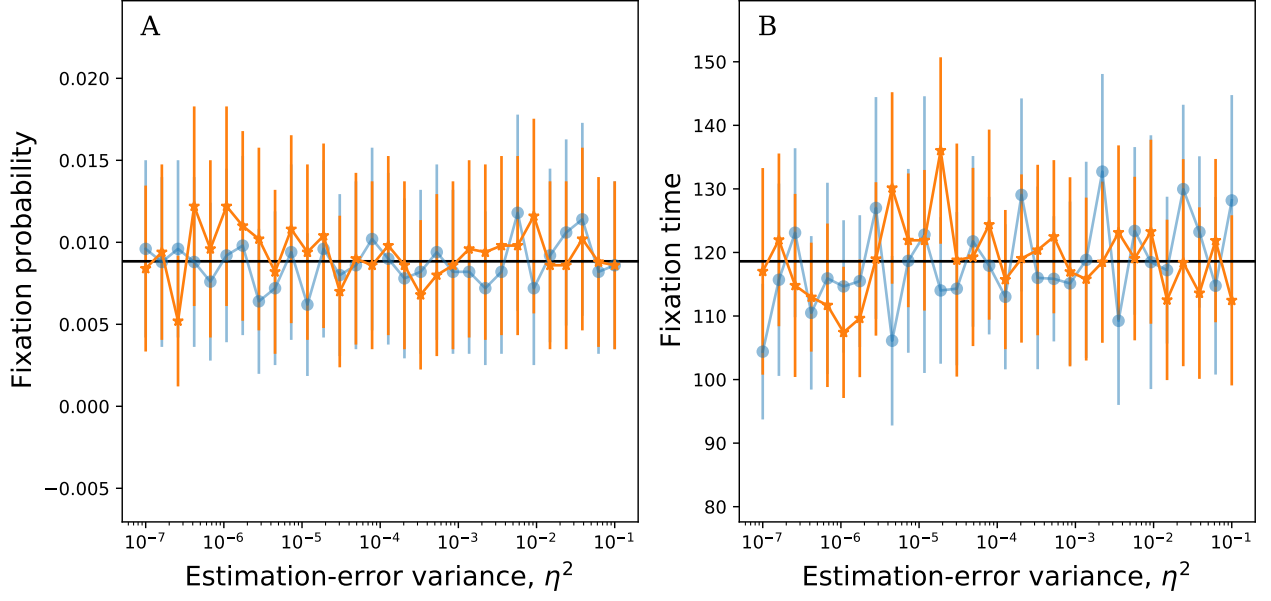


Figure 4: Robustness of DM approximations to success estimation error. Both the DM approximation (orange) and our approximation (black) agree with the stochastic simulations (blue), even with a high estimation error. Markers are averages across simulations, error bars are 95% confidence intervals. 5,000 simulations per data point; population size, $N = 1000$; success-bias weight, $\alpha = 0.1$; phenotype values, $\hat{A} = 1, A = 0.7$; bias strength parameter $J \sim N(1, \eta^2)$ where $\eta \in [10^{-7}, 10^{-1}]$.

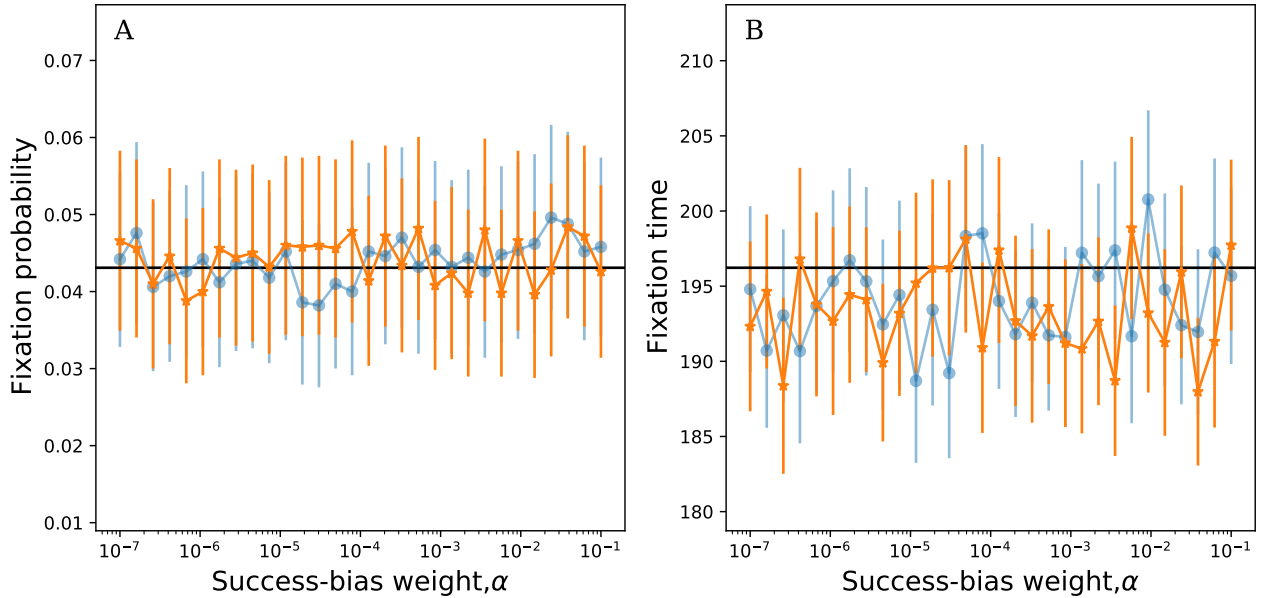


Figure 5: Robustness of DM approximations to variation in the bias weight α . Both the DM approximation (orange) and Kimura's equation (black line) fit the stochastic simulations (blue) well even with a high variation in success bias weight. Markers for average across 5,000 simulations, error bars are 95% confidence intervals. Here, population size, $N = 1000$; success bias weight normally distributed, $\alpha \sim N(0.5, x^2)$; phenotype values $\hat{A} = 1, A = 0.7$; success bias value, $\beta(A) = 0.956$.

eq. 7). Therefore, we have the following result.

Result 4 (Effective selection coefficient and population size). *The effective selection coefficient s and effective population size N_e can be written in terms of the success coefficient β (eq. (5)), the bias weight α (eq. (10)), and the population size N as*

$$s = 1 - \beta(A), \quad N_e = \alpha N + (1 - \alpha). \quad (18)$$

Note that when $N \gg 1$, $N_e \approx \alpha N$, resulting in a very convenient expression.

Using our effective selection coefficient, $1 - \beta$, and effective population size, N_e , with the population-genetics fixation probability approximation given by Kimura (1962, eq. 8), we get the following result:

Result 5 (Fixation probability). *The fixation probability is approximately*

$$\pi(x) = \frac{1 - e^{-2(1-\beta)N_e x}}{1 - e^{-2(1-\beta)N_e}} \quad (19)$$

where x is the initial frequency of the favored phenotype \hat{A} .

Similarly, we can use $1 - \beta$ and N_e in the population-genetics fixation time approximation given by Kimura and Ohta (1969, eq. 17).

Result 6 (Fixation time). *The fixation time (conditioned on fixation) from an initial frequency x is approximately*

$$T(x) = J_1(x) + \frac{1 - \pi(x)}{\pi(x)} \cdot J_2(x), \quad (20)$$

where $N_e = \alpha N + (1 - \alpha)$, $S = N_e(1 - \beta)$, and

$$\begin{aligned} J_1(x) &= \frac{2}{(1 - \beta)(1 - e^{-2S})} \int_x^1 \frac{(e^{2S\xi} - 1)(e^{-2S\xi} - e^{-2S})}{\xi(1 - \xi)} d\xi, \\ J_2(x) &= \frac{2}{(1 - \beta)(1 - e^{-2S})} \int_0^x \frac{(e^{2S\xi} - 1)(1 - e^{-2S\xi})}{\xi(1 - \xi)} d\xi. \end{aligned} \quad (21)$$

Note that these integrals cannot be solved in closed form, so we can only estimate them numerically.

