

# Prestige bias in cultural evolutionary dynamics

Saar Egozi<sup>1,3</sup> and Yoav Ram<sup>2,3,†</sup>

<sup>1</sup>School of Computer Science, Reichman University, Herzliya, Israel

<sup>3</sup>School of Zoology, Faculty of Life Sciences, Tel Aviv University, Tel Aviv, Israel

<sup>3</sup>Sagol School of Neuroscience, Tel Aviv University, Tel Aviv, Israel

<sup>†</sup>Corresponding author: yoav@yoavram.com

April 16, 2023

## Abstract

If the traits of more successful individuals are more likely to be adopted, the resulting cultural transmission is described as being success-biased. Prestige may be used as a proxy for success. 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 that role-model. The effect of prestige on cultural evolutionary dynamics is analyzed using mathematical analysis and stochastic simulations. Analytic approximations to the stochastic role-model choice process facilitate the mathematical analysis and reduce the computational complexity of simulations. Approximations are given to the fixation probability and the fixation time of an invading cultural trait in different environments. We show that indirect success bias effectively plays the role of natural selection, whereas influence bias effectively plays the role of genetic drift. Influence bias, which may be strong in communities where social media are prevalent, also accelerates the evolutionary dynamics, as can be expected in a rich-get-richer process.

# Introduction

22 Cultural transmission of attitudes, preferences, beliefs, norms, and behaviors may combine vertical  
transmission, in which parents transmit to their offspring; oblique transmission, in which adults  
24 (teachers, leaders, and even strangers) transmit to unrelated offspring; and horizontal transmission, in  
which individuals from the same age cohort transmit to one another [4]. It has been demonstrated that  
26 non-vertical cultural transmission can maintain maladaptive traits, which can be beneficial in changing  
environments [21, 55].

28 Transmission biases may cause a cultural trait to have a higher rate of transmission than its frequency in  
the population. *Success bias* occurs when individuals prefer to copy from role models that demonstrate  
30 success in some activity, such as fishing, growing yams, using medicinal plants [24], or hunting [54],  
and it can increase the probability of learning a trait that is present in those successful individuals [53].  
32 Indeed, in a tournament between learning strategies [7], most winning strategies included a mixture  
of success-biased social learning and individual learning, implying that success-biased learning is a  
34 good strategy, but that by itself it is not enough to best other strategies, even when success is measured  
accurately.

36 Boyd and Richerson [5, Ch. 5] suggested that the assessment of success can be divided to three  
categories: *direct bias*, *indirect bias* and *frequency-dependent bias*. Direct bias occurs when one  
38 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 try the observed paddle grips to determine  
40 which grip is better. Frequency-dependent bias occurs when the probability of copying a phenotype  
is higher or lower than the frequency of the phenotype among demonstrators. For example, suppose  
42 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  
44 by 40% of copiers, then transmission is under negative frequency bias, or *anti-conformity* [15]. The  
effects of conformity and anti-conformity on cultural evolution have been studied with both models  
46 [36, 37, 52] and experiments [3]. Indirect bias occurs when a copier uses some observed phenotype to  
evaluate the attractiveness of a potential role-model. For example, an observer may copy the paddle  
48 grip of the ping-pong player who scored more points in the match, thus indirectly evaluating the grip  
by the points scored. However, this may cause mismatches between the copied trait and the rest of  
50 the cultural or genetic repertoire of the individual [51]. Furthermore, Boyd and Richerson [5, Ch. 8]  
suggest that maladaptive traits may spread widely in a population if indirect bias is strong enough,  
52 e.g., by a runaway process caused by a cultural equivalent of sexual selection [8]. Indeed, helping  
behaviors can evolve due to horizontal transmission bias even without any benefit to the recipient, or  
54 when the benefit is much larger than the cost [35].

Heinrich and Broesch [24] studied such indirect success biases, which they call cross-domain success  
56 bias or *prestige bias* (e.g., great fishermen may be chosen as role-models for growing yams). They  
suggested that such biases, over generations, can lead to cultural adaptations, and that although prestige  
58 can lead to maladaptive traits spreading in the population, it can also accelerate the spread of adaptive

traits [5, Ch. 8]. Prestige bias may be more common in humans than success bias [6], and prestige is often mentioned in the cultural evolution literature, although there are few models of it.

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. [31] 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. Henrich and Gil-White [26] gave a similar definition of prestige as “freely conferred deference”, in contrast to *dominance*, and provided examples from the anthropological literature.

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-dependent bias, rather than a content-dependent bias: it does not depend on the phenotype itself but rather on the number of copiers that have already copied each role-model, which may be easier and more accurate to estimate than success. Influence bias is also frequency independent (see Corollary 1 below), and thus it differs from 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 make it especially easy to estimate the social and cultural influence individuals have over others, which can have an effect on decision making. Online social networks such as *Facebook* and *Instagram* are known to affect the influence of individuals [39, 40, 41], and specific marketing practices have been invented to capitalize on this effect [38].

In the following, we develop a stochastic model of cultural transmission with prestige bias that combines both cross-domain indirect success bias and influence bias. We develop analytic approximations for this model and analyze its dynamics. We also find approximations for the probability and time to fixation of a ‘successful’ phenotype (i.e., that is subject to success bias). Comparing these approximations to Kimura’s approximations for the fixation of a favorable allele [19, 45], we demonstrate that success and influence bias play the role of natural selection and genetic drift, respectively.

## Models

We begin with a continuous trait model with indirect bias, previously suggested by Boyd and Richerson [5]. We extend this model to include influence bias. We then develop a dichotomous trait model. Note that the indirect bias is due to an indirect evaluation, in which a certain phenotype is used to evaluate the success of potential role-models.

We implement our stochastic models and approximations, perform statistical analyses, and produce figures using Python [42] with NumPy [43] and Matplotlib [44]. Source code is available at <https://github.com/yoavram-lab/PrestigeBias>.

## 92 Continuous trait

We follow the Boyd and Richerson model [5], assuming only oblique transmission of a single trait. Consider a population of  $N$  individuals, described by a single trait that takes continuous values. Each generation,  $N$  naive individuals, or copiers, each chooses a single role-model from the entire previous generation. The copier then copies its trait value from the chosen role-model. Note that our transmission models is slightly different than previous those modeled before, e.g. [5, 37, 57], in which the population is infinite and each copier samples  $n$  role-models and then copies its trait from one or more of the sampled role-models.

Similar to a Wright-Fisher model, generations are non-overlapping, and the entire population is replaced in each generation. The population at time  $t$  can be described by  $\mathbf{A}(t) = (A_1(t), \dots, A_N(t))$  where  $A_i(t)$  is trait value of individual  $i$  at time  $t$ , and the initial population is drawn from a standard normal distribution,  $\mathbf{A}(0) \sim N(0, 1)$ . Cultural transmission is modeled by a function  $F$  such that

$$104 \quad A_i(t + 1) = F_i(\mathbf{A}(t)) . \quad (1)$$

**Success bias.** Boyd and Richerson [5, Ch. 8, p. 247-249] describe a transmission algorithm by defining  $F$ , a weighted average of the traits of all role-models, as

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

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

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

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

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

112 and  $\beta(\cdot)$  is the bias function that quantifies the success bias of a role-model [5, eq. 5.11],

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

114 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  $j$ .

116 Boyd and Richerson [5] note that the deterministic blended transmission algorithm they use has alternatives. We can develop a similar stochastic model with transmission from a single random role-model where instead of eq. (2), we define the transmission function  $F$  as a random variable with

its distribution given by

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

here  $G_{i,j}$  is the probability that copier  $i$  chooses to copy the trait of role-model  $j$ .

**Influence bias.** Here we introduce a new element to the model by assuming that in each generation, copiers choose their role-models one by one. Denote by  $K_{i,j}$  the number of copiers that choose role-model  $j$  after copier  $i$  chose a role-model. Thus,  $i$  out of  $N$  copiers had already chosen 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) with

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

where  $W_i$  is a normalizing factor that sums the numerator over all role-models ( $1 \leq j \leq N$ ) to ensure  $\sum_{j=1}^N G_{i,j} = 1$ .

Here, the bias weight  $\alpha_{i,j}$  determines the relative weighting of success and influence bias. It 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. Different individuals may evaluate the importance of success and influence differently. Additionally, we assume each role-model displays its influence and success individually. For example, individuals with more followers but lacking skill may emphasize the number of their followers rather than their skill. Finally, the trait of role-model  $j$  estimated by copier  $i$ ,  $A_{i,j}$ , remains as in eq. (4).

