



This article appeared in a journal published by Elsevier. The attached copy is furnished to the author for internal non-commercial research and education use, including for instruction at the authors institution and sharing with colleagues.

Other uses, including reproduction and distribution, or selling or licensing copies, or posting to personal, institutional or third party websites are prohibited.

In most cases authors are permitted to post their version of the article (e.g. in Word or Tex form) to their personal website or institutional repository. Authors requiring further information regarding Elsevier's archiving and manuscript policies are encouraged to visit:

<http://www.elsevier.com/copyright>



Contents lists available at ScienceDirect

# Theoretical Population Biology

journal homepage: [www.elsevier.com/locate/tpb](http://www.elsevier.com/locate/tpb)



## Rates of cultural change and patterns of cultural accumulation in stochastic models of social transmission

Kenichi Aoki<sup>a,\*</sup>, Laurent Lehmann<sup>b</sup>, Marcus W. Feldman<sup>c</sup>

<sup>a</sup> Department of Biological Sciences, The University of Tokyo, Hongo 7-3-1, Bunkyo-ku, Tokyo 113-0033, Japan

<sup>b</sup> University of Neuchâtel, Batiment UniMail, Rue Emile-Argand 11, CP 158, 2009, Neuchâtel, Switzerland

<sup>c</sup> Department of Biological Sciences, Stanford University, Stanford CA 94305, USA

### ARTICLE INFO

#### Article history:

Received 25 October 2010

Available online 16 February 2011

#### Keywords:

Infinite site model  
One-to-many transmission  
Fixation probability  
Cultural heterogeneity  
Archaeology

### ABSTRACT

Cultural variation in a population is affected by the rate of occurrence of cultural innovations, whether such innovations are preferred or eschewed, how they are transmitted between individuals in the population, and the size of the population. An innovation, such as a modification in an attribute of a handaxe, may be lost or may become a property of all handaxes, which we call “fixation of the innovation.” Alternatively, several innovations may attain appreciable frequencies, in which case properties of the frequency distribution—for example, of handaxe measurements—is important. Here we apply the Moran model from the stochastic theory of population genetics to study the evolution of cultural innovations. We obtain the probability that an initially rare innovation becomes fixed, and the expected time this takes. When variation in cultural traits is due to recurrent innovation, copy error, and sampling from generation to generation, we describe properties of this variation, such as the level of heterogeneity expected in the population. For all of these, we determine the effect of the mode of social transmission: conformist, where there is a tendency for each naïve newborn to copy the most popular variant; pro-novelty bias, where the newborn prefers a specific variant if it exists among those it samples; one-to-many transmission, where the variant one individual carries is copied by all newborns while that individual remains alive. We compare our findings with those predicted by prevailing theories for rates of cultural change and the distribution of cultural variation.

© 2011 Elsevier Inc. All rights reserved.

### 1. Introduction

The rate of genetic (molecular) evolution viewed over a long time is estimated by counting the (large) number of mutant substitutions and dividing by the elapsed time. Similarly, for studies of cultural evolution based on empirical observations in the archaeological record or inferences from differentiation of contemporary populations, the accumulated differences in many cultural traits (e.g., morphometric data on handaxes, design traits of canoes) are examined to obtain an estimate for the rate of evolution (Guglielmino et al., 1995; Eerkens and Lipo, 2007; Lycett and Gowlett, 2008; Rogers and Ehrlich, 2008; Rogers et al., 2009).

Drawing on this analogy, we define the long-term theoretical rate of cultural change,  $R$ , as

$$R = Nu\pi_1. \quad (1.1)$$

Eq. (1.1) is based on the “infinite site model” (Kimura, 1969) of population genetics, where in the current context  $N$  is the

population size,  $u$  is the innovation rate per individual per generation, and  $\pi_1$  is the fixation probability of an innovation that is initially made by a single individual. Although this analogy clearly has its limitations (e.g., Brown and Feldman, 2009), Eq. (1.1) is conceptually useful because it provides a summary statement relating three important factors that contribute to the cultural evolutionary rate (Cavalli-Sforza and Feldman, 1981; see also Eq. (1.2)).

More generally, if there are  $m$  different social roles (e.g., male and female, more than one age class, teacher and non-teacher), the long-term cultural evolutionary rate should be written as

$$R = \sum_{i=1}^m N_i u_i \pi_{1i}, \quad (1.2)$$

where  $N_i$  is the number of individuals in role  $i$ ,  $u_i$  is the innovation rate of an individual in role  $i$ , and  $\pi_{1i}$  is the fixation probability of an innovation made by an individual in role  $i$ .

In short-term cultural evolution, on the other hand, as illustrated by the “diffusion of hybrid corn” (Ryan and Gross, 1943; Rogers, 1995), interest attaches to the pattern and timing of the spread of a particular innovation. Here, an appropriate theoretical measure of the rate of cultural evolution is the mean

\* Corresponding author.

E-mail address: [kenaoki@biol.s.u-tokyo.ac.jp](mailto:kenaoki@biol.s.u-tokyo.ac.jp) (K. Aoki).

time until fixation (Kimura and Ohta, 1969). Results on the mean fixation time and mean absorption time are presented for the sake of completeness, but the emphasis in the present paper is on the fixation probability, which determines the long-term rate of cultural evolution.

It is also of interest to investigate the patterns of accumulation of cultural traits and the amount of cultural variation that are expected to be maintained within a population when the same innovations to a cultural trait occur repeatedly. Using the methods of Lehmann et al. (2011), we evaluate here the mean number of distinct cultural traits carried by an individual and segregating in the population, as well as the expected level of cultural heterogeneity (trait diversity) among individuals at steady state of the cultural accumulation dynamics.

A major difference between a gene and a cultural trait lies in their modes of transmission. Whereas genetic information (in higher organisms) flows solely from parent to child (vertical transmission), the inheritance of a cultural trait is not necessarily limited by the degree of relatedness. In addition, some cultural traits may be acquired under the concerted influence of a group of individuals (many-to-one transmission), while for others one particular individual (e.g., teacher, authority, prestigious individual) may serve as the exemplar (cultural parent) for a large number of individuals (one-to-many transmission).

Semi-quantitative arguments have been invoked to predict the theoretical rates of cultural evolution under various modes of social transmission (Guglielmino et al., 1995; Lycett and Gowlett, 2008; Hewlett and Cavalli-Sforza, 1986; MacDonald, 1998). For example, cultural changes under many-to-one (MtO) transmission and one-to-many (OtM) transmission are described as being “very slow” and “very rapid”, respectively. Archaeologists also discuss the amount of cultural variation within a population (Lycett and Gowlett, 2008; MacDonald, 1998). In particular, they agree that MtO transmission would result in very low “cultural variation between individuals”. A major motivation for the present paper is to reevaluate these claims.

Theoretical studies have also addressed the consequences of choosing the most adaptive or attractive variant of a cultural trait from among  $k$  exemplars (best-of- $k$  transmission) (Shennan, 2001; Henrich, 2004). Rates of popular culture change have been investigated, assuming oblique transmission and recurrent innovation in a finite population with discrete generations and simultaneous reproduction using a statistical framework that originated in population genetics (Bentley et al., 2004, 2007).

An important parameter in studies of cultural evolution is the size of the population or social group in which the transmission of cultural traits occurs. Here we are principally interested in interpreting the rates of cultural change and amounts of cultural variation in Pleistocene hominid populations. Hence, we use band size in current hunter–gatherer societies, which averages perhaps a total of 25 individuals of both sexes and all age groups (Ichikawa, 1978; Lee, 1979; Terashima, 1985; Kelly, 1995). However, bands do not exist in isolation, and a “local community” may be formed by neighboring bands among which networks of social transmission presumably extend. Perhaps five bands comprising approximately 125 individuals are tightly linked into a local community (Terashima, 1985; Dunbar, 1992 proposed 150, which is now called Dunbar’s number.) These two estimates of social group size (i.e., 25 and 125) are used to compare the rates of cultural change under different modes of social transmission (Table 5).

In this paper, we obtain detailed quantitative predictions on cultural evolutionary rates and on cultural variation within a population. Stochastic models based on the Moran model (Moran, 1958) are suitable vehicles for this analysis (see also Lehmann et al., 2011). There are several reasons why this approach is preferable

to one based on the Wright–Fisher model (Wright, 1931; Fisher, 1930). First, Moran-type models are mathematically simpler since they involve asynchronous as opposed to synchronous updating. Second, they result in overlapping generations. Third, they allow an in-depth treatment of OtM transmission (cf. Cavalli-Sforza and Feldman, 1981, pp. 192–202).

Our major findings are as follows. First, the type of MtO transmission considered in this paper is conformist transmission (Lumsden and Wilson, 1981; Boyd and Richerson, 1985). With the specific functional form of conformist transmission we assume (see Eqs. (3.1)–(3.5)), the fixation probability of an innovation (Table 2) is extremely small. In other words, cultural change under MtO transmission is predicted to be very slow (Eq. (1.1)). In addition, this social learning rule results in a very low number of distinct traits carried by an individual and a moderately low number present in the population, but a high level of cultural heterogeneity under the process of cultural accumulation (in contradiction to prevailing theory).