Numerical validation. We compare our approximations (eqs. (19) and (20)) with results of simulations of our dichotomous model using various α and β values, as well as simulations of the Wright-Fisher model, using the effective selection coefficient, $1 - \beta$, and effective population size, $N_e = \alpha N + (1 - \alpha)$. We find see that the two models have similar dynamics, and both are well approximated our approximations (Figure 6).

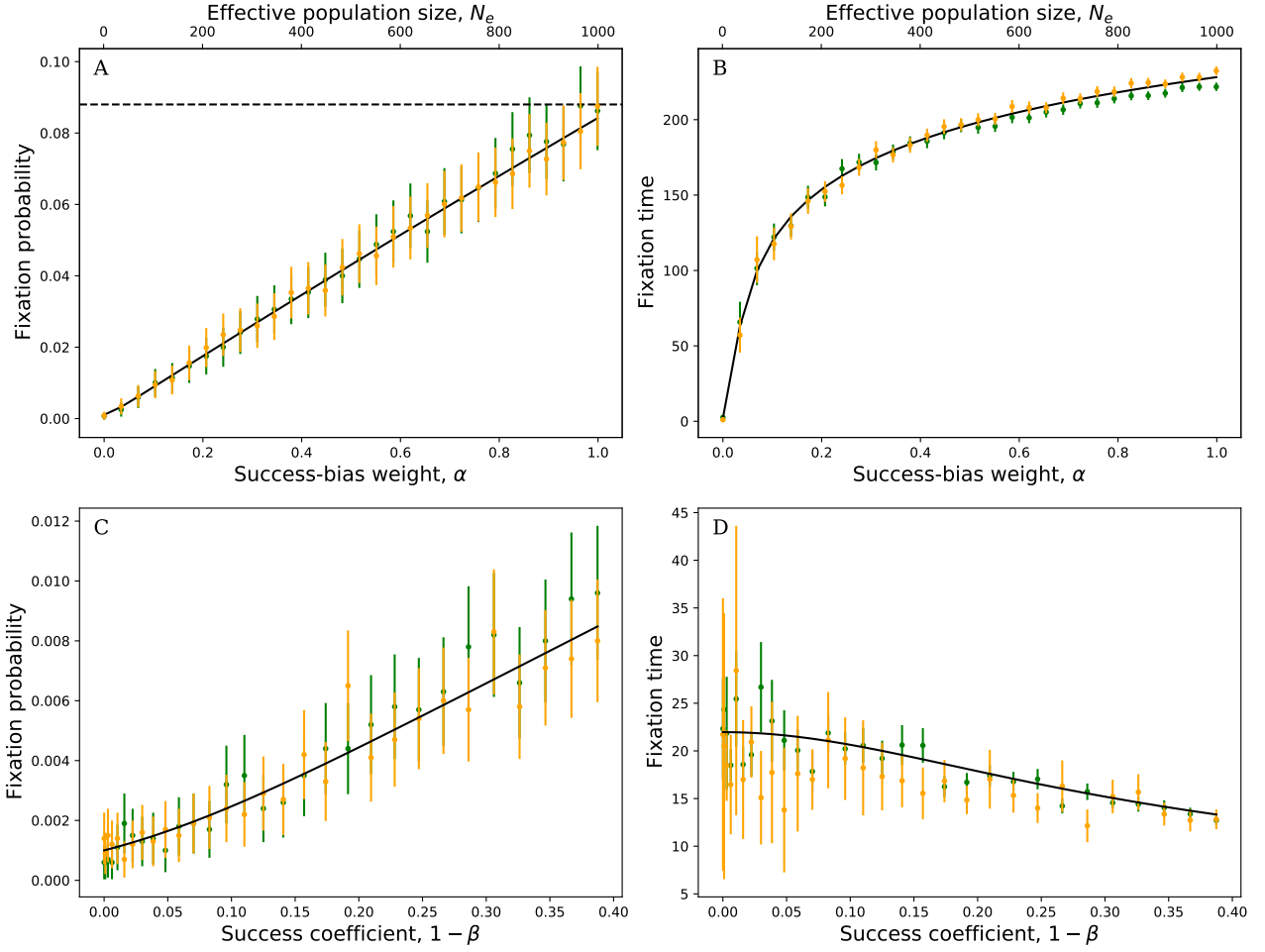


Figure 6: Fixation probability and time in a constant environment. Fixation probability and time (in generations) as a function of the success-bias weight α (bottom x-axis), or effective population size N_e (top x-axis) in the top row, and as a function of the success coefficient, $1 - \beta$, on the bottom row. The approximation (black; eq. (19)) agrees with both DM simulations (green) and Wright-Fisher simulation (orange). Fixation probability (A) is bounded by $2(1 - \beta)$ (blue). Markers are averages of 10,000 simulations, error bars show 95% confidence intervals for panels A and B and 75% for panels C and D. Here, Population size, $N = 1,000$; phenotype values, $\hat{A} = 1$, $A = 0.7$ (A and B), $A = a \cdot \hat{A}$ with $0.01 \leq a \leq 0.99$ (panels C and D); success coefficient, $1 - \beta = s = 0.044$ (A and B); success-bias weight, $\alpha = 0.01$ (panels C and D).

344 **Changing environment** . After finding a good approximation in constant environment, where the
favorable trait is always \hat{A} , we proceeded to find an approximation for a changing environment. Fol-
346 lowing Ram et al. (2018), we denote k as the number of generations in which the invading phenotype
is favored by success bias, and l as the number of generations in which the resident phenotype is
348 favored by success bias. We then proceed to find expressions for the expectation and variance of the
change in frequency between $n = k + l$ generations. The proof is in Appendix C.

350

Result 7 (Drift and diffusion terms in a changing environment). *Let x be the initial frequency of the
352 invading phenotype and X_t is the number of individuals with the phenotype after n generation. Then,*

$$E[X_n/N - x] \simeq x(1 - x)S_n/N_e, \quad \text{and} \quad V(X_n/N - x) \simeq nx(1 - x)/N_e, \quad (22)$$

354 where $S_n = \sum_{t=1}^n N(1 - \beta_t)$.

Note that here, we have the average selection coefficient during a cycle of n generations as the
356 selection coefficient eq. (19). Using the drift and diffusion terms and following Ram et al. (2018), we
can approximate the fixation probability in a changing environment using

358
$$\tilde{\pi}(x) = \frac{1 - e^{-2\frac{S_n}{n}N_e x}}{1 - e^{-2\frac{S_n}{n}N_e}} . \quad (23)$$

Numerical validation. Comparing our approximation (eq. (23)) to simulations, we find that the
360 approximation fits simulation results well for variable bias weights, α , which corresponds to the
effective population size (Figure 7A).

362 However, the approximation is more sensitive to the value of the success bias coefficient β (Figure 7B).
We suspect that when β is too small, there will not be many cycles in the simulations, because either
364 the population reaches a high frequency of the fitter phenotype after just a few cycles, or the fitter
phenotype becomes extinct very quickly. For such β values (below 0.65), the fixation probability
366 exceeds even the constant environment approximation (which is the upper limit). We note that the
diffusion-equation approximation assumes weak selection, or in our case, weak success bias.

