

# Prestige as a Driving Force in Cultural Transmission

Saar Egozi

together with Yoav Ram

Efi Arazi School of Computer Science, IDC Herzliya

*March 9, 2022*

saartk@gmail.com

# Contents

2	<b>Abstract</b>	<b>2</b>
	<b>Introduction</b>	<b>2</b>
4	<b>Models and Methods</b>	<b>5</b>
	Continuous Model . . . . .	5
6	Binary model . . . . .	7
	Methods . . . . .	7
8	<b>Results</b>	<b>8</b>
	Approximations . . . . .	8
10	General Binomial Distribution Approximation . . . . .	8
	Dirichlet-Multinomial Distribution Approximation . . . . .	10
12	Numeric comparisons . . . . .	11
	Fixation probability and time - binary model . . . . .	12
14	Changing environment . . . . .	16
	<b>Discussion</b>	<b>21</b>
16	<b>Summaries</b>	<b>23</b>
	Animal leadership (King and Cowlshaw, 2009) . . . . .	23
18	Prestige as a type of leadership (Van Vugt and Smith, 2019) . . . . .	23
	Evidence from Fijian villages for adaptive learning biases (Henrich and Broesch, 2011a) . . . . .	24
20	Using social influence to reverse harmful traditions (Efferson et al., 2020) . . . . .	24
	The evolution of prestige (Henrich and Gil-White, 2001) . . . . .	25

## 22 Abstract

Copying our role-models have always been an efficient method of acquiring knowledge. Copying  
24 successful role-model is one of the methods of cultural transmission of traits. In this paper, we  
study the various biases and their effects of the population and it's evolution. The most common  
26 bias when choosing a role-model to copy is *success bias*, i.e copying whoever appears successful to  
us. This estimation is based on the performance of the role-model alone, without any other factors.  
28 Here, we study another factor we believe aids to better model the cultural inheritance of traits  
in a large population. *Influence bias* is a bias evaluated by the number of copiers a role-model  
30 already has. In our model we combine these components to what we call the **Prestige bias** and  
analyze its relationship to the dynamics of the population. We successfully found mathematical  
32 approximations to our model, easing the mathematical analysis and the computation power required  
for simulations. We show the value of these approximations using simulations, and their robustness  
34 to variations such as mutation and other relaxations required for the mathematical proofs. In the  
binary form of our model, we found alternatives to *Kimura's* equations for approximating fixation  
36 probability and time to fixation of an invading advantageous trait, in both a constant and changing  
environment. We show that *Influence* solely affects the effective population size. We found that  
38 influence acts as an accelerator for a state of the population, matching the *rich getting richer* it  
was based on. We believe such model better describes how humans acquire knowledge from one  
40 another, mainly in the last years where social networks are very popular. Social networks allow  
easy access to estimate number of copiers a role-model has, with little to no effort.

## 42 Introduction

Traits transmission is when an individual passes on a trait, genetic or behavioral, to another  
44 individual. Transmission in nature manifests in two main ways: genetic and cultural. Genetic  
transmission is when an individual, or several, transmit their genes to their offspring by duplication  
46 of their own cells. Cultural transmission is the way individuals transmit cultural traits (i.e behavior)  
from one another, typically via teaching and demonstrating. Cultural transmission is most common  
48 in humans (Cavalli-Sforza and Feldman, 1981, pg. 3) and in primates like chimpanzees (Horner  
et al., 2010; Kendal et al., 2015). The common cultural traits in humans are behavioral patterns,  
50 like personalities and habits, transmitted via observations and verbal discussions. Henrich and  
McElreath (2007) suggest that cultural learning may be particular to humans, but McComb et al.  
52 (2001) suggest that it appears in other mammals as well, elephants for example:

... the possession of enhanced discriminatory abilities by the oldest individual [matri-  
54 arch] in a group can influence the social knowledge of the group as a whole.

They showed that once a matriarch is removed from the group, the group's survival instincts are  
56 inferior. They support their hypothesis by exacting an experiment: playing audio recordings of  
African elephants, showing that groups with a matriarch recognize and react better to hostile or

58 friendly calls than the groups without one. Moreover, cultural transmission appears in other species,  
even simpler than mammals, such as *Drosophila*. Battesti et al. (2012) show that oviposition site  
60 choice in fruit flies is culturally transmitted. They showed that flies without experience in choosing  
sites, after spending some time with "experienced" flies, chose the same type of site without directly  
62 observing this behavior. Battesti et al. (2012) mention that how the information is transferred is  
still an open question, but suggest that the flies may use olfactory cues, like other animals such as  
64 rodents and bees.

66 Cultural transmission is similar to genetic transmission in many ways, while different in others.  
Similar to genetic transmission, the effects of culturally transmitted traits can be physiological  
68 rather than behavioral, and transmitted from parents to offspring. For example, parents can teach  
their children to be strong or tall, within some biological limits, by instructing them to maintain a  
70 healthy diet and engage in physical activity. Contrary to genetic transmission, the sources of the  
traits can be many, and not only parents. They can even be unrelated, like teachers, celebrities,  
72 coaches, the media, or any stranger that comes in contact with them. Cultural transmission can  
be vertical, where parents transmit to their children, but also oblique, where other adults transmit  
74 traits to children (not their own). Horizontal transmission is also possible, where peers transmit  
traits to one another. Lastly, vertical transmission in the opposite direction is possible too, where  
76 parents copy traits from their children (e.g. playing video games) as Cavalli-Sforza and Feldman  
(1981) and Creanza et al. (2017) suggest. In addition, even when a cultural trait is disfavored by  
78 natural selection, it still may spread across a population given transmission biases strong enough  
to negate the selection bias (Boyd and Richerson, 1988, Ch. 8 pg. 279).

80

Transmission bias occurs when a trait has a disproportionate probability from its frequency in  
82 the population to be transmitted. For example, Eickbush et al. (2019) show that there are genes  
of yeast called *wtf genes*, that bias their transmission to the gametes. They secrete a long life ex-  
84 pectancy poison, together with a short life expectancy antidote, so a gamete without the gene will  
perish (the poison will outlive the antidote). Transmission biases, though exist in genetic transmis-  
86 sion, are probably more common in cultural transmission. Much like mutation in genetic evolution,  
one could learn behavioral patterns or traits on his own, usually referred to as *innovation*, also called  
88 individual learning, and just like mutation, without it humans might have been remained at the  
stone age, or even go extinct. Rendell et al. (2010) suggest that success biased social transmission  
90 contribute more to the general success of the population than individual learning. They conducted  
a tournament for developing learning strategies of a population, where each participant need to  
92 devise a strategy. Each strategy must define when individuals should observe and copy from others,  
and when to engage in individual learning. The best strategies contained a high percentage of social  
94 learning relative to individual learning, even when the error when copying was as high as almost  
0.5. It is important to add that all of the strategies include some percentage of individual learning,  
96 and without it the results would be a lot worse. In addition to Rendell et al. (2010), Fogarty et al.

(2017) define different types of transmission biases based on success. They define several types of

role-model choosing methods, all assuming that the copier correctly identifies the successful ones.

Both studies assume that individuals can successfully evaluate successful individuals. 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 is 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 between two others can try both of the pad-

dle grips it observed, and decide what grip is better for it. An indirect bias is 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. Continuing with the example, a bystander could copy the paddle

grip of the Ping-Pong player who scored more points in the match. A frequency-dependent bias is

when an individual has a probability to copy a variant of the trait that is nonlinear to the trait's

frequency in the parent's generation. Continuing with the 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.

Frequency bias could be negative too. Aljadeff et al. (2020) show that societies under competitive

conditions are likely to develop diversity in foraging specialization rather than uniformity.

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) claim 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 the indirect bias is strong enough. They claim the bias could lead to

a *runaway process*, caused by a cultural equivalent of *sexual selection* (Andersson, 1994). On the

other hand, Henrich and Broesch (2011b) claim that prestige biases, over generations, can lead

to cultural adaptations. According to them, prestige can make a maladaptive trait spread in the

population, but can also accelerate the spread of adaptive traits as well. *Prestige bias* is often

mentioned in the literature, but seldom modeled. Boyd and Richerson (1988) have modeled the

prestige bias, but didn't include the effects the copiers of a role-model has on the probability of

other individuals to choose the same role model.

This effect is similar to *conformity* (Denton et al., 2020), which is usually modeled as a different bias.

*Conformist learning* (imitating locally common behaviors) is a known bias in cultural transmission

(Molleman et al., 2013), and we suggest that prestige bias is made up by both indirect bias and

a new type of conformity. Our new component, *influence*, is assigned to a role-model, contrary to

conformity, which refers to the frequency of a trait in the population, regardless which individuals

posses it. **The goal of this study is to define a more realistic model for prestige bias and**

**analyze the dynamics of the population it causes.**

Today, due to social media, it is easier than ever to estimate the influence individuals have

over others, therefore it is probably a major part of humans decision-making process. For example,

the number of *followers* a person has in the mobile application *Instagram* may significantly affect

136 how his beliefs are perceived by the population. We want to create a model that better fits reality  
 and simulate scenarios that better mimic cultural transmission dynamics. With a more accurate  
 138 model of prestige bias, we may understand better how cultural traits are transmitted, and why.  
 Moreover, we could better explain the cause for the spread of maladaptive traits, or the acceleration  
 140 of adaptive traits often seen in humans.

