

Cultural Evolutionary Dynamics with Success and Influence biases

Saar Egozi¹ and Yoav Ram^{2,3,*}

¹School of Computer Science, Reichman University, Herzliya 4610101, Israel

²School of Zoology, Faculty of Life Sciences, Tel Aviv University, Tel Aviv 6997801, Israel

³Sagol School of Neuroscience, Tel Aviv University, Tel Aviv 6997801, Israel

*Corresponding author: yoav@yoavram.com, ORCID 0000-0002-9653-4458

June 13, 2022

Abstract

Copying role-models can be an efficient method for acquiring knowledge. A common bias when choosing a role-model to copy is success bias: copying whoever appears more successful. This bias depends on the performance of the role-model alone, with no other factors. We propose an additional bias that may be prevalent in cultural transmission: influence bias, in which role-model choice is affected by the number of individuals that have already copied each potential role-model. We combine success and influence bias together to a “prestige bias” and analyze its effects on cultural-evolutionary dynamics using mathematical analysis and stochastic simulations. We find analytic approximations to our stochastic model, facilitating further mathematical analysis and reducing the computational complexity of simulations. We validate these approximations using simulations, and demonstrate their robustness to model assumptions. We also find approximations to the fixation probability and the fixation time of an invading advantageous cultural trait, in both constant and changing environments, which resemble Kimura’s approximations for population-genetic models. These approximations show that success bias effectively plays the part of natural selection, whereas influence bias effectively reduces the population size. Influence bias also accelerates the evolutionary dynamics, as can be expected in a *rich-getting-richer* process. Our model may provide a good description of cultural transmission, especially in human societies where social media is popular. Further work is needed to test if this model could predict various phenomena in cultural evolution when extended with the effects of selection and innovation.

28 Introduction

Cultural transmission. In cultural transmission, individuals transmit cultural traits (i.e., behavior, beliefs, norms) to one another, typically via learning and demonstrating (Cavalli-Sforza and Feldman, 1981). Examples for cultural traits in humans are behavioral patterns, such as personalities and habits, transmitted via both verbally and by observations (O'Brien et al., 2010). Although cultural transmission is most common in humans, it is also observed in other animals such as chimpanzees (Horner et al., 2010; Kendal et al., 2015), dolphins and whales (Whitehead, 2017). In elephants, McComb et al. (2001) showed that once a matriarch is removed from the group, the group's survival instincts are inferior and that "the possession of enhanced discriminatory abilities by the oldest individual [matriarch] in a group can influence the social knowledge of the group as a whole." By playing audio recordings of African elephants, they showed that groups with a matriarch recognize and react better to hostile or friendly calls than groups without a matriarch. Battesti et al. (2012) showed that choice of oviposition site in fruit flies is culturally transmitted: inexperienced flies that spent some time with experienced flies chose the same type of oviposition site even without directly observing this behavior. How the information is transmitted is still an open question, but it has been suggested that flies may use olfactory cues like other animals, such as rodents and bees.

Direction of transmission. Similar to genetic transmission, culturally transmitted traits can be transmitted from parents to offspring, and their effects of can be physiological rather than behavioral. For example, parents can "teach" their children to be strong or tall, within some biological limits, by instructing them to maintain a specific diet and engage in physical activity. Contrary to genetic transmission, cultural transmission can be non-vertical, that is, traits may be transmitted via social learning from non-parental individuals, and even unrelated individuals such as teachers, leaders, media, or any stranger that interacts with the learning individual. Thus, cultural transmission may combine vertical transmission, where parents transmit to their offspring; oblique transmission, where adults transmit traits to unrelated offspring; and horizontal transmission, where peers from the same age cohort transmit to one another. Vertical transmission is also possible in the opposite direction: parents may copy traits from their offspring (Cavalli-Sforza and Feldman, 1981; Creanza et al., 2017).

Transmission biases. In social learning, transmission biases cause a trait to have a disproportionate probability to be transmitted compared to its frequency in the population. Although more common in cultural transmission, transmission biases do occur in genetic transmission. For example, *wtf* genes in yeast bias their transmission to the gamete by secreting a long life-expectancy poison together with a short life-expectancy antidote, so that a gamete without the gene will perish because the poison will outlive the antidote (Eickbush et al., 2019). Importantly, even when a trait is disfavored by natural selection, it may still spread in a population due to transmission biases that are strong enough to overcome selection (Boyd and Richerson, 1988, Ch. 8 pg. 279). Cohen et al. (2021) show that cooperative behavior can evolve in an individual due to horizontal transmission bias even when there is no benefit to it, or when it benefits its competitors.

Success bias. Rendell et al. (2010) have conducted a tournament between learning strategies. Each strategy defines when individuals observe and copy from others, and when they engage in individual learning, in which an individual learns a cultural trait on his own. The best strategies had a high frequency of social learning relative to individual learning, even when the transmission error was almost 50%. From these results we understand that all the winning strategies were mostly based on success biased social learning, meaning it contributed more to the general success of the population than individual learning. However, all winning strategies included individual learning to some degree, implying that success-biased learning alone isn't the best way to advance, even when success is measured accurately.

Evaluating success. Boyd and Richerson (1988, Ch. 5) suggest that the evaluation of success can be divided into three groups: *direct bias*, *indirect bias* and *frequency-dependent bias*. A direct bias occurs when a variation of a trait is more attractive than others, and is evaluated by *directly* testing the variation of the trait. For example, an individual observing a Ping-Pong match can attempt both of the observed paddle grips to determine which grip is better. An indirect bias occurs when an individual uses the value of one trait to determine the attractiveness of another, so it *indirectly* evaluates the attractiveness of the role-model. 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. A frequency-dependent bias occurs when an individual has a probability to copy a variant of the trait that higher or lower than trait's frequency among demonstrators. For example, when an individual is 80% likely to copy the common paddle grip even when only 60% of the population is using it, it is said to be frequency-biased, or in this case, conformist. Frequency bias could be negative, i.e., non-conformist bias. Conformity and non-conformity are well-known biases in cultural transmission (Molleman et al., 2013), and its effect on cultural evolution have been studied with both models (Denton et al., 2020a,b) and experiments (Aljadeff et al., 2020).

Prestige. Prestige means having a good reputation or high esteem, therefore does not directly signify success (although it may imply it), making it an indirect bias. Both Boyd and Richerson (1988, Ch. 8) and Fogarty et al. (2017) suggest that prestige biases are probably more common in humans than success biases. Boyd and Richerson (1988, Ch. 8) add that maladaptive traits may spread widely in a population if indirect biases are strong enough. They suggest that such biases could lead to a runaway process caused by a cultural equivalent of sexual selection (Andersson, 1994). On the other hand, Henrich and Broesch (2011) suggest that prestige 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. Prestige is often mentioned in the cultural-evolution literature, but seldom modeled.

Influence bias. Today, social media provides an easy way to estimate the social and cultural influence individuals have over others, and therefore may have an effect on decision making. Online social networks such as *Facebook* or *Instagram* are known to affect the influence of individuals (Anagnostopoulos et al., 2008; Peng et al., 2018; Diga and Kelleher, 2009), and specific marketing practices were invented to capitalize on this effect (Lee et al., 2012). Here, we model such influence as an

indirect bias in cultural transmission, in which the choice of a role-model depends on the choices made by other individuals that have already chosen a role-model. In addition to using influence bias as an easy way to approximate role-models worth, it can also be used to approximate it better than the copier's own estimation. This bias can save the copier costs: time, effort etc, and it can also improve its choice of role-model.

This bias depends on the state of a role-model rather than on its trait, in contrast to frequency biases such as conformity (Hamilton, 1964), which depend on the frequency of a trait in the population or in a sample of role-models. We define a model of cultural transmission with prestige bias that combines both success and influence biases, provide analytic approximations for this model, and analyze its dynamics.

Models and methods

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 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 (Hunter, 2007). Source code is available at <https://github.com/yoavram-lab/PrestigeBias>.

Continuous trait

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 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 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 $\mathbf{A}(t) = (A_1(t), \dots, A_N(t))$ 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 modeled by a function F such that

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

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

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

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

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

136 $A_{i,j}$ is the absolute trait value copier i estimates for role-model j with some error $e_i \sim N(0, \eta^2)$,

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

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

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

140 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
142 fitness in evolutionary-genetic transmission models.

Boyd and Richerson (1988) note that the deterministic blended transmission algorithm they use has
144 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
146 distribution given by

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

148 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,
150 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,
152 $\sum_{j=1}^N K_{i,j} = i$, 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 $K_{0,j} =$
158 0.

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

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

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

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

166 Here, the bias weight α_{ij} is a characteristic of the interaction of role-model j with copier i , that
 168 determines the relative significance of success and influence within the role-model's overall prestige,
 in the eyes of the copier. The trait of role-model j estimated by copier i , $A_{i,j}$, remains as in eq. (4),
 and W_i is a normalizing factor that sums the numerator over all role-models ($1 \leq j \leq N$) to ensure
 170 $\sum_{j=1}^N G_{i,j} = 1$.