## Dichotomous trait

We introduce a simplified version where the trait has only two phenotypes: an optimal phenotype  $\hat{A}$  and a sub-optimal phenotype  $A$ . All role-models with the same phenotype will contribute to the

probability that that phenotype is transmitted, and thus influence is determined by the number of  
 152 copiers that have already chosen a role-model with either phenotype. In addition, for simplicity and  
 for easier mathematical analysis, we assume  $\alpha$  is homogeneous, which entails exchangeability between  
 154 role-models. Therefore, the probability that the  $i$ -th copier copies phenotype  $A$  is

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

156 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; assuming (without loss of generality) that  $\beta(\hat{A}) = 1$ ; and defining  
 158  $\alpha' = \frac{\alpha}{1-\alpha}$  as the odds ratio of the bias weight. 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.

## 160 Results

Our models are defined by two nested stochastic processes. Change over multiple generations is  
 162 described by the dynamics of the phenotype distribution at each generation,  $\{\mathbf{A}(t)\}_t$ , see eq. (1). The  
 transition from one generation to the next is described by the number of copiers each role-model has  
 164 after  $i$  copiers have chosen a role-model,  $\{\mathbf{K}_i\}_{i=1}^N$ , see eq. (7). We emphasize that the models are  
 nested:  $\mathbf{A}(t + 1)$  can be computed from  $\mathbf{A}(t)$  by evaluating  $\mathbf{K}_N$ , where  $K_{N,j}$  is the number of copiers  
 166 that chose role-model  $j$  after all copiers chose a role model. However, the former requires iterating  
 over eqs. (8) and (9). Thus, we sought to find an equivalent stochastic process that has the same joint  
 168 distribution as  $\mathbf{K}_N$ . We found two approximations for the distribution of  $\mathbf{K}_N$ , summarized here and  
 explained in detail below. In both we assume that the bias weight is either completely homogeneous,  
 170  $\alpha_{i,j} = \alpha$ , or that  $\alpha_{i,j} = \alpha_j$  is a bias of role-model  $j$  only, meaning it does not vary between copiers. Note  
 that these approximations apply for both the and the continuous trait (eq. (10)) and the dichotomous  
 172 trait (eq. (11)) models.

**Generalized binomial approximation.** The number of copiers of a specific role-model at each step,  
 174  $K_{i,j}$ , follows the *generalized binomial distribution* [16] and therefore,

(i) the expected number of copiers of role-model  $j$  equals its prestige in the eyes of the first copier,  
 176 multiplied by the total number of copiers, that is,  $E[K_{N,j}] = N \cdot G_{1,j}$  if trait estimation error is uniform  
 for all copiers ( $e = e_i$  for  $i = 1, \dots, N$ ); and (ii) the expected number of copiers of each role-model  
 178 equals its relative biased trait value, similar to the role of relative fitness in population-genetic models,  
 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  $j = 1, \dots, N$ ),  
 180 where  $\bar{\beta} = 1/N \sum_{j=1}^N \beta(A_j + e)$  is the population mean estimated trait value.

**Dirichlet-Multinomial approximation.** The role-model choice process,  $\{\mathbf{K}_i\}_{i=1}^N$ , is equivalent to a  
 182 *Pólya urn* model if trait estimation error is uniform for all copiers ( $e = e_i$  for all  $i = 1, \dots, N$ ), meaning  
 that the order of copiers does not matter.

## 184 Generalized binomial distribution

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 denoted by  $GBD(n, p, \theta)$  where  $n$  is the number of trials,  $p$  is the probability of success of the first trial, and  $\theta$  is the correlation between trials (the latter can be estimated from data, but its value is insignificant for our approximation). Note that  $\theta = 0$  gives the standard binomial distribution.

190

**Result 1** (Generalized binomial approximation). *The number of copiers of role-model  $j$  after  $i$  copiers have chosen a role-model follows the generalized binomial distribution,  $K_{i,j} \sim GBD(i, \alpha_j \cdot \beta(A_j + e), \theta)$  if  $e_i = e$  for all role-models  $i = 1, \dots, N$ , and  $\theta$  is the correlation between successive role-model choices.*

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

$$198 \quad Q_j(k, i) = P_j(S_{i,j} = 1 \mid k - 1, i - 1) \cdot Q_j(k - 1, i - 1) + P_j(S_{i,j} = 0 \mid k, i - 1) \cdot Q_j(k, i - 1), \quad (12)$$

where  $S_{i,j} = 1$  when the  $i$ -th copier chooses role-model  $j$ . Equation (12) is equivalent to eq. (2.1) in [16], which completes the proof.

This result gives the following corollary on the expected number of followers of a given role-model  $j$  by the end of the role-model choice process,  $K_{N,j}$ .

**Corollary 1.** *The expected number of copiers of role-model  $j$  after all copiers have chosen a role-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  $j$ . In addition,  $E[K_{N,j}] = \alpha_j \cdot \beta(A_j + e) / \overline{\alpha \cdot \beta(A + e)}$ , where the averaging in the denominator is over the role-models index,  $j$ .*

*Proof.* The expected value of the GBD is  $E[K_{N,j}] = N \cdot Q_j(1, 1)$ , see Drezner and Farnum [16, eq. (2.3)]. Here,  $Q_j(1, 1)$  is the initial probability to choose role-model  $j$ , before any role-model choices are made, such that  $Q_j(1, 1) = G_{1,j}$  by definition. The rest of the proof is in Appendix A.

From Corollary 1 we see that if there is no success bias, meaning  $Q_j(1, 1) = q$  for  $j$  in  $1, \dots, N$ , then the expected number of copiers will be Binomial distributed. This proves that influence bias is not frequency dependent.

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

$$216 \quad E[K_{N,j}] = N \cdot \frac{\alpha \cdot \beta(A_j + e)}{\sum_{m=1}^N \alpha \cdot \beta(A_m + e)} = \beta(A_j + e) \left/ \overline{\beta(A + e)} \right., \quad (13)$$

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

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

220 **Numerical validation.** To validate that GBD approximation for the number of copiers of a role-  
 model is correct (eq. (13)), we ran 1,000 simulations of the full model, and compared the results with  
 222 Corollary 1. We compare the distribution of number of copiers by plotting the histograms of both our  
 simulations results and the expected values based on Corollary 1.

224 Although basic, Figure S1 shows good fit of the GB approximation. This validation is initial, and we  
 do more extensive validations on the Dirichlet-Multinomial approximation, because that is what we  
 226 will use in our analysis.

## Dirichlet-Multinomial distribution approximation

228 **Pólya urn model.** This stochastic process consists of  $N$  draws from an urn with an initial number of  
 colored balls of  $M$  colors. When a ball is drawn, it is then placed back in the urn together with an  
 230 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  
 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  
 232 0 otherwise. The probability that  $S_{i,j} = 1$  given  $\mathbf{U}_{i-1}$  is

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

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

236

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



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

$$242 \quad G_{i,j} = \frac{\alpha' \beta(A'_j) + K_{i-1,j}}{\sum_{m=1}^N [\alpha' \beta(A'_m) + K_{i-1,m}]} = \frac{\alpha' \beta(A'_j) + K_{i-1,j}}{i - 1 + \sum_{m=1}^N \alpha' \beta(A'_m)}. \quad (16)$$

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  
244 proof.

Frigyik et al. [17, section 2] prove that the proportion of different colored balls in a *Pólya urn model*  
246 converges to the Dirichlet distribution as the number of draws (the population size,  $N$ , in our model) approaches infinity, based on the *Martingale Convergence Theorem* [18]. We therefore have an approximation for the relative prestige each role-model has when evaluated by copiers. Thus, choosing  
248 the role-models for 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.  
250

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

254 **Numerical validation.** We next validated the DM approximation of our model and tested its sensitivity to the assumptions ( $e_i = e$  and  $\alpha_i = \alpha$  for  $i = 1, \dots, N$ ) by comparing results of stochastic simulations  
256 of our model (eq. (11)) with the DM approximation (Corollary 2). We used a relatively small population size,  $N = 100$ , thus validating that the approximation is in good agreement even for  
258 small  $N$ , despite the assumption of large  $N$  in the proof by Frigyik et al. [17, section 2]. First, we computed an observed distribution of the number of copiers from the average empirical distribution  
260 of multiple simulations. We then compared this observed distribution with the expected theoretical DM distribution (Figure S2A). The difference in distributions was not perceived when plotting both  
262 distributions on the same figure, so we used the difference instead. The maximum difference is 0.5 role-models, which indicates a very good fit. In addition, we tested the likelihood of the observed data  
264 to be drawn from the DM distribution, against a shuffle of the parameters vector of the DM distribution itself (Figure S2B). We see that the negative log likelihood of the observed data is much higher than  
266 any other shuffled version of the parameters vector, strongly supporting our approximation.