Second, our best-of- $k$  (Bok) model posits that each newborn prefers the new cultural variant over the old and adopts the former (pro-novelty bias) provided it is carried by at least one individual among  $k$  exemplars sampled from the population. Fixation probabilities of the innovation are quite large and depend little on whether it is selectively neutral, adaptive, or maladaptive (Table 3). With this social learning rule, individuals and the population are both predicted to accumulate a high number of different traits on average, and the cultural heterogeneity is relatively low.

Third, in our OtM transmission model, one individual has the special status of teacher and – as long as it remains alive – is copied by each newborn individual. The fixation probability of an innovation then depends on whether it is made by the teacher or some other individual. In the former case the fixation probability is high, but in the latter it is lower than with vertical, oblique, or horizontal transmission (Eqs. (5.7)–(5.8), and Table 4). When Eq. (1.2) is applied, we do not observe a substantial acceleration of the long-term cultural evolutionary rate unless teachers are highly innovative (again in contradiction to prevailing theory). For cultural accumulation, OtM transmission results in individuals and the population carrying a low number of distinct traits and a high level of cultural homogeneity.

Before proceeding, we wish to point out that the specific models of cultural transmission we introduce and analyze in this paper make extreme simplifying assumptions. This allows us to rigorously demonstrate the large effect that different modes of social transmission can have on cultural evolutionary rates and cultural variation (in addition to the effect of natural selection often considered in genetic models). Predictions based on extreme functional forms do not give realistic cultural evolutionary rates, for example, even if accurate estimates of the innovation rate are available. Hence, the values derived from our models should be used only in a comparative fashion.

## 2. The Moran model and vertical, oblique, or horizontal transmission

Moran (1958) proposed a model of random genetic drift, which we modify as follows to describe cultural evolution. Assume a finite population comprising  $N$  individuals. A cultural trait exists in two variant forms  $A$  and  $B$ , with  $i$  individuals of type  $A$  and  $N - i$  of type  $B$ . At each arbitrary time step, an individual is born and acquires one or the other of these types by social learning (e.g., imitation), followed by the death of an individual other than this newborn (birth–death event). If the newborn copies a randomly chosen member of the population—oblique or horizontal transmission, there being no conceptual distinction between the two in the

Moran model—it acquires type  $A$  with probability  $b_A(i) = i/N$  and type  $B$  with probability  $b_B(i) = (N-i)/N$ . Similarly, if death occurs at random, the dying individual will be of type  $A$  with probability  $d_A(i) = i/N$  and type  $B$  with probability  $d_B(i) = (N-i)/N$ . The probability that after this birth–death event the number of type  $A$  individuals has increased by one (up-transition) is

$$p_i = b_A(i)d_B(i) = i(N-i)/N^2, \quad (2.1)$$

the probability it has decreased by one (down-transition) is

$$q_i = b_B(i)d_A(i) = i(N-i)/N^2, \quad (2.2)$$

and the probability it has not changed is  $1 - p_i - q_i = (i/N)^2 + [(N-i)/N]^2$ .

Interestingly, these transition probabilities also apply when social learning is vertical – the newborn acquires the cultural type of one of its parents – provided we assume equal fertilities. For brevity, we will refer to this model of social transmission as the oblique transmission model.

Differential mortality (viability selection) can be incorporated by setting the mortality of type  $B$  to be  $v$  times that of type  $A$  ( $0 < v < \infty$ ). Then, the probabilities of death for type  $A$  and  $B$  individuals are

$$d_A(i) = i/[i + (N-i)v], \quad (2.3)$$

and

$$d_B(i) = (N-i)v/[i + (N-i)v], \quad (2.4)$$

respectively.

In Appendix A we present general formulas from Ewens (2004, p. 90) for: (i) the fixation probability of type  $A$ ,  $\pi_i$  (the probability that type  $A$  will ultimately spread through the population, eliminating type  $B$ ); (ii) the mean absorption time,  $\bar{t}_i$  (the average time until type  $A$  is either lost or fixed); and (iii) the mean fixation time of type  $A$ ,  $\bar{t}_i^*$  (the average time until type  $A$  is fixed, excluding the cases when it is lost), if there are initially  $i$  individuals of type  $A$ . These values can be computed for the oblique transmission model from the transition probabilities Eqs. (2.1)–(2.4).

In particular, the fixation probability of type  $A$  is

$$\pi_i = i/N \quad (2.5)$$

in the selectively neutral case ( $v = 1$ ), and

$$\pi_i = [1 - (1/v)^i]/[1 - (1/v)^N] \quad (2.6)$$

when there is viability selection ( $v \neq 1$ ; Ewens, 2004, p. 109). The appropriate formula to be used in computing the long-term rate of cultural evolution is Eq. (1.1), since social roles are not distinguished in this oblique transmission model. In the selectively neutral case, Eq. (1.1) reduces to  $R = u$ .

Of particular interest is the case where type  $A$  is initially represented once in the population, as when it arises by innovation or is introduced by one individual from an external source. The tables illustrate the fixation probabilities, mean absorption times, and mean fixation times for this special but important instance. Table 1 gives results for the oblique transmission model. These values serve as baselines against which the accelerating or decelerating effects of the alternative modes of social transmission we investigate can be compared. Note the equivalence of  $\bar{t}_1^*$  for  $v = 49/50$  and  $50/49$  and also for  $v = 9/10$  and  $10/9$ . This result – the mean fixation time of type  $A$ , initially represented once in the population, depends on the intensity but not the direction of selection – is proved in general in Appendix B.

### 3. Many-to-one (conformist) transmission

As before, let variant cultural types  $A$  and  $B$  exist in numbers  $i$  and  $N - i$ , respectively. Suppose each newborn samples

**Table 1**

Numerical examples of the fixation probability ( $\pi_1$ ), the mean absorption time ( $\bar{t}_1$ ), and the mean fixation time ( $\bar{t}_1^*$ ) in the oblique transmission model. The population size is  $N$ , and the mortality of type  $B$  is  $v$  times that of type  $A$ . One individual of type  $A$  is assumed to be initially present. Units for  $\bar{t}_1$  and  $\bar{t}_1^*$  are  $N$  birth–death events, the equivalent of one generation when  $v = 1$ .

$N$	$v$	$\pi_1$	$\bar{t}_1$	$\bar{t}_1^*$
25	1	0.04	3.776	24.000
	49/50	0.0311	3.552	23.913
	50/49	0.0504	3.996	23.913
	9/10	0.00859	2.672	21.947
	10/9	0.108	4.770	21.947
125	1	0.008	5.402	124.000
	49/50	0.00178	4.256	114.355
	50/49	0.0217	6.524	114.355
	9/10	$2.12 \times 10^{-7}$	2.390	57.380
	10/9	0.100	8.112	57.380

**Table 2**

Numerical examples of the fixation probability ( $\pi_1$ ), the mean absorption time ( $\bar{t}_1$ ), and the mean fixation time ( $\bar{t}_1^*$ ) in the many-to-one (conformist) transmission model. The population size is  $N$ , the mortality of type  $B$  is  $v$  times that of type  $A$ , and the number of exemplars sampled is  $k$ . The error rate is fixed at  $\varepsilon = 0.001$ . One individual of type  $A$  is initially present. Units for  $\bar{t}_1$  and  $\bar{t}_1^*$  are  $N$  birth–death events, the equivalent of one generation when  $v = 1$ .

$N$	$v$	$k$	$\pi_1$	$\tilde{t}_1$	$\tilde{t}_1^*$
25	1	3	$9.98 \times 10^{-8}$	1.002	9.373
		4	$1.35 \times 10^{-7}$	1.002	9.379
		5	$1.63 \times 10^{-11}$	1.003	7.761
	9/10	3	$2.67 \times 10^{-8}$	0.905	9.383
		4	$3.62 \times 10^{-8}$	0.906	9.389
		5	$4.46 \times 10^{-12}$	0.906	7.786
	10/9	3	$3.35 \times 10^{-7}$	1.109	9.383
		4	$4.54 \times 10^{-7}$	1.109	9.389
		5	$5.59 \times 10^{-11}$	1.110	7.786
125	1	3	$2.59 \times 10^{-26}$	1.002	16.834
		4	$3.98 \times 10^{-26}$	1.002	16.847
		5	$6.14 \times 10^{-44}$	1.003	13.342
	9/10	3	$2.71 \times 10^{-29}$	0.902	16.875
		4	$4.17 \times 10^{-29}$	0.902	16.888
		5	$7.40 \times 10^{-47}$	0.903	13.394
	10/9	3	$1.28 \times 10^{-23}$	1.112	16.875
		4	$1.97 \times 10^{-23}$	1.112	16.888
		5	$3.49 \times 10^{-41}$	1.113	13.394

$k$  exemplars without replacement from the population of  $N$  individuals ( $2 \leq k \leq N$ ). The probability that there are  $j$  individuals of type  $A$  and  $k - j$  individuals of type  $B$  in this sample is given by the hypergeometric distribution,

$$h(j; k, N, i) = \binom{i}{j} \binom{N-i}{k-j} / \binom{N}{k}, \quad (3.1)$$

where  $h(j; k, N, i) = 0$  if  $j < 0$ ,  $j > i$ ,  $j > k$ , or  $j < k - (N - i)$ .