Dichotomous trait

172 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
 174 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
 176 either phenotype. In addition, we assume homogeneous α for simplicity and for easier mathematical
 analysis. Therefore, the probability of the i -th copier to copy phenotype A is

$$178 \quad G_{i,A} = \frac{(N - X)\alpha'\beta(A) + K_{i,A}}{i - 1 + (N - X)\alpha'\beta(A) + X\alpha'\beta(\hat{A})} = \frac{(N - X)\alpha'\beta(A) + K_{i,A}}{i - 1 + (N - X)\alpha'\beta(A) + \alpha'X} \quad (11)$$

where X is the number of role-models with trait \hat{A} and $K_{i,A}$ is the number of copiers that already chose
 180 A when copier i chooses a role-model and α' is the odd ratio of the bias weight $\alpha' = \frac{\alpha}{1-\alpha}$.

We prove this equation later, based on the proof of the Dirichlet-Multinomial approximation.

182 Initially, we assume the population has a single individual with phenotype \hat{A} and $N - 1$ individuals
 with phenotype A . The rest of the details follow the continuous trait model.

184 Results

Approximations

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

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

(a) 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 for all copiers ($e = e_i$ for $i = 1, \dots, N$).

(b) 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, 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.

2. 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 there's no meaning for the order of copiers. Therefore, the number of copiers of all role-models, $\mathbf{K}_i = (K_{i,1}, \dots, K_{i,N})$, follows a Dirichlet-Multinomial distribution, $\mathbf{K}_i \sim DM(N, \mathbf{G}_1)$, where $\mathbf{G}_1 = (G_{1,1}, \dots, G_{1,N})$ is the prestige vector of all role-models in the "eyes" of the first copier. 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 error is uniform, meaning all copiers are identical.

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 parameterized by $GBD(n, p, \theta)$ 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).

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

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 $K_{i-1,j}$ out of $i-1$ copiers chose role-model j . Using conditional probability and eq. (8),

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

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.

Corollary 1. $E[K_{N,j}] = N \cdot G_{1,j}$.

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 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 the role-models index, j .

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

$$G_{1,j} = \frac{\alpha_j \cdot \beta(A_j + e)}{\sum_{m=1}^N \alpha_m \cdot \beta(A_m + e)} . \quad (13)$$

The denominator of eq. (13) can also be formulated as:

$$\sum_{m=1}^N \alpha_m \beta(A_m + e) = N \cdot \overline{\alpha \cdot \beta(A + e)} , \quad (14)$$

where $\overline{\alpha \cdot \beta(A + e)}$ is the mean value of $\alpha_m \cdot \beta(A_m + e)$ for all m . Using eq. (14) and **Corollary 1** we get,

$$E[K_{Nj}] = \alpha_j \cdot \beta(A_j + e) \left/ \overline{\alpha \cdot \beta(A + e)} \right. , \quad (15)$$

completing the proof.

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

$$E[K_{Nj}] = N \cdot \frac{\alpha \cdot \beta(A_j + e)}{\sum_{m=1}^N \alpha \cdot \beta(A_m + e)} = \beta(A_j + e) \left/ \overline{\beta(A + e)} \right. , \quad (16)$$

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

$$E[K_{Nj}] = L \cdot \beta(A_j + e) . \quad (17)$$

Numerical validation. To validate our results we ran 1,000 simulations of the full model, and 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 the average of the 1,000 simulations to eliminate drift errors, as can be seen in fig. 1.

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

Dirichlet-Multinomial distribution

Pólya urn model. This stochastic process consists of N draws from an urn with an initial amount of 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 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,

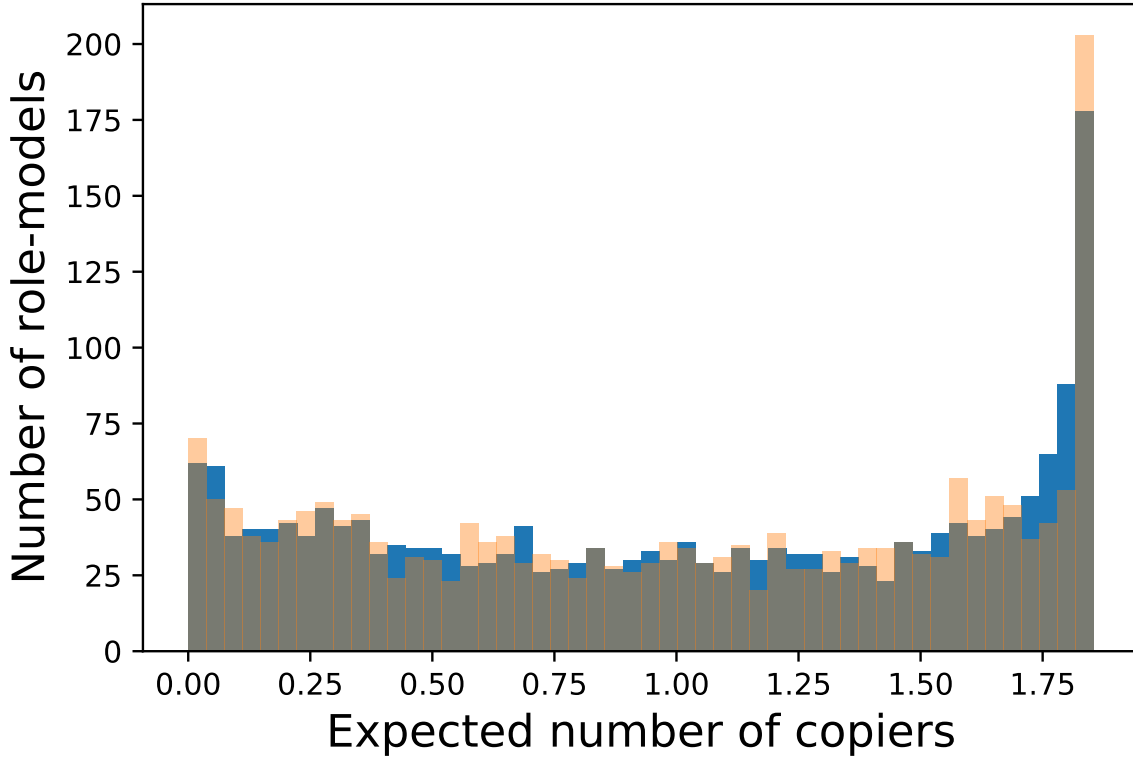


Figure 1: Numerical validation for the GBD approximation. The approximation (orange) fits simulation results (blue) well when using 1,000 simulations for both models. Here, population size, $N = 2,000$; bias weight, $\alpha = 0.1$; idea phenotype value, $\hat{A} = 1$; role-model traits $\mathbf{A} \sim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

258 and 0 otherwise. The probability that $S_{i,j} = 1$ given \mathbf{U}_{i-1} is

$$P(S_{i,j} = 1 \mid \mathbf{U}_{i-1}) = \frac{u_{i-1,j}}{\sum_{m=1}^M u_{i-1,m}} = \frac{o_j + w_{i-1,j}}{\sum_{m=1}^M o_m + w_{i-1,m}} = \frac{o_j + w_{i-1,j}}{i - 1 + \sum_{m=1}^M o_m}, \quad (18)$$

260 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
 262 new balls that were added to the urn after i draws of the color j .

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

266 *Proof.* Denote $\alpha' = \frac{\alpha}{1-\alpha}$ 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

$$G_{i,j} = \frac{\alpha' \beta(A'_j) + K_{i-1,j}}{\sum_{m=1}^N \alpha' \beta(A'_m) + K_{i-1,m}} = \frac{\alpha' \beta(A'_j) + K_{i-1,j}}{i - 1 + \sum_{m=1}^N \alpha' \beta(A'_m)}. \quad (19)$$

Substituting $M = N$, $o_j = \alpha' \beta(A'_j)$, and $w_{i,j} = K_{i,j}$ in eq. (18) gives eq. (19), thus completing the proof.

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 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 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.

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

Numerical validation. To validate our analytical result (corollary 3) and test its sensitivity to the 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 empirical distribution of multiple simulations. We then compared this observed distribution with the expected theoretical DM distribution as can be seen in fig. 2 (a). The difference in distributions 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 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 fig. 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.