Next, we examined the fixation probability and fixation time of a favored phenotype  $\hat{A}$  when invading  
268 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  
270 with phenotype  $A$ . That the number of simulations needed to sufficiently approximate our model with the DM approximation is roughly 1,000 (Figure S3). We examined the robustness of the DM  
272 approximation to relaxing the following 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  
274 trait value before evaluation of the bias ( $A_{i,j} = A_j + e_i$ ). When estimation error is applied, we sample

$e_i$  for each copier  $i$  from a normal distribution with expected value zero and variance  $\eta^2$ . Even when  
 276 the estimation error variance is 0.1, both the fixation probability and fixation time approximations  
 are accurate (Figure S4). We also relaxed our assumption of a uniform bias weight  $\alpha$  (i.e.,  $\alpha_i = \alpha$ ).  
 278 We allowed  $\alpha$  to vary in the population, drawing  $\alpha_j$  for each role-model  $j$  from a normal distribution  
 such that  $\alpha_j \sim N(0.5, q)$  where  $q$  is between  $10^{-7}$  and  $10^{-1}$ . We found again that results of the DM  
 280 approximation are similar to those from stochastic simulations of the full model (Figure S5).

## Fixation probability and time

282 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  
 284 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  
 286 population-genetic models [19, 45, 46]. We are mainly interested in the effect of the bias weight,  $\alpha$ ,  
 which determines the relative effect of success and influence on prestige bias, given by eq. (10). For  
 288 simplicity, we use the dichotomous model, and we do not include role-model estimation error in this  
 analysis, i.e.  $e_i = 0$  for every copier  $i$ . As shown above, transmission in our model is approximately  
 290 DM distributed (Corollary 2 and eq. (16)).

**Drift and diffusion terms in a constant environment.** We start by finding the expectation and  
 292 variance of the change in frequency from one generation to the next, which are the drift and diffusion  
 terms of the diffusion equation. Let  $x$  and  $x'$  be the frequency of phenotype  $\hat{A}$  in a population with  
 294  $N$  individuals in the current and next generation, respectively. We set  $\beta$  to be the success bias of  
 phenotype  $A$  relative to phenotype  $\hat{A}$ , such that  $\beta = \beta(A)/\beta(\hat{A}) < 1$ . Then (see Appendix B for  
 296 derivation),

$$\begin{aligned} E[x' - x] &= x(1 - x)(1 - \beta) + o(1 - \beta), \\ V(x' - x) &= x(1 - x) \left( \frac{1}{\alpha N + (1 - \alpha)} \right) + o \left( \frac{1}{\alpha N + (1 - \alpha)} \right). \end{aligned} \quad (17)$$

298 This analysis gives a surprising result relating the parameters  $\alpha$  and  $\beta$  to parameters of the classical  
 Wright-Fisher model from population genetics: the selection coefficient  $s$ , a measure of the effect of  
 300 natural selection on the change in frequency of genotypes, and the effective population size,  $N_e$ , a  
 measure of the effect of random genetic drift on the change in frequency of genotypes. In a diffusion-  
 302 equation approximation of the classical Wright-Fisher model, the expectation and variance of the  
 change in frequency are  $E[x' - x] = x + x(1 - x)s + o(s)$  and  $V[x' - x] = x(1 - x)/N_e$  [19, eq. 7],  
 304 respectively. Therefore, we have the following result.

306 **Result 3** (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  $\beta$  (eq. (5)), the bias*

308 weight  $\alpha$  (eq. (10)), and the population size  $N$  as

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

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

312 Using our effective selection coefficient,  $s = 1 - \beta$ , and effective population size,  $N_e$ , with the  
population-genetics fixation probability approximation given by Kimura and Ohta [19, eq. 8], we  
314 obtain:

**Result 4** (Fixation probability). *The fixation probability of an invading phenotype favored by success  
316 bias is approximately*

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

318 where  $x$  is the initial frequency of the invading phenotype.

Similarly, we can use  $1 - \beta$  and  $N_e$  in the population-genetics fixation time approximation given by  
320 [45, eq. 17].

**Result 5** (Fixation time). *The expected fixation time (conditioned on fixation) from an initial frequency  
322  $x$  is approximately*

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

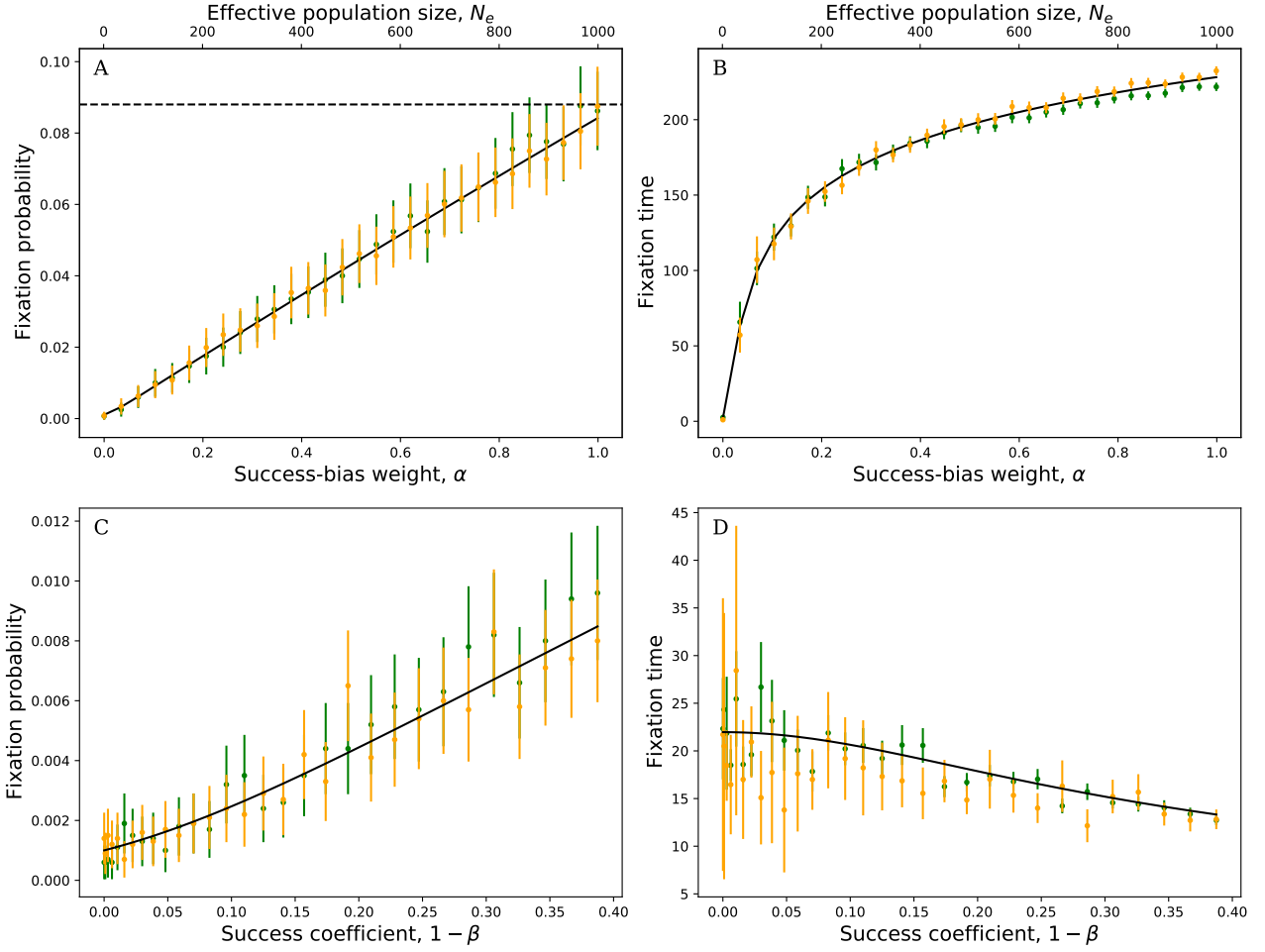
where  $N_e = \alpha N + (1 - \alpha)$ ,  $S = N_e(1 - \beta)$ , and

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

Note that these integrals cannot be solved in closed form, and are estimated numerically.

**Numerical validation.** We compare our approximations (eqs. (19) and (20)) with results of sim-  
ulations of our dichotomous model using various  $\alpha$  and  $\beta$  values, as well as simulations of the  
330 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  
332 approximated by our approximations (Figure 1).