Pure conformist transmission entails that each newborn adopt the majority cultural type among its  $k$  exemplars. Here we modify this condition slightly by introducing a small probability proportional to  $\varepsilon$  (error rate) of adopting the minority cultural type provided there is at least one individual of this minority type among the  $k$  exemplars. Moreover, when both cultural types are equally represented in the sample, it is natural to assume that the newborn will adopt either type with probability one-half.

Then, when  $k \geq 2$  is even, the newborn will copy type  $A$  with probability

$$b_A(i) = \sum_{j=0}^{k/2-1} h(j; k, N, i)j\varepsilon + h(k/2; k, N, i)/2 + \sum_{j=k/2+1}^k h(j; k, N, i)[1 - (k-j)\varepsilon], \quad (3.2)$$

**Table 3**

Numerical examples of the fixation probability ( $\pi_1$ ), the mean absorption time ( $\bar{t}_1$ ), and the mean fixation time ( $\bar{t}_1^*$ ) in the best-of- $k$  (pro-novelty) model. The population size is  $N$ , the mortality of type  $B$  is  $v$  times that of type  $A$ , and the number of exemplars sampled is  $k$ . One individual of type  $A$  is initially present. Units for  $\bar{t}_1$  and  $\bar{t}_1^*$  are  $N$  birth–death events, the equivalent of one generation when  $v = 1$ .

$N$	$v$	$k$	$\pi_1$	$\bar{t}_1$	$\bar{t}_1^*$
25	1	2	0.531	3.719	6.399
		3	0.701	3.577	4.933
		4	0.785	3.559	4.453
		5	0.836	3.570	4.224
		5	0.836	3.570	4.224
	9/10	2	0.481	3.719	6.957
		3	0.668	3.686	5.310
		4	0.762	3.723	4.792
		5	0.819	3.767	4.551
		5	0.819	3.767	4.551
	10/9	2	0.576	3.689	5.914
		3	0.730	3.462	4.597
		4	0.806	3.400	4.148
		5	0.853	3.384	3.931
		5	0.853	3.384	3.931
125	1	2	0.506	5.384	9.963
		3	0.673	5.182	7.498
		4	0.757	5.144	6.703
		5	0.807	5.144	6.320
		5	0.807	5.144	6.320
	9/10	2	0.452	5.357	10.977
		3	0.637	5.326	8.121
		4	0.730	5.373	7.250
		5	0.786	5.425	6.841
		5	0.786	5.425	6.841
	10/9	2	0.555	5.354	9.112
		3	0.706	5.020	6.947
		4	0.781	4.915	6.213
		5	0.826	4.874	5.852
		5	0.826	4.874	5.852

and type  $B$  with probability

$$b_B(i) = \sum_{j=0}^{k/2-1} h(j; k, N, i)(1 - j\varepsilon) + h(k/2; k, N, i)/2 + \sum_{j=k/2+1}^k h(j; k, N, i)(k - j)\varepsilon. \quad (3.3)$$

Similarly, when  $k \geq 3$  is odd, the newborn will copy type  $A$  with probability

$$b_A(i) = \sum_{j=0}^{(k-1)/2} h(j; k, N, i)j\varepsilon + \sum_{j=(k+1)/2}^k h(j; k, N, i)[1 - (k - j)\varepsilon], \quad (3.4)$$

and type  $B$  with probability

$$b_B(i) = \sum_{j=0}^{(k-1)/2} h(j; k, N, i)(1 - j\varepsilon) + \sum_{j=(k+1)/2}^k h(j; k, N, i)(k - j)\varepsilon. \quad (3.5)$$

Note that Eqs. (3.1)–(3.5) entail a more extreme form of conformity than the often-used sigmoid function (e.g., Boyd and Richerson, 1985).

If the mortality of type  $B$  is  $v$  times that of type  $A$ , the up- and down-transition probabilities for this model are  $p_i = b_A(i)d_B(i)$  and  $q_i = b_B(i)d_A(i)$ , respectively, where  $b_A(i)$  and  $b_B(i)$  are defined by Eqs. (3.1)–(3.5), and  $d_A(i)$  and  $d_B(i)$  by Eqs. (2.3)–(2.4). When  $k = 2$ , or when  $k = N$  and  $\varepsilon = 1/N$ , this model reduces to the oblique transmission model.

The fixation probability, mean absorption time, and mean fixation time of type  $A$ , when there are initially  $i$  individuals of this type, can be computed from the formulas of Appendix A.

**Table 4**

Numerical examples of the fixation probability ( $\pi_{1\alpha}$ ), the mean absorption time ( $\bar{t}_{1\alpha}$ ), and the mean fixation time ( $\bar{t}_{1\alpha}^*$ ) in the one-to-many (teacher) transmission model. The population size is  $N$ . Types  $A$  and  $B$  are equally viable. Initially, there is one individual of type  $A$ , and the teacher may be type  $A$  ( $\alpha = A$ ) or type  $B$  ( $\alpha = B$ ). Units for  $\bar{t}_{1\alpha}$  and  $\bar{t}_{1\alpha}^*$  are  $N$  birth–death events, the equivalent of one generation when  $v = 1$ .

$N$	$\alpha$	$\pi_{1\alpha}$	$\bar{t}_{1\alpha}$	$\bar{t}_{1\alpha}^*$
25	$A$	0.52	3.776	4.092
	$B$	0.02	1.107	5.054
125	$A$	0.504	5.402	5.851
	$B$	0.004	1.035	6.843

**Table 5**

Numerical examples of the selectively neutral cultural evolutionary rate under various modes of social transmission. Innovation rate of  $u = 1$  per individual per generation is assumed in all models except the OtM teacher model, where we have set  $u_t = 2$ ,  $u_{nt} = 1$ . Where applicable,  $k = 4$  and  $\varepsilon = 0.001$ . See text for details.

Mode of transmission	$N = 25$	$N = 125$
Oblique	1	1
Conformist	$3.38 \times 10^{-6}$	$9.95 \times 10^{-25}$
Best-of- $k$	19.6	94.6
Teacher	1.52	1.504

Illustrative results are shown in Table 2 for the case of  $i = 1$ . The values 3–5 for parameter  $k$  represent the likely number of (same-sex) adults in a band who may serve as exemplars for the newborn. The most salient feature of Table 2 is the low fixation probabilities. Conformist transmission entails that a newborn is unlikely to adopt the minority cultural type. Hence, an innovation that is initially represented once in the population ( $i = 1$ ) has only a small chance of spreading through the population. Table 2 shows a sharp drop in the fixation probabilities as the number of exemplars,  $k$ , increases from 4 to 5. There is an even larger effect of the population size,  $N$ . In this model also, the long-term evolutionary rate is given by Eq. (1.1).

The mean fixation times are shorter than with oblique transmission, indicating that in the rare cases when fixation occurs, spread of the innovation is fast. When viability selection is imposed, the mean fixation time is identical whether cultural type  $A$  is selected for or against (e.g., compare 4th and 7th lines of Table 2). This was also observed in the oblique transmission model.

#### 4. Best-of- $k$ (pro-novelty) transmission

As in the previous model, each newborn samples  $k$  exemplars from the population of  $N$  individuals with  $i$  of type  $A$  and  $N - i$  of type  $B$ . Suppose that type  $A$  is preferred over type  $B$ . Specifically, type  $A$  is adopted by the newborn provided there is at least one individual of type  $A$  among the  $k$  exemplars. Since we regard type  $A$  to be the newly arisen cultural type (i.e., the innovation), this implies a pro-novelty bias. The probabilities of adoption of type  $A$  and  $B$  can be written, respectively, as

$$b_A(i) = 1 - \binom{N-i}{k} / \binom{N}{k}, \quad (4.1)$$

$$b_B(i) = \binom{N-i}{k} / \binom{N}{k}, \quad (4.2)$$

if  $i \leq N - k$ , and

$$b_A(i) = 1, \quad (4.3)$$

$$b_B(i) = 0, \quad (4.4)$$

if  $i \geq N - k + 1$ . These assumptions entail an extreme form of pro-novelty bias.



Eq. (4.4) implies that the down-transition probability,  $q_i = b_B(i)d_A(i)$ , is 0 when  $i \geq N - k + 1$ . As a result, Eqs. (A.3)–(A.7) of Appendix A are ill-defined. Revised formulas applicable to this model are derived in Appendix C and used to calculate the entries in Table 3. Fixation probabilities are quite high whether the preferred cultural type is selectively neutral, adaptive, or maladaptive. Moreover, in contrast to the oblique and MtO transmission models, fixation probabilities are not substantially lower for the larger population size. Mean fixation times are also relatively short.