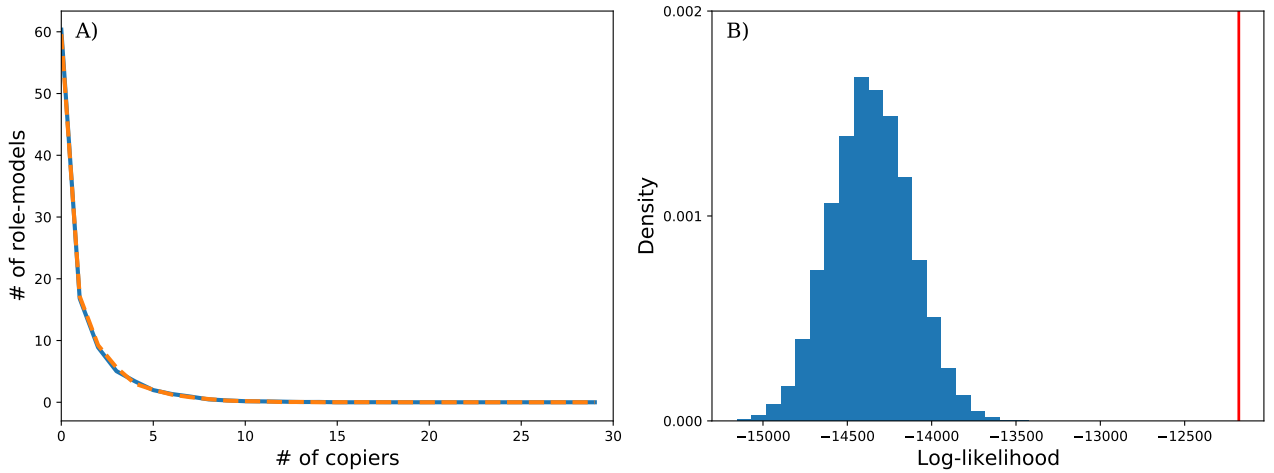


Figure 2: Goodness of fit of Dirichlet-Multinomial approximation to the model. The difference between the DM approximation distribution to observed model distribution (a) is very small. The DM approximation (orange) and the simulation results (blue) are almost indistinguishable. The optimal negative log likelihood (b) is given when using the DM approximation (black vertical line), and is much better when compared to a shuffled vector of the DM distribution (blue bars). Here, population size, $N = 100$; number of distributions averaged, $m = 100$; phenotype values, $\hat{A} = 1$, $A \sim N(0, 1)$; homogeneous success bias weight, $\alpha = 0.5$.

Next, we examined the fixation probability and fixation time of an advantageous phenotype \hat{A} when invading a population of phenotype A and compared results from the full model and the DM approximation. We find that the number of simulations needed to sufficiently approximate our model with the DM approximation is roughly 1,000 (Figure 3). Next, 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 (fig. 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 $\alpha_j \sim N(0.5, x)$ where $x \in [10^{-7}, 10^{-1}]$. We found again that results of the DM approximation are similar to those from stochastic simulations of the full model (fig. 5).

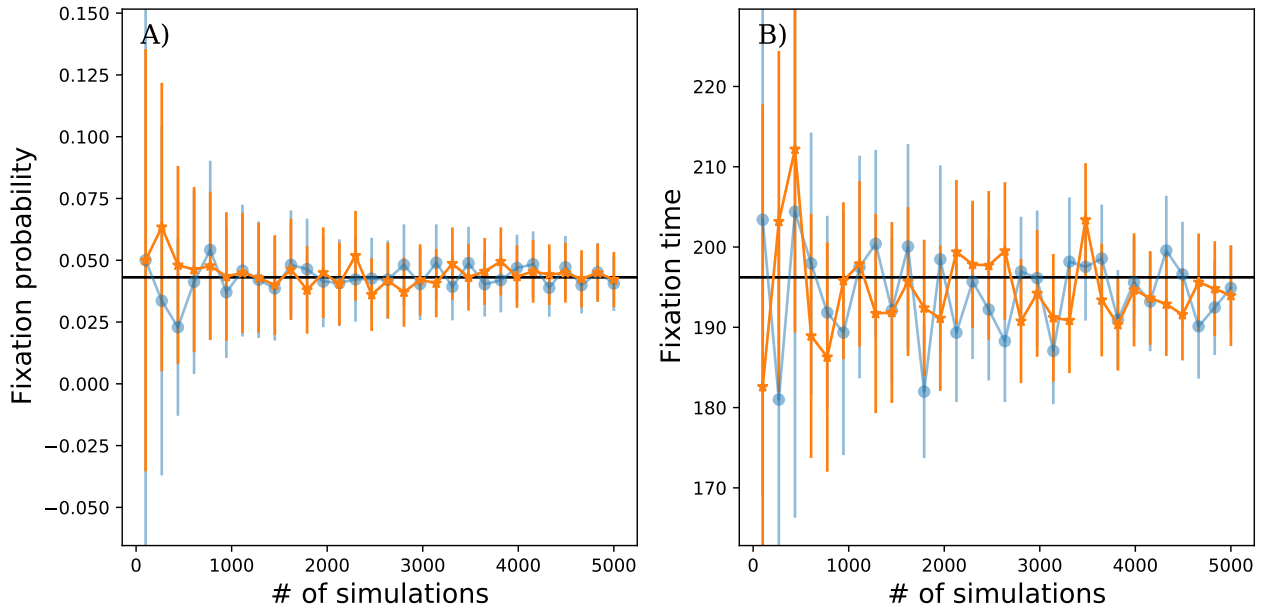


Figure 3: Minimum number of simulations required to get a good approximation. The approximation (orange) fits simulation results (blue) well when using 1,000 simulations for both models. The approximated value (black) is given by equation eq. (32). Markers for average value across simulations, error bars for 95% confidence interval. Here, population size, $N = 1000$; bias weight, $\alpha = 0.5$; phenotype values, $\hat{A} = 1$, $A = 0.7$; success bias value, $\beta(A) = 0.956$.

304 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 fixation time of an advantageous phenotype, using a diffusion-equation approximation approach, similar to analyses of population-genetic models (Kimura, 1962; Kimura and Ohta, 1969; Otto and Whitlock, 2006). We are 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 error in this analysis, i.e $e_i = 0$ for every

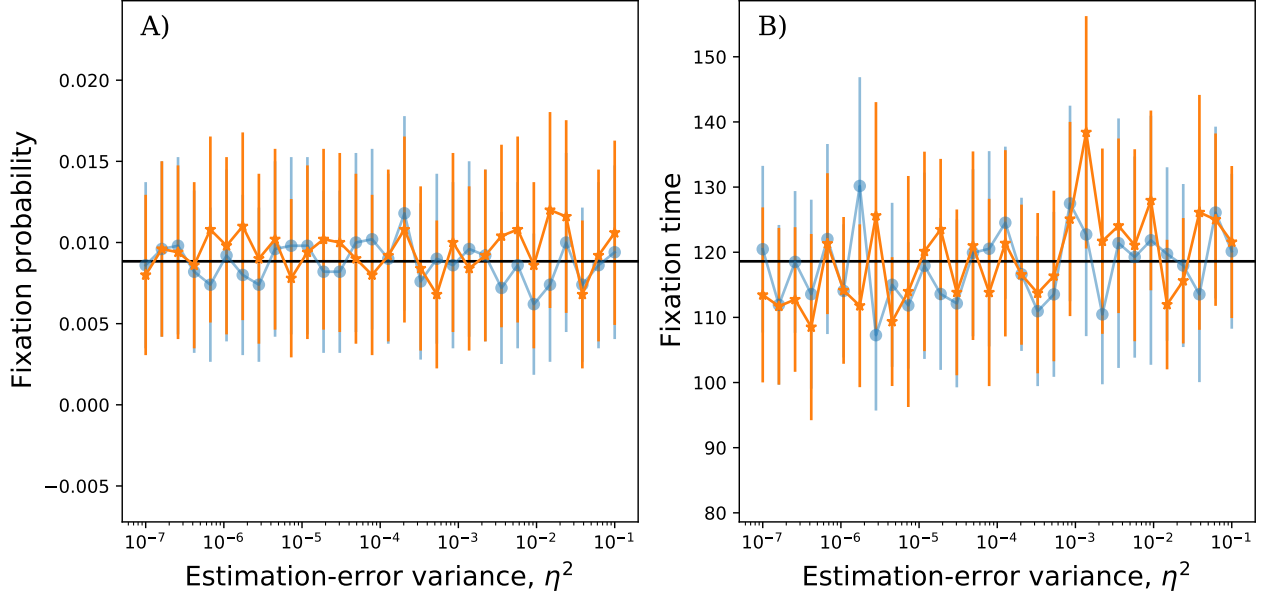


Figure 4: Robustness of DM approximations to inclusion of estimation error. Both the DM approximation (orange) and Kimura's equation (black line) fit the stochastic simulations (blue) well even with a high estimation error rate. Markers for average across simulations, error bars for 95% confidence intervals.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