**Changing environment** . After finding a good approximation in a constant environment, where the  
334 favorable trait is always  $\hat{A}$ , we proceeded to find an approximation for a periodically changing environ-  
ment. Following [21], we denote  $k$  as the number of generations in which the invading phenotype is  
336 favored by success bias, and  $l$  as the number of generations in which the resident phenotype is favored  
by success bias. Thus, during the first  $k$  generations of the environmental cycle,  $\beta = \frac{\beta(A)}{\beta(\hat{A})} < 1$ , where  $\hat{A}$  is



**Figure 1: Fixation probability and time in a constant environment.** Fixation probability and time (in generations) as a function of the success-bias weight  $\alpha$  (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,  $A = 0.7$  (panels A and B),  $A/\hat{A}$  varies between 0.01 and 0.99 (panels C and D), and  $\hat{A} = 1$ , which affects  $\beta$  via eq. (17); success coefficient,  $1 - \beta = s = 0.044$  (panels A and B); success-bias weight,  $\alpha = 0.01$  (panels C and D).

the invading phenotype. During the following  $l$  generations of the environmental cycle, the phenotype favored by success bias is switched, such that  $\frac{\beta(A)}{\beta(\hat{A})} > 1$ . We then proceed to find expressions for the expectation and variance of the change in the frequency of phenotype  $\hat{A}$  after  $n = k + l$  generations. The proof is in Appendix C.

**Drift and diffusion terms in a changing environment.** Let  $x$  be the initial frequency of the invading phenotype and  $X_t$  the number of individuals with that phenotype after  $n$  generations. Then,

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

where  $S_n = \sum_{t=1}^n N(1 - \beta_t)$  and  $\beta_t$  is  $\beta(A)$  at generation  $t$ .

Note that here, we have the average selection coefficient during a cycle of  $n$  generations as the selection coefficient in eq. (19). Using the drift and diffusion terms and following [21], we can approximate the fixation probability in a changing environment.

**Result 6** (Fixation probability in a changing environment). *The fixation probability of an invading phenotype under periodical environmental changes is approximately*

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

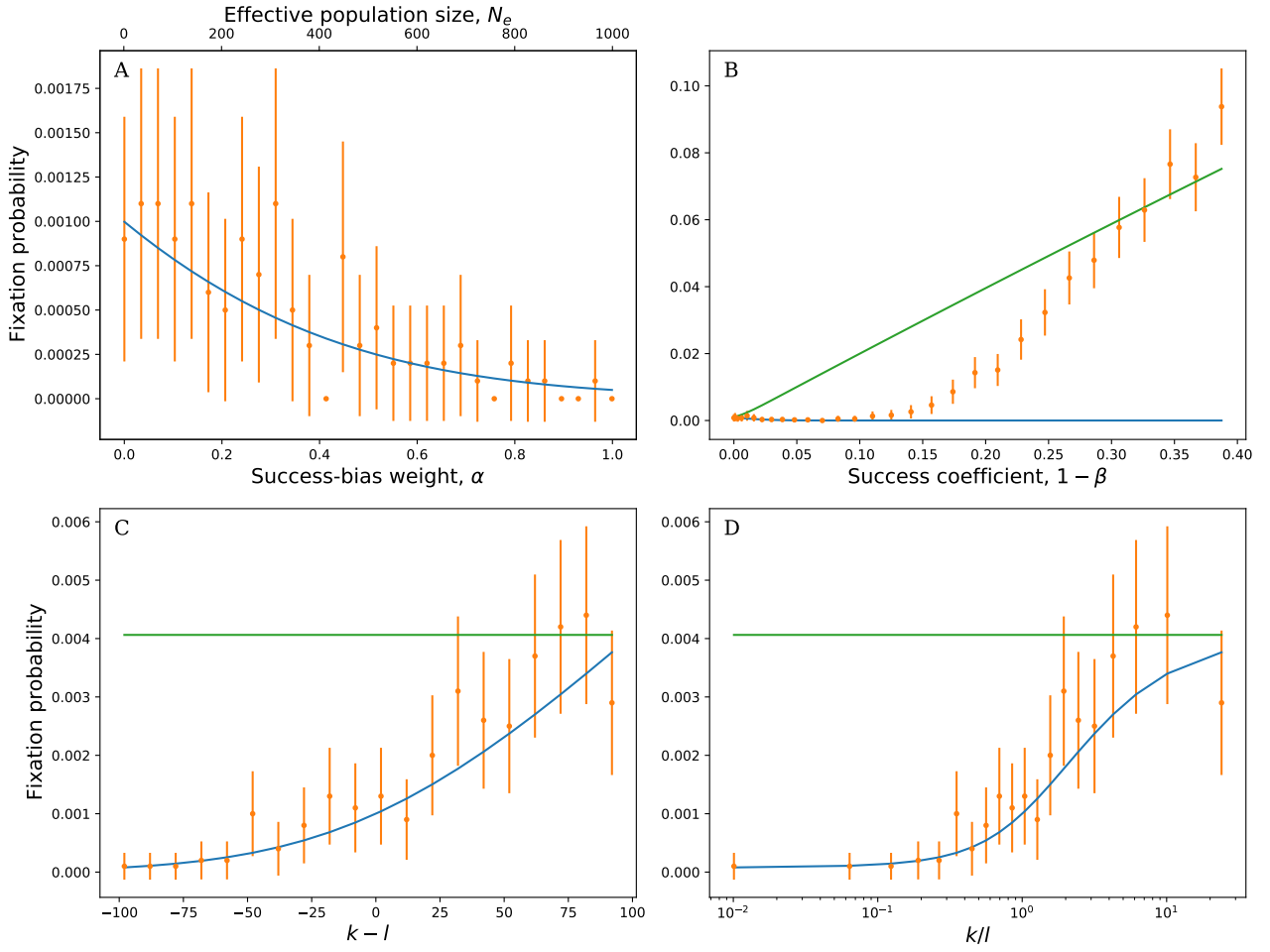
where  $x$  is the initial frequency of the invading phenotype.

**Numerical validation.** To validate the the approximation for the fixation probability in a changing environment (eq. (23)), we compare it to results of simulations that use the DM approximation (Corollary 2). We find that the approximation fits the simulation results well for variable bias weights,  $\alpha$ , which corresponds to the effective population size (Figure 2A).

However, the approximation is more sensitive to the value of the success bias coefficient  $\beta$  (Figure 2B). We suspect that when  $\beta$  is too small, there will not be many cycles in the simulations, because either the population reaches a high frequency of the fitter phenotype after just a few cycles, or the fitter phenotype goes extinct very quickly. For such  $\beta$  values (below 0.65), the fixation probability 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.

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)), see Figure 2C and Figure 2D.

The approximation follows the trend of the simulation results for all  $\alpha$  values. On increasing the success coefficient  $\alpha$  to more than 0.15, the simulation results were located above the changing environment approximation, and below the constant environment approximation. We believe the reason is the structure of the cycle. Our proof and approximation in the changing environment are for a large number of cycles, and when the success coefficient  $\alpha$  is too high, there might be very few cycles. 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 happens, and the ideal trait goes extinct before there are 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 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 2: Fixation probability in a changing environment.** (A) Fixation probability decreases with 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,  $\beta$ . 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% (panels A, C, and D) and 95% (panel B) confidence intervals. Here, population size,  $N = 1,000$ ; phenotype values,  $\hat{A} = 1$  with  $A = 0.9$  (panels A and B) and  $A = 0.8$  (panels C and D); In panel A, the success coefficient is  $1 - \beta = s = 0.005$ ; In panels B, C, and D, the success-bias weight is  $\alpha = 0.1$ .