## Models and Methods

142 **Reminder:** A *Wright–Fisher model* is a mathematical model meant to describe a genetic drift  
 process. This model assumes that generations do not overlap and that each copy of the gene found  
 144 in the new generation is drawn independently at random from all copies of the gene in the old  
 generation.

146 A *Moran model* assumes overlapping generations. At each time step, one individual is chosen  
 to reproduce and one individual is chosen to die. In our models we harness these two models and  
 148 modify them to describe new mathematical models that we use to expand the basic indirect bias  
 model Boyd and Richerson (1988) suggest.

### 150 Continuous Model

Consider a population of  $N$  individuals, each individual has one trait on a continuous scale. Every  
 152 generation,  $N$  naive individuals (*copiers*) must choose a trait to copy from one of the individuals of  
 the previous generation (*role-models*). Similar to a Wright–Fisher model, we assume the generations  
 154 don’t overlap. We base our model on the model of Boyd and Richerson (1988), by assuming only  
 oblique transmission of the traits (*Indicator trait* -  $A$ ). Unlike their model, we omit a second trait  
 156 called **Indirectly biased trait** to lower complexity. The model’s state at time  $t$  can be described  
 by:

$$158 \quad \vec{A}_t = (A_{t,1}, \dots, A_{t,N}) \quad (1)$$

where  $\vec{A}_t$  is a vector describing the indicator traits at time  $t$ , and  $\vec{A}_0$  is drawn from a standard  
 160 normal distribution. Each individual from generation  $t + 1$ , a *copier*, inherits traits like so:

$$A'_i = F_i(\vec{A}_t) \quad (2)$$

162 where  $A'_i$  is the indicator and indirect trait values correspondingly, that copier  $i$  acquires. We use  
 $A'_i$  as an alias for  $A_{i,(t+1)}$  for simplicity for the transition between generations  $t \rightarrow t + 1$ .  $F$  is a  
 164 function over the  $t$  generation traits vector, and is defined differently for every implementation of  
 the **Generic model**.

166 **Success bias.** Boyd and Richerson (1988, Ch.8, p.247-249) describe a method of inheritance using  
 a *blend*, i.e weighted average of the trait of the entire generation. They define  $F$  as a weighted

168 average of the role-models' traits in a single generation:

$$F_i(\vec{X}) = \sum_{j=1}^N (G_{ij} \cdot X_{ij}) \quad (3)$$

170 where  $G_{i,j}$  is:

$$G_{ij} = \frac{\beta(A_{ij})}{\sum_{l=1}^N \beta(A_{il})} \quad (4)$$

172 We define  $G_{ij}$  to be the *Success bias* of role-model  $j$  in the eyes of copier  $i$ .  $A_{i,j}$  is the absolute indicator trait value copier  $i$  estimates role-model  $j$  has:

$$174 \quad A_{i,j} = A_j + e_i, \quad (5)$$

where  $e_i$  is the copier's error of estimation,  $\vec{e} \sim N(0, \frac{1}{\eta^2})$ .  $\beta(X)$  is the bias function, meant to  
176 quantify the success bias of a role-model:

$$\beta(A_{i,j}) = b \cdot \exp\left(-\frac{(A_{i,j} - \hat{A})^2}{2J}\right), \quad (6)$$

178 where  $\hat{A}$  is the optimal indicator value and  $J, b$  are model parameters to control the "strength"  
of the bias.  $G_{i,j}$  is therefore the relative success score copier  $i$  assigns to role-model  $j$ , resembling  
180 *relative fitness* in genetic transmission models.

**Random choice transmission.** Boyd and Richerson (1988) note that the method of transmis-  
182 sion they use in their model has alternatives. We follow their suggestion and create a model similar  
to theirs, with random choice as a transmission method: The probability of copier  $i$  to choose  
184 role-model  $j$  as his role-model to copy its traits from is  $G_{i,j}$ . Once a copier chose its role-model,  
it will copy both its traits only from his role-model, instead of a weighted average of the entire  
186 role-model generation:

$$A'_i = A_{i,j} \quad (7)$$

188 **Influence bias.** Copiers choose their role-models one by one. After copier  $i$  chose a role-model, we  
denote  $K_{ij}$  as the number of copiers that chose role-model  $j$  until that point, such that  $\sum_{j=1}^N K_{i,j} =$   
190  $i$ . The stochastic process of role-model choice,

$$\{\vec{K}_i\}_{i=1}^N, \quad \vec{K}_i = (K_{i1}, \dots, K_{iN}), \quad (8)$$

192 is described by the recurrence equation

$$K_{i,j} = K_{i-1,j} + S_{i,j}, \quad i, j = 1, 2, \dots, N \quad (9)$$

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

The probability that the  $i$ -th copier chose role-model  $j$

$$G_{i,j} = P(S_{i,j} = 1 | S_{1,j}, S_{2,j}, \dots, S_{i-1,j}) \quad (10)$$

is the prestige of role-model  $j$  in the eyes of copier  $i$ . This prestige  $G_{i,j}$  is determined as follows.  
 198 First, role-model  $j$  is characterized by its indicator value  $A_j$  as before, and the estimated indicator  
 value by copier  $i$ ,  $A_{i,j}$  remains as eq. (5). Finally, the prestige  $G_{i,j}$  of role-model  $j$  in the eyes of  
 200 copier  $i$  is determined by the estimated biased indicator value  $\beta(A_{i,j})$  and the number of copiers  
 that chose role-model  $j$  before copier  $i$ ,  $K_{i-1,j}$ ,

$$G_{i,j} = \frac{\alpha_j \cdot \beta(A_{i,j}) + (1 - \alpha_j) \cdot K_{i-1,j}}{W_i}, \quad (11)$$

where the weight  $\alpha_j$  is a characteristic of role-model  $j$  that determines the relative significance of the  
 204 indicator and the influence in the prestige, and  $W_i$  is a normalizing factor to ensure  $\sum_{j=1}^N G_{i,j} = 1$ ,

$$W_i = \sum_{j=1}^N \left( \alpha_j \cdot \beta(A_{i,j}) + (1 - \alpha_j) \cdot K_{i-1,j} \right). \quad (12)$$

## 206 Binary model

The indicator trait can now manifest in only two phenotypes, and for simplicity we define they  
 208 can be either  $\hat{A}$  or  $A$ . In the binary model, the influence is determined by the number of copiers  
 already chosen **any** role-model with either  $A$  or  $\hat{A}$ , as all role-models with  $A$  will contribute to the  
 210 probability of the trait to be inherited just the same (can be proved with simple induction). Simply  
 put, assuming there are two role-models with the  $A$  trait, the probability a copier will copy from  
 212 either role-model will be the same, and the probability the  $A$  trait will be inherited is the sum of  
 both role-models. In the general case, the probability of the  $i$ -th individual to inherit trait  $A$ , based  
 214 on eq. (22) is:

$$P_{i,A} = \frac{(N - X)\alpha'\beta(A) + K_A}{i - 1 + (N - X)\alpha'\beta(A) + X\alpha'\beta(\hat{A})} = \frac{(N - X)\alpha'\beta(A) + K_A}{i - 1 + (N - X)\alpha'\beta(A) + \alpha'X} \quad (13)$$

216 where  $X$  is the number of role-models with trait  $\hat{A}$  and  $K_A$  is the number of copiers that already  
 chose  $A$ .

218 The model begins with the first generation having a single individual with  $\hat{A}$ , and the rest have  $A$ .  
 The process itself is the same stochastic process as the continuous model.

## 220 Methods

The main methods we used to experiment and compare our models is using computer generated  
 222 simulations. In order to establish our claims and base our mathematical approximations of our



models, we used the  $\chi^2$  test for the full continuous model, and the Kimura's equations of fixation  
 224 probability and time to fixation for the binary model.

## Results

### 226 Approximations

Currently  $\{\vec{K}_i\}_{i=1}^N$  is a stochastic process where each state depends on the previous state, i.e a  
 228 Markov chain. We wanted to find an equivalent stochastic process that has the same joint distribution on  $\{\vec{K}_i\}_{i=1}^N$ , but it is possible to evaluate the joint distribution directly without evaluating  
 230 all the marginal conditional distributions: eq. (9), eq. (10).