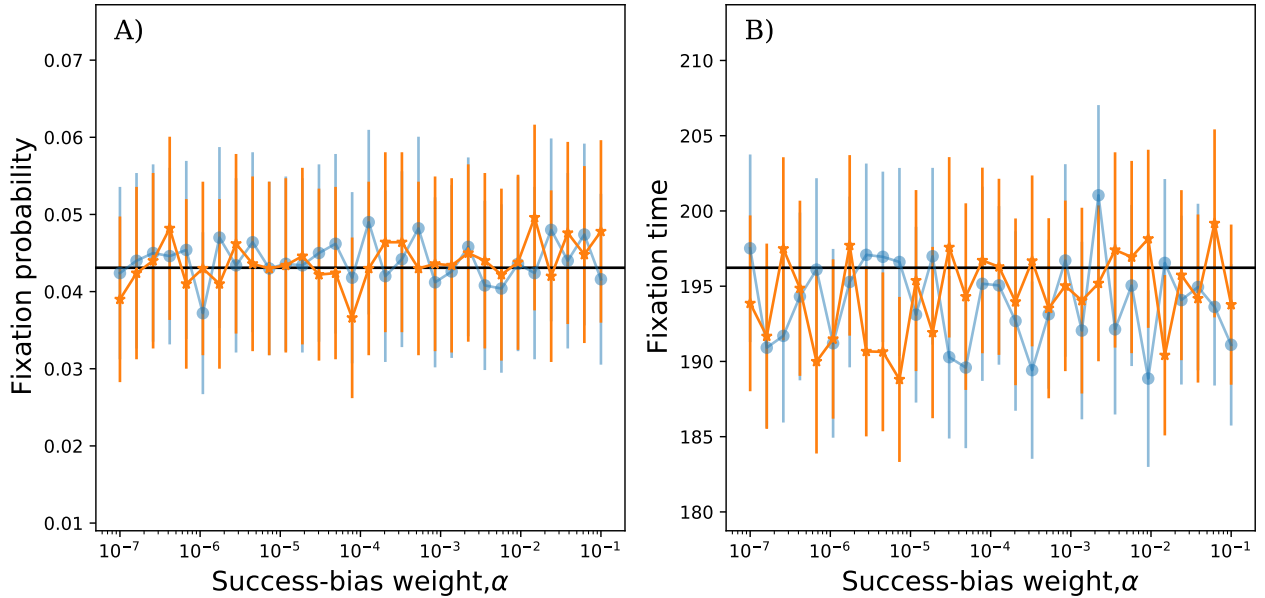


Figure 5: Robustness of DM approximations to variation in the bias weight α . Both the DM approximation (orange) and Kimura's equation (black line) fit the stochastic simulations (blue) well even with a high variation in success bias weight. Markers for average across 5,000 simulations, error bars are 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, $\beta(A) = 0.956$.

312 copier i . As shown above, transmission in our model is approximately DM distributed (corollary 3
and eq. (19)).

314 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.

316

Result 3 (Drift and diffusion terms in a constant environment). *Let x and x' be the frequency of type \hat{A}
318 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,*

$$320 \quad E[x' - x] \approx x(1 - x)(1 - \beta), \quad V(x' - x) \approx x(1 - x) \left(\frac{1}{\alpha N + (1 - \alpha)} \right). \quad (20)$$

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

$$324 \quad E[X'] = N \frac{\alpha_1}{\alpha_1 + \alpha_2}, \quad (21)$$

where $\alpha_1 = \alpha'X$ and $\alpha_2 = \alpha'(N - X)\beta$, from eq. (11). To use frequencies instead of counts,
326 $E[x'] = E[X'/N] = \frac{1}{N}E[X']$. Putting it together,

$$\begin{aligned} E[x'] &= \frac{1}{N} N \frac{\alpha'xN}{\alpha'xN + \alpha'N(1 - x)\beta} = \frac{x}{x + (1 - x)\beta} \\ &= \frac{x}{x + (1 - x) - (1 - x) + (1 - x)\beta} = x \frac{1}{1 - (1 - x)(1 - \beta)} \\ &= x(1 + (1 - x)(1 - \beta) + o(\beta)) = x + x(1 - x)(1 - \beta) + o(\beta), \end{aligned} \quad (22)$$

328 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

$$330 \quad E[x' - x] = E[x'] - E[x] = x(1 - x)(1 - \beta) + o(1 - \beta), \quad (23)$$

which gives us the drift term of the diffusion equation.

332 Using the variance of the DM distribution,

$$V(X') = N \frac{\alpha_1}{\alpha_1 + \alpha_2} \left(1 - \frac{\alpha_1}{\alpha_1 + \alpha_2} \right) \left(\frac{N + \alpha_1 + \alpha_2}{1 + \alpha_1 + \alpha_2} \right). \quad (24)$$

334 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,

$$336 \quad V(x') = \frac{1}{N^2} N \frac{x}{x + (1 - x)\beta} \left(1 - \frac{x}{x + (1 - x)\beta} \right) \left(\frac{N + \alpha'xN + \alpha'N(1 - x)\beta}{1 + \alpha'xN + \alpha'N(1 - x)\beta} \right). \quad (25)$$

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

$$\frac{x}{x + (1-x)\beta} \approx x \quad (26)$$

and for the entire variance expression we get

$$V(x') \approx \frac{1}{N}x(1-x) \left(\frac{N + \alpha'xN + \alpha'N - \alpha'xN}{1 + \alpha'xN + \alpha'N - \alpha'xN} \right) = x(1-x) \left(\frac{1 + \alpha'}{1 + \alpha'N} \right). \quad (27)$$

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

$$V(x' - x) = V(x') - V(x) \approx x(1-x) \left(\frac{1 + \alpha'}{1 + \alpha'N} \right). \quad (28)$$

α' is the odds ratio of the bias weight,

$$\alpha' = \frac{\alpha}{1 - \alpha}. \quad (29)$$

Combining eq. (28) and eq. (29) we get:

$$V(x' - x) \approx x(1-x) \left(\frac{1 + \frac{\alpha}{1-\alpha}}{1 + \frac{\alpha}{1-\alpha}N} \right) = x(1-x) \left(\frac{1}{\alpha N + (1-\alpha)} \right). \quad (30)$$

This gives the diffusion term of the diffusion equation, which completes the proof.

This analysis gives a surprising result relating the parameters α and β to parameters of the classical 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 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 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, 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*

$$s = 1 - \beta(A), \quad N_e = \alpha N + (1 - \alpha). \quad (31)$$

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*

$$\pi = \frac{1 - e^{-2(1-\beta)N_e x}}{1 - e^{-2(1-\beta)N_e}} \quad (32)$$

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

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

368

Result 6 (Fixation time approximation).

$$T = J_1 + \frac{1 - u(x)}{u(x)} \cdot J_2, \quad (33)$$

370 where, given $S = Ne(1 - \beta)$ and $x = \frac{1}{N_e}$,

$$u(x) = \frac{1 - e^{-2Sx}}{1 - e^{-2S}} \quad (34)$$

372 and,

$$J_1 = \frac{2}{(1 - \beta)(1 - e^{-2S})} \int_x^1 \frac{(e^{2S\xi} - 1)(e^{-2S\xi} - e^{-2S})}{\xi(1 - \xi)} d\xi, \quad (35)$$

374

$$J_2 = \frac{2}{(1 - \beta)(1 - e^{-2S})} \int_0^x \frac{(e^{2S\xi} - 1)(1 - e^{-2S\xi})}{\xi(1 - \xi)} d\xi \quad (36)$$

$$376 \quad T = \frac{1 - \pi}{1 - \beta} \int_0^x \frac{e^{2(1-\beta)\xi} - 1}{\xi(1 - \xi)} d\xi + \frac{\pi}{1 - \beta} \int_x^1 \frac{1 - e^{-2(1-\beta)(1-\xi)}}{\xi(1 - \xi)} d\xi \quad (37)$$

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

Numerical validation. To validate our analysis we performed multiple simulations comparing our binary model with the classic Wright-Fisher model, using different α and β each time, and using the corresponding values of $s = 1 - \beta$ and $N_e = \alpha N + (1 - \alpha)$ for the Wright-Fisher simulations.

382 We find see that the two models have similar dynamics, and both are well approximated the the approximations in eqs. (32) and (37).

384 After finding a good approximation for our model in a constant environment, when the favorable trait is always \hat{A} , we continued to find an approximation for our model in a changing environment.

386 Following Ram et al. (2018), we find the following expressions for the expectation and variance of the change in frequency between t generations.