## Adaptive success-bias weight

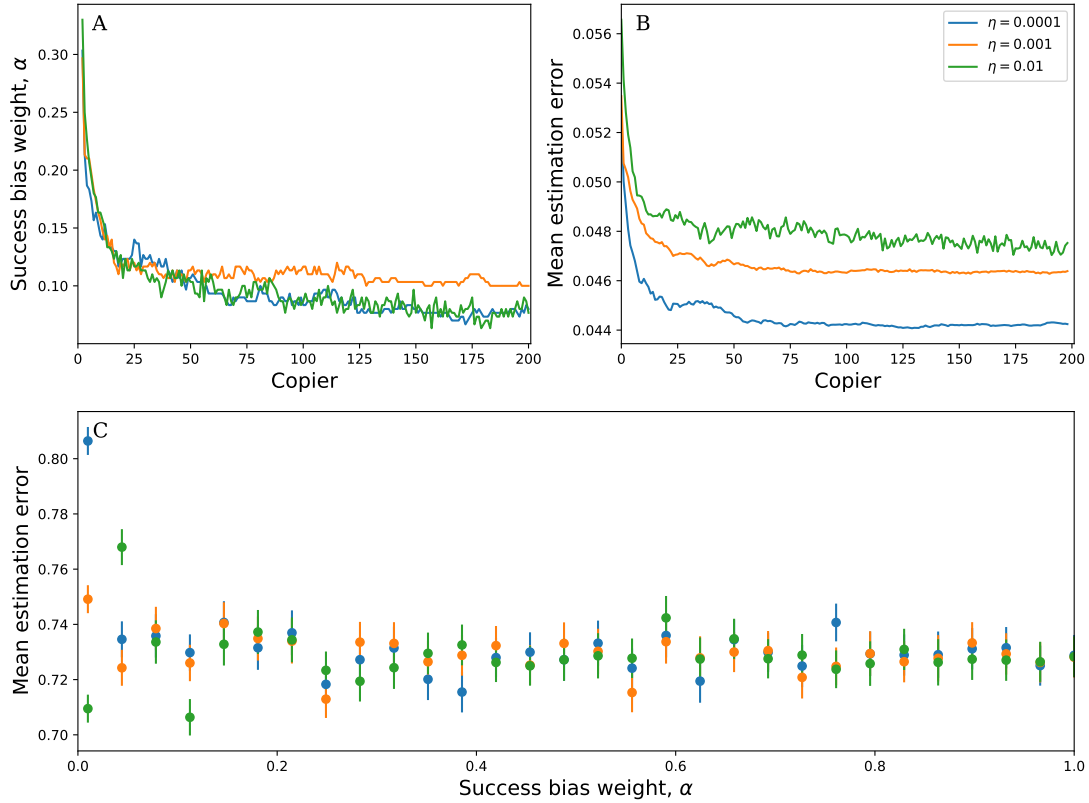
380 We ran simulations of the role-model choice process during a single generation in which every copier evaluates its own optimal success-bias weight,  $\alpha^*$ , which minimizes the expected squared error

382 between the estimated and the ideal trait values,

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

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

386 We find that when copiers choose their success-bias weight, it decreases with the number of copiers  
 388 that have already chosen a role-model (Figure 3). Moreover, their estimation error is much lower  
 compared to a constant success-bias weight, which gives roughly the same high estimation error to all  
 390 copiers (compare Figure 3B and C): in this example, the adaptive weight estimation error converges  
 to 0.046, whereas a constant weight gives values  $> 0.74$ .



**Figure 3: Advantage of an adaptive success-bias weight.** Both success-bias weight  $\alpha$  (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  $\alpha$  is homogeneous (C), the mean estimation error doesn't decrease, regardless of  $\alpha$  or  $\eta$ . The mean estimation error in the homogeneous  $\alpha$  model is larger by a factor of 10 than the adaptive  $\alpha$  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.

## Discussion

Some cultural traits or cultural role-models may be copied more often than others due to transmission biases. One such bias is success bias, in which copiers are more likely to copy a successful role-model. Although many models assume that success can be accurately estimated, it has been suggested that because it is hard to estimate success, a more common bias is *prestige bias*—a bias towards role-models perceived to be successful. This perceived success can be determined by performance with respect to another trait (indirect) [5, 24], or by the influence an individual has on others [26, 31].

We developed a cultural-evolution model with prestige bias that includes both indirect success and influence biases, where the latter is a bias towards role-models with many copiers and hence is the same as conformity bias. We model these biases using a stochastic role-model choice process: each copier, in turn, randomly chooses a role-model, and this choice is affected both by the estimated success of each potential role-model and the number of copiers that already chose each role-model (eq. 10).

Hence, our model has two “nested” stochastic processes: the role-model choice process 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 choice process using the *generalized binomial distribution* (GBD, Result 1) and the *Dirichlet-Multinomial distribution* (DM, Corollary 2). The latter is especially useful, as it approximates the 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 and not the copier.

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 approximations are similar to Kimura’s evolutionary-genetic approximations, in that (i) the difference between the resident and invading cultural trait values,  $1 - \beta(A)$ , is equivalent to the selection coefficient in favor of a beneficial allele,  $s$ , and (ii) increasing the relative weight of influence versus success bias,  $\alpha$ , decreases the effective population size,  $N_e$  (Figure 1).

We also analyzed a cyclic changing environment in which the identity of the success-biased trait switches after a fixed amount of generations (Figure 2). We find that, similarly to the constant environment approximation, a change in the success-bias weight  $\alpha$  has no negative effects on the goodness-of-fit of the approximation to simulation results. We also showed that this approximation is more sensitive to changes in the success coefficient  $\beta$  than the constant environment approximation, and a lower value is required to have a good fit. The ratio between the number of generations in which 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.

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



428 copiers) and the estimation error decrease. The latter is significantly lower than in a population using  
a constant, fixed success-bias weight, regardless of the value of the constant weight (Figure 3). This  
430 suggests that the later a copier makes its choice, the more it should rely on choices of previous copiers,  
and the less it should rely on its own estimation. The rationale, then, is that the more information a  
432 copier has, e.g., by using others as information sources, the more informative and effective his choice  
can be.

434 Chudek et al. [31] report the first direct tests in children that suggest the existence of prestige bias,  
defined as the tendency to learn from individuals to whom others have preferentially attended, learned,  
436 or deferred. Their definition of prestige is similar to our influence bias. They showed that the odds  
of 3-4 years-old children learning from an adult role-model to whom bystanders had previously  
438 preferentially attended for 10 seconds were more than twice those of learning from a role model  
whom bystanders ignored. They also note that prestige effects are domain sensitive: they found that  
440 prestigious role-models were attended more when demonstrating artifact use, whereas role-models  
presenting food preferences had less attendants, suggesting that the domain itself (artifact use vs.  
442 food preference) can affect the attendance, and hence the prestige of the role-model. This led to the  
suggestion that when the trait is costly to learn individually, prestige will have a stronger bias [31]. It  
444 would be interesting to include costs in our model to try and observe these effects and dynamics in a  
large population.

446 According to [24], natural selection has favored the emergence of psychological biases for learning  
from those individuals most likely to possess adaptive information. They studied Fijian villages to  
448 examine if and how such biases 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  
450 across domains. Their research thus suggests that copying from those perceived as successful, rather  
than who are actually successful, is a common phenomenon. They show that the social networks  
452 representing copier–role-model relationships are centralized, suggesting that it is consistent with  
the prediction that people substantially share notions about who is a good cultural model, but that  
454 individuals’ role-model selections are influenced by multiple factors.

Prestige bias also occurs in more modern domains such as western medicine. Norredam and Al-  
456 bum [28] 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  
458 in the upper part of the bodies of young and middle-aged people were accorded high levels of prestige,  
whereas medicine and practices that were not of these types had low levels of prestige. For example,  
460 they found that surgery counts as the most prestigious specialty, while psychiatry is the least presti-  
gious. In addition, doctors tend to rank practices that require more time to master as more prestigious,  
462 while 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  
464 differences in prestige may affect for actual priority setting in healthcare systems.

Prestige bias can help to cheaply estimate and acquire knowledge, which may facilitate survival and  
466 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 [27] mention that social learning  
468 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  
470 samples  $k$  role-models and chooses the one he deems most "successful". They mention that a previous  
mathematical analysis has shown that it may sometimes result in maladaptive cultural evolution when  
472 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  
474 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  
476 an on average inferior variant that sometimes gives a very high payoff (see Fogarty et al. [56] and  
references there). Our model, which includes influence bias, is consistent with this claim. When a  
478 maladaptive trait is "piggybacking" on a role-model with high influence, the former could spread in  
the population. In addition, best-of-k learning can be stable against invasion by unbiased learning if  
480 social learning is sometimes combined with individual learning [27]. Our model includes only social  
learning, and not individual learning, but it could be interesting to combine it with individual learning  
482 and see how it affects the dynamics.

Prestige bias can also accelerate reversal of harmful traditions such as child marriage and domestic  
484 violence. Efferson et al. [25] 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  
486 behavior. They find that there are individuals who act as *agents*, who are often observed, and therefore  
they are ideal targets for interventions. This is similar to influential role-models in our model, where a  
488 prestigious individual will be copied more often, and will therefore spread their trait faster and wider  
in the population. They also suggest a way to use this phenomenon to change existing traditions in  
490 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 [30] 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 this, Muthukrishnan  
494 and Henrich [29] suggested 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.  
496 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  
498 *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  
500 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  
502 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.