368 We found that for a large k/l ratio (with a constant cycle length, $n = k + l = 100$), the changing
environment approximation (eq. (23)) converges to the constant environment approximation (eq. (19)),
370 see Figure 7C and Figure 7D.

The approximation follows the trend of the simulation results for all α values. When increasing
372 the success coefficient to more than 0.15, the simulation results were located above the changing
environment approximation, and below the constant environment approximation. We believe the
374 reason is the structure of the cycle. Our proof and approximation in the changing environment are for
a large amount of cycles, and when the success coefficient is too high, there might be very few cycles.
376 Either the ideal trait is copied by enough copiers so that the influence is sufficient to negate the success
bias when the cycle changes (and the trait favored by the bias becomes the disfavored), or the opposite
378 happens, and the ideal trait gets extinct before there were enough copiers that copied it. We then tried
to change the ratio between the number of cycles where \hat{A} is favored and disfavored. We showed
380 that the approximation fits well regardless of the ratio, but when the ratio of favored generations to
disfavored ones is very high, it is very similar to a constant environment model.

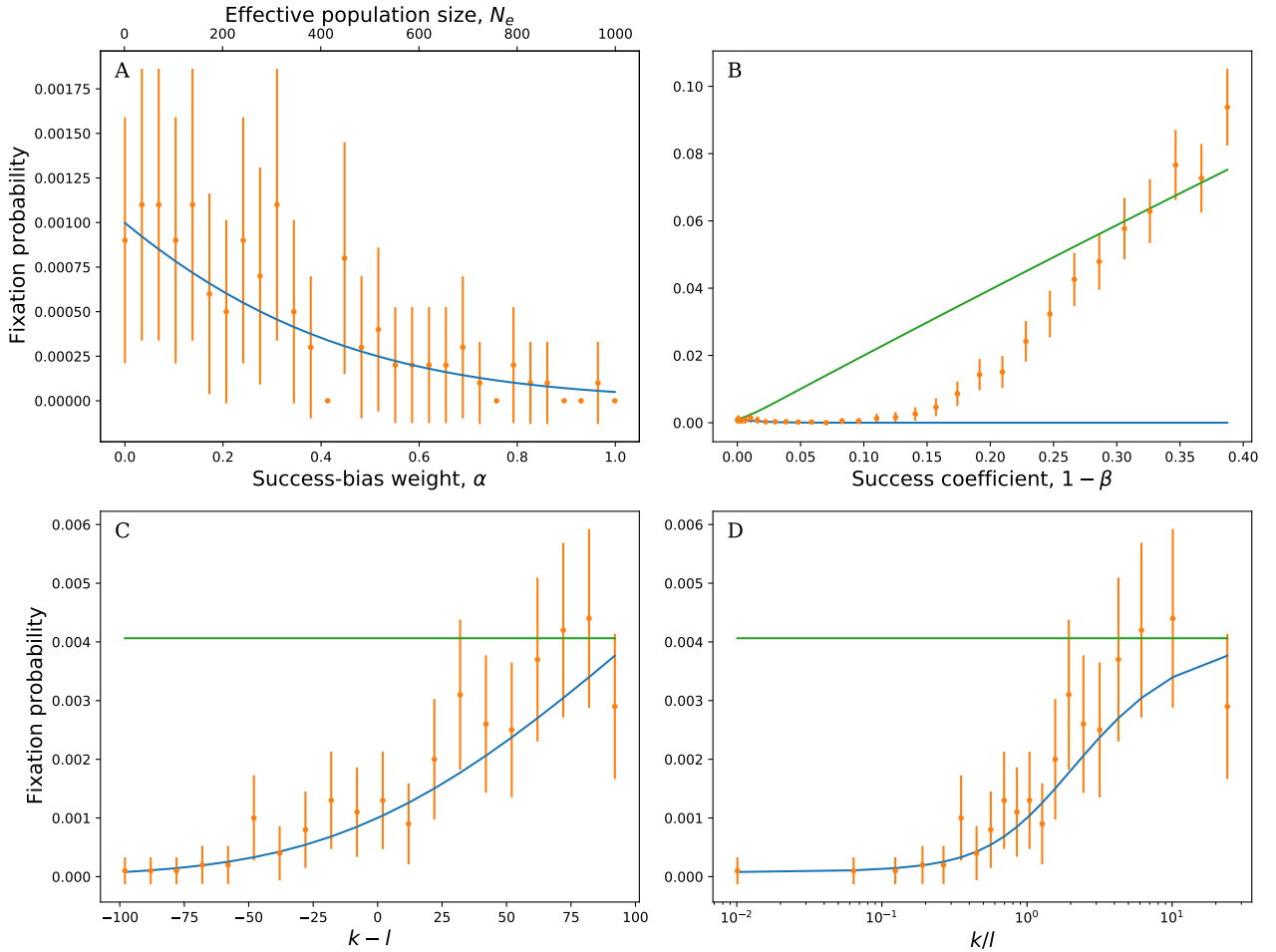


Figure 7: Fixation probability in a changing environment. (A) Fixation probability decreases with the the success-bias weight (bottom x-axis) and effective population size (top x-axis). The approximation (blue; eq. (23)) agrees with simulation results (orange). (B) Fixation probability increases with the success coefficient, β . When success bias is large ($1 - \beta > 0.1$), simulation results (orange) are underestimated by the changing environment approximation (blue; eq. (23)). With even larger success bias ($1 - \beta > 0.35$), even the constant environment approximation (green; eq. (19)) slightly underestimates simulation results, likely because the diffusion equation approximation assumes weak "selection". (C,D) The approximation (blue) is robust to changes in environmental cycle length, as it agrees with simulations (orange) for different sizes of the changing environment cycle, where k and l are the number of generations each trait value is under success bias. When $k > l$, the approximation and the simulations are both very close to the constant environment approximation (green), because the more generations the rare phenotype is favored, the more similar it is to the constant environment model, where it is always favored by the success bias. Markers show average of 10,000 simulations, error bars show 75% (A, C, and D) and 95% (B) confidence intervals. Here, population size, $N = 1,000$; phenotype values, $\hat{A} = 1$, $A = 0.9$ (A and B), $A = 0.8$ (C and D); In (A), the success coefficient is: $1 - \beta = s = 0.005$; In (B, C, and D) the success-bias weight is $\alpha = 0.1$.

382 Adaptive success-bias weight

We ran simulations of the role-model choice process during a single generation in which every
 384 copier evaluates its own optimal success-bias weight, α^* , which minimizes the expected squared error

between the estimated and the ideal trait values,

$$\alpha^* = \underset{j}{\operatorname{argmin}} \sum_{j=1}^N \frac{\alpha A_j + (1 - \alpha) K_j}{\sum_{l=1}^N \alpha A_l + (1 - \alpha) K_l} (\hat{A} - A_j)^2, \quad (24)$$

where A_j is the trait of role-model j and K_j the number of copiers that already chose role-model j .

We find that when copiers adapt their success-bias weight, it decreases with the number of copiers that have already chosen a role-model (Figure 8). Moreover, their estimation error is much lower compared to a constant success-bias weight, which gives roughly the same high estimation error to all copiers (compare Figure 8B and C): in this example, the adaptive weight estimation error converges to 0.046, whereas a constant weight gives values > 0.74 .