## 5. One-to-many (teacher) transmission

Assume that the population of size  $N$  contains two selectively and preferentially neutral variant cultural types  $A$  and  $B$ . One individual in this population has the special status of teacher. At each time step, one newborn is produced and adopts the cultural type of the current teacher, followed by the death of one random individual excluding the newborn. If the teacher is the one to die, its social role is taken by another individual, randomly chosen from among the survivors including the newborn.

As with the other models considered in this paper, the OtM transmission model entails some extreme assumptions. For example, there is just one teacher that is copied by all newborns while it remains alive. Nevertheless, this assumption receives some empirical support in the case of the transmission of stone-tool-making. Based on the evidence from refitted débitage at the Magdalenian site of Etiolles (France), Pigeot (1990) argues that novices may have turned to skilled knappers for instruction. Moreover, she suggests that there may have been just one specialist who served as teacher to all other individuals at this site.

The state of the population is completely described by the number  $i$  of type  $A$  individuals and the type  $\alpha$  of the current teacher. We require  $1 \leq i \leq N$  when  $\alpha = A$  and  $0 \leq i \leq N - 1$  when  $\alpha = B$ . States  $OB$  and  $NA$  are absorbing, whereas the remaining  $2(N - 1)$  states are transient. Let  $p_{i\alpha, j\beta}$  be the transition probability from state  $i\alpha$  to state  $j\beta$ . Then, the assumptions of the paragraph before last entail that the positive transition probabilities are

$$p_{iB, (i-1)B} = i/N, \quad (5.1)$$

$$p_{iB, iB} = (N - i - 1)/N + (N - i)/N^2, \quad (5.2)$$

$$p_{iB, iA} = i/N^2, \quad (5.3)$$

$$p_{iA, iB} = (N - i)/N^2, \quad (5.4)$$

$$p_{iA, iA} = (i - 1)/N + i/N^2, \quad (5.5)$$

$$p_{iA, (i+1)A} = (N - i)/N, \quad (5.6)$$

when  $1 \leq i \leq N - 1$ , and moreover  $p_{OB, OB} = p_{NA, NA} = 1$ . For example,  $p_{iA, iA}$  (Eq. (5.5)) is obtained by noting that (i) the newborn will be type  $A$  since the teacher is type  $A$  (probability 1), and (ii) one of the  $i - 1$  type  $A$  individuals that is not the teacher dies (probability  $(i - 1)/N$ ), or (iii) the teacher dies (probability  $1/N$ ) and its role is adopted by a type  $A$  individual (probability  $i/N$ ).

The fixation probabilities for this model are

$$\pi_{iB} = i/2N \quad (5.7)$$

if initially there are  $i$  individuals of type  $A$  and the teacher is type  $B$ , and

$$\pi_{iA} = (N + i)/2N \quad (5.8)$$

if initially there are  $i$  individuals of type  $A$  including the teacher (Appendix D). In the latter case, the fixation probability exceeds one-half for all population sizes (Eq. (5.8)). With a type  $B$  teacher, on the other hand, the fixation probability is one-half that for the oblique transmission model (compare Eq. (5.7) with Eq. (2.5)). The

mean absorption and mean fixation times from state  $1A$  and  $1B$  given in Table 4 are computed using the formulas in Appendix E.

The OtM model posits two social roles, teacher and non-teacher. If the one teacher and each of the  $N - 1$  non-teachers innovate at rates  $u_t$  and  $u_{nt}$ , respectively, then from Eq. (1.2) the long-term cultural evolutionary rate is given by

$$R = [u_t(N + 1) + u_{nt}(N - 1)] / (2N), \quad (5.9)$$

where we have used Eqs. (5.7) and (5.8) with  $i = 1$ . Clearly, we recover  $R = u$  if we substitute  $u_t = u_{nt} = u$  in Eq. (5.9), which parallels the standard result for oblique transmission that the rate of change of a selectively neutral cultural trait is independent of population size.

## 6. Accumulation and maintenance of cultural variation with oblique transmission

In order to address the question of the amount of cultural variation maintained in a finite population, we introduce a different model of cultural innovation which allows for imperfect copying. As before, in a population of size  $N$ , assume two cultural types  $A$  and  $B$ , where we now interpret  $A$  as the presence of some trait and  $B$  as its absence. At each birth-death event, a newborn may innovate with probability  $u$  ( $0 < u < 1$ ), resulting in its acquisition of cultural type  $A$ . When it does not innovate, it acquires cultural type  $A$  by social learning with probability  $\beta$  ( $0 < \beta < 1$ , imperfect copying). Then, with oblique transmission (random copying) and  $i$  individuals of type  $A$ , the probability that a newborn is of cultural type  $A$  can be expressed as

$$\tilde{b}_A(i) = u + (1 - u)\beta i/N, \quad (6.1)$$

and the probability it is of cultural type  $B$  is

$$\tilde{b}_B(i) = (1 - u)[(N - i)/N + (1 - \beta)i/N]. \quad (6.2)$$

The probabilities of death,  $d_A(i)$  and  $d_B(i)$ , are unchanged (Eqs. (2.3)–(2.4)).

If  $0 < u < 1$  and  $0 < \beta < 1$ , the stochastic process describing the evolution of the frequency of cultural type  $A$  (Eqs. (2.3)–(2.4), (6.1)–(6.2)) has a stationary distribution whose properties refer to the cultural variation in the population. Considering a collection of  $c$  independent traits, from this stationary distribution we can evaluate the mean number of distinct traits carried by an individual,  $\lambda_f$ ; the mean number of traits segregating in the population,  $\lambda_p$ ; and the mean number of traits shared between two randomly chosen distinct individuals,  $\lambda_s$ . We can also calculate the proportion of shared traits between two randomly sampled individuals,  $\varphi = \lambda_s/\lambda_f$  (see Lehmann et al., 2011). More precisely,  $\varphi$  is the conditional probability that a trait carried by an individual is also carried by another individual. Hence, we may regard  $\varphi$  as a measure of the cultural homogeneity of the population (similarity between individuals in the traits that they carry), whereas  $1 - \varphi$  is a measure of cultural heterogeneity (trait diversity).

With a very large number of independent and selectively neutral traits ( $c$  large and  $v = 1$ ), we have for large  $N$  (see Lehmann et al., 2011)

$$\lambda_f = U/(1 - \beta), \quad (6.3)$$

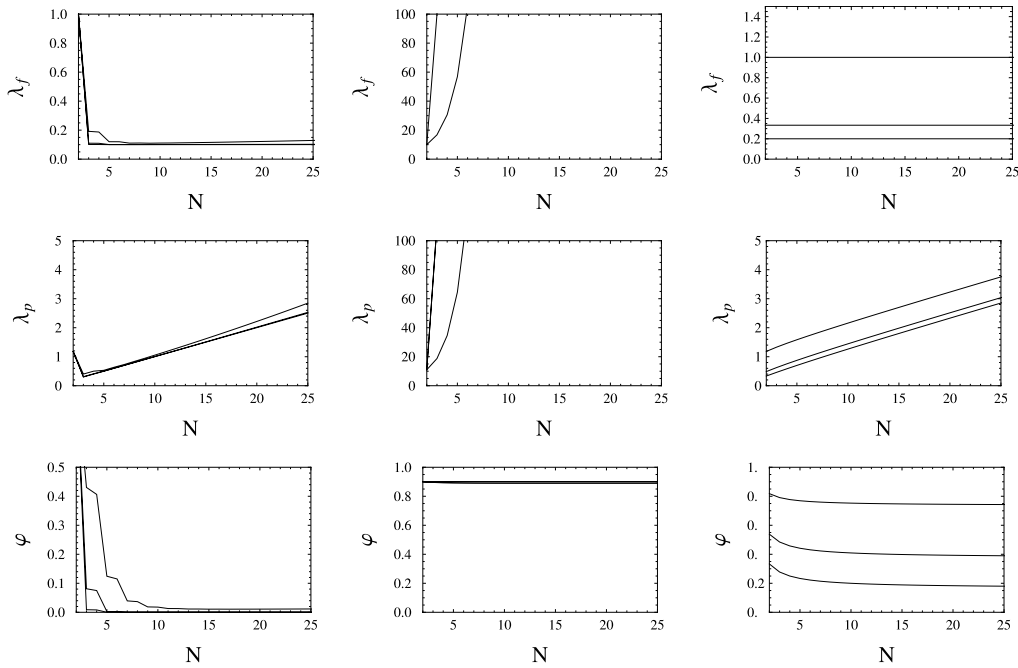
where  $U = cu$  is the total innovation rate. Moreover,

$$\lambda_p = \frac{U}{1 - \beta} - \frac{NU}{\beta} \left[ \log(1 - \beta) + \frac{\varphi}{2} \right], \quad (6.4)$$

where

$$\varphi = \beta/N(1 - \beta), \quad (6.5)$$

which we use as benchmarks for comparison with the other modes of social transmission.