504 One path forward is an analysis of the dynamics of the adaptive success-bias weight model, in which  
every copier chooses its  $\alpha$ . It would be interesting to see the if the mean estimation error and the

506 adaptive weight,  $\alpha^*$ , converge 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  
508 estimation error and success-bias weight. Lastly, it would be interesting to analyze the continuous  
model and determine how much it differs from the dichotomous model.

510 Another way to expand our model is to account for the two types of prestige or leadership suggested  
by Van Vugt and Smith [23] that are attributed to Confucius and Machiavelli. Confucius viewed  
512 leaders as role-models who exercise influence through possessing superior knowledge, skills, and  
(outstanding) personal qualities. This fits the success bias in our model. In contrast, Machiavelli  
514 viewed leaders as rulers who exercise influence by imposing costs through (the threat of) punishment  
and violence. Van Vugt and Smith suggest that these opposing views are both partially supported by  
516 the available evidence but each one on its own offers an incomplete view of the complex and dynamic  
concept of leadership. Several adjustments could be made so that our model reflects these leadership  
518 styles, such as assuming there is a correlation between phenotype and leadership style. The resulting  
cultural-evolutionary dynamics and their dependence on the costs and benefits are intriguing.

520 **Conclusions.** 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. We found  
522 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 random  
524 genetic drift. We also find a clear advantage to individuals that can choose the relative weight of the  
two biases.

## 526 Acknowledgements

We thank Marc Feldman, Martin Pontz, Gili Greenbaum, Alon Rozen, and Tal Simon for discussions and  
528 comments. This work was supported in part by Minerva Stiftung Center for Lab Evolution (YR) and the John  
Templeton Foundation (YR 61809).

## 530 References

- [1] Battesti, Marine, et al. "Spread of social information and dynamics of social transmission within  
Drosophila groups." *Current biology* 22.4 (2012): 309-313.
- [2] Creanza, Nicole, Oren Kolodny, and Marcus W. Feldman. "Cultural evolutionary theory: How  
culture evolves and why it matters." *Proceedings of the National Academy of Sciences* 114.30  
(2017): 7782-7789.
- [3] Aljadeff, Naama, Luc-Alain Giraldeau, and Arnon Lotem. "Competitive advantage of rare be-  
haviours induces adaptive diversity rather than social conformity in skill learning." *Proceedings of  
the Royal Society B* 287.1933 (2020): 20201259.

- [4] Cavalli-Sforza, Luigi Luca, and Marcus W. Feldman. Cultural transmission and evolution: A quantitative approach. No. 16. Princeton University Press, 1981.
- [5] Boyd, Robert, and Peter J. Richerson. Culture and the evolutionary process. University of Chicago press, 1988.
- [6] Fogarty, Laurel, et al. "The driving forces of cultural complexity: neanderthals, modern humans, and the question of population size." *Human Nature* 28 (2017): 39-52.
- [7] Rendell, Luke, et al. "Why copy others? Insights from the social learning strategies tournament." *Science* 328.5975 (2010): 208-213.
- [8] Andersson, Malte, and Yoh Iwasa. "Sexual selection." *Trends in ecology & evolution* 11.2 (1996): 53-58.
- [9] Kendal, Rachel, et al. "Chimpanzees copy dominant and knowledgeable individuals: implications for cultural diversity." *Evolution and Human Behavior* 36.1 (2015): 65-72.
- [10] Horner, Victoria, et al. "Prestige affects cultural learning in chimpanzees." *PloS one* 5.5 (2010): e10625.
- [11] Henrich, Joseph, and Richard McElreath. "Dual-inheritance theory: the evolution of human cultural capacities and cultural evolution." (2007).
- [12] McComb, Karen, et al. "Matriarchs as repositories of social knowledge in African elephants." *Science* 292.5516 (2001): 491-494.
- [13] Eickbush, Michael T., Janet M. Young, and Sarah E. Zanders. "Killer meiotic drive and dynamic evolution of the wtf gene family." *Molecular Biology and Evolution* 36.6 (2019): 1201-1214.
- [14] Xing, Wenpu, and Ali Ghorbani. "Weighted pagerank algorithm." *Proceedings. Second Annual Conference on Communication Networks and Services Research*, 2004.. IEEE, 2004.
- [15] Molleman, Lucas, Ido Pen, and Franz J. Weissing. "Effects of conformism on the cultural evolution of social behaviour." *PloS one* 8.7 (2013): e68153.
- [16] Drezner, Zvi, and Nicholas Farnum. "A generalized binomial distribution." *Communications in Statistics-Theory and Methods* 22.11 (1993): 3051-3063.
- [17] Frigyik, Bela A., Amol Kapila, and Maya R. Gupta. "Introduction to the Dirichlet distribution and related processes." *Department of Electrical Engineering, University of Washington, UWEETR-2010-0006 6* (2010): 1-27.
- [18] Durrett, Richard, and R. Durrett. *Essentials of stochastic processes*. Vol. 1. New York: Springer, 1999.
- [19] Kimura, Motoo. "On the probability of fixation of mutant genes in a population." *Genetics* 47.6 (1962): 713.

- [20] Durrett, Richard, and Richard Durrett. Probability models for DNA sequence evolution. Vol. 2. New York: Springer, 2008.
- [21] Ram, Yoav, Uri Liberman, and Marcus W. Feldman. "Evolution of vertical and oblique transmission under fluctuating selection." *Proceedings of the National Academy of Sciences* 115.6 (2018): E1174-E1183.
- [22] King, Andrew J., and Guy Cowlshaw. "Leaders, followers, and group decision-making." *Communicative & integrative biology* 2.2 (2009): 147-150.
- [23] Van Vugt, Mark, and Jennifer E. Smith. "A dual model of leadership and hierarchy: Evolutionary synthesis." *Trends in Cognitive Sciences* 23.11 (2019): 952-967.
- [24] Henrich, Joseph, and James Broesch. "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 (2011): 1139-1148.
- [25] Efferson, Charles, Sonja Vogt, and Ernst Fehr. "The promise and the peril of using social influence to reverse harmful traditions." *Nature human behaviour* 4.1 (2020): 55-68.
- [26] Henrich, Joseph, and Francisco J. Gil-White. "The evolution of prestige: Freely conferred deference as a mechanism for enhancing the benefits of cultural transmission." *Evolution and human behavior* 22.3 (2001): 165-196.
- [27] Takahashi, Takuya, and Yasuo Ihara. "Cultural and evolutionary dynamics with best-of-k learning when payoffs are uncertain." *Theoretical Population Biology* 128 (2019): 27-38.
- [28] Norredam, Marie, and Dag Album. "Prestige and its significance for medical specialties and diseases." *Scandinavian journal of public health* 35.6 (2007): 655-661.
- [29] Muthukrishna, Michael, and Joseph Henrich. "Innovation in the collective brain." *Philosophical Transactions of the Royal Society B: Biological Sciences* 371.1690 (2016): 20150192.
- [30] Dunbar, Robin IM. "The social brain hypothesis and its implications for social evolution." *Annals of human biology* 36.5 (2009): 562-572.
- [31] Chudek, Maciej, et al. "Prestige-biased cultural learning: Bystander's differential attention to potential models influences children's learning." *Evolution and human behavior* 33.1 (2012): 46-56.
- [32] Henrich, Joseph, and Richard McElreath. "The evolution of cultural evolution." *Evolutionary Anthropology: Issues, News, and Reviews: Issues, News, and Reviews* 12.3 (2003): 123-135.
- [33] O'Brien, Michael J., et al. "Cultural traits as units of analysis." *Philosophical Transactions of the Royal Society B: Biological Sciences* 365.1559 (2010): 3797-3806.
- [34] Whitehead, Hal. "Gene–culture coevolution in whales and dolphins." *Proceedings of the National Academy of Sciences* 114.30 (2017): 7814-7821.