We found two approximations to our process, which are summarized here and explained in  
 232 detail later on:

1.  $K_{i,j}$  follows the general binomial distribution defined by Drezner and Farnum (1993). More-  
 234 over,  $\mathbb{E}[K_{N,j}] = N \cdot G_{1,j}$  if  $e = e_l = e_m$  for all  $l, m$ . That is, the expected number of copiers of  
 role-model  $j$  equals its prestige in the eyes of the first copier, multiplied by the total number  
 236 of copiers. In addition, we find that when  $\alpha$  is homogeneous,  $\alpha_l = \alpha_m$  for all  $l, m$ , then  
 $\mathbb{E}[K_{N,j}] = \beta(A'_j) / \overline{\beta(A')}$ , where  $A'_j$  is the estimated indicator value  $A'_j = A_j + e$ , and  $\overline{\beta(A')}$   
 238 is the population mean estimated indicator value. That is, the expected number of copiers of  
 a role-model equals its relative biased indicator value, similar to the role of relative fitness in  
 240 population-genetic models.
2. The role-model choice process eq. (8) is equivalent to a Pólya urn model if  $e_l = e_m$  for all  
 242  $l, m$ . Therefore,  $\vec{K}_i = (K_{i,1}, \dots, K_{i,N})$  follows a Dirichlet-Multinomial distribution,

$$\vec{K}_i \sim DM(N, \vec{G}_1), \quad (14)$$

244 where  $\vec{G}_1 = (G_{1,1}, \dots, G_{1,N})$ . Note that here  $G_{i,j}$  is only a function of the indicator values  
 $A_j$  and the weights  $\alpha_j$ .

### 246 General Binomial Distribution Approximation

The general binomial distribution (GBD) is achieved by a series of Bernoulli experiments, with  
 248 possible dependency between experiments.

**Proposition:** The number of copiers  $K_{i,j}$  follows the GBD,  $K_{i,j} \sim GBD(i, \alpha_i \cdot \beta(A'_j))$ , when  
 250  $e_l = e_m$  for all  $l, m \in N$  and  $A'_j = A_j + e$

**Proof:** We'll denote  $Q_j(k, i) = P(K_{i,j} = k | K_{i-1,j})$  as the probability that exactly  $k$  out of  $i$   
 252 copiers choose role-model  $j$ , using conditional probability and eq. (9):

$$Q_j(k, i) = P_j(S_{i,j} = 1 | k-1, i-1) \cdot Q_j(k-1, i-1) + P_j(S_{i,j} = 0 | k, i-1) \cdot Q_j(k, i-1) \quad (15)$$

254 where  $S_{i,j} = 1$  when the  $i$ -th copier chooses role-model  $j$ .

We see that eq. (15) is equivalent to eq. (2.1) that Drezner and Farnum (1993) define.  $Q_j(k, N)$  is the probability that  $k$  out of  $N$  copiers choose role-model  $j$  at the end of the process, which by our previous notation is  $k = K_{N,j}$ . By describing the process of eq. (8) as (Drezner and Farnum, 1993) did, we've completed the proof.

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

260 In (Drezner and Farnum, 1993, equation 2.3), they show that the expected value of  $k$  is:

$\mathbb{E}[k] = N \cdot Q_j(1, 1)$  (using different notations).  $Q_j(1, 1)$  is the initial probability to choose role-model  $j$ , before any choices are made.  $Q_j(1, 1) = G_{1,j}$  by definition, therefore we can say that  $\mathbb{E}[K_{N,j}] = N \cdot G_{1,j}$ .

264

**Corollary 2:**  $\mathbb{E}[K_{Nj}] = \alpha_j \cdot \beta(A'_j) / \overline{\alpha \cdot \beta(A')}$ .

266 **Proof:** The initial prestige of role-model  $j$  based on eq. (11) is:

$$G_{1,j} = \frac{\alpha_j \cdot \beta(A'_j)}{\sum_{m=1}^N \alpha_m \cdot \beta(A'_m)} \quad (16)$$

268 The denominator of eq. (16) can also be formulated as:

$$\sum_{m=1}^N \alpha_m \beta(A'_m) = N \cdot \overline{\alpha \cdot \beta(A')} \quad (17)$$

270 where  $\overline{\alpha \beta(A')}$  is the mean value of  $\alpha_m \cdot \beta(A'_m)$  for all  $m$ . Using eq. (17) we get:

$$\mathbb{E}[K_{Nj}] = \alpha_j \cdot \beta(A'_j) / \overline{\alpha \cdot \beta(A')} \quad (18)$$

272 , completing our proof.

The special case where  $\alpha = \alpha_l = \alpha_m$  for all  $l, m \in N$  is interesting, because we can evaluate the expected number of copiers using a linear equation:

$$\mathbb{E}[K_{Nj}] = N \cdot \frac{\alpha \cdot \beta(A'_j)}{\sum_{m=1}^N \alpha \cdot \beta(A'_m)} = \beta(A'_j) / \overline{\beta(A')} \quad (19)$$

276 where the only variable is  $A'_j$ , because  $\overline{\beta(A')}$  is the mean of the distribution we draw the indicator values from, modified by some constant parameters of  $\beta$ . We can then denote  $L = 1/\overline{\beta(A')}$  and write:

$$\mathbb{E}[K_{Nj}] = L \cdot \beta(A'_j) \quad (20)$$

## 280 Dirichlet-Multinomial Distribution Approximation

**Reminder:** *Pólya urn model* is a stochastic process that is defined as such: The 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 colour. Let  $\vec{U}_i = \{u_{i,1}, u_{i,2}, \dots, u_{i,M}\}$  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 0 otherwise. The probability that  $S_{i,j} = 1$  given  $\vec{U}_{i-1}$  is:

$$\begin{aligned} P(S_{i,j} = 1 | \vec{U}_{i-1}) &= \frac{u_{i-1,j}}{\sum_{m=1}^M u_{i-1,m}} = \frac{o_j + w_{i-1,j}}{\sum_{m=1}^M o_m + w_{i-1,m}} \\ &= \frac{o_j + w_{i-1,j}}{i-1 + \sum_{m=1}^M o_m} \end{aligned} \quad (21)$$

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

**Proposition:** process  $\{\vec{K}_i\}_{i=1}^N$  is equivalent to a *Pólya urn model* when  $e = e_i = e_j$  and  $\alpha = \alpha_j = \alpha_i$  for all  $i, j \in N$ .

**Proof:** We denote  $\alpha'$  as the odds ratio between the weights of the indicator and the influence ( $\alpha' = \frac{\alpha}{1-\alpha}$ ). Using eq. (11) we get:

$$\begin{aligned} G_{i,j} &= \frac{\alpha \cdot \beta(A'_j) + (1-\alpha) \cdot K_{i-1,j}}{W_i} \cdot \frac{1-\alpha}{1-\alpha} \\ &= \frac{\alpha' \beta(A'_j) + K_{i-1,j}}{\sum_{m=1}^N \alpha' \beta(A'_m) + K_{i-1,m}} \\ &= \frac{\alpha' \beta(A'_j) + K_{i-1,j}}{i-1 + \sum_{m=1}^N \alpha' \beta(A'_m)} \end{aligned} \quad (22)$$

We see that eq. (21) and eq. (22) are equivalent when setting  $M = N$ ,  $o_j = \alpha' \beta(A'_j)$ ,  $w_{i,j} = K_{i,j}$ , completing the proof.

**Corollary 1:** In their paper, Frigiyik et al. (2010, section 2) prove that the proportion of different colored balls in a *Pólya urn model* will converge to the Dirichlet distribution as the number of draws approaches infinity, based on *Martingale Convergence Theorem* (Durrett, 1999). We therefore have

300 an approximation for the relative "weight" or the proportion each role-model has when evaluated  
 as a role-model. Drawing from a Multinomial distribution where the parameters are the modified  
 302 weights gained from the Dirichlet distribution is viable for selecting the role-model for the next  
 generation. We can therefore sample from a Dirichlet-Multinomial distribution to approximate  
 304 how many copiers each of the role-models will have:  $\vec{K}_i \sim DM(N, \vec{G}_1)$ .

**Numeric validation:** We showed our process is DM (Dirichlet-Multinomial) distributed when  
 306 there are no errors when copying or evaluating the traits, and when  $\alpha$  is homogeneous in the popu-  
 lation. To support our proof, we tested our approximation empirically using computer simulations.  
 308 To test our hypothesis, we used a *goodness of fit* method known as *Pearson's chi-squared test*. In  
 this test, one can reject or accept the null hypothesis, which in our case is the hypothesis that the  
 310 simulations results were drawn from a DM distribution.

To use this test, we ran many simulations of our original model, and used the mean distribution  
 312 of copiers. This mean distribution is our observed distribution, and we tested it with the DM  
 expected distribution, using said chi-squared test. We tested multiple variations of the trait weight  
 314 parameter ( $\alpha$ ).

In all our tests, the p-value was 1. This means that the probability to reject the null hypothesis  
 316 is essentially nonexistent (the usual threshold for a p-value needed to reject  $H_0$  is 0.05 or lower).  
 In addition, we found out that for high  $\alpha$  values (above 0.5), very few simulations are needed to  
 318 reach p-value 1. (less than five simulations for  $\alpha = 0.9$ , and less than 20 for  $\alpha = 0.7$ ) For very  
 low  $\alpha$  values, which means very high influence weight, the number of simulations needed was 100,  
 320 which is still a relatively small amount. To verify our codes results, we also ran the test for different  
 distributions, for example the uniform distribution (all role-models have exactly one copier).

322 All these tests resulted in a p-value 0, which means we can likely reject  $H_0$  for these distributions,  
 as expected.