388

Result 7 (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,*

$$E[X_t/N - x] \simeq x(1 - x)S_t/N, \quad V(X_t/N - x) = V(X_t/N) \simeq x(1 - x)t/N_e, \quad (38)$$

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

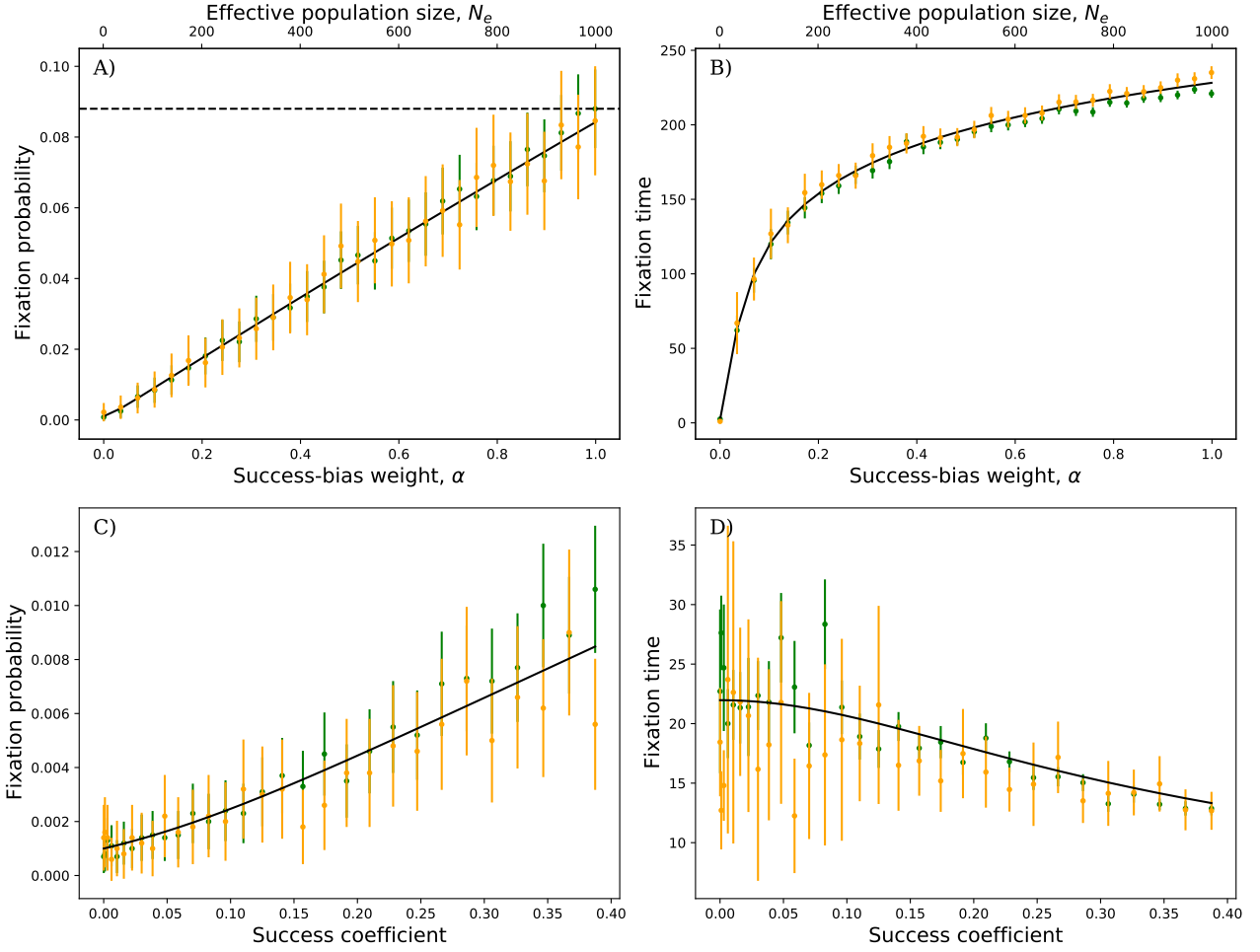


Figure 6: Fixation probability and time approximation for varying bias weight and success coefficient. Fixation probability (A) and time (B) for varying bias weight on the bottom x-axis. The corresponds effective population size is on the top x-axis. The approximation (black) fits DM simulations results (green) and Wright-Fisher simulation results (orange) when using eq. (32). Fixation probability (A) is limited at approximately 2s (blue), as in the classic WF model. Markers are averages of 10,000 simulations, error bars show 95% confidence intervals for (A,B) and 75% for (C,D). Here, Population size, $N = 1,000$; phenotype values, $\hat{A} = 1$, $A = 0.7$ (A,B), $A = a \cdot \hat{A}$, $a \in [0.01, 0.99]$ (C,D); success coefficient, $1 - \beta = s = 0.044$ (A,B); bias weight, $\alpha = 0.01$ (C,D).

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
394 both terms in eq. (38). From eq. (23) we know that

$$E \left[\frac{X_{t+1}}{N} - \frac{X_t}{N} \middle| X_t \right] = \frac{X_t}{N} \left(1 - \frac{X_t}{N} \right) (1 - \beta_{t+1}) = \frac{1}{N} \frac{X_t}{N} \left(1 - \frac{X_t}{N} \right) s_{t+1} . \quad (39)$$

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

$$\begin{aligned} E \left[\frac{X_t}{N} \left(1 - \frac{X_t}{N} \right) \right] &= E \left[\frac{X_t}{N} - \left(\frac{X_t}{N} \right)^2 \right] = E \left[\frac{X_t}{N} \right] - E \left[\left(\frac{X_t}{N} \right)^2 \right] \\ &= E \left[\frac{X_t}{N} \right] - V \left(\frac{X_t}{N} \right) - \left(E \left[\frac{X_t}{N} \right] \right)^2 . \end{aligned} \quad (40)$$

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

$$E \left[\frac{X_t}{N} \left(1 - \frac{X_t}{N} \right) \right] \simeq E \left[\frac{X_t}{N} \right] \left(1 - E \left[\frac{X_t}{N} \right] \right) - \frac{1}{N_e} tx(1-x) . \quad (41)$$

400 From eq. (39) we know that

$$\begin{aligned} E \left[\frac{X_{t+1}}{N} - \frac{X_t}{N} \right] &= \frac{1}{N} s_{t+1} E \left[\frac{X_t}{N} \left(1 - \frac{X_t}{N} \right) \right] \simeq \frac{1}{N} s_{t+1} \left(E \left[\frac{X_t}{N} \right] \left(1 - E \left[\frac{X_t}{N} \right] \right) - \frac{1}{N_e} tx(1-x) \right) \\ &\simeq \frac{1}{N} s_{t+1} \cdot E \left[\frac{X_t}{N} \right] \left(1 - E \left[\frac{X_t}{N} \right] \right) - \frac{1}{N_e N} s_{t+1} tx(1-x) . \end{aligned} \quad (42)$$

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

$$E \left[\frac{X_{t+1}}{N} - \frac{X_t}{N} \right] \simeq \frac{1}{N} s_{t+1} \cdot E \left[\frac{X_t}{N} \right] \left(1 - E \left[\frac{X_t}{N} \right] \right) . \quad (43)$$

404 We now look at the induction assumption to see that

$$E \left[\frac{X_t}{N} - x \right] = E \left[\frac{X_t}{N} \right] - E[x] = E \left[\frac{X_t}{N} \right] - x , \quad (44)$$

406 so using the assumption we get

$$\begin{aligned} E \left[\frac{X_t}{N} \right] &\simeq \frac{1}{N} S_t x(1-x) + x \\ 1 - E \left[\frac{X_t}{N} \right] &\simeq 1 - \frac{1}{N} S_t x(1-x) + x . \end{aligned} \quad (45)$$

408 We use both expressions in eq. (43) and get

$$\begin{aligned} E \left[\frac{X_{t+1}}{N} - \frac{X_t}{N} \right] &\simeq \frac{1}{N} s_{t+1} \left(\frac{1}{N} S_t x(1-x) + x \right) \left(1 - \frac{1}{N} S_t x(1-x) + x \right) \\ &\simeq \frac{1}{N} s_{t+1} \cdot x(1-x) , \end{aligned} \quad (46)$$

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

$$E \left[\frac{X_{t+1}}{N} - x \right] = E \left[\frac{X_{t+1}}{N} - \frac{X_t}{N} \right] + E \left[\frac{X_t}{N} - x \right] , \quad (47)$$

412 so again using the induction assumption, together with eq. (46) we get

$$\begin{aligned} E \left[\frac{X_{t+1}}{N} - x \right] &\simeq \frac{1}{N} s_{t+1} \cdot x(1-x) + \frac{1}{N} S_t \cdot x(1-x) \\ &\simeq \frac{1}{N} x(1-x) (S_t + s_{t+1}) \simeq \frac{1}{N} S_{t+1} x(1-x) , \end{aligned} \quad (48)$$

414 which proves the drift term.