- [35] Cohen, Dor, et al. "Non-vertical cultural transmission, assortment and the evolution of cooperation." *Proceedings of the Royal Society B* 288.1951 (2021): 20203162.
- [36] Denton, Kaleda K., Yoav Ram, and Marcus W. Feldman. "Conformity and content-biased cultural transmission in the evolution of altruism." *Theoretical Population Biology* 143 (2022): 52-61.
- [37] Denton, Kaleda Krebs, et al. "Cultural evolution of conformity and anticonformity." *Proceedings of the National Academy of Sciences* 117.24 (2020): 13603-13614.
- [38] Lee, Woojin, Lina Xiong, and Clark Hu. "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 (2012): 819-827.
- [39] Anagnostopoulos, Aris, Ravi Kumar, and Mohammad Mahdian. "Influence and correlation in social networks." *Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*. 2008.
- [40] Peng, Sancheng, et al. "Influence analysis in social networks: A survey." *Journal of Network and Computer Applications* 106 (2018): 17-32.
- [41] Diga, Marichris, and Tom Kelleher. "Social media use, perceptions of decision-making power, and public relations roles." *Public Relations Review* 35.4 (2009): 440-442.
- [42] Van Rossum, Guido. "Python Programming Language." *USENIX annual technical conference*. Vol. 41. No. 1. 2007.
- [43] Van Der Walt, Stefan, S. Chris Colbert, and Gael Varoquaux. "The NumPy array: a structure for efficient numerical computation." *Computing in science & engineering* 13.2 (2011): 22-30.
- [44] Hunter, John D. "Matplotlib: A 2D graphics environment." *Computing in science & engineering* 9.03 (2007): 90-95.
- [45] Kimura, Motoo, and Tomoko Ohta. "The average number of generations until fixation of a mutant gene in a finite population." *Genetics* 61.3 (1969): 763.
- [46] Slatkin, Montgomery. "Fixation probabilities and fixation times in a subdivided population." *Evolution* (1981): 477-488.
- [47] Erhart, Elizabeth M., and Deborah J. Overdorff. "Female coordination of group travel in wild *Propithecus* and *Eulemur*." *International Journal of Primatology* 20 (1999): 927-940.
- [48] Boyd, Robert, and Joseph Henrich. "On modeling cognition and culture: Why cultural evolution does not require replication of representations." *Journal of cognition and culture* 2.2 (2002): 87-112.
- [49] Gardner, Andy, Stuart A. West, and Geoff Wild. "The genetical theory of kin selection." *Journal of evolutionary biology* 24.5 (2011): 1020-1043.
- [50] Truskanov, Noa, Yasmin Emery, and Redouan Bshary. "Juvenile cleaner fish can socially learn the consequences of cheating." *Nature communications* 11.1 (2020): 1159.

- [51] Kolodny, Oren, et al. "Differential application of cultural practices at the family and individual levels may alter heritability estimates." *Behavioral and Brain Sciences* 45 (2022): e167.
- [52] Denton, Kaleda K., Uri Liberman, and Marcus W. Feldman. "On randomly changing conformity bias in cultural transmission." *Proceedings of the National Academy of Sciences* 118.34 (2021): e2107204118.
- [53] Borofsky, Talia, and Marcus W. Feldman. "Success-biased social learning in a one-consumer, two-resource model." *Theoretical Population Biology* 146 (2022): 29-35.
- [54] The cultural transmission of great basin projectile-point technology II: An agent-based computer simulation
- [55] Lehmann, L., and M.W. Feldman. Coevolution of adaptive technology, maladaptive culture, and population size in a producer-scrounger game. *Proc. Roy. Soc. B* 276 (2009): 3853-3862.
- [56] Fogarty, L., J. Y. Wakano, M. W. Feldman, and K. Aoki. The driving forces of cultural complexity: Neanderthals, modern humans, and the question of population size. *Hum. Nat.* (2017): 28: 39-52.
- [57] Denton, Kaleda K, Yoav Ram, and Marcus W. Feldman. 2023. "Conditions That Favour Cumulative Cultural Evolution." *Philosophical Transactions of the Royal Society B: Biological Sciences* 378 (1872). <https://doi.org/10.1098/rstb.2021.0400>.

# 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 average 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  $\alpha_m \cdot \beta(A_m + e)$ . Using eq. (A2) and **Corollary 1** we get,

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

## 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  $\beta$  be 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,

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

following Durrett [20, p. 253, ch 7.2] and because  $1/(1-y) = 1 + y + y^2 + \dots$

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 [20, 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

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

The current frequency  $x$  is a given, such that  $V(x) = 0$ , and therefore

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

$\alpha'$  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  $\beta_t$  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

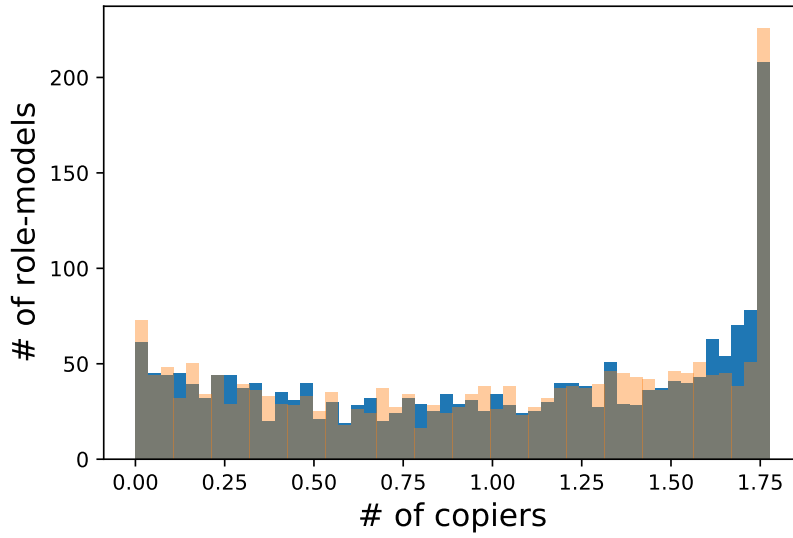
$$V\left(E\left[\frac{X_{t+1}}{N}\middle|X_t\right]\right) = V\left(\frac{X_t}{N}\right) + O\left(\frac{1}{N^2}\right) \simeq V\left(\frac{X_t}{N}\right). \quad (\text{C17})$$

Using the induction assumption and eq. (C14),

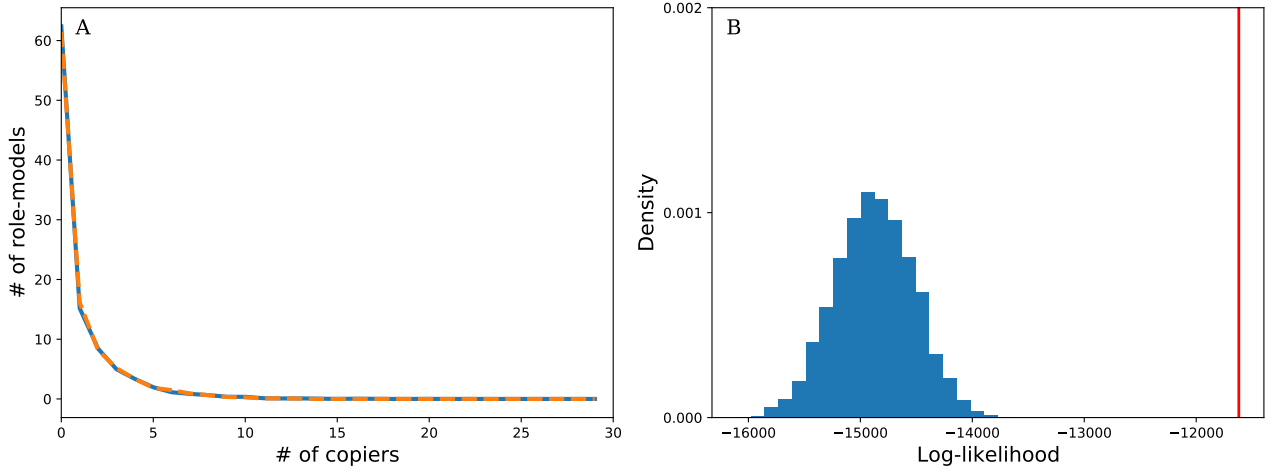
$$V\left(\frac{X_{t+1}}{N}\right) \simeq \frac{1}{N_e} x(1 - x) + \frac{1}{N_e} t x(1 - x) \simeq \frac{1}{N_e} x(1 - x)(t + 1) \quad (\text{C18})$$

which proves the diffusion term.