324 Once we validated our proof for a single iteration of the model, we went on to more complex  
 validations for the entire model.

## 326 Numeric comparisons

We're interesting in studying the difference between the real binary model as we defined in eq. (13),  
 328 and the Dirichlet-Multinomial approximation. Specifically, we're interesting in the fixation proba-  
 bility of the favored trait ( $\hat{A}$ ) and its time to fixation.

330 The first step was to find the number of simulations needed to sufficiently approximate the real  
 model with the DM approximation. From fig. 1 we see that 1000 simulations or higher is enough.

332 The next step was to see how the observed metrics (fixation probability and time) varies when  
 relaxing our assumptions we used to prove the DM approximation. First we relaxed our assumption  
 334 of no mutation. To include mutation in the binary model, it needs to be redefined, since in the  
 original model it was based on the fact the traits are drawn from a continuous scale. In the binary

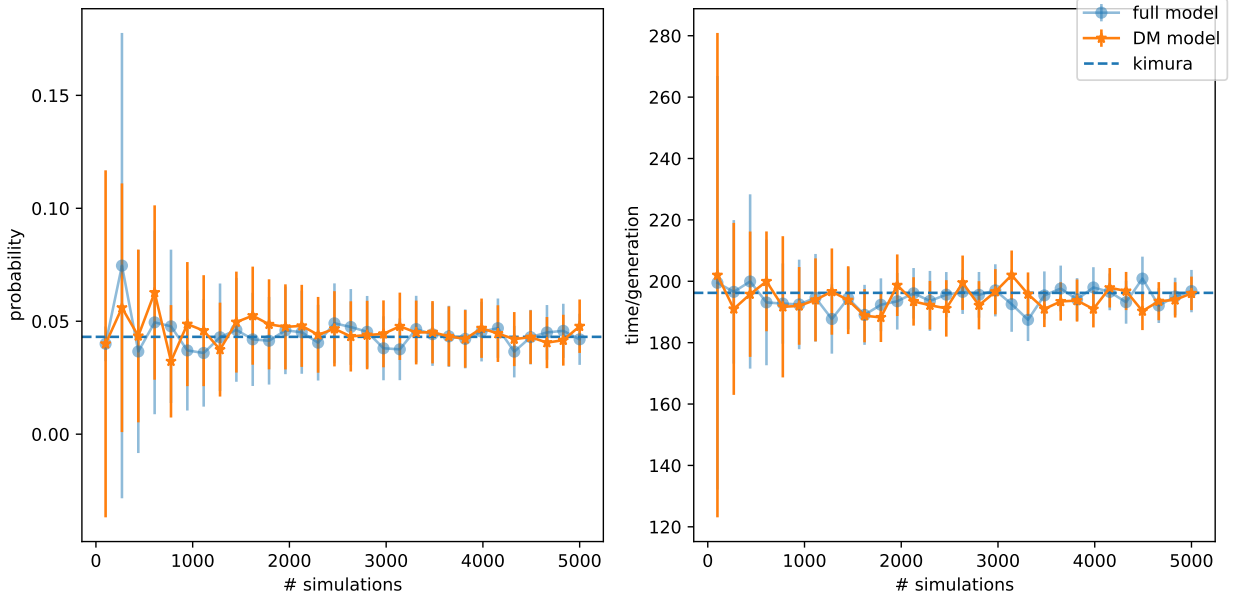


Figure 1: The number of simulations needed to get a good approximation. At 1,000 the approximation is good enough. Error bars represent 95% confidence interval. Population size  $N = 1000$ ,  $\alpha = 0.5$ ,  $J = 1$ ,  $\hat{A} = 1$ ,  $A = 0.7$ ,  $\beta(A) = 0.956$ .

model mutation will be manifested as an error when evaluating the bias itself. This is easily done by using a heterogeneous  $J$  parameter, which controls the strength of the success bias in eq. (6).

In fig. 2 we see the comparison when heterogeneous mutation is applied to both models. When mutation is applied, we sample  $J_i$  for each copier  $i$  from a normal distribution with varying scale (variance). We can see that even when the standard deviation is 0.1, the metrics of both models are both similar, and close to the Kimura approximation (more details in the next section).

In fig. 3 we relaxed our assumption of a homogeneous  $\alpha$ , and used a heterogeneous  $\alpha$  instead. Similar to the mutation comparison, we drew  $\alpha_j$  for each role-model  $j$  from a normal distribution with varying scale. We again see that the metrics of both models are similar in the entire spectrum of our x-axis, and the Kimura approximation is within both confidence intervals.

## Fixation probability and time - binary model

**Kimura's approximation:** After establishing a case in the favor of our DM approximation, we wanted to use it to examine the behavior of the population. Specifically, we wanted to analyze the influence of the indicator weight ( $\alpha$ ) on the fixation probability and time to fixation of the favored phenotype in a binary model. For simplicity, we don't include mutation rate in the binary model approximations. Following Durrett (2008), we used our DM approximation of the model to find the effective population size. From eq. (13) we can derive that the process of inheritance in our binary model is DM distributed with a parameters vector of size two:  $\vec{V} = (\alpha'X, (N - X)\alpha'\beta(A))$ .

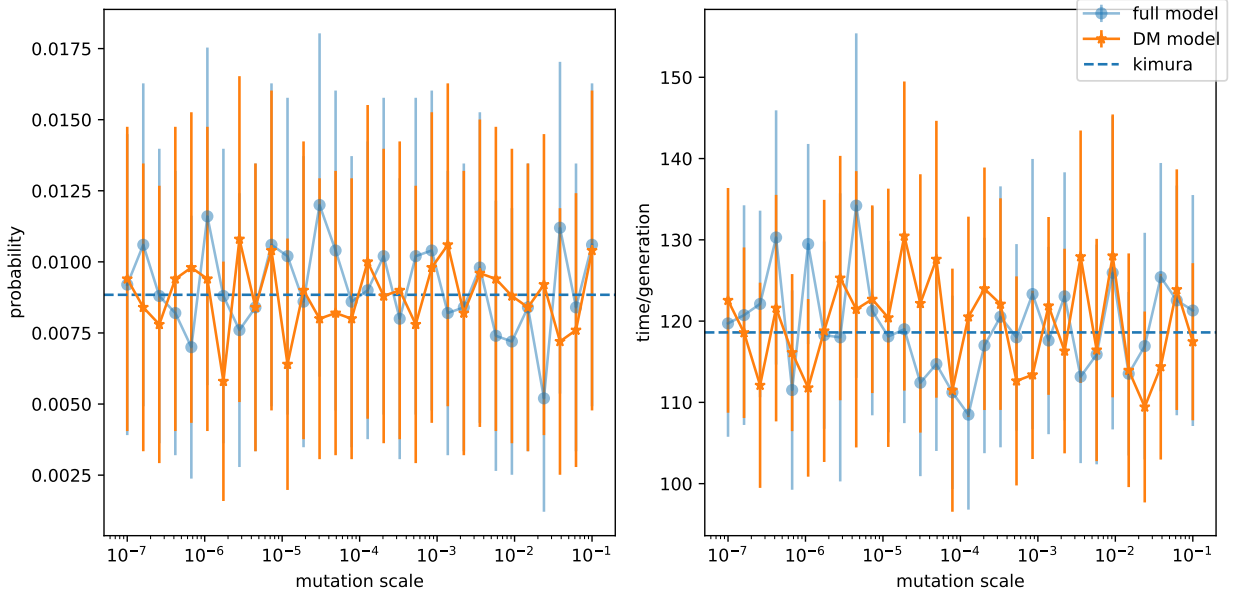


Figure 2: Comparison of the DM approximation and the full model when mutation is included. Even high mutation rate doesn't worsen the approximation, and the data points are close to the mathematical estimation (Kimura's). Error bars are 95% confidence intervals, and are condensed ( $\pm 0.01$  probability and  $\pm 40$  generations) 5000 simulations per data point,  $N = 1000$ ,  $\alpha = 0.1$ ,  $\hat{A} = 1$ ,  $A = 0.7$ ,  $J \sim N(1, x^2)$  where  $x$  is the mutation scale in the x-axis.