**Fig. 1.** Measures of variation for a selectively neutral cultural trait. In the first column of panels,  $\lambda_f$ ,  $\lambda_p$ , and  $\varphi$  are graphed as functions of population size  $N$  for the many-to-one (conformist) transmission model with  $U = 0.1$  and  $\beta = 0.9$ . Curves from top to bottom correspond to  $\varepsilon = 0.01, 0.001$ , and  $0.0001$  (the latter two are nearly indistinguishable), respectively. Similarly, in the second column of panels,  $\lambda_f$ ,  $\lambda_p$ , and  $\varphi$  are plotted against  $N$  for the best-of- $k$  (pro-novelty) transmission model, again with  $U = 0.1$  and  $\beta = 0.9$ . From top to bottom, curves are drawn for  $k = 2, 4$ , and  $8$  (when  $k > N$ , we set  $k = N$ ). When  $N$  becomes large we consistently observe that  $\varphi$  converges to  $\beta$ . Finally, in the third column of panels, values of  $\lambda_f$ ,  $\lambda_p$ , and  $\varphi$  are given for the one-to-many (teacher) transmission model, where in this case the curves from top to bottom are for  $\beta = 0.9, 0.7$ , and  $0.5$ . All curves are graphed as functions of  $N$  for  $1 \leq N \leq 25$  because this allows for a good compromise between distinguishing the different curves in the panels and letting the variable  $\varphi$  approximate its limiting value. The qualitative increase of the remaining variables for larger values of  $N$  can be inferred from the panels. By comparison, for the oblique transmission model with  $U = 0.1$  and  $\beta = 0.9$ , we have for large  $N$ ,  $\lambda_f = 1$ ,  $\lambda_p = 0.3 + 0.26N$ , and  $\varphi = 9/N$ .

## 7. Cultural variation at equilibrium with alternative modes of social transmission

With MtO (conformist) transmission, the probabilities of birth of  $A$  and  $B$ , respectively, are

$$\tilde{b}_A(i) = u + (1 - u)\beta b_A(i), \quad (7.1)$$

and

$$\tilde{b}_B(i) = (1 - u)[b_B(i) + (1 - \beta)b_A(i)], \quad (7.2)$$

where  $b_A(i)$  and  $b_B(i)$  are given by Eqs. (3.1)–(3.3) (for  $k \geq 2$  even) or by Eqs. (3.1), (3.4)–(3.5) (for  $k \geq 3$  odd).

The stationary distribution is obtained numerically for various parameter sets, which allows computation of  $\lambda_f$ ,  $\lambda_p$ , and  $\varphi$ . Examples are presented in the first column of Fig. 1, where we see that the average number of traits carried by an individual,  $\lambda_f$ , is very small (top panel). In addition, values of  $\lambda_p$  are moderately low (middle panel) while values of  $\varphi$  are extremely low (bottom panel) for, say,  $N > 10$ , where the latter entails that most individuals carry different traits, implying high cultural heterogeneity (trait diversity) among individuals. These results are in qualitative agreement with our previous work (Fig. 3 of Lehmann et al., 2011).

For Bok transmission, we substitute Eqs. (4.1)–(4.4) in Eqs. (7.1)–(7.2). The stationary distribution is again obtained numerically for various parameter sets, with an example shown in the middle column of Fig. 1. We find that individuals carry a large number of traits on average (high  $\lambda_f$ ), the population has a large number of segregating traits (high  $\lambda_p$ ), and the proportion of shared traits is relatively high (high  $\varphi$ ). Hence, the effects of pro-novelty bias on the accumulation of cultural variation are similar to sensitivity to the minority (Lehmann et al., 2011). Note also that  $\varphi$  appears to approach  $\beta$  as  $N$  becomes large (bottom panel), which was observed for all parameter sets tested.

Finally, for the model of OtM (teacher) transmission, the positive transition probabilities,  $\tilde{p}_{i\alpha,j\beta}$ , incorporating recurrent innovation and imperfect copying are as shown in Appendix F. From the stationary distribution of the frequency of type  $A$ , we can compute the statistics  $\lambda_f$ ,  $\lambda_p$ , and  $\varphi$  for any values of  $N$ ,  $u$ , and  $\beta$ . An example is given in the third column of Fig. 1: the average number of traits per individual,  $\lambda_f$ , is low, while the average number of traits in the population,  $\lambda_p$ , increases with  $N$  in an almost linear fashion. The bottom panel shows that the proportion of shared traits between two randomly sampled distinct individuals,  $\varphi$ , is high – hence the cultural heterogeneity  $1 - \varphi$  is low – unless the copying fidelity,  $\beta$ , is low. This last result is an intuitively reasonable consequence of one-to-many transmission.

## 8. Discussion

In Table 5 we compare the long-term evolutionary rates for *selectively neutral* cultural variants expected under the various modes of social transmission. In computing  $R$  as defined by Eq. (1.1), we have arbitrarily assumed that each individual produces, or introduces from an external source, one innovation during a generation ( $u = 1$ ). In this case, the cultural evolutionary rate for oblique transmission is  $R = 1$  regardless of the population size. Also, we have set  $k = 4$  and  $\varepsilon = 0.001$  where applicable. For the OtM transmission model where Eq. (1.2) is appropriate, we have assumed that the teacher innovates at rate  $u_t = 2$ , while each non-teacher innovates at rate  $u_{nt} = 1$ .

Oblique transmission here includes vertical and horizontal transmission, since in the Moran model either there is no conceptual distinction or their effects are identical. It is often claimed that more rapid cultural evolution is possible with horizontal than vertical transmission (Guglielmino et al., 1995; Lycett and Gowlett, 2008; MacDonald, 1998). This prediction follows from the assumption of discrete generations, which

entails that horizontal transmission can occur repeatedly during a generation while vertical transmission occurs just once.

With conformist transmission, the evolutionary rates are exceedingly low. The value of  $R = 3.38 \times 10^{-6}$  for  $N = 25$  means that an innovation that ultimately becomes fixed arises, on average, once every  $1 \div (3.38 \times 10^{-6}) = 2.96 \times 10^5$  generations. This is approximately how long hominids have walked the earth. When the population size is  $N = 125$ , the mean interval between successful innovations is even longer. The functional form of Eqs. (3.1)–(3.5) and the setting for parameter  $\varepsilon$  (Table 2) result in a low probability that a minority cultural variant will spread. Larger values of  $\varepsilon$  or alternative functional forms for conformist transmission (e.g., Boyd and Richerson, 1985) would produce higher evolutionary rates. Our numerical examples show how extreme the decelerating effect of conformist transmission can be and support the suggestion of Lycett and Gowlett (2008) that “a many-to-one system could have operated ... within Acheulean populations, at least ... with handaxe manufacture”, which they made to explain the extremely conservative nature of this lithic tradition.

Table 5 shows that Bok transmission, on the other hand, accelerates cultural evolution relative to oblique transmission. In the Bok model, each newborn has a preference for a newly introduced cultural variant (type A), which it adopts regardless of its fitness effects, provided there is at least one individual of this type among the  $k$  exemplars. Table 3 records the results of a more detailed numerical analysis. We see that fixation probabilities are high (about 1/2) even when only two exemplars are sampled ( $k = 2$ ) and when type A is selected against ( $v = 9/10$ ). Hence, the cultural evolutionary rates are also high. (However, with even stronger selection against type A, the fixation probability may be substantially reduced). Tables 3 and 5 also reveal that a larger population size implies a higher cultural evolutionary rate. This is because the fixation probability in the Bok model decreases with  $N$ , but at a rate slower than  $1/N$ .

OtM transmission has different consequences depending on whether the innovator is the teacher or some other individual. In the former case, the fixation probability exceeds one-half for all population sizes ( $\pi_{1A} = (N + 1)/2N$ ), whereas in the latter case, the fixation probability is exactly one-half that with oblique transmission ( $\pi_{1B} = 1/2N$ ).

If the one teacher and each of the  $N - 1$  non-teachers innovate at rates  $u_t$  and  $u_{nt}$ , respectively, then the long-term cultural evolutionary rate is given by Eq. (5.9)—a weighted average of the innovation rates. In particular, if we make the arbitrary assumption as in Table 5 that  $u_t = 2$  and  $u_{nt} = 1$ , we obtain

$$R = 3/2 + 1/(2N) \approx 3/2. \quad (8.1)$$

It is arguable whether teachers are more innovative than non-teachers. One may be a teacher by virtue of a large fund of knowledge, which may facilitate innovation. Similarly, the teacher may have many contacts outside the population and hence be more aware of externally occurring innovations. On the other hand, a non-teacher may be more willing to experiment with new techniques and styles. If all individuals are equally likely to innovate, then substituting  $u_t = u_{nt} = 1$  yields  $R = 1$ , i.e., the OtM model predicts the same cultural evolutionary rate as the oblique transmission model. Thus, evidence for teaching such as has been found at the Magdalenian archaeological site of Etioilles (Pigeot, 1990) cannot necessarily be invoked to explain the rapid turnover of lithic traditions during the Upper Paleolithic.