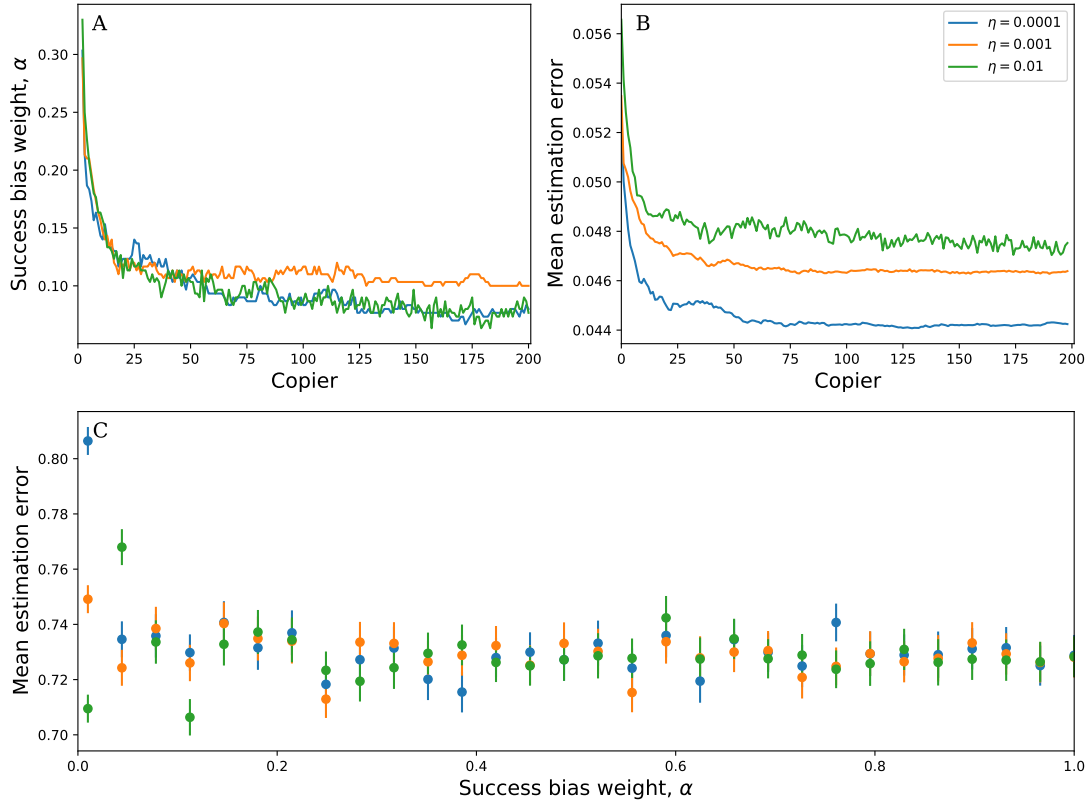


Figure 8: Advantage of an adaptive success-bias weight. Both success-bias weight α (A) and estimation error (B) decrease during the role-model choosing process, demonstrating that influence becomes more favored as more copiers have made their choice. However, when α is homogeneous (C), the mean estimation error doesn't decrease, regardless of α or η . The mean estimation error in the homogeneous α model is larger by a factor of 10 than the adaptive α model. Here, population size $N = 200$; estimation error is normally distributed $e \sim N(0, \eta^2)$ with standard deviation $\eta = 0.0001$ (blue), 0.001 (orange), 0.01 (green), plots are average of 300 simulations.

394 Discussion

During cultural transmission, cultural traits such as attitudes, values, beliefs, and behavioral patterns
396 are transmitted between individuals, for example via copying and social learning. Some cultural traits
or cultural role-models may be copied more often due to transmission biases. A common bias is
398 success bias, in which copiers are more likely to copy a successful role-model. Many models assume
that success can be accurately estimated. However, it has been suggested that because it is hard to
400 estimate success, a more common bias is *prestige bias*—a bias towards role-models perceived to be
successful in other traits (Henrich and Broesch, 2011).

402 Inspired by a model by Boyd and Richerson (1988), we developed a cultural-evolution model with
prestige bias that includes both success and influence biases, where the latter is a bias towards role-
404 models with many copiers. This is accomplished by a role-model choice process: each copier, in its
turn, chooses a role-model, and that choice is affected both by the estimated success of each potential
406 role-model and the number of copiers that already chose each role-mode (eq. 10).

This resulted in a model with two "nested" stochastic processes: the role-model choice process
408 within each generation, and the cultural-evolutionary process between generations. To simplify the
mathematical and computational analysis, we developed analytic approximations for the role-model
410 choice process using the *generalized binomial distribution* (GBD, Result 1) and the *Dirichlet-*
Multinomial distribution (DM, Corollary 3). The latter is especially useful, as it approximates the
412 entire role-model choice process and only requires us to assume that the relative effect of success and
influence is a characteristic of the role-model, rather than the copier.

414 Analyzing the model with the DM distribution, we found approximations for the fixation probability
and fixation time of a cultural trait under biased transmission in a constant environment. Our ap-
416 proximations are similar to Kimura's evolutionary-genetic approximations, such that (i) the difference
between the resident and invading cultural trait values, $1 - \beta(A)$, is equivalent to the selection coeffi-
418 cient in favor of a beneficial allele, s , and (ii) increasing the relative weight of influence versus success
bias, α , decreases the effective population size, N_e (Figure 6).

420 We also analyzed a cyclic changing environment in which the identity of the success-biased trait
switches after a fixed amount of generations (Figure 7). We find that, similarly to the constant
422 environment approximation, a change in the success-bias weight α has no negative effects on the
goodness-of-fit of the approximation to simulation results. We also showed that this approximation is
424 more sensitive to changes in the success coefficient β than the constant environment approximation,
and a lower value is required to have a good fit. The ratio between the number of generation in which
426 the rare phenotype is under positive transmission bias and the number of generations in which it is
under negative bias does not affect the goodness-of-fit of the approximation.

428 We also examined a scenario in which copiers can adapt their success-weight bias, α , to minimize
their copying error, i.e., copy trait values closer to the optimal value. We found that as the role-model
430 choice process proceeds (that is, more copiers make their choices), both the success-bias weight

(adapted by copiers) and the estimation error decrease. The latter is significantly lower compared to a
432 population using a constant, fixed success-bias weight, regardless of the value of the constant weight
(Figure 8). This suggests that the later a copier makes its choice, the more it should rely on choices
434 of previous copiers, and the less it should rely on its own estimation. The rationale, then, is that the
more information a copier has, e.g. by using others as information sources, the more informative and
436 effective his choice can be.

Prestige in the literature. According to King and Cowlishaw (2009), there are two main approaches
438 to group decision making in nature: leadership and consensus. Leaders would usually be high-ranking
members of the group: elders, individuals with many kin relations, or individuals possessing other
440 dominant traits. King and Cowlishaw (2009) describe benefits for the closest associates of a dominant
baboon, such as protection from predators. In some species, like the females of *Eulemur fulvus rufus*
442 (Red lemur), leaders may arise due to nutritional needs, and not due to possessing superior traits
(Erhart and Overdorff, 1999). In humans, leadership also has its costs and benefits. Leaders can
444 make decisions that would most benefit them and their closest followers, while still maintaining group
cohesion. However, wrong decision making that harms the group could result in negative effects
446 for the leader. In modern society, many humans strive for prestigious positions, as they may reap
rewards greater than the risk and costs to achieve them, or due to individual personality and pressure
448 or education from the family.

Henrich and Gil-White (2001) suggested that there are two types of leadership: prestige-based and
450 dominance-based leadership. In the latter, social status is acquired through aggression, intimidation
and violence. It is also more common than prestige in non-human animals. Their definition of prestige
452 seems to overlap with ours. They suggest prestige is composed of estimation in the eyes of other people
(similar to our success bias) and commanding position in people's minds, i.e., the number of followers
454 other people think one has, which they define as *influence*, similar to our influence bias. Henrich and
Gil-White (2001) show that prestige can evolve via natural selection as an efficient process to extract
456 reproductive benefit from the flow of socially transmitted information. They back up their claims by
creating testable predictions and using evidence from the literature and experiments of social sciences.
458 Simply put, prestige can naturally evolve when social learning already exists to save costs of individual
learning.