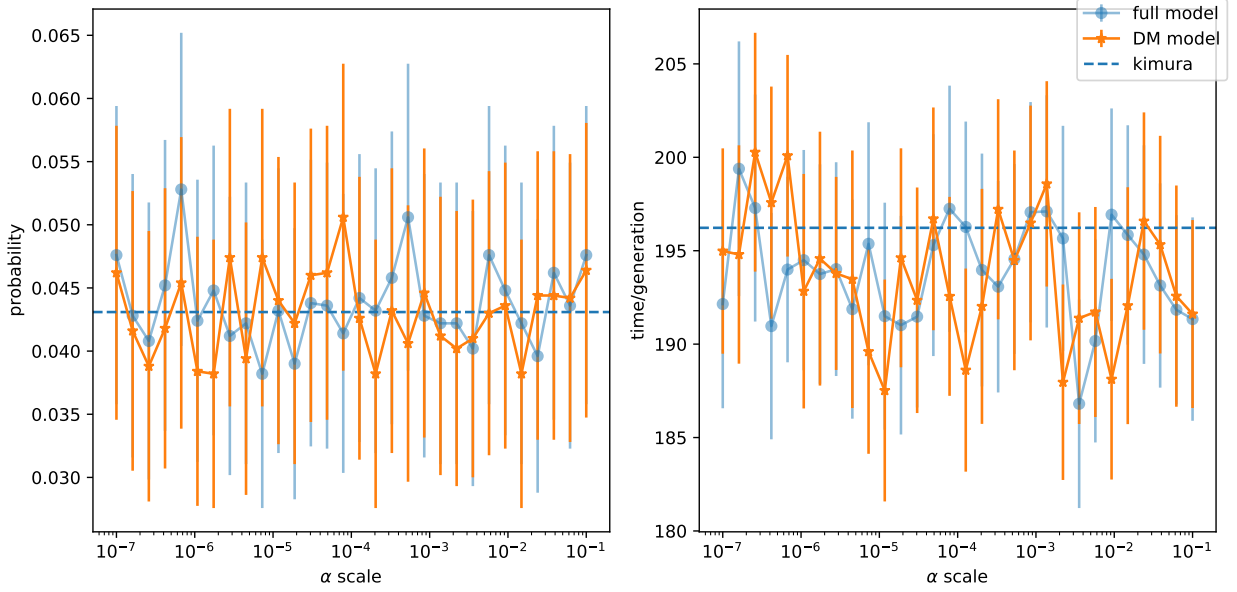


Figure 3: Comparison of the DM approximation and the full model when success weight is heterogeneous. High success weight variance distances the approximation and the full model of generations to fixation from the Kimura's approximation, but not by much (confidence intervals still cover it). Error bars are 95% confidence intervals, and are less condensed ( $\pm 0.03$  probability and  $\pm 40$  generations) 5000 simulations per data point,  $N = 1000$ ,  $\alpha \sim N(0.5, x^2)$ ,  $\hat{A} = 1$ ,  $A = 0.7$ ,  $J = 1$ ,  $\beta(A) = 0.956$ .

354 **Proposition:**  $1 - \beta(A)$  is equivalent to the selection coefficient  $s$  in a classic Wright-Fisher model  
 356 in the diffusion equations meant to approximate the fixation probability and time of the advantageous trait.

**Proof:** Let  $x$  be the frequency of type  $\hat{A}$  in the population with  $N$  individuals. Let  $X$  be the  
 358 number of individuals of type  $\hat{A}$  so  $x = X/N$ .  $X'$  is the number of individuals with  $\hat{A}$  in the next  
 generation and  $x'$  their frequency. By definition  $\beta(\hat{A}) = 1$ , and for simplicity we'll denote  $\beta(A) = \beta$   
 360 ( $\beta < 1$ ).

The expected number of individuals of a DM distribution is:

$$362 \quad E[X'] = N \frac{\alpha_1}{\alpha_1 + \alpha_2}, \quad (23)$$

where  $\alpha_1 = \alpha'X$  and  $\alpha_2 = \alpha'(N - X)\beta$ , from eq. (13). We want to use frequencies instead of  
 364 quantities to follow Durrett's process so:

$$E[x'] = E\left[\frac{X'}{N}\right] = \frac{1}{N}E[X'] \quad (24)$$

366 Putting it together we get:

$$\begin{aligned} E[x'] &= \frac{1}{N}N \frac{\alpha'xN}{\alpha'xN + \alpha'N(1-x)\beta} \\ &= \frac{x}{x + (1-x)\beta} \end{aligned} \quad (25)$$

368 which is identical to the equation in the top of page 253, chap 7.2 in Durrett (2008). We  
 therefore use the same approximation and say that:

$$\begin{aligned} E[x'] &= \frac{x}{x + (1-x)\beta} = \frac{x}{x + (1-x)(1-s)} = \\ 370 \quad &= x + x(1-x)s + o(s) \\ &= x + x(1-x)(1-\beta) + o(1-\beta) \end{aligned} \quad (26)$$

By definition,  $x$  is constant, so  $E[x] = x$ . We continue to calculate  $E[x' - x]$ :

$$372 \quad E[x' - x] = E[x'] - E[x] = x(1-x)(1-\beta) + o(1-\beta) \quad (27)$$

where when substituting  $1 - \beta$  with  $s$ , we get the same result as Durrett (2008) which is the desired  
 374 result.

**Proposition:**  $Ne = \alpha N + (1 - \alpha)$ , where  $Ne$  is the effective population size of our binary model.

376 **Proof:** The variance of a DM distribution is:

$$V(X') = N \frac{\alpha_1}{\alpha_1 + \alpha_2} (1 - \frac{\alpha_1}{\alpha_1 + \alpha_2}) (\frac{N + \alpha_1 + \alpha_2}{1 + \alpha_1 + \alpha_2}) \quad (28)$$

378 And again, we want to use frequencies so:

$$V(\frac{X'}{N}) = \frac{1}{N^2} V(x') \quad (29)$$

380 Putting it together with our model's notations:

$$V(x') = \frac{1}{N^2} N \frac{x}{x + (1-x)\beta} (1 - \frac{x}{x + (1-x)\beta}) (\frac{N + \alpha' x N + \alpha' N(1-x)\beta}{1 + \alpha' x N + \alpha' N(1-x)\beta}) \quad (30)$$

382 Like Durret, we'll use the zero order of the approximation when  $\beta \approx 1$ , so:

$$\frac{x}{x + (1-x)\beta} \approx x \quad (31)$$

384 and we also use  $\beta \approx 1$  for the entire variance expression and get:

$$\begin{aligned} V(x') &\approx \frac{1}{N} x(1-x) (\frac{N + \alpha' x N + \alpha' N - \alpha' x N}{1 + \alpha' x N + \alpha' N - \alpha' x N}) \\ &= x(1-x) (\frac{1 + \alpha'}{1 + \alpha' N}) \end{aligned} \quad (32)$$

386 Again following Durret we want to calculate:

$$V(x' - x) = V(x') - V(x) \approx x(1-x) (\frac{1 + \alpha'}{1 + \alpha' N}) \quad (33)$$

388 because  $x$  is a constant so  $V(x) = 0$

In our model,  $\alpha'$  is the odds ratio of a parameter we called "indicator weight", denoted in our  
390 model as  $\alpha$ , so:

$$\alpha' = \frac{\alpha}{1 - \alpha} \quad (34)$$



392 Combining eq. (33) and eq. (34) we get:

$$\begin{aligned}
V(x' - x) &\approx x(1 - x) \left( \frac{1 + \frac{\alpha}{1-\alpha}}{1 + \frac{\alpha}{1-\alpha}N} \right) \\
&= x(1 - x) \left( \frac{\frac{1-\alpha+\alpha}{1-\alpha}}{\frac{1-\alpha+\alpha N}{1-\alpha}} \right) \\
&= x(1 - x) \left( \frac{1}{1 - \alpha(1 - N)} \right) \\
&= x(1 - x) \left( \frac{1}{\alpha N + (1 - \alpha)} \right) \\
&= x(1 - x) \frac{1}{N_e}
\end{aligned} \tag{35}$$

394 Using our substitute for a selection coefficient,  $1 - \beta$ , and the effective population size  $N_e$ , we  
can estimate the fixation probability and time of our binary model.

396 The fixation probability derived from Kimura is therefore:

$$P_{fix} = \frac{1 - e^{-2(1-\beta)N_e x}}{1 - e^{-2(1-\beta)N_e}} \tag{36}$$

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

The time to fixation can be estimated by:

$$T_{fix} = \frac{1 - P_{fix}}{1 - \beta} \int_0^x \frac{e^{2(1-\beta)\xi} - 1}{\xi(1 - \xi)} d\xi + \frac{P_{fix}}{1 - \beta} \int_x^1 \frac{1 - e^{-2(1-\beta)(1-\xi)}}{\xi(1 - \xi)} d\xi \tag{37}$$

where the integrals cannot be solved in closed form, so we can only estimate them numerically.

402 To validate our math we ran multiple simulations comparing our binary model with the classic  
Wright-Fisher model, using different  $\alpha$  and  $\beta$  each time, and using the corresponding values of  $s$   
404 and  $N_e$  for the WF simulations. In fig. 4 we changed  $\alpha$  (and  $N_e$  accordingly) and used a constant  
 $\beta$ . In fig. 5 we changed  $\beta$  and used a constant  $\alpha$ . In both cases we can see that the two models  
406 behave similarly, and both are approximated well by the Kimura's equations: eq. (36) and eq. (37).

## Changing environment

408 After finding good estimations for our model in a constant environment, when the favorable trait  
is always  $\hat{A}$ , we want to find an estimation for our model in a changing environment.

410 For that we will find an expression for the expected and variance of the change in frequency  
between  $t$  generations. Let  $s_t = N(1 - \beta_t)$ , and  $S_n = \sum_{i=1}^n s_i$ , where  $\beta_t$  is  $\beta(A)$  at time/generation  $t$ .