Let us now turn to the question of the amount of cultural variation at equilibrium when there is recurrent innovation and imperfect copying of each trait. The various social transmission models we investigated produce remarkably different predictions. For example, the cultural heterogeneity or trait diversity,  $1 -$

$\varphi$ , expected under MtO (conformist) transmission is appreciably greater than with oblique (random copying) transmission. This is because a moderately but not negligibly small number of traits accumulate in the equilibrium population (middle panel of first column of Fig. 1), but since each exists at a low frequency (top panel), they are unlikely to be copied. The result is a decrease in  $\varphi$  (bottom panel), or equivalently an increase in the trait diversity,  $1 - \varphi$ .

This prediction would appear to contradict prevailing theory in archaeology, where the intuitive claim is often made that MtO transmission will produce little cultural variation among individuals (Lycett and Gowlett, 2008; MacDonald, 1998). If by little variation among individuals it is meant that artifacts produced by different individuals tend to share the same traits, then this claim requires revision, as we have shown. Lehmann et al. (2011) discuss in more detail why intuition should fail in this case. On the other hand, Fig. 1 (middle panel of first column) shows that the mean number of traits segregating in the population,  $\lambda_p$ , will be smaller than with oblique (random copying) transmission. Hence, MtO transmission should result in reduced intra-site variability, in the sense that fewer cultural traits (e.g., identifiable attributes of stone tools) are present and variable at one location.

By contrast, the Bok (pro-novelty) transmission yields large values of  $\lambda_f$ ,  $\lambda_p$ , and  $\varphi$  (second column of Fig. 1) relative to oblique transmission. Hence, the trait diversity,  $1 - \varphi$ , will be low. This model of social transmission is similar to sensitivity to minority studied by Lehmann et al. (2011). It also effectively constitutes a stochastic version of the model suggested by Enquist et al. (2010), in which each individual has more than one ( $k$ ) cultural parent and traits are copied with probability  $\beta$  (see their Eq. (5)). Fig. 2 shows the results for  $k = 2$  (first column) and  $k = 5$  (second column) and with  $\beta = 0.9, 0.7, 0.5$  (top to bottom curves). The predictions of the two models are consistent; in particular the cultural homogeneity remains substantial for  $\beta \geq 0.5$  and is very high for  $\beta = 0.9$ .

The other transmission scheme that we consider, OtM (teacher), represents the maximum possible level of cultural inbreeding (Cavalli-Sforza and Feldman, 1981), giving here an effective size of approximately 2.26, which is independent of population size (see Appendix G for the derivation). In other words, one and the same individual, i.e., the teacher, acts as the cultural parent for all individuals that are born during its lifetime. We obtain the intuitively reasonable result that a substantial level of cultural homogeneity is maintained for each of the three values of  $\beta$  (bottom panel of third column of Fig. 1). Thus, our teacher model (Eqs. (F.1)–(F.16) of Appendix F) does not require multiple cultural parents in order for cultural homogeneity to evolve. On the other hand, when just one cultural parent is chosen *at random* as in our oblique transmission model, cultural homogeneity will be low in a large population at equilibrium (Eq. (6.5)), which agrees with Enquist et al. (2010).

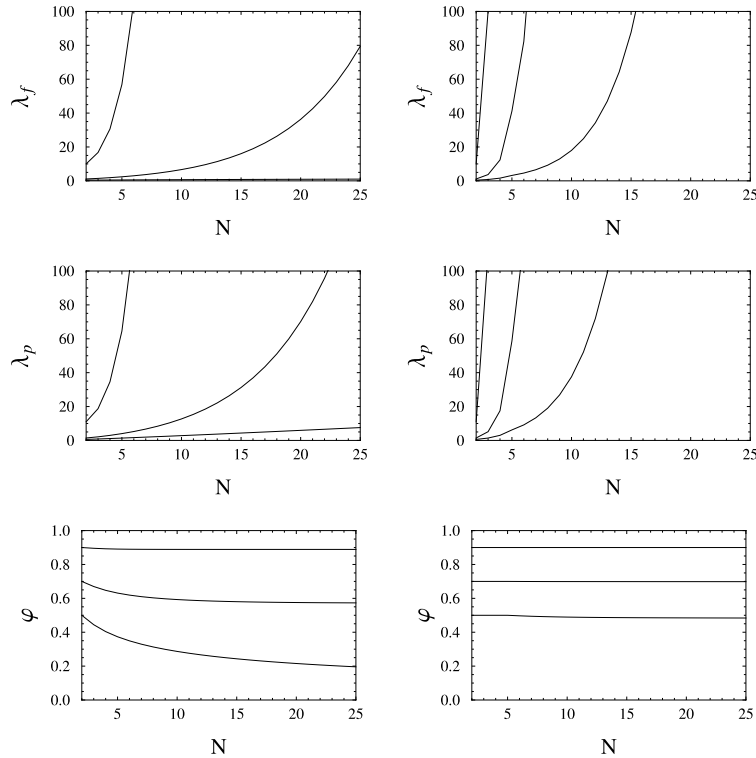
## Acknowledgments

We thank H. Terashima and Y. Nishiaki for information, discussion, and references, and the referees for suggestions. M. Seki and S. Kurokawa pointed out that  $\bar{t}_{1A}$  in the one-to-many transmission model and  $\bar{t}_1$  in the oblique transmission model are equivalent. This research is supported in part by Monbukagakusho grants 17102002 and 22101004 to KA and NIH grant GM28016 to MWF.

## Appendix A

The fixation probability and mean times for a finite birth–death chain with two absorbing states ( $i = 0$  and  $i = N$ ) and with





**Fig. 2.** Measures of variation for a selectively neutral trait in the best-of- $k$  transmission model. In the first column of panels,  $\lambda_f$ ,  $\lambda_p$ , and  $\varphi$  are graphed as functions of population size  $N$  for the case of  $U = 0.1$  and  $k = 2$ . The top to bottom curves correspond to transmission rates  $\beta = 0.9, 0.7$ , and  $0.5$ , respectively. The second column of panels is the same as the first except that  $k = 5$ .

positive probabilities of transition from each of the transient states ( $p_i > 0$  and  $q_i > 0$  for  $1 \leq i \leq N-1$ ).  
Let

$$\rho_0 = 1, \quad \rho_l = \prod_{j=1}^l (q_j/p_j), \quad (\text{A.1})$$

and assume there are initially  $i$  type  $A$  individuals. Then the fixation probability of type  $A$  is

$$\pi_i = \sum_{l=0}^{i-1} \rho_l / \sum_{l=0}^{N-1} \rho_l. \quad (\text{A.2})$$

In terms of the number of birth–death events, the mean sojourn time in state  $j$  is

$$\bar{t}_{ij} = \sum_{l=i}^{N-1} \rho_l \sum_{l=0}^{j-1} \rho_l / \left( \rho_{j-1} q_j \sum_{l=0}^{N-1} \rho_l \right) \quad (\text{A.3})$$

if  $j \leq i$  and

$$\bar{t}_{ij} = \sum_{l=0}^{i-1} \rho_l \sum_{l=j}^{N-1} \rho_l / \left( \rho_j p_j \sum_{l=0}^{N-1} \rho_l \right) \quad (\text{A.4})$$

if  $j \geq i$ . The mean sojourn time in state  $j$  conditional on fixation (averaging over the cases where type  $A$  is fixed) is

$$\bar{t}_{ij}^* = \bar{t}_{ij} \pi_j / \pi_i, \quad (\text{A.5})$$

and similarly the mean sojourn time in state  $j$  conditional on loss (averaging over the cases where type  $A$  is lost) is

$$\bar{t}_{ij}^{**} = \bar{t}_{ij} (1 - \pi_j) / (1 - \pi_i). \quad (\text{A.6})$$

Finally, the mean fixation time, mean loss time, and mean absorption time are, respectively,

$$\bar{t}_i^* = \sum_{j=1}^{N-1} \bar{t}_{ij}^*, \bar{t}_i^{**} = \sum_{j=1}^{N-1} \bar{t}_{ij}^{**}, \bar{t}_i = \sum_{j=1}^{N-1} \bar{t}_{ij} = \pi_i \bar{t}_i^* + (1 - \pi_i) \bar{t}_i^{**}. \quad (\text{A.7})$$

## Appendix B

Equality of  $\bar{t}_1^*$  with positive and negative selection in the oblique transmission model.