460 Furthermore, according to Boyd and Henrich (2002), the process of cultural evolution doesn't require
accurate replication of cultural traits, addressed as "representation" in their paper. They base their
462 assumption on three claims: mental representations are non-discrete, cultural transmission is highly
inaccurate, and mental representations are not replicated, but rather are 'reconstructed' through an
464 inferential process that is strongly affected by cognitive 'attractors.' They describe three different
models to support their points. We see a high similarity between their model and ours. Like them, we
466 treat the cultural trait as non-discrete (in the main model, before simplifying it to facilitate analysis).
We also assume error in estimation, such that copiers do not precisely replicate their role-models, but
468 rather reconstruct them to create a potentially different trait, which is their representation of the role-

model's trait. In addition, the inferential process that they describe as strongly affected by cognitive
470 attractors may include our definition of influence bias.

Empirical evidence of prestige bias Chudek et al. (2012) report the first direct tests in children that
472 suggest the existence of prestige bias, defined as the tendency to learn from individuals to whom
others have preferentially attended, learned, or deferred. Their definition of prestige is similar to our
474 influence bias. They showed that the odds of 3-4 years-old children learning from an adult role-model
to whom bystanders had previously preferentially attended for 10 seconds were more than twice those
476 of their learning from a role-models whom bystanders ignored. They also note that prestige effects are
domain sensitive: they found that prestigious role-models were attended more when demonstrating
478 artifact use, whereas role-models presenting food preferences had less attendants, suggesting that the
domain itself (artifact use vs. food preference) can affect the attendance, and hence the prestige of the
480 role-model. This lead Chudek et al. (2012) to suggest that when the trait is costly to learn individually,
prestige will have a stronger bias. It would be interesting to include costs in our model to try and
482 observe these effects and dynamics in a large population.

According to Henrich and Broesch (2011), evolutionary theorists propose that natural selection has
484 favored the emergence of psychological biases for learning from those individuals most likely to
possess adaptive information. Thus, they studied Fijian villages to examine if and how such biases
486 emerge in a small-scale society. They found that Fijian villagers are more likely to learn from role-
models perceived as more successful/knowledgeable, both within and across domains. Their research
488 thus suggests that copying from those perceived as successful, rather than actually are successful, is a
common phenomena. They show that the social networks representing copier–role-model relationships
490 are centralized, suggesting that it is consistent with the prediction that people substantially share
notions about who is a good cultural model, but that individuals' role-model selections are influenced
492 by multiple factors.

We can also find prestige bias in more modern domains such as western medicine. Norredam and
494 Album (2007) examined literature from 1950 to 2005 on the effects of prestige on medicinal practices.
They found that active, specialized, biomedical, and high-technological types of medicine on organs
496 in the upper part of the bodies of young and middle-aged people were accorded high levels of prestige,
whereas medicine and practices with opposite characteristics had low levels of prestige. For example,
498 they found that surgery counts as the most prestigious specialty, while psychiatry is the less prestigious.
In addition, doctors tend to rank practices that require more time to master as more prestigious, while
500 other procedures that are considered easier to master are less prestigious. This means that there may be
very important practices that are neglected due to prestige bias. They concluded that such differences
502 in prestige may bear consequences for actual priority setting in healthcare systems.

Prestige bias can help to cheaply estimate and acquire knowledge, which may facilitate survival and
504 reproduction. However, it is not always the case, and there could be negative repercussions to this
bias, such as invasion of maladaptive traits. Takahashi and Ihara (2019) mention that social learning
506 not only takes the form of random copying of other individuals, but also involves learners' choice of

what to learn and from whom to learn. They suggest a best-of-k model where an individual samples k role-models and chooses the one he deems most "successful". They mentioned that a previous mathematical analysis has shown that it may sometimes result in maladaptive cultural evolution when the payoffs associated with cultural variants vary stochastically. In such a case, learners may be selectively disfavored and in the long run replaced by unbiased learners, who simply copy someone chosen at random. They developed new mathematical models that are simpler and mathematically tractable. They found that best-of-k learning, unlike unbiased learning, can facilitate the invasion of an on average inferior variant that sometimes gives a very high payoff. Our model, which includes influence bias, is consistent with this claim. When a maladaptive trait is "piggybacking" a role-model with high influence, this trait could spread in the population. In addition, they show that best-of-k learning can be stable against invasion by unbiased learning if social learning is sometimes combined with individual learning. Our model only includes social learning, and not individual learning, but it could be interesting to combine it with individual learning and see how it affects the dynamics.

Prestige bias can also accelerate reversal of harmful traditions such as child marriage and domestic violence. Efferson et al. (2020) suggest a *spillover* mechanism, in which an intervention affects a large enough group in a target population, so that others not included in the intervention also change their behavior. They found that there are individuals who act as *agents*, who are often looked upon, and therefore they are ideal targets for interventions. This is similar to influential role-models in our model, where a prestigious individual will be copied more often, and will therefore spread his trait faster and wider in the population. They also suggest a way to use this phenomena to change existing traditions in a population. It is very clear however, that just as it can be used to end harmful traditions, the same agents could start harmful traditions.

Dunbar (2009) hypothesized that larger, more complex brains can store and manage more information and in turn, this information can support the costs of a larger brain. Following up on this, Muthukrishna and Henrich (2016) offered that prestige can directly affect human physical evolution. They present a concept called *cultural brains*—brains that evolved primarily for the acquisition of adaptive knowledge. They then develop a model that predicts a strong relationship between brain size and group size, because group size also provides access to more adaptive knowledge. They also presented the *cumulative cultural brain* hypothesis, which proposes that human brains have evolved with an ability and tendency for selective, high-fidelity social learning. As part of this process, there are a variety of strategies and biases that have evolved to hone in on the most adaptive knowledge. These strategies and biases include direct and indirect cues of the popularity of cultural traits (e.g. success and prestige biases). They suggest that one of the reasons for the extreme increase in brain size in humans is the ability to "cheaply" acquire adaptive knowledge via transmission biases such as prestige.

Further work. One path forward is an analysis of the dynamics of the adaptive success-bias weight model, in which every copier chooses its α . It would be interesting to see if the mean estimation error and the adaptive weight, α^* , are converging to specific values, and how they are affected by the model parameters. It may also be possible to relax the assumptions required for our approximations,

such as homogeneous estimation error and success-bias weight. Lastly, it would be interesting to
546 analyze the continuous model and determine how much it differs from the dichotomous model.

Another way to expand our model is to account for the two types of prestige or leadership suggested
548 by Van Vugt and Smith (2019) that are attributed to Confucius and Machiavelli. Confucius viewed
leaders as role-models who exercise influence through possessing superior knowledge, skills, and
550 (outstanding) personal qualities. This fits the success bias in our model. In contrast, Machiavelli
viewed leaders as rulers who exercise influence by imposing costs through (the threat of) punishment
552 and violence. Van Vugt and Smith (2019) suggest that these two opposing views are both partially
supported by the available evidence but each one on its own offers an incomplete view of the complex
554 and dynamic concept of leadership. Several adjustments could be made so that our model reflects
these leadership styles, such as assuming there is a correlation between phenotype to leadership style.
556 The emerging cultural-evolutionary dynamics and their dependence on the costs and benefits are
intriguing.