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

$$V\left(\frac{X_{t+1}}{N}\right) = E\left[V\left(\frac{X_{t+1}}{N}\middle|X_t\right)\right] + V\left(E\left[\frac{X_{t+1}}{N}\middle|X_t\right]\right). \quad (49)$$

Using eq. (39) we see that

$$\begin{aligned} E\left[\frac{X_{t+1}}{N}\middle|X_t\right] - E\left[\frac{X_t}{N}\middle|X_t\right] &= \frac{1}{N}s_{t+1}\frac{X_t}{N}\left(1 - \frac{X_t}{N}\right) \\ E\left[\frac{X_{t+1}}{N}\middle|X_t\right] &= \frac{X_t}{N} + \frac{1}{N}s_{t+1}\frac{X_t}{N}\left(1 - \frac{X_t}{N}\right). \end{aligned} \quad (50)$$

Using eq. (30) we get

$$V\left(\frac{X_{t+1}}{N}\middle|X_t\right) = \frac{1}{N_e}\frac{X_t}{N}\left(1 - \frac{X_t}{N}\right), \quad (51)$$

and using the equation $y'(1 - y') \simeq y(1 - y)$ on the first part of eq. (49) we get

$$E\left[V\left(\frac{X_{t+1}}{N}\middle|X_t\right)\right] = \frac{1}{N_e}E\left[\frac{X_t}{N}\left(1 - \frac{X_t}{N}\right)\right] \simeq \frac{1}{N_e}x(1 - x). \quad (52)$$

Moving on to simplify the second part of eq. (49) using eq. (50),

$$V\left(E\left[\frac{X_{t+1}}{N}\middle|X_t\right]\right) = V\left(\frac{X_t}{N} + \frac{1}{N}s_{t+1}\frac{X_t}{N}\left(1 - \frac{X_t}{N}\right)\right) \quad (53)$$

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

$$V\left(\frac{1}{N}s_{t+1}\frac{X_t}{N}\left(1 - \frac{X_t}{N}\right)\right) \leq \frac{1}{4N^2}s_{t+1}^2, \quad (54)$$

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

$$V\left(E\left[\frac{X_{t+1}}{N}\middle|X_t\right]\right) = V\left(\frac{X_t}{N}\right) + O\left(\frac{1}{N^2}\right) \simeq V\left(\frac{X_t}{N}\right). \quad (55)$$

Using the induction assumption and eq. (52),

$$V\left(\frac{X_{t+1}}{N}\right) \simeq \frac{1}{N_e}x(1 - x) + \frac{1}{N_e}tx(1 - x) \simeq \frac{1}{N_e}x(1 - x)(t + 1) \quad (56)$$

which proves the diffusion term.

Using the drift and diffusion terms and following Ram et al. (2018), we can approximate the fixation probability in a changing environment using

$$\tilde{\pi} = \frac{1 - e^{-2\frac{S_n}{n}N_e x}}{1 - e^{-2\frac{S_n}{n}N_e}} \quad (57)$$

436 where $\frac{S_n}{n} = \frac{k-l}{k+l}(1 - \beta)$ and $n = k + l$. That is, we use the average selection coefficient during a cycle of $k + l$ generations as the selection coefficient eq. (32).

438 **Numerical validation.** Comparing our approximation (eq. (57)) to numerical simulations, we find that the approximation fits simulations results well for variable bias weights, α , which corresponds to
440 the effective population size (Figure 7A).

However, the approximation is more sensitive to the value of the success bias coefficient β (Figure 7B).
442 We suspect that when β 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
444 phenotype becomes extinct very quickly. For the β values (0.65 and below), the fixation probability exceeds even the constant environment approximation (which is the upper limit). However, this is to
446 be expected, because the diffusion-equation approximations assume weak selection (i.e., low selection coefficient s).

448 When k is the number of generations where the invading phenotype is advantageous, and l when the original trait is, we found that for large k -to- l ratio (while keeping constant total cycle length,
450 $n = k + l = 100$), the changing environment approximation (eq. (57)) converges to the constant environment approximation (eq. (32)), see Figure 7 (C,D).

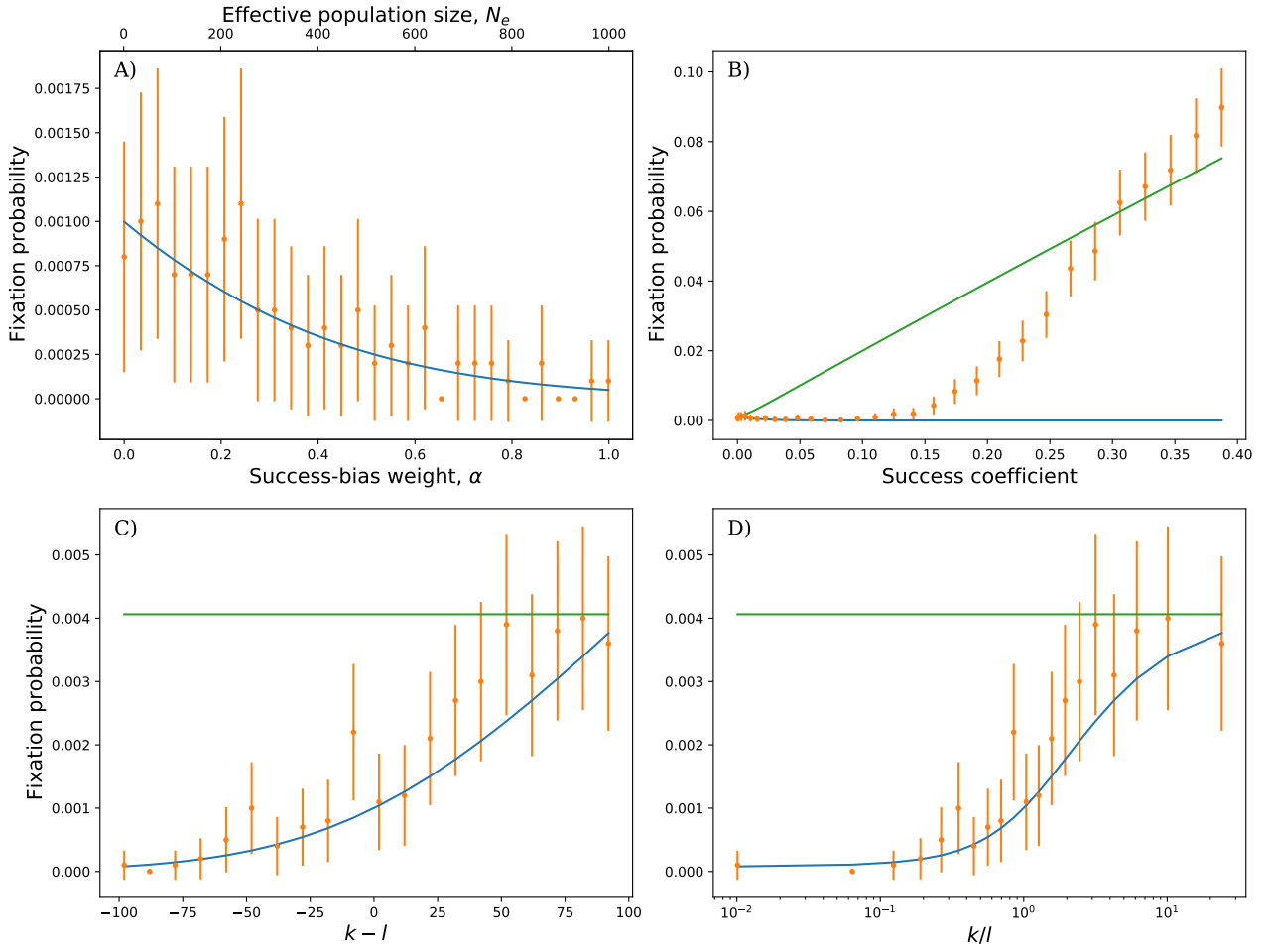


Figure 7: Sensitivity of fixation probability approximation in a changing environment to varying success coefficient and bias weight and robustness to varying ratios of cycles. (A) When the bias weight, or effective population size varies, the approximation (blue; eq. (57)) fits simulation results (orange). (B) When success bias is large ($1 - \beta > 0.1$), simulation results (orange) are underestimated by the approximation (blue; eq. (57)). With even larger success bias ($1 - \beta > 0.35$), even the constant environment approximation (green; eq. (32)) underestimate simulation results, likely because the diffusion equation approximation assumes weak selection. We can also see the robustness of the approximation to changes in cycle ratios (C,D): The approximation (blue) fits the simulations (orange) well for different sizes of the changing environment cycle, where k is the number of generations where the bias is advantageous, and l are the generations it is disadvantageous. When $k < l$ the approximation is good. When $k > l$, the approximation and the simulations are both very close to the constant environment approximation (green). Markers show average of 10,000 simulations, error bars show are 75% (A,C,D) and 95% (B) confidence intervals. Here, population size, $N = 1,000$; phenotype values, $\hat{A} = 1$, $A = 0.9$ (A,B), $A = 0.8$ (C,D); In (A) success bias/selection coefficient is: $1 - \beta = s = 0.005$; In (B,C,D) success bias weight is: $\alpha = 0.1$.

452 Adaptive success-bias weight

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

$$456 \quad \alpha^* = \operatorname{argmin} \sum_{j=1}^N \frac{\alpha A_j + (1 - \alpha) K_j}{\sum_{l=1}^N \alpha A_l + (1 - \alpha) K_l} \cdot (\hat{A} - A_j)^2, \quad (58)$$

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 (fig. 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 fig. 8b and c): in this example, the adaptive weight estimation error converges to 0.046, whereas a constant weight gives values > 0.74 .