Applying the formulas of Appendix A with  $\rho_0 = 1$  and  $\rho_l = (1/v)^l$ , we can write the mean sojourn time at  $j \geq i$  conditional on fixation as

$$\bar{t}_{ij}^*(v) = \left\{ \left[ 1 - (1/v)^j \right] \left[ 1 - (1/v)^{N-j} \right] \right\} / \left\{ \left[ 1 - (1/v) \right] \right. \\ \left. \times \left[ 1 - (1/v)^N \right] N \left[ j + (N-j)v \right] / \left[ j(N-j)v \right] \right\}, \quad (\text{B.1})$$

which does not depend on  $i$ . When selection is reversed – mortality of type  $B$  is  $1/v$  times that of type  $A$  – we have

$$\bar{t}_{ij}^*(1/v) = \left\{ \left( 1 - v^j \right) \left( 1 - v^{N-j} \right) \right\} / \left\{ \left( 1 - v \right) \left( 1 - v^N \right) \right\} \\ \times N \left[ vj + (N-j) \right] / \left[ j(N-j) \right] \\ = \left\{ \left[ 1 - (1/v)^j \right] \left[ 1 - (1/v)^{N-j} \right] \right\} / \left\{ \left[ 1 - (1/v) \right] \right. \\ \left. \times \left[ 1 - (1/v)^N \right] N \left[ vj + (N-j) \right] / \left[ j(N-j)v \right] \right\}, \quad (\text{B.2})$$

which differs subtly from Eq. (B.1). Hence, the equality of mean sojourn times for positive and negative genic selection as predicted by diffusion theory (Maruyama, 1974) does not hold in the Moran model.

Nevertheless, substituting the dummy variable  $l = N - j$  and summing over the transient states gives

$$\bar{t}_1^*(1/v) = \sum_{j=1}^{N-1} \bar{t}_{1j}^*(1/v) \\ = \sum_{l=1}^{N-1} \left\{ \left[ 1 - (1/v)^{N-l} \right] \left[ 1 - (1/v)^l \right] \right\} / \left\{ \left[ 1 - (1/v) \right] \right. \\ \left. \times \left[ 1 - (1/v)^N \right] N \left[ v(N-l) + l \right] / \left[ (N-l)lv \right] \right\} \\ = \sum_{j=1}^{N-1} \bar{t}_{1j}^*(v). \quad (\text{B.3})$$

## Appendix C

Derivation of mean times in the best-of- $k$  transmission model.

It is convenient to view the state space as comprising two overlapping subspaces,  $S_1 = \{0, 1, \dots, N-k, N-k+1\}$  and  $S_2 = \{N-k+1, N-k+2, \dots, N-1, N\}$ . Eqs. (4.1)–(4.4) entail that  $S_1$  has two absorbing states, 0 and  $N-k+1$ , either of which can be reached from any of the remaining,  $N-k$  transient, states within  $S_1$ . On the other hand, only up-transitions are possible in  $S_2$ , and state  $N$  will eventually be reached (fixation of type  $A$  will occur).

Let the initial state  $i$  lie between 1 and  $N-k$ . For  $j$  such that  $1 \leq j \leq N-k$ , the mean sojourn time at state  $j$  conditional on fixation of type  $A$  can be computed from Eqs. (A.3)–(A.5) with  $N-1$  replaced by  $N-k$ . Similarly, the mean sojourn time conditional on loss can be obtained from Eqs. (A.3), (A.4) and (A.6). When  $N-k+1 \leq j \leq N-1$ , the mean sojourn time conditional on fixation is equal to the mean waiting time until an up-transition, since down-transitions are impossible. Hence,

$$\bar{t}_{ij}^* = p_j^{-1}. \quad (\text{C.1})$$

On the other hand, given that type  $A$  is eventually lost, we clearly have

$$\bar{t}_{ij}^{**} = 0 \quad (\text{C.2})$$

when  $N-k+1 \leq j \leq N-1$ .

Thus, the mean fixation time from state  $i \leq N-k$  has the form

$$\begin{aligned} \bar{t}_i^* &= \sum_{j=1}^i \left[ \pi_j \sum_{l=0}^{j-1} \rho_l / (\rho_{j-1} q_j) \right] \cdot \sum_{l=i}^{N-k} \rho_l / \left( \pi_i \sum_{l=0}^{N-k} \rho_l \right) \\ &\quad + \sum_{j=i+1}^{N-k} \left[ \pi_j \sum_{l=j}^{N-k} \rho_l / (\rho_j p_j) \right] \\ &\quad \times \sum_{l=0}^{i-1} \rho_l / \left( \pi_i \sum_{l=0}^{N-k} \rho_l \right) + \sum_{j=N-k+1}^{N-1} p_j^{-1}, \end{aligned} \quad (\text{C.3})$$

and similarly the mean loss time can be expressed as

$$\begin{aligned} \bar{t}_i^{**} &= \sum_{j=1}^i \left[ (1 - \pi_j) \sum_{l=0}^{j-1} \rho_l / (\rho_{j-1} q_j) \right] \\ &\quad \times \sum_{l=i}^{N-k} \rho_l / \left[ (1 - \pi_i) \sum_{l=0}^{N-k} \rho_l \right] \\ &\quad + \sum_{j=i+1}^{N-k} \left[ (1 - \pi_j) \sum_{l=j}^{N-k} \rho_l / (\rho_j p_j) \right] \\ &\quad \times \sum_{l=0}^{i-1} \rho_l / \left[ (1 - \pi_i) \sum_{l=0}^{N-k} \rho_l \right]. \end{aligned} \quad (\text{C.4})$$

The mean absorption time can then be obtained from Eq. (A.7).

## Appendix D

Recursions satisfied by the fixation probability of type  $A$  in the one-to-many transmission model.

From Gale (1990, p. 155), the fixation probabilities satisfy the recursions

$$\pi_{iA} = [(N-i)/N] \pi_{(i+1)A} + [(i-1)/N + i/N^2] \pi_{iA} + [(N-i)/N^2] \pi_{iB}, \quad (\text{D.1})$$

$$\pi_{iB} = (i/N^2) \pi_{iA} + (i/N) \pi_{(i-1)B} + [(N-i-1)/N + (N-i)/N^2] \pi_{iB} \quad (\text{D.2})$$

for  $1 \leq i \leq N-1$ , where we have used Eqs. (5.1)–(5.6). The boundary conditions are  $\pi_{0B} = 0$  and  $\pi_{NA} = 1$ . Truth of Eqs. (5.7)–(5.8) can be confirmed by direct substitution into Eqs. (D.1)–(D.2).

## Appendix E

Mean sojourn times in the one-to-many transmission model.

Order the transient states  $1B, 1A, 2B, 2A, \dots, (N-1)B, (N-1)A$ . Let  $\mathbf{P} = \{p_{i\alpha, j\beta}\}$  be the  $2(N-1) \times 2(N-1)$  submatrix of transition probabilities between the transient states (see Eqs. (5.1)–(5.6)). Also, let  $\mathbf{I}$  be the  $2(N-1) \times 2(N-1)$  identity matrix, and let  $\mathbf{T} = \{\bar{t}_{i\alpha, j\beta}\}$  be the  $2(N-1) \times 2(N-1)$  matrix of mean sojourn times (in numbers of birth–death events).

State  $1B$  must be visited when type  $A$  is ultimately lost. Similarly, state  $(N-1)A$  must be visited when type  $A$  is ultimately fixed. Hence,

$$\bar{t}_{i\alpha, 1B} = \bar{t}_{1B, 1B} \frac{1 - \pi_{i\alpha}}{1 - \pi_{1B}}, \quad (\text{E.1})$$

$$\bar{t}_{i\alpha, (N-1)A} = \bar{t}_{(N-1)A, (N-1)A} \frac{\pi_{i\alpha}}{\pi_{(N-1)A}}, \quad (\text{E.2})$$

where  $\pi_{i\alpha}$  is the fixation probability from state  $i\alpha$  (see Eqs. (5.7)–(5.8)) (Gale, 1990, p. 90, p. 214).

We have  $(\mathbf{I} - \mathbf{P})\mathbf{T} = \mathbf{I}$  (Gale, 1990, p. 217). Equating the  $1B, 1B$  elements on both sides yields

$$\frac{N+1}{N^2} \bar{t}_{1B, 1B} - \frac{1}{N^2} \bar{t}_{1A, 1B} = 1.$$

Then, using Eq. (E.1) gives

$$\frac{N+1}{N^2} \bar{t}_{1B, 1B} - \frac{1}{N^2} \cdot \frac{N-1}{2N-1} \bar{t}_{1B, 1B} = 1.$$

Hence,  $\bar{t}_{1B, 1B} = (2N-1)/2$  and using Eq. (E.1) again gives

$$\bar{t}_{iA, 1B} = (N-i)/2, \quad (\text{E.3})$$

$$\bar{t}_{iB, 1B} = (2N-i)/2. \quad (\text{E.4})$$

By symmetry,  $\bar{t}_{(N-1)A, (N-1)A} = \bar{t}_{1B, 1B} = (2N-1)/2$  and hence

$$\bar{t}_{iA, (N-1)A} = (N+i)/2,$$

$$\bar{t}_{iB, (N-1)A} = i/2.$$

Thus, the first and last columns of matrix  $\mathbf{T}$  are readily obtained.

The other elements can be obtained by reversing the order of matrix multiplication—from  $\mathbf{T}(\mathbf{I} - \mathbf{P}) = \mathbf{I}$  (Gale, 1990, p. 217). The mean sojourn times can be defined recursively by this method, but closed formulas are not generally forthcoming. In what follows, we first derive recursions in  $\bar{t}_{1B, iB}$  and  $\bar{t}_{1B, iA}$  and then in  $\bar{t}_{iA, iB}$  and  $\bar{t}_{iA, iA}$ . Using the former permits us to evaluate the mean absorption time and the mean fixation time from initial state  $1B$ , and the latter gives us the corresponding values for state  $1A$ .