558 So far we examined different variations of our model using a Wright-Fisher model as our basis, mainly
because we wanted to base our model on a known one from the literature, namely Boyd and Richerson
560 (1988) indirect bias model. When approximating our model by comparing it to a Polya Urn process,
we noticed that our model would be identical to a model where every copier that chose a role-model
562 would copy its trait and join the role-models during the selection process. Based on this, we theorize
that when using a Moran model as basis instead of a Wright-Fisher mode, we would get very similar,
564 and perhaps identical results. It would be interesting to run simulations and analyze that model, to
prove or disprove our hypothesis.

566 **Conclusions.** Here, we studied a model of cultural evolution under two transmission biases: the
commonly studied success bias, together with influence bias, which has so far received less attention.
568 We found approximations for this complex dynamics. We then showed that success bias affects the
evolutionary dynamics much like natural selection does, whereas influence bias has a similar effect to
570 random genetic drift. We also find a clear advantage to individuals that can choose the relative weight
of the two biases.

572 **Acknowledgements**

We thank Marcus Feldman, Martin Pontz, and Tal Simon for discussions and comments. This work was
574 supported in part by the Israel Science Foundation (YR 552/19), Minerva Stiftung Center for Lab Evolution
(YR), and the John Templeton Foundation (YR 61809).

- Aljadeff, N., Giraldeau, L.-A., and Lotem, A. (2020). Competitive advantage of rare behaviours induces adaptive diversity rather than social conformity in skill learning. *Proceedings of the Royal Society B*, 287(1933):20201259.
- Anagnostopoulos, A., Kumar, R., and Mahdian, M. (2008). Influence and correlation in social networks. In *Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 7–15.
- Andersson, M. B. (1994). *Sexual selection*. Princeton University Press.
- Battesti, M., Moreno, C., Joly, D., and Mery, F. (2012). Spread of social information and dynamics of social transmission within drosophila groups. *Current Biology*, 22(4):309 – 313.
- Borofsky, T. M. and Feldman, M. W. (2022). Success-biased social learning in a one-consumer, two-resource model. *Theor. Popul. Biol.*, 146:29–35.
- Boyd, R. and Henrich, J. (2002). On modeling cognition and culture: Why cultural evolution does not require replication of representations. *Journal of Cognition and Culture*, 2(2):87 – 112.
- Boyd, R. and Richerson, P. J. (1988). *Culture and the evolutionary process*. University of Chicago press.
- Cavalli-Sforza, L. L. and Feldman, M. W. (1981). *Cultural transmission and evolution: A quantitative approach*. Number 16. Princeton University Press.
- Chudek, M., Heller, S., Birch, S., and Henrich, J. (2012). Prestige-biased cultural learning: bystander’s differential attention to potential models influences children’s learning. *Evolution and Human Behavior*, 33(1):46–56.
- Cohen, D., Lewin-Epstein, O., Feldman, M. W., and Ram, Y. (2021). Non-vertical cultural transmission, assortment and the evolution of cooperation. *Proceedings of the Royal Society B*, 288(1951):20203162.
- Creanza, N., Kolodny, O., and Feldman, M. W. (2017). Cultural evolutionary theory: How culture evolves and why it matters. *Proceedings of the National Academy of Sciences USA*, 114(30):7782–7789.
- Denton, K. K., Liberman, U., and Feldman, M. W. (2021). On randomly changing conformity bias in cultural transmission. *Proceedings of the National Academy of Sciences USA*, 118(34).
- Denton, K. K., Ram, Y., and Feldman, M. W. (2022). Conformity and content-biased cultural transmission in the evolution of altruism. *Theoretical population biology*, 143:52–61.
- Denton, K. K., Ram, Y., Liberman, U., and Feldman, M. W. (2020). Cultural evolution of conformity and anticonformity. *Proceedings of the National Academy of Sciences USA*, 117(24):13603–13614.

- Diga, M. and Kelleher, T. (2009). Social media use, perceptions of decision-making power, and public relations roles. *Public Relations Review*, 35(4):440–442.
- Drezner, Z. and Farnum, N. (1993). A generalized binomial distribution. *Communications in Statistics - Theory and Methods*, 22(11):3051–3063.
- Dunbar, R. (2009). The social brain hypothesis and its implications for social evolution. *Annals of Human Biology*, 36(5):562–572.
- Durrett, R. (1999). *Essentials of stochastic processes*, volume 1. Springer.
- Durrett, R. (2008). *Probability models for DNA sequence evolution*, volume 2. Springer.
- Efferson, C., Vogt, S., and Fehr, E. (2020). The promise and the peril of using social influence to reverse harmful traditions. *Nature human behaviour*, 4(1):55–68.
- Erhart, E. M. and Overdorff, D. J. (1999). Female coordination of group travel in wild propithecus and eulemur. *International Journal of Primatology*, 20(6):927–940.
- Fogarty, L., Wakano, J. Y., Feldman, M. W., and Aoki, K. (2017). The driving forces of cultural complexity. *Human Nature*, 28(1):39–52.
- Frigyik, B. A., Kapila, A., and Gupta, M. R. (2010). Introduction to the dirichlet distribution and related processes. *Department of Electrical Engineering, University of Washington, UWEETR-2010-0006*, (0006):1–27.
- Henrich, J. and Broesch, J. (2011). On the nature of cultural transmission networks: evidence from fijian villages for adaptive learning biases. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 366(1567):1139–1148.
- Henrich, J. and Gil-White, F. J. (2001). The evolution of prestige: Freely conferred deference as a mechanism for enhancing the benefits of cultural transmission. *Evolution and human behavior*, 22(3):165–196.
- Horner, V., Proctor, D., Bonnie, K. E., Whiten, A., and de Waal, F. B. (2010). Prestige affects cultural learning in chimpanzees. *PloS one*, 5(5):e10625.
- Hunter, J. D. (2007). Matplotlib: A 2d graphics environment. *Computing in science & engineering*, 9(03):90–95.
- Kendal, R., Hopper, L. M., Whiten, A., Brosnan, S. F., Lambeth, S. P., Schapiro, S. J., and Hoppitt, W. (2015). Chimpanzees copy dominant and knowledgeable individuals: implications for cultural diversity. *Evolution and Human Behavior*, 36(1):65–72.
- Kimura, M. (1962). On the probability of fixation of mutant genes in a population. *Genetics*, 47(6):713.
- Kimura, M. and Ohta, T. (1969). The average number of generations until fixation of a mutant gene in a finite population. *Genetics*, 61(3):763.

- King, A. J. and Cowlshaw, G. (2009). Leaders, followers, and group decision-making. *Communicative & Integrative Biology*, 2(2):147–150.
- Kolodny, O., Feldman, M. W., Lotem, A., and Ram, Y. (2022). Differential application of cultural practices at the family and individual levels may alter heritability estimates. *Behav. Brain Sci.*, 45:e167.
- Lee, W., Xiong, L., and Hu, C. (2012). The effect of facebook users’ arousal and valence on intention to go to the festival: Applying an extension of the technology acceptance model. *International Journal of Hospitality Management*, 31(3):819–827.
- McComb, K., Moss, C., Durant, S. M., Baker, L., and Sayialel, S. (2001). Matriarchs as repositories of social knowledge in african elephants. *Science*, 292(5516):491–494.
- Mesoudi, A. and O’Brien, M. J. (2008). The cultural transmission of great basin projectile-point technology II: An agent-based computer simulation. *Am. Antiq.*, 73(4):627–644.
- Molleman, L., Pen, I., and Weissing, F. J. (2013). Effects of conformism on the cultural evolution of social behaviour. *PloS one*, 8(7):e68153.
- Muthukrishna, M. and Henrich, J. (2016). Innovation in the collective brain. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 371(1690):20150192.
- Norredam, M. and Album, D. (2007). Prestige and its significance for medical specialties and diseases. *Scandinavian journal of public health*, 35(6):655–661.
- Otto, S. P. and Whitlock, M. C. (2006). Fixation probabilities and times. *Encyclopedia of Life Sciences*.
- Peng, S., Zhou, Y., Cao, L., Yu, S., Niu, J., and Jia, W. (2018). Influence analysis in social networks: A survey. *Journal of Network and Computer Applications*, 106:17–32.
- Ram, Y., Liberman, U., and Feldman, M. W. (2018). Evolution of vertical and oblique transmission under fluctuating selection. *Proceedings of the National Academy of Sciences USA*, 115(6):E1174–E1183.
- Rendell, L., Boyd, R., Cownden, D., Enquist, M., Eriksson, K., Feldman, M. W., Fogarty, L., Ghirlanda, S., Lillicrap, T., and Laland, K. N. (2010). Why copy others? insights from the social learning strategies tournament. *Science*, 328(5975):208–213.
- Takahashi, T. and Ihara, Y. (2019). Cultural and evolutionary dynamics with best-of-k learning when payoffs are uncertain. *Theoretical Population Biology*, 128:27–38.
- Truskanov, N., Emery, Y., and Bshary, R. (2020). Juvenile cleaner fish can socially learn the consequences of cheating. *Nat. Commun.*, 11(1):1–9.
- Van Der Walt, S., Colbert, S. C., and Varoquaux, G. (2011). The numpy array: a structure for efficient numerical computation. *Computing in science & engineering*, 13(2):22–30.