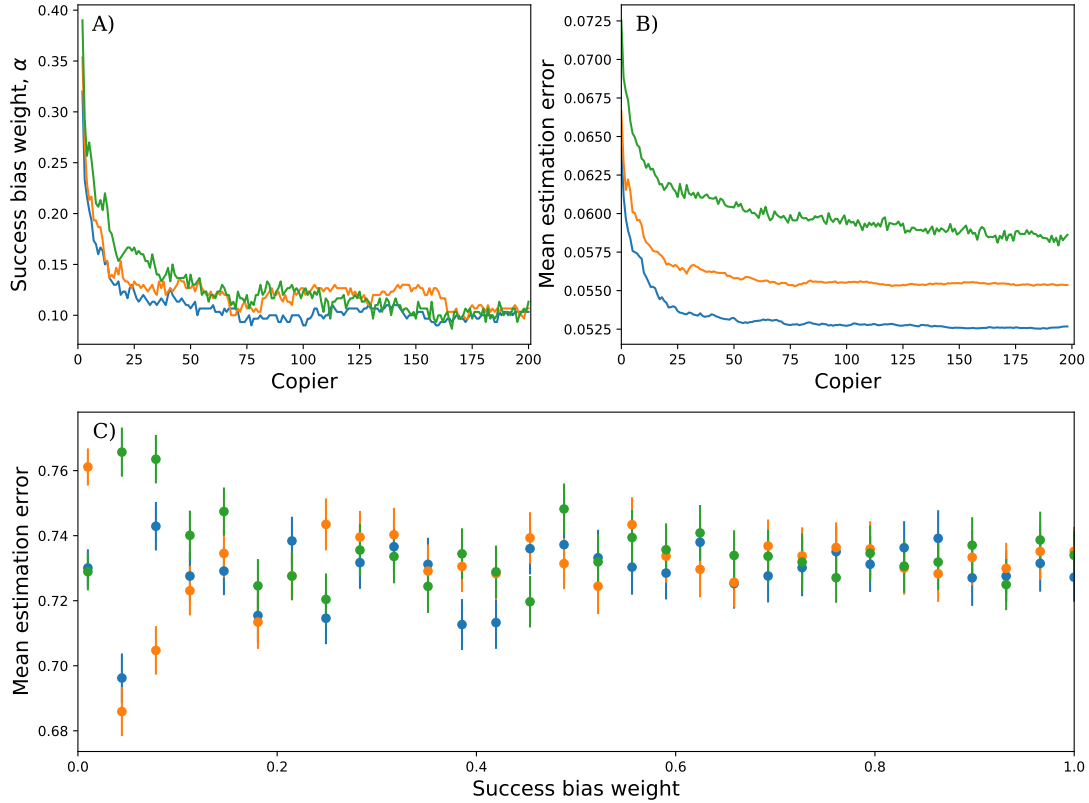


Figure 8: Advantage of an adaptive success-bias weight. Both success-bias weight α (a) and estimation error (b) decrease during the role-model choosing process, demonstrating that influence becomes more advantageous as more copiers have made their choice. However, when α is homogeneous, the mean estimation 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.

464 Discussion

Summary. During cultural transmission, cultural traits such as attitudes, values, beliefs, and behavioral patterns are transmitted between individuals, for example via copying and social learning. Some cultural traits or cultural role-models may be more likely to be copied due to transmission biases. A common one is 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 (Fogarty et al., 2017) that because estimating success is hard, a more common bias is *prestige bias*—a bias towards role-models perceived to be successful, portrayed by other traits or means.

Inspired by a model by Boyd and Richerson (1988), we developed a cultural-evolution model with prestige bias, which included both success and influence biases, where the latter is a bias towards role-models with many copiers. As shown in fig. 8, given the option to choose the ratio of success and influence, copiers will minimize their estimation error, hence will copy the traits that are closer to the ideal phenotype. We see that throughout the choosing process, the estimation error decreases, and the success bias rate also decreases. This suggests that the later a copier gets to choose, the less he should rely on his own estimation, and more on previous choices. This behavior fits the intuitive thought - the more information you have, using others as an information source, the more informative your choice can be, therefore minimizing the distance from your choice of trait to the ideal one.

In addition, we see that the process of choosing itself is necessary, and when using a constant ratio (α), the mean estimation error in the population is greater by a factor of 10, even when using the α the population converged to when being able to choose it (fig. 8). After we showing that choosing one's success bias weight is beneficial to the copier, we continued to create our models- continuous and dichotomous, that incorporated both success and influence. However, running simulations for a nested iterative model is costly, and analyzing such model mathematically is very hard. We therefor found approximations for the role-model choice process: the generalized binomial distribution (result 1) and the Dirichlet-Multinomial distribution (corollary 3). To use approximate our model with the GBD distribution, we must assume homogeneous estimation error, e , and homogeneous success bias weight, α . Both assumptions are necessary to this approximation, because it requires all copiers to be identical (order of copiers must not be a factor). For the DM approximation we can relax the assumption for α , and allow a varying α_j for each role-model j . This means that a copier is still not able to choose a success bias weight for itself, and instead it is a characteristic of the role-model only. Using the DM approximation, because it is easier to analyze, we found equivalents for Kimura's equations, for our dichotomous model. Following result 4 we found an approximation for the fixation probability of the ideal trait, and an approximation for the time of fixation. These approximations hold only when the equivalent for the selection coefficient, in our model $1 - \beta(A)$, is low (0.4 or lower). We also discovered that changing the success bias weight doesn't hurt the goodness of the approximations, and only affects the effective population size N_e . Once we found the equations for a constant environment, we moved on to a cyclic changing environment. Based on result 7, we found the approximations for the fixation probability and time to fixation. We discovered that as in the constant environment, the

effective population size doesn't affect the goodness of fit of the approximation. The success coefficient however, must be even lower than before for the approximation to fit. When the success coefficient is higher 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 amount of cycles, and when the success coefficient 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 beneficial trait becomes the maladaptive), or the opposite 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 beneficial and maladaptive. We showed that the approximation fits well regardless of the ratio, but when the ratio of adaptive generations to maladaptive ones is very high, it is very similar to a constant environment model.

Prestige in the literature. According to King and Cowlishaw (2009), there are two main approaches to group decision-making in nature: leadership and consensus. Leaders may lead the group when traveling, decide on a nesting site, or choose foraging patches. They found that leadership is mostly observed when there are a lot of connections in the social network of the group, and when there are individuals that present leadership behavior. Leaders would usually be high-ranking members of the group: elders, individuals with many kin relations, or individuals possessing other dominant traits. When no individual possesses such traits, or when there are not enough connections and relations in the social network, decisions are more likely to be made by consensus. Leaders may carry selection costs, e.g. due to higher predation risk, poisoning when exploring new foraging patches. In some cases, the leader may be the hungriest or the weakest individual, while the others would prefer to follow the leader and minimizing their risks and costs. However, leadership may also carry selection benefits. For example, given the route to the foraging site was successful, the leader and his closest followers would gain most or the best of the food, whereas in consensus decision making, the food would be shared more equally. Or, sexual selection may also provide a selection benefit advantage to leaders, who might gain more sexual partners due to their leadership of the group. 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* (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 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 harm.

In modern society humans strive for the prestigious positions, as they may reap rewards greater than the risk and costs to achieve them, or due to individual personality and pressure/education from the family.

Van Vugt and Smith (2019) suggest a different view of leadership. They note that most discussions assume there is one type of leadership, as seen above, and so they differ in their definitions. Van Vugt and Smith (2019) suggest a way to solve said contradiction by defining two types of leaderships:

prestige-based and dominance-based. They present classical views of leaderships by Confucius and
542 Machiavelli. Confucius views leaders as role-models who exercise influence through possessing
superior knowledge, skills, and (outstanding) personal qualities. This description is very similar to
544 success bias in our model. By contrast, Machiavelli views leaders as rulers who exercise influence by
imposing costs through (the threat of) punishment and violence. They say that these two opposing
546 views are both partially supported by the available evidence but each one on its own offers an
incomplete view into the complex and dynamic processes of leadership. Our model does not reflect
548 these leadership styles, but several adjustments could be made in order to match it, such as assuming
there is a correlation between phenotype to leadership style. The emerging cultural-evolutionary
550 dynamics and their dependence on the costs and benefits are intriguing.