412 **Proposition:**  $E[\frac{X_t}{N} - x] \simeq \frac{1}{N} S_t x(1 - x)$ ,  $V(\frac{X_t}{N}) \simeq \frac{1}{N_e} t x(1 - x)$ , where  $x$  is the initial frequency  
of the favorable/invading trait and  $X_t$  is the number of individuals with the trait at time  $t$ .

414 The proof is based on the proof of Ram et al. (2018), proving a similar scenario.

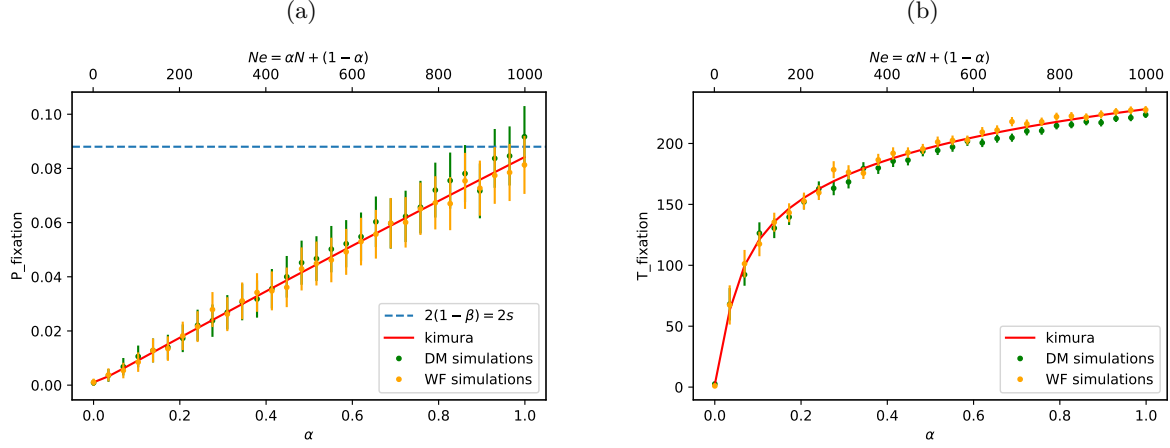


Figure 4: Comparison of the DM approximation and the WF model for different values of the effective population size. The approximation seems very good, and is also condensed around the mathematical equation expectancy. Error bars are 95% confidence intervals. Effective population calculated by  $N_e = \alpha N + (1 - \alpha)$ . 5,000 simulations per data point,  $N = 1,000$ ,  $\hat{A} = 1$ ,  $A = 0.7$ ,  $J = 1$ ,  $1 - \beta = s = 0.044$ .

**Proof by induction:** From eq. (27) we know that

$$\begin{aligned}
 E \left[ \frac{X_{t+1}}{N} - \frac{X_t}{N} \middle| X_t \right] &= \frac{X_t}{N} \left( 1 - \frac{X_t}{N} \right) (1 - \beta_{t+1}) \\
 &= \frac{1}{N} \frac{X_t}{N} \left( 1 - \frac{X_t}{N} \right) s_{t+1}
 \end{aligned} \tag{38}$$

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

$$\begin{aligned}
 E \left[ \frac{X_t}{N} \left( 1 - \frac{X_t}{N} \right) \right] &= E \left[ \frac{X_t}{N} - \left( \frac{X_t}{N} \right)^2 \right] \\
 &= E \left[ \frac{X_t}{N} \right] - E \left[ \left( \frac{X_t}{N} \right)^2 \right] \\
 &= E \left[ \frac{X_t}{N} \right] - V \left( \frac{X_t}{N} \right) - \left( E \left[ \frac{X_t}{N} \right] \right)^2
 \end{aligned} \tag{39}$$

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

$$E \left[ \frac{X_t}{N} \left( 1 - \frac{X_t}{N} \right) \right] \simeq E \left[ \frac{X_t}{N} \right] \left( 1 - E \left[ \frac{X_t}{N} \right] \right) - \frac{1}{N_e} t x (1 - x) \tag{40}$$

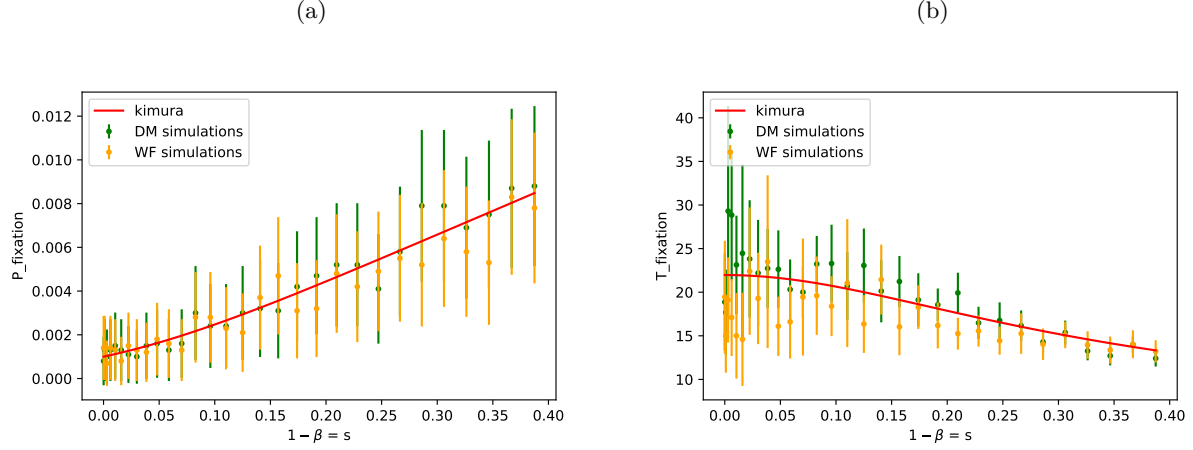


Figure 5: Comparison of the DM approximation and the WF model for different values of the selection coefficient, manifested as success bias in our model. The approximation seems good, and is also condensed around the mathematical equation expectancy. Error bars are 95% confidence intervals. Effective population calculated by  $N_e = \alpha N + (1 - \alpha)$ . 5,000 simulations per data point,  $N = 1,000$ ,  $\hat{A} = 1$ ,  $A = 0.7, J = 1, \alpha = 0.01$ .

From eq. (38) we know that

$$\begin{aligned}
 E \left[ \frac{X_{t+1}}{N} - \frac{X_t}{N} \right] &= \frac{1}{N} s_{t+1} E \left[ \frac{X_t}{N} \left( 1 - \frac{X_t}{N} \right) \right] \\
 &\simeq \frac{1}{N} s_{t+1} \left( E \left[ \frac{X_t}{N} \right] \left( 1 - E \left[ \frac{X_t}{N} \right] \right) - \frac{1}{N_e} t x (1 - x) \right) \\
 &\simeq \frac{1}{N} s_{t+1} \cdot E \left[ \frac{X_t}{N} \right] \left( 1 - E \left[ \frac{X_t}{N} \right] \right) - \frac{1}{N_e N} s_{t+1} t x (1 - x)
 \end{aligned} \tag{41}$$

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

$$E \left[ \frac{X_{t+1}}{N} - \frac{X_t}{N} \right] \simeq \frac{1}{N} s_{t+1} \cdot E \left[ \frac{X_t}{N} \right] \left( 1 - E \left[ \frac{X_t}{N} \right] \right) \tag{42}$$

We'll now look at the induction assumption to see that

$$E \left[ \frac{X_t}{N} - x \right] = E \left[ \frac{X_t}{N} \right] - E[x] = E \left[ \frac{X_t}{N} \right] - x, \tag{43}$$

so using the assumption we get

$$\begin{aligned}
 E \left[ \frac{X_t}{N} \right] &\simeq \frac{1}{N} S_t x (1 - x) + x \\
 1 - E \left[ \frac{X_t}{N} \right] &\simeq 1 - \frac{1}{N} S_t x (1 - x) + x
 \end{aligned} \tag{44}$$

we'll use both expressions in eq. (42) and get

$$\begin{aligned}
E \left[ \frac{X_{t+1}}{N} - \frac{X_t}{N} \right] &\simeq \frac{1}{N} s_{t+1} \left( \frac{1}{N} S_t x (1-x) + x \right) \left( 1 - \frac{1}{N} S_t x (1-x) + x \right) \\
&\simeq \frac{1}{N} s_{t+1} \cdot x (1-x)
\end{aligned} \tag{45}$$

after again omitting  $O(\frac{1}{N^2})$  parts of the equation. To conclude our proof, we see that

$$E \left[ \frac{X_{t+1}}{N} - x \right] = E \left[ \frac{X_{t+1}}{N} - \frac{X_t}{N} \right] + E \left[ \frac{X_t}{N} - x \right] \tag{46}$$

so again using the induction assumption, together with eq. (45) we get

$$\begin{aligned}
E \left[ \frac{X_{t+1}}{N} - x \right] &\simeq \frac{1}{N} s_{t+1} \cdot x (1-x) + \frac{1}{N} S_t \cdot x (1-x) \\
&\simeq \frac{1}{N} x (1-x) (S_t + s_{t+1}) \\
&\simeq \frac{1}{N} S_{t+1} x (1-x)
\end{aligned} \tag{47}$$

which proves the first part of our preposition.