- Van Rossum, G. et al. (2007). Python programming language. In *USENIX annual technical conference*, volume 41, pages 1–36.
- Van Vugt, M. and Smith, J. E. (2019). A dual model of leadership and hierarchy: Evolutionary synthesis. *Trends in Cognitive Sciences*, 23(11):952–967.
- Whitehead, H. (2017). Gene–culture coevolution in whales and dolphins. *Proceedings of the National Academy of Sciences USA*, 114(30):7814–7821.

Appendices

Appendix A General binomial distribution approximation

Proving $E[K_{Nj}] = \alpha_j \cdot \beta(A_j + e) / \overline{\alpha \cdot \beta(A + e)}$, where the averaging in the denominator is over the role-models index, j .

Proof. The initial prestige of role-model j based on eq. (10) is

$$G_{1,j} = \frac{\alpha_j \cdot \beta(A_j + e)}{\sum_{m=1}^N \alpha_m \cdot \beta(A_m + e)} . \quad (\text{A1})$$

The denominator of eq. (A1) can also be formulated as:

$$\sum_{m=1}^N \alpha_m \beta(A_m + e) = N \cdot \overline{\alpha \cdot \beta(A + e)} , \quad (\text{A2})$$

where $\overline{\alpha \beta(A + e)}$ is the mean value of $\alpha_m \cdot \beta(A_m + e)$ for all m . Using eq. (A2) and **Corollary 1** we get,

$$E[K_{Nj}] = \alpha_j \cdot \beta(A_j + e) \left/ \overline{\alpha \cdot \beta(A + e)} \right. , \quad (\text{A3})$$

Appendix B Drift and diffusion in a constant environment

Proving drift and diffusion terms in a constant environment. Let x and x' be the frequency of type \hat{A} in a population with N individuals in the current and next generation, and β is the success coefficient of phenotype A , $\beta = \beta(A) < \beta(\hat{A}) = 1$. Then,

$$E[x' - x] \approx x(1 - x)(1 - \beta) , \quad V(x' - x) \approx x(1 - x) \left(\frac{1}{\alpha N + (1 - \alpha)} \right) .$$

Proof. Let X be the number of individuals of type \hat{A} such that $x = X/N$. X' is the number of individuals with \hat{A} in the next generation. The expected number of individuals is (due to the DM approximation),

$$E[X'] = N \frac{\alpha_1}{\alpha_1 + \alpha_2} , \quad (\text{B1})$$

where $\alpha_1 = \alpha'X$ and $\alpha_2 = \alpha'(N - X)\beta$, from eq. (11). To use frequencies instead of counts, $E[x'] = E[X'/N] = \frac{1}{N}E[X']$. Putting it together,

$$\begin{aligned} E[x'] &= \frac{1}{N}N \frac{\alpha'xN}{\alpha'xN + \alpha'N(1-x)\beta} = \frac{x}{x + (1-x)\beta} \\ &= \frac{x}{x + (1-x) - (1-x) + (1-x)\beta} = x \frac{1}{1 - (1-x)(1-\beta)} \\ &= x(1 + (1-x)(1-\beta) + o(\beta)) = x + x(1-x)(1-\beta) + o(\beta), \end{aligned} \quad (\text{B2})$$

following Durrett (2008, p. 253, ch 7.2) and because $1/(1-y) = 1 + y + y^2 + \dots$

By definition, x is constant, so $E[x] = x$. We therefore have

$$E[x' - x] = E[x'] - E[x] = x(1-x)(1-\beta) + o(1-\beta), \quad (\text{B3})$$

which gives us the drift term of the diffusion equation.

Using the variance of the DM distribution,

$$V(X') = N \frac{\alpha_1}{\alpha_1 + \alpha_2} \left(1 - \frac{\alpha_1}{\alpha_1 + \alpha_2}\right) \left(\frac{N + \alpha_1 + \alpha_2}{1 + \alpha_1 + \alpha_2}\right). \quad (\text{B4})$$

Again, we want to use frequencies so we have $V(X'/N) = \frac{1}{N^2}V(X')$. Putting it together with our model notations,

$$V(x') = \frac{1}{N^2}N \frac{x}{x + (1-x)\beta} \left(1 - \frac{x}{x + (1-x)\beta}\right) \left(\frac{N + \alpha'xN + \alpha'N(1-x)\beta}{1 + \alpha'xN + \alpha'N(1-x)\beta}\right). \quad (\text{B5})$$

Following Durrett (2008, ch 7.2), we assume $\beta \approx 1$, such that

$$\frac{x}{x + (1-x)\beta} \approx x \quad (\text{B6})$$

and for the entire variance expression we get

$$V(x') \approx \frac{1}{N}x(1-x) \left(\frac{N + \alpha'xN + \alpha'N - \alpha'xN}{1 + \alpha'xN + \alpha'N - \alpha'xN}\right) = x(1-x) \left(\frac{1 + \alpha'}{1 + \alpha'N}\right). \quad (\text{B7})$$

Now because x is a constant, $V(x) = 0$, and therefore

$$V(x' - x) = V(x') - V(x) \approx x(1-x) \left(\frac{1 + \alpha'}{1 + \alpha'N}\right). \quad (\text{B8})$$

α' is the odds ratio of the bias weight,

$$\alpha' = \frac{\alpha}{1 - \alpha}. \quad (\text{B9})$$

Combining eq. (B8) and eq. (B9) we get:

$$V(x' - x) \approx x(1 - x) \left(\frac{1 + \frac{\alpha}{1-\alpha}}{1 + \frac{\alpha}{1-\alpha}N} \right) = x(1 - x) \left(\frac{1}{\alpha N + (1 - \alpha)} \right). \quad (\text{B10})$$

This gives the diffusion term of the diffusion equation.

Appendix C Drift and diffusion in a changing environment

Proving drift and diffusion terms in a changing environment. Let x be the initial frequency of the invading phenotype and X_t is the number of individuals with the phenotype at time t . Then,

$$E[X_t/N - x] \simeq x(1 - x)S_t/N_e, \quad \text{and} \quad V(X_t/N - x) \simeq tx(1 - x)/N_e,$$

where $S_t = \sum_{i=1}^t N(1 - \beta_i)$.