Henrich and Gil-White (2001) support said claim that there are two types of leadership, and also
552 define the two as prestige based and dominance based leadership types. By their definition, the latter
is defined by acquiring social status by using aggression, intimidation and violence. It is also more
554 common than prestige in non humans. Their definition of prestige is somewhat synonymous with
ours. According to their manuscript, prestige is composed both of estimation in the eyes of people
556 (our success trait) and commanding position in people's minds, i.e number of copiers people think
they have, which they define as *influence* (similar to our definition for influence). In their paper, they
558 show that prestige evolved from natural selection, as an efficient process to extract reproductive benefit
from the flow of socially transmitted information. Simply put, prestige is a natural step where social
560 learning exists, due to saving costs of individual learning. It could be interesting in the future to expand
our model using this idea: observing the copier trait of *evaluation*, rather than only observing the
562 evolution of the trait copied. Henrich and Gil-White (2001) suggest that the most skilled role-models
will, on average, end up with most copiers. Their research, definitions and results, is consistent with
564 ours.

Furthermore, according to Boyd and Henrich (2002), the process of cultural evolution doesn't require
566 replication of representations. They base their assumption on three claims - mental representations are
non-discrete, cultural transmission is highly inaccurate, and mental representations are not replicated,
568 but rather are 'reconstructed' through an inferential process that is strongly affected by cognitive
'attractors.' In their paper they describe three different models to support their points. We see a high
570 similarity between their model and ours. Like them, we treat the cultural trait as non-discrete (in
the main model, before simplifying it to dichotomous in order to analyze it). We also assume error
572 in estimation, which the copiers aren't really replicating their role-models, but rather reconstructing
them to create a different trait, which is their representation of the role-model's trait. In addition, the
574 inferential process that they describe as strongly affected by cognitive attractors match our influence
bias, resulting in a very similar base assumptions in our models.

576 So far we presented the theory behind prestige, and it's appearance in nature. The following will show
the appearances of prestige biases in humans, and in recent times.

578 Chudek et al. (2012), for example, tested the existence of prestige in young children. Chudek et al.
(2012) report the first direct tests in children that suggest the existence of *prestige bias*, a tendency

580 to learn from individuals to whom others have preferentially attended, learned or deferred. Their
definition of prestige is similar to our *influence bias*, and brings concrete proof of its existence and
582 effects. Their study showed that the odds of 3-4 years-old children learning from an adult model to
whom bystanders had previously preferentially attended for 10 seconds were over twice those of their
584 learning from a model whom bystanders ignored. In addition to this first study, they also discovered
prestige effects are domain-sensitive. They saw that prestigious models were listened to by most when
586 demonstrating artifact-use, but not as much as when presenting food preferences. It lead Chudek et al.
(2012) to believe that when the trait is costly to learn individually, prestige will have a higher effect. It
588 would be interesting to include costs in our model to try and observe these effects and their dynamics
in the simulations of a larger population than this study.

590 Henrich and Broesch (2011) researched Fijian villages, looking for evidence of social learning biases
and their origins. They mention that:

592 evolutionary theorists propose that natural selection has favored the emergence of psy-
chological biases for learning from those individuals most likely to possess adaptive
594 information.

Their goal is to bridge from the laboratory to the field by examining if and how these biases emerge
596 in a small-scale society. During their research they found that:

Fijian villagers (ages 10 and up) are biased to learn from others perceived as more
598 successful/knowledgeable, both within and across domains (prestige effects).

Their research shows promising evidence for our prestige model, suggesting that copying from others
600 who are *perceived* as successful, rather than actually are successful. In their paper, they show that the
social networks representing copier-role-model relationships are centralized, suggesting:

602 This degree of centralization is consistent with the prediction that people substantially
share notions about who is a good cultural model (network centrality), but that individuals'
604 model selections are influenced by multiple factors.

We see here support for both our trait and our influence bias in their data.

606 Aside from children's learning biases and small villages in a relatively primitive population, we can
see prestige in more advanced domains as well, like western medicine. Norredam and Album (2007)
608 present a specific and important effect of prestige - its significance for medical specialties and diseases.
They examined literature from 1950 to 2005 regarding the effects of prestige on medicinal practices.
610 They discovered that active, specialized, biomedical, and high-technological types of medicine on
organs in the upper part of the bodies of young and middle-aged people were accorded high levels of
612 prestige, while medicine with opposite characteristics had low levels of prestige. They have concluded
that such differences in prestige bear consequences for actual priority setting in healthcare systems.
614 They discovered 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
616 prestigious, while other procedures that are considered *easier* are less prestigious. Simply put, they

found that the advance in technology and research was in accordance with the prestige rankings. This
618 means that there may be very important practices that are neglected due to the prestige bias.

As we seen so far, prestige can explain many behaviors and evolution of cultural traits. It is a tool to
620 cheaply estimate and acquire knowledge, which helps an individual to survive and breed. However,
it is not always the case, and there could be negative repercussions to this bias, such as invasion of
622 maladaptive traits.

Takahashi and Ihara (2019) mention that social learning not only takes the form of random copying
624 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 choose the one he deems
626 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
628 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 develop new mathematical
630 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
632 a very high payoff. Our model, which includes influence bias, is consistent with that claim. When
a maladaptive trait is "piggybacking" a role-model with high influence, said trait could spread in the
634 population, as mentioned. 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 is
636 based on copying based learning only, but it could be interesting to combine it with individual learning
and see how it affects the dynamics of the population.

We discussed prestige in depth, and provided several proofs for its existence in nature, humans, and
638 even medicine. We saw it could aid invasion of maladaptive traits, or neglect of important medicinal
specialities. But, it can also accelerate reversal of harmful traditions. Harmful traditions can be child
640 marriage, open defecation, and domestic violence, to name a few. Efferson et al. (2020) suggest a
mechanism called *spillover*. By their definition, a spillover is when an intervention affect a large
642 enough group in a target population, so that others not included in the intervention starts changing
their behavior as well. In their research, they found that there are individuals who act as *agents*, who
644 are often looked upon, and therefore they are ideal targets for interventions. This is the same concept
as our role-models, where a more prestigious individual will be copied more, therefore spreading his
646 trait wider in the population. Their research support therefore in our assumption that there are social
biases, conformist influence specifically. They also suggest a way to use this phenomena to change
648 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 be used for any negative way that comes to mind.
650

Up until now, we showed that cultural transmission is a process that manifests in many species, with
652 emphasis on humans. We also displayed similarities between this process and genetic transmission,
while presenting differences between them, specifically selective biases such as influence and prestige.
654 We also presented examples of good and bad usages of such biases. All of these are mainly presented
as a parallel process to the natural selection process in regards to physical anatomy, or at least have

an indirect effect on it. Muthukrishna and Henrich (2016) offer a take on prestige as a factor of human physical evolution directly. They present a concept called *cultural brains* | brains that evolved primarily for the acquisition of adaptive knowledge. They build on the hypothesis of Dunbar (2009) that shows that larger, more complex brains can store and manage more information and in turn, this information can support the costs of a larger brain. Muthukrishna and Henrich (2016) built a model that predicts a strong relationship between brain size and group size, because group size also provides access to more adaptive knowledge. They later present their *cumulative cultural brain* hypothesis, an approach which proposes that human brains have evolved with an ability and proclivity 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). In short, they suggest that some of the reasons for the extreme increase in brain size in humans, are the ability to "cheaply" acquire adaptive knowledge, i.e transmission biases, such as prestige.

Further work. In this paper we described several models and processes, analyzed them, and discovered some interesting approximations and results. But even though, there's much more to research in the direction we took, which can help us better model, understand, and maybe even predict, human behavior, specifically in this era where social media is very accessible and informative. Specifically, a deeper analysis of the first model, where every copier chooses its α , is preferred. It would be interesting to see if the mean error value and the α are converging to specific values, and what they depend on. Following that, it is possible to relax some assumptions we took in our main model in order to find approximations for it - like homogeneous estimation error and success bias weight. Lastly, it would be interesting to analyze the continuous model, instead of its dichotomous version. As we've discussed it is probably a much more accurate model of cultural evolution.

Acknowledgements

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

References

- 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.
- 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*, 114(30):7782–7789.
- Denton, K. K., Ram, Y., Liberman, U., and Feldman, M. W. (2020a). Cultural evolution of conformity and anticonformity. *Proceedings of the National Academy of Sciences*.
- Denton, K. K., Ram, Y., Liberman, U., and Feldman, M. W. (2020b). Cultural evolution of conformity and anticonformity. *Proceedings of the National Academy of Sciences*, 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. PMID: 19575315.
- 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.
- Eickbush, M. T., Young, J. M., and Zanders, S. E. (2019). Killer meiotic drive and dynamic evolution of the wtf gene family. *Molecular biology and evolution*, 36(6):1201–1214.
- 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.
- Hamilton, W. (1964). The genetical theory of kin selection. *J. Theor. Biol.*, 7(7):1–52.
- 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.
- 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.
- 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.
- O’Brien, M. J., Lyman, R. L., Mesoudi, A., and VanPool, T. L. (2010). Cultural traits as units of analysis. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 365(1559):3797–3806.
- 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*, 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.
- 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*, 114(30):7814–7821.