For the second part, we'll use a property of variance:

$$V \left( \frac{X_{t+1}}{N} \right) = E \left[ V \left( \frac{X_{t+1}}{N} \middle| X_t \right) \right] + V \left( E \left[ \frac{X_{t+1}}{N} \middle| X_t \right] \right) \tag{48}$$

using eq. (38) we see that:

$$\begin{aligned}
E \left[ \frac{X_{t+1}}{N} \middle| X_t \right] - E \left[ \frac{X_t}{N} \middle| X_t \right] &= \frac{1}{N} s_{t+1} \frac{X_t}{N} \left( 1 - \frac{X_t}{N} \right) \\
E \left[ \frac{X_{t+1}}{N} \middle| X_t \right] &= \frac{X_t}{N} + \frac{1}{N} s_{t+1} \frac{X_t}{N} \left( 1 - \frac{X_t}{N} \right)
\end{aligned} \tag{49}$$

Using eq. (35) we get:

$$V \left( \frac{X_{t+1}}{N} \middle| X_t \right) = \frac{1}{N_e} \frac{X_t}{N} \left( 1 - \frac{X_t}{N} \right) \tag{50}$$

and using the equation  $y'(1-y') \simeq y(1-y)$  on the first part of eq. (48) we get:

$$E \left[ V \left( \frac{X_{t+1}}{N} \middle| X_t \right) \right] = \frac{1}{N_e} E \left[ \frac{X_t}{N} \left( 1 - \frac{X_t}{N} \right) \right] \simeq \frac{1}{N_e} x (1-x) \tag{51}$$

and moving on to simplify the second part of eq. (48) using eq. (49):

$$V \left( E \left[ \frac{X_{t+1}}{N} \middle| X_t \right] \right) = V \left( \frac{X_t}{N} + \frac{1}{N} s_{t+1} \frac{X_t}{N} \left( 1 - \frac{X_t}{N} \right) \right) \tag{52}$$

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

$$V\left(\frac{1}{N}s_{t+1}\frac{X_t}{N}\left(1 - \frac{X_t}{N}\right)\right) \leq \frac{1}{4N^2}s_{t+1}^2 \quad (53)$$

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

$$V\left(E\left[\frac{X_{t+1}}{N}\middle|X_t\right]\right) = V\left(\frac{X_t}{N}\right) + O\left(\frac{1}{N^2}\right) \simeq V\left(\frac{X_t}{N}\right) \quad (54)$$

Using the induction assumption and eq. (51):

$$V\left(\frac{X_{t+1}}{N}\right) \simeq \frac{1}{N_e}x(1-x) + \frac{1}{N_e}tx(1-x) \simeq \frac{1}{N_e}x(1-x)(t+1) \quad (55)$$

proving the second part of our preposition.

Following our proof, we can say that after many cycles, we can use a modified version of our fixation probability:

$$P_{fix} = \frac{1 - e^{-2\frac{S_n}{n}N_ex}}{1 - e^{-2\frac{S_n}{n}N_e}} \quad (56)$$

where  $\frac{S_n}{n} = \frac{k-l}{k+l}(1 - \text{beta})$ ,  $n = k + l$ . Put into words, we use the average selection coefficient of a cycle  $(k + l)$  as the selection coefficient in our original equation. In our proof we showed that the expected change in frequency and variance is only manifested in the selection coefficient  $S_n$ , and that we can use those modified equation as a base for Kimura's equation.

We wanted again to validate our results, using simulations. This time, the number of parameters increased: in addition to  $\alpha, \beta$ , there are also  $k, l$  as model parameters.

We again tried different variations of the parameters, changing only one of them at a time. In fig. 6 we can see that  $\alpha$  on it's own doesn't cause any deviation for the the estimation.  $\beta$  however affects the results greatly.

We plotted along the modified estimation the original Kimura's estimation, as a limiter. We suspect that when  $\beta$  is too large, there won't be many cycles in the simulations. This might happen if either the population reaches a high frequency of the ideal trait after only a few cycles, or it get extinct very quickly, because the advantage it had in the  $k$  generations wasn't sufficient, and the same  $s$  becomes a greater disadvantage when the environment changes, resulting in a fast extinction.

In the larger values of  $\text{beta}$  we even see a deviation from the original estimation environment, but it's to be expected, because Kimura's equations are only viable for small  $s$  values.

We then also tried changing the composition of the cycle, by keeping a constant  $n = 40$ , but changing  $k, l$  accordingly.

In fig. 7 we see that the larger  $k$  relative to  $l$ , the closer the modified equation is to the original estimation of the constant environment. When using higher values of  $n$ , the simulation results doesn't fit the equation result as with lower values. This is due to the fact that our proof, and

(a) success bias/selection coefficient is:  $1 - \beta = s = 0.005$

(b) success weight is:  $\alpha = 0.1$

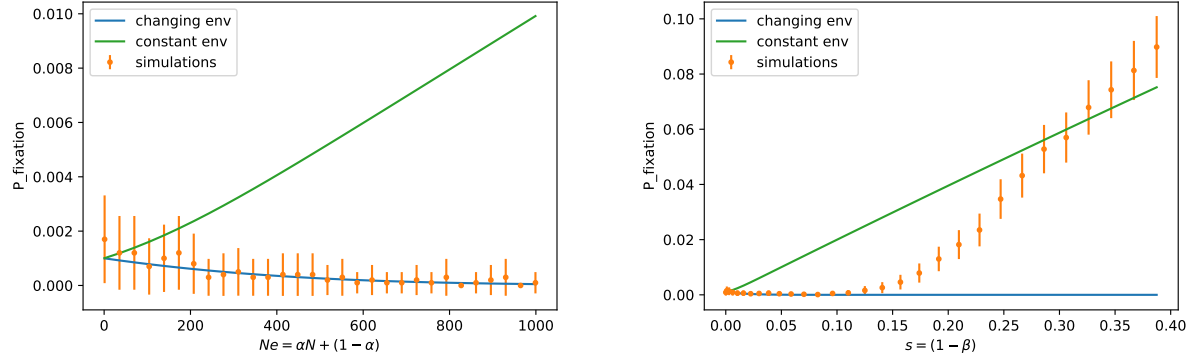


Figure 6: Model simulations compared with both the constant environment and the changing environment equations with different effective populations sizes and selection coefficients. Changing the effective population size doesn't affect the approximation, and it is condensed the mathematical expected values across all values. High values of success bias ( $s > 0.1$ ) will distance the simulations from the changing environment expected values. Very high values ( $s > 0.35$ ) will even deviate from the constant environment expected values. This is expected because Kimura's approximation are only viable for low selection coefficient values. 10,000 simulations per data point,  $N = 1,000$ ,  $\hat{A} = 1$ ,  $A = 0.9$ ,  $J = 1$ .

therefore our equation is more accurate when more cycles occur. When  $n$  is high, there will be less  
480 cycles, and the simulations will get closer to the constant environment equation.

## Discussion

482 Cultural transmission is the phenomenon of which cultural elements, in the form of attitudes,  
values, beliefs, and behavioral patterns, are transmitted between individuals, typically via copying.  
484 Some cultural traits can be more likely to be copied by others, regardless of their frequency in the  
population. Such transmission biases are common in cultural transmission processes. Many models  
486 are based on the assumption that success can be correctly identified, and easily copied. Here we  
assume that success isn't correctly identified, therefore individuals may use other indicators to try  
488 and estimate the success of potential role-models. We believe, as Fogarty et al. (2017) suggest,  
that *prestige biases* are more common in nature than success biases, since estimating success is  
490 harder. We believe prestige is composed of two main components: a trait that indicates success  
(but doesn't guarantee it), and the influence the individual already has on others, i.e number of  
492 individuals already chose him as a role-model. We suggest a model for *prestige bias*, inspired by  
the model Boyd and Richerson (1988) have suggested, and added the *influence bias* to it. We  
494 approximated our models using various distributions, and compared them to the original model  
using simulations. We showed that a *Rich getting richer* type of model can be approximated well  
496 by the general binomial distribution and the dirichlet multinomial distribution. We experimented  
with constant and changing environment in our model, and created a variation of a binary model