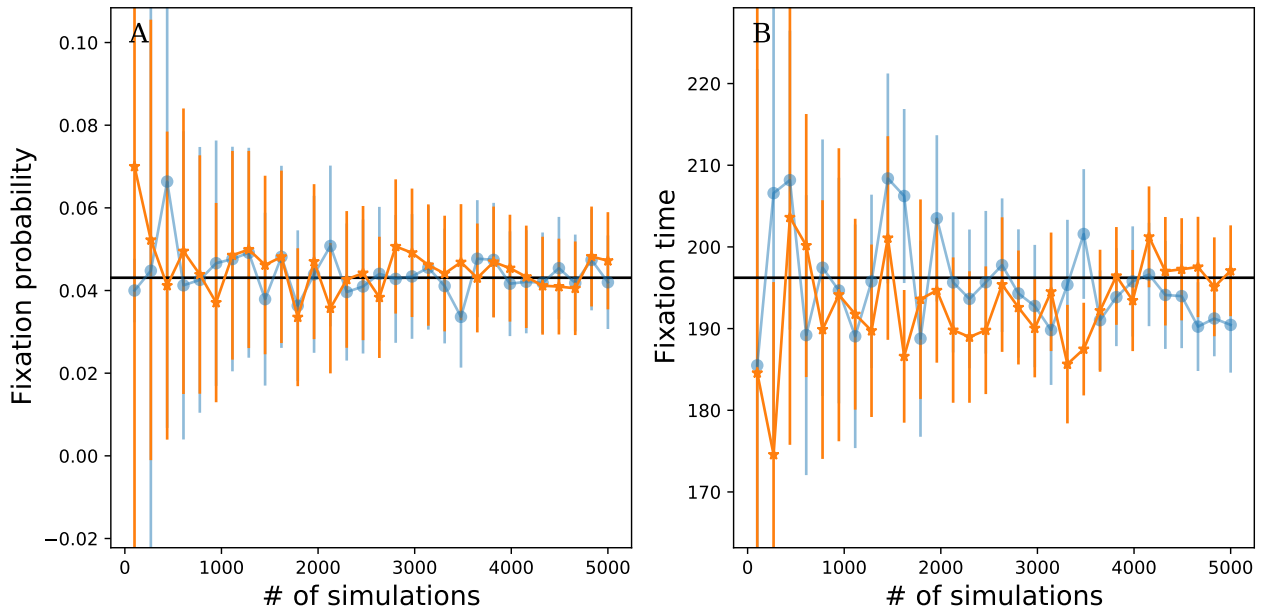
## Supplementary Figures



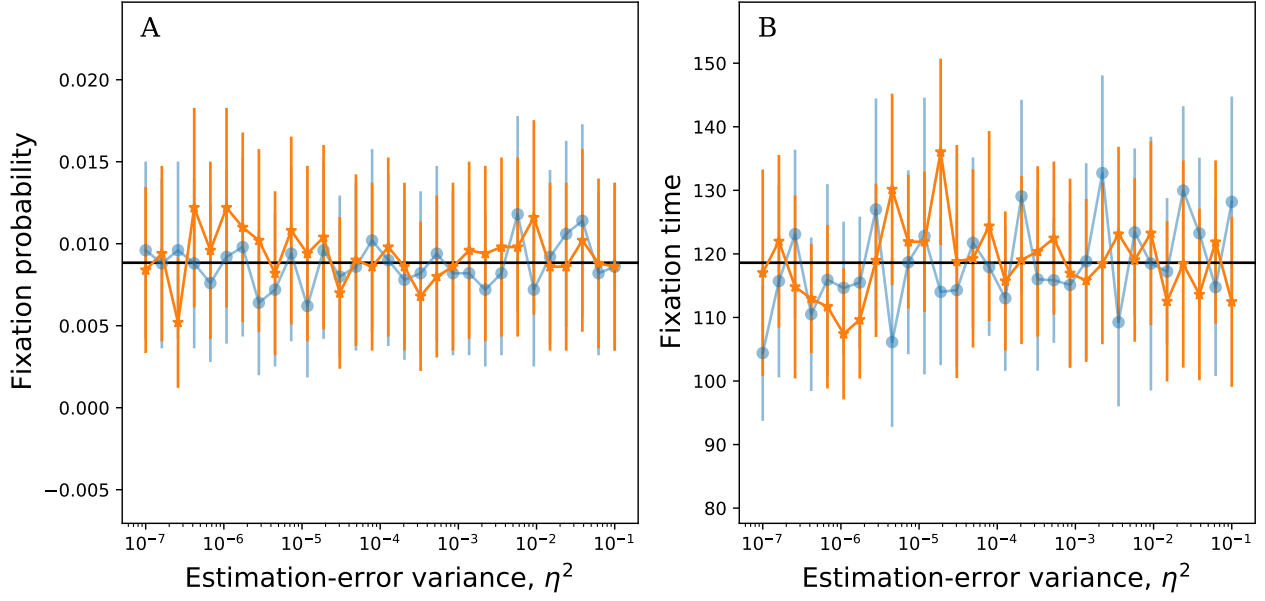
**Figure S1: Numerical validation of the GB approximation.** The approximation (orange) fits simulation results (blue) well when using 1,000 simulations. Here, population size,  $N = 2,000$ ; bias weight,  $\alpha = 0.1$ ; ideal phenotype value,  $\hat{A} = 1$ ; role-model traits  $\mathbf{A} \sim N(0, 1)$ ; success bias value,  $\beta(A) = 0.956$ .



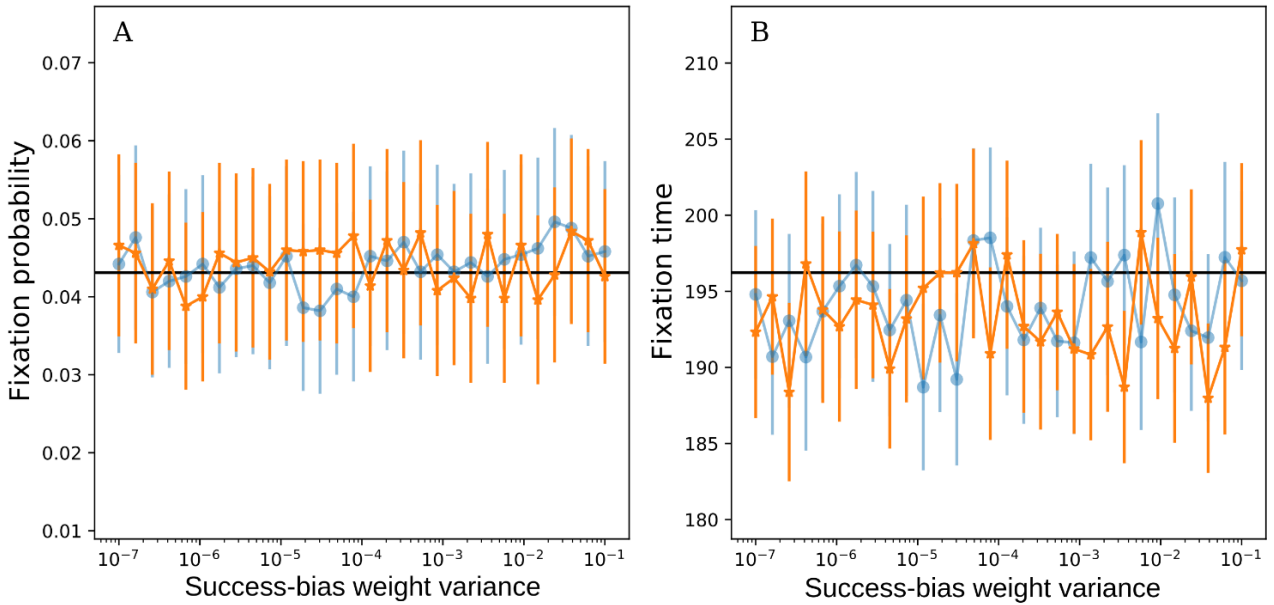
**Figure S2: Numerical validation of the DM approximation.** We performed computational simulations of the role-model choice process (Equation (10)) and compared them to simulations when using the DM distribution approximation (Corollary 2). **(A)** The difference between the DM distribution (orange) and the empirical distribution of the simulations (blue) is very small. **(B)** The log-likelihood of the DM distribution for results of the simulations (red vertical line) is much higher than the log-likelihood of permutations of simulations (blue histogram). Here, population size,  $N = 100$ ; number of simulations,  $m = 100$ ; 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.



**Figure S3: 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,  $\beta(A) = 0.956$ .



**Figure S4: 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  $\eta^2$  in on the x-axis.



**Figure S5: Robustness of DM approximations to variation in the bias weight  $\alpha$ .** Fixation probability does not seem to be effected by variation in success bias weight between individuals. Thus, both the DM approximation (orange) and Kimura's equation (black line) have a good fit to results of stochastic simulations (blue). Markers for average across 5,000 simulations, error bars are 95% confidence intervals. Here, population size,  $N = 1000$ ; success bias weight is normally distributed,  $\alpha \sim N(0.5, \epsilon^2)$  where  $10^{-7} \leq \epsilon \leq 10^{-1}$ ; phenotype values,  $\hat{A} = 1, A = 0.7$ ; success bias value,  $\beta(A) = 0.956$ .