Proof. Let $s_t = N(1 - \beta_t)$, and $S_n = \sum_{i=1}^n s_i$, where β_t is $\beta(A)$ at generation t . We prove by induction both terms in eq. (22). From eq. (B3) we know that

$$E \left[\frac{X_{t+1}}{N} - \frac{X_t}{N} \middle| X_t \right] = \frac{X_t}{N} \left(1 - \frac{X_t}{N} \right) (1 - \beta_{t+1}) = \frac{1}{N} \frac{X_t}{N} \left(1 - \frac{X_t}{N} \right) s_{t+1}. \quad (\text{C1})$$

Also note that using the definition of $V(y) = E[y^2] - (E[y])^2$

$$\begin{aligned} E \left[\frac{X_t}{N} \left(1 - \frac{X_t}{N} \right) \right] &= E \left[\frac{X_t}{N} - \left(\frac{X_t}{N} \right)^2 \right] = E \left[\frac{X_t}{N} \right] - E \left[\left(\frac{X_t}{N} \right)^2 \right] \\ &= E \left[\frac{X_t}{N} \right] - V \left(\frac{X_t}{N} \right) - \left(E \left[\frac{X_t}{N} \right] \right)^2. \end{aligned} \quad (\text{C2})$$

We can now use the induction assumption of $V(\frac{X_t}{N})$ to get

$$E \left[\frac{X_t}{N} \left(1 - \frac{X_t}{N} \right) \right] \simeq E \left[\frac{X_t}{N} \right] \left(1 - E \left[\frac{X_t}{N} \right] \right) - \frac{1}{N_e} tx(1 - x). \quad (\text{C3})$$

From eq. (C1) we know that

$$\begin{aligned} E \left[\frac{X_{t+1}}{N} - \frac{X_t}{N} \right] &= \frac{1}{N} s_{t+1} E \left[\frac{X_t}{N} \left(1 - \frac{X_t}{N} \right) \right] \simeq \frac{1}{N} s_{t+1} \left(E \left[\frac{X_t}{N} \right] \left(1 - E \left[\frac{X_t}{N} \right] \right) - \frac{1}{N_e} tx(1 - x) \right) \\ &\simeq \frac{1}{N} s_{t+1} \cdot E \left[\frac{X_t}{N} \right] \left(1 - E \left[\frac{X_t}{N} \right] \right) - \frac{1}{N_e N} s_{t+1} tx(1 - x). \end{aligned} \quad (\text{C4})$$

Now we omit $O(\frac{1}{Ne \cdot N})$ and get

$$E \left[\frac{X_{t+1}}{N} - \frac{X_t}{N} \right] \simeq \frac{1}{N} s_{t+1} \cdot E \left[\frac{X_t}{N} \right] \left(1 - E \left[\frac{X_t}{N} \right] \right) . \quad (C5)$$

We now look at the induction assumption to see that

$$E \left[\frac{X_t}{N} - x \right] = E \left[\frac{X_t}{N} \right] - E[x] = E \left[\frac{X_t}{N} \right] - x , \quad (C6)$$

so using the assumption we get

$$\begin{aligned} E \left[\frac{X_t}{N} \right] &\simeq \frac{1}{N} S_t x (1 - x) + x , \\ 1 - E \left[\frac{X_t}{N} \right] &\simeq 1 - \frac{1}{N} S_t x (1 - x) + x . \end{aligned} \quad (C7)$$

We use both expressions in eq. (C5) and get

$$\begin{aligned} E \left[\frac{X_{t+1}}{N} - \frac{X_t}{N} \right] &\simeq \frac{1}{N} s_{t+1} \left(\frac{1}{N} S_t x (1 - x) + x \right) \left(1 - \frac{1}{N} S_t x (1 - x) + x \right) \\ &\simeq \frac{1}{N} s_{t+1} \cdot x (1 - x) , \end{aligned} \quad (C8)$$

after again omitting $O(\frac{1}{N^2})$ terms. To conclude the proof, we note that

$$E \left[\frac{X_{t+1}}{N} - x \right] = E \left[\frac{X_{t+1}}{N} - \frac{X_t}{N} \right] + E \left[\frac{X_t}{N} - x \right] , \quad (C9)$$

so again using the induction assumption, together with eq. (C8) we get

$$\begin{aligned} E \left[\frac{X_{t+1}}{N} - x \right] &\simeq \frac{1}{N} s_{t+1} \cdot x (1 - x) + \frac{1}{N} S_t \cdot x (1 - x) \\ &\simeq \frac{1}{N} x (1 - x) (S_t + s_{t+1}) \simeq \frac{1}{N} S_{t+1} x (1 - x) , \end{aligned} \quad (C10)$$

which proves the drift term.

For the diffusion term, we use a property of variance,

$$V \left(\frac{X_{t+1}}{N} \right) = E \left[V \left(\frac{X_{t+1}}{N} \middle| X_t \right) \right] + V \left(E \left[\frac{X_{t+1}}{N} \middle| X_t \right] \right) . \quad (C11)$$

Using eq. (C1) we see that

$$\begin{aligned} E \left[\frac{X_{t+1}}{N} \middle| X_t \right] - E \left[\frac{X_t}{N} \middle| X_t \right] &= \frac{1}{N} s_{t+1} \frac{X_t}{N} \left(1 - \frac{X_t}{N} \right) \\ E \left[\frac{X_{t+1}}{N} \middle| X_t \right] &= \frac{X_t}{N} + \frac{1}{N} s_{t+1} \frac{X_t}{N} \left(1 - \frac{X_t}{N} \right) . \end{aligned} \quad (C12)$$

Using eq. (B10) we get

$$V\left(\frac{X_{t+1}}{N}\middle|X_t\right) = \frac{1}{N_e} \frac{X_t}{N} \left(1 - \frac{X_t}{N}\right), \quad (\text{C13})$$

and using the equation $y'(1 - y') \simeq y(1 - y)$ on the first part of eq. (C11) we get

$$E\left[V\left(\frac{X_{t+1}}{N}\middle|X_t\right)\right] = \frac{1}{N_e} E\left[\frac{X_t}{N} \left(1 - \frac{X_t}{N}\right)\right] \simeq \frac{1}{N_e} x(1 - x). \quad (\text{C14})$$

Moving on to simplify the second part of eq. (C11) using eq. (C12),

$$V\left(E\left[\frac{X_{t+1}}{N}\middle|X_t\right]\right) = V\left(\frac{X_t}{N} + \frac{1}{N} s_{t+1} \frac{X_t}{N} \left(1 - \frac{X_t}{N}\right)\right) \quad (\text{C15})$$

Now, because $\frac{X_t}{N}$ is a frequency, i.e $0 \leq X_t/N \leq 1$, we know that $V\left(\frac{X_t}{N} \left(1 - \frac{X_t}{N}\right)\right) \leq \frac{1}{4}$. We therefore find that

$$V\left(\frac{1}{N} s_{t+1} \frac{X_t}{N} \left(1 - \frac{X_t}{N}\right)\right) \leq \frac{1}{4N^2} s_{t+1}^2, \quad (\text{C16})$$

and so it can be ignored. Combining our equations we get

$$V\left(E\left[\frac{X_{t+1}}{N}\middle|X_t\right]\right) = V\left(\frac{X_t}{N}\right) + O\left(\frac{1}{N^2}\right) \simeq V\left(\frac{X_t}{N}\right). \quad (\text{C17})$$

Using the induction assumption and eq. (C14),

$$V\left(\frac{X_{t+1}}{N}\right) \simeq \frac{1}{N_e} x(1 - x) + \frac{1}{N_e} t x(1 - x) \simeq \frac{1}{N_e} x(1 - x)(t + 1) \quad (\text{C18})$$

which proves the diffusion term.