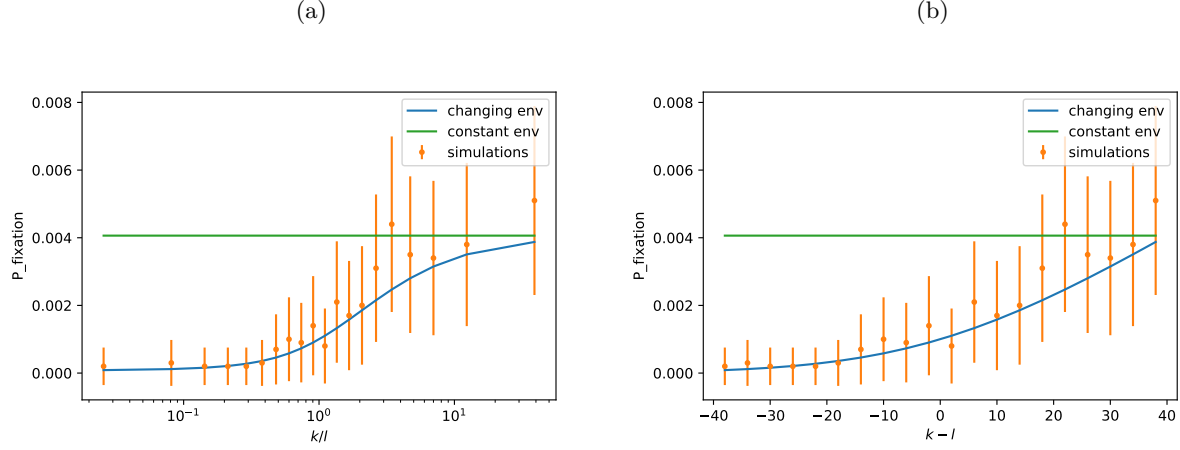


Figure 7: Model simulations compared with both the constant environment and the changing environment equations for different compositions of the environment cycle. When  $k < l$  the approximation is good. When  $k > l$ , the approximation and the simulations are both very close to the constant environment approximation. 10,000 simulations per data point,  $N = 1,000$ ,  $\hat{A} = 1$ ,  $A = 0.8$ ,  $J = 1$ ,  $1 - \beta = s = 0.02$ ,  $\alpha = 0.1$ .

498 for easier mathematical and computational analysis. We believe that in this era of social media it  
is easy to estimate one's influence over others. It is therefore crucial to model the cultural biases  
500 more realistically than success bias based model, and we believe including influence is crucial for  
that purpose.

502 With a more realistic model of a common cultural transmission bias, we may be able to better  
understand decision-making processes in humans, including life-changing choices such as occupation  
504 or a life partner. Our model can be expanded in multiple ways: observing the effects of different  
bias functions, including errors in estimating the influence, combining factors of cost when copying  
506 from an influential role model (not all could afford to copy from the most popular role-model),  
and analyzing the differences when including several optimal values for the indicator trait (multiple  
508 preference traits in the population).

## Summaries

### 510 **Animal leadership (King and Cowlshaw, 2009)**

There are two main approaches to decision makings of groups in nature: leadership and consensus.  
512 Prey leaders would lead the pack when traveling, while other animal group leaders will decide on  
a nesting site or foraging patches. They found out that leadership is observed mostly when there  
514 is a profound social network in the group, and when there are individuals that present leadership  
behavior. Leaders would usually be high ranking members in the group, such as elders, individuals  
516 with many kin relations, or possess other dominant traits. When no individual possesses such traits, or  
when the social network is lacking, a consensus is more likely to occur. When a leader is present,  
518 they will have greater selection costs, such as higher risk for predation, being poisoned by unknown  
experimental patch, but also greater benefits. For example, given the route to the foraging site was  
520 successful, the leader and his closest followers would gain most of the food, unlike in a consensus,  
where the food would be shared more equally. It appears leaders appear in simple organisms as well,  
522 like fish. In these organisms however, the leader would usually be the hungriest or the weakest, while  
the rest would prefer to follow, minimizing their costs. In baboons however, King and Cowlshaw  
524 (2009) describe many benefits for the closest associates of the dominant male, such as protection  
from predators. This is an instance of sexual-selection, where the leader, whose survival chances  
526 are lower, gains more sexual partners due to the benefits. ("The greater the risk, the greater the  
reward") In humans, leadership also has its perks and costs. Leaders can make decisions that would  
528 benefit them and their closest followers the most, while still maintaining group cohesion. However,  
wrong decision making that would harm the group could result in harm (media, social status, even  
530 violent behavior of subjects on certain cases).

### **Prestige as a type of leadership (Van Vugt and Smith, 2019)**

532 In their paper, Van Vugt and Smith (2019) suggest a different view of leadership. They say that  
most discussions assume there is one type of leadership, and so they differ in their definition.  
534 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  
536 Confucius and Machiavelli. Confucius views leaders as role models who exercise influence through  
possessing superior knowledge, skills, and (outstanding) personal qualities. This description is very  
538 similar to our indicator trait. By contrast, Machiavelli views leaders as rulers who exercise influence  
by imposing costs through (the threat of) punishment. They say that these two opposing views  
540 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.

542 Our current model doesn't reflect the model described in this article, but several adjustments  
could be made in order to match it. If we assume there's a correlation between trait value to a type  
544 of leadership (so in our binary model, one trait would be of prestige, and the other of dominance)  
we can implement their suggested model. For that we would need to add cost-benefit parameters,



546 so the ones choosing prestige will be rewarded, but pay more, while the ones choosing dominance  
would pay less, but gain less benefits. It could be interesting to see the dynamics and relations  
548 between our model parameters and these cost-benefit parameters.

## 550 **Evidence from Fijian villages for adaptive learning biases (Henrich and Broesch, 2011a)**

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

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

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

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

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

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

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

## **Using social influence to reverse harmful traditions (Efferson et al., 2020)**

568 Social influence could be a powerful tool to end traditions, harmful or advantageous. Harmful  
traditions can be child marriage, open defecation, and domestic violence, to name a few. Efferson  
570 et al. (2020) suggest a mechanism called *spillover*. By their definition, a spillover is when an  
intervention affect a large enough group in a target population, so that others not included in the  
572 intervention starts changing their behavior as well. In their research, they found that there are  
individuals who act as *agents*, who are often looked upon, and therefore they are ideal targets for  
574 interventions. This is the same concept as our role-models, where a more prestigious individual  
will be copied more, therefore spreading his trait wider in the population. Their research support  
576 therefore in our assumption that there are social biases, conformist influence specifically. They also  
suggest a way to use this phenomena to change existing traditions in a population. It is very clear  
578 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.

## 580 The evolution of prestige (Henrich and Gil-White, 2001)

One cannot mention prestige and cultural transmission without the work of Henrich and Gil-White  
582 (2001). They as well distinguish between prestige and dominance. The latter is defined by ac-  
quiring social status by using aggression, intimidation and violence. They also mention dominance  
584 is more common than prestige in non humans. Their definition of prestige is somewhat synonymous  
with ours. Specifically prestige is composed both of estimation in the eyes of people (our indica-  
586 tor/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,  
588 they show that prestige evolved from natural selection, as an efficient process to extract reproduc-  
tive benefit from the flow of socially transmitted information. Simply put, prestige is a natural step  
590 where social learning exists, due to saving costs of individual learning. It could be interesting in the  
future to enhance our model with this idea, i.e observing the copier trait of *evaluation*, rather than  
592 only observing the evolution of the indicator trait copied. Henrich and Gil-White (2001) suggest  
that the most skilled role-models will, on average, end up with most copiers. This is consistent  
594 with our findings, and in a way our model is a mathematical equivalent to their suggested model.

## 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.
- 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 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.
- 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. (2020). Cultural evolution of conformity and anticonformity. *Proceedings of the National Academy of Sciences*.
- Drezner, Z. and Farnum, N. (1993). A generalized binomial distribution. *Communications in Statistics - Theory and Methods*, 22(11):3051–3063.
- 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.
- 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 Washignton, UWEETR-2010-0006*, (0006):1–27.
- Henrich, J. and Broesch, J. (2011a). 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.

- 626 Henrich, J. and Broesch, J. (2011b). On the nature of cultural transmission networks: evidence  
from fijian villages for adaptive learning biases. *Philosophical Transactions of the Royal Society*  
628 *B: Biological Sciences*, 366(1567):1139–1148.
- Henrich, J. and Gil-White, F. J. (2001). The evolution of prestige: Freely conferred deference as a  
630 mechanism for enhancing the benefits of cultural transmission. *Evolution and human behavior*,  
22(3):165–196.
- 632 Henrich, J. and McElreath, R. (2007). Dual-inheritance theory: the evolution of human cultural  
capacities and cultural evolution. In *Oxford handbook of evolutionary psychology*.
- 634 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.
- 636 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  
638 for cultural diversity. *Evolution and Human Behavior*, 36(1):65–72.
- King, A. J. and Cowlshaw, G. (2009). Leaders, followers, and group decision-making. *Communi-*  
640 *cative & Integrative Biology*, 2(2):147–150.
- McComb, K., Moss, C., Durant, S. M., Baker, L., and Sayialel, S. (2001). Matriarchs as repositories  
642 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  
644 of social behaviour. *PloS one*, 8(7):e68153.
- Ram, Y., Liberman, U., and Feldman, M. W. (2018). Evolution of vertical and oblique transmission  
646 under fluctuating selection. *Proceedings of the National Academy of Sciences*, 115(6):E1174–  
E1183.
- 648 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  
650 learning strategies tournament. *Science*, 328(5975):208–213.
- Van Vugt, M. and Smith, J. E. (2019). A dual model of leadership and hierarchy: Evolutionary  
652 synthesis. *Trends in Cognitive Sciences*, 23(11):952–967.