The elements of the first row of  $\mathbf{T}(\mathbf{I} - \mathbf{P}) = \mathbf{I}$  are

$$\bar{t}_{1B, 1B} \frac{N+1}{N^2} + \bar{t}_{1B, 1A} \left( -\frac{N-1}{N^2} \right) + \bar{t}_{1B, 2B} \left( -\frac{2}{N} \right) = 1$$

$$\bar{t}_{1B, 1B} \left( -\frac{1}{N^2} \right) + \bar{t}_{1B, 1A} \frac{(N-1)(N+1)}{N^2} = 0$$

$$\bar{t}_{1B, iB} \frac{i(N+1)}{N^2} + \bar{t}_{1B, iA} \left( -\frac{N-i}{N^2} \right) + \bar{t}_{1B, (i+1)B} \left( -\frac{i+1}{N} \right) = 0 \quad (\text{E.5})$$

$$\begin{aligned} \bar{t}_{1B, (i-1)A} \left( -\frac{N-i+1}{N^2} \right) + \bar{t}_{1B, iB} \left( -\frac{i}{N} \right) \\ + \bar{t}_{1B, iA} \frac{(N-i)(N+1)}{N^2} = 0 \end{aligned} \quad (\text{E.6})$$

$$\bar{t}_{1B, (N-1)B} \frac{(N-1)(N+1)}{N^2} + \bar{t}_{1B, (N-1)A} \left( -\frac{1}{N^2} \right) = 0$$

$$\bar{t}_{1B, (N-2)A} \left( -\frac{2}{N^2} \right) + \bar{t}_{1B, (N-1)B} \left( -\frac{N-1}{N} \right)$$

$$+ \bar{t}_{1B, (N-1)A} \frac{N+1}{N^2} = 0.$$

Eq. (E.5) holds for  $2 \leq i \leq N-2$  whereas Eq. (E.6) holds for  $2 \leq i \leq N-1$ .

Rearranging Eq. (E.5) and renumbering ( $i + 1 \rightarrow i$ ) yields

$$\bar{t}_{1B,iB} = \bar{t}_{1B,(i-1)B} \frac{(N+1)(i-1)}{Ni} - \bar{t}_{1B,(i-1)A} \frac{N-i+1}{Ni} \quad (\text{E.7})$$

valid for  $3 \leq i \leq N-1$ . Using Eq. (E.7), we rewrite Eq. (E.6) as

$$\bar{t}_{1B,iA} = \bar{t}_{1B,(i-1)B} \frac{i-1}{N(N-i)} + \bar{t}_{1B,(i-1)A} \frac{(N-1)(N-i+1)}{N(N-i)} \quad (\text{E.8})$$

for  $2 \leq i \leq N-1$ .

Eqs. (E.7)–(E.8), together with  $\bar{t}_{1B,1B} = (2N-1)/2$ ,  $\bar{t}_{1B,1A} = (2N-1)/[2(N-1)(N+1)]$ ,  $\bar{t}_{1B,2B} = (N-2)/4(N+1)$ , and  $\bar{t}_{1B,2A} = (N-1)/[(N-2)(N+1)]$ , are sufficient to compute the mean absorption time from state 1B

$$\bar{t}_{1B} = \sum_{i=1}^{N-1} (\bar{t}_{1B,iB} + \bar{t}_{1B,iA}). \quad (\text{E.9})$$

When the initial state is 1A, we equate the elements of the second row on both sides of  $\bar{T}(\mathbf{I} - \mathbf{P}) = \mathbf{I}$ . Then, we obtain on repeating the above analysis,

$$\bar{t}_{1A,iB} = \bar{t}_{1A,(i-1)B} \frac{(N+1)(i-1)}{Ni} - \bar{t}_{1A,(i-1)A} \frac{N-i+1}{Ni}, \quad (\text{E.10})$$

$$\bar{t}_{1A,iA} = \bar{t}_{1A,(i-1)B} \frac{i-1}{N(N-i)} + \bar{t}_{1A,(i-1)A} \frac{(N-1)(N-i+1)}{N(N-i)}, \quad (\text{E.11})$$

which are valid for  $2 \leq i \leq N-1$ . The coefficients in Eqs. (E.10) and (E.11) are identical to those in Eqs. (E.7) and (E.8), respectively. The mean absorption time can be computed from

$$\bar{t}_{1A} = \sum_{i=1}^{N-1} (\bar{t}_{1A,iB} + \bar{t}_{1A,iA}) \quad (\text{E.12})$$

where  $\bar{t}_{1A,1B} = (N-1)/2$  and  $\bar{t}_{1A,1A} = (2N-1)/2(N-1)$ .

## Appendix F

Positive transition probabilities,  $\tilde{p}_{i\alpha j\beta}$ , for the one-to-many transmission model, incorporating recurrent innovation (with probability  $u$ ) and imperfect copying (with probability  $\beta$ ).

$$\tilde{p}_{iB,(i-1)B} = (1-u)i/N, \quad (\text{F.1})$$

$$\tilde{p}_{iB,iB} = ui/N + (1-u)[(N-i-1)/N + (N-i)/N^2], \quad (\text{F.2})$$

$$\tilde{p}_{iB,(i+1)B} = u[(N-i-1)/N + (N-i-1)/N^2], \quad (\text{F.3})$$

$$\tilde{p}_{iB,iA} = (1-u)i/N^2, \quad (\text{F.4})$$

$$\tilde{p}_{iB,(i+1)A} = u(i+1)/N^2, \quad (\text{F.5})$$

$$\tilde{p}_{iA,(i-1)B} = (1-u)(1-\beta)(N-i+1)/N^2, \quad (\text{F.6})$$

$$\tilde{p}_{iA,iB} = [u + (1-u)\beta](N-i)/N^2, \quad (\text{F.7})$$

$$\tilde{p}_{iA,(i-1)A} = (1-u)(1-\beta)[(i-1)/N + (i-1)/N^2], \quad (\text{F.8})$$

$$\tilde{p}_{iA,iA} = [u + (1-u)\beta] \times [(i-1)/N + i/N^2] + (1-u)(1-\beta)(N-i)/N, \quad (\text{F.9})$$

$$\tilde{p}_{iA,(i+1)A} = [u + (1-u)\beta](N-i)/N, \quad (\text{F.10})$$

for  $1 \leq i \leq N-1$ . From the terminal states, 0B and NA, we have

$$\tilde{p}_{0B,0B} = 1-u, \quad (\text{F.11})$$

$$\tilde{p}_{0B,1B} = u(N-1)[1/N + 1/N^2], \quad (\text{F.12})$$

$$\tilde{p}_{0B,1A} = u/N^2, \quad (\text{F.13})$$

$$\tilde{p}_{NA,NA} = u + (1-u)\beta, \quad (\text{F.14})$$

$$\tilde{p}_{NA,(N-1)A} = (1-u)(1-\beta)(N-1)[1/N + 1/N^2], \quad (\text{F.15})$$

$$\tilde{p}_{NA,(N-1)B} = (1-u)(1-\beta)/N^2. \quad (\text{F.16})$$

## Appendix G

Largest non-unit eigenvalue of the one-to-many transmission model

In order to find the rate of approach to either fixation ( $u = 0$ ,  $\beta = 1$ ) or the stationary distribution ( $0 < u < 1$ ,  $0 < \beta < 1$ ), we examined the numerically obtained largest non-unit eigenvalue of the transition matrix specified by Eqs. (F.1)–(F.16). For each of a large number of values of  $N$ , we obtained this eigenvalue as a function of  $u$  and  $\beta$  using Mathematica. The inductively determined eigenvalue has the form

$$\lambda^* = 1 - \frac{2}{N} - \frac{C}{N^2} + \left\{ \frac{2N(N-1)(1-C) + C^2}{N^4} \right\}^{1/2}, \quad (\text{G.1})$$

where

$$C = u + (1-u)(1-\beta). \quad (\text{G.2})$$

In particular, when  $u = 0$  and  $\beta = 1$  this eigenvalue reduces to

$$\lambda_0^* = 1 - \frac{2}{N} + \sqrt{\frac{2(N-1)}{N^3}}. \quad (\text{G.3})$$

From Eq. (G.3), we may estimate the corresponding effective population size,  $N_e$ , as follows. Clearly, after  $N$  birth–death events the cultural variation declines to  $(\lambda_0^*)^N$  of its previous value (Gale, 1990, p. 49). Then, because  $N$  birth–death events in this Moran-type model are analogous to one generation in the Wright–Fisher model, we can set  $1 - \frac{1}{N_e} = (\lambda_0^*)^N$  from which we obtain  $N_e = [1 - (\lambda_0^*)^N]^{-1}$ . Note that we cannot use the standard (approximate) formula given by Felsenstein (1971), since  $1 - \lambda_0^*$  is of order of magnitude  $1/N$ . Hence, as  $N$  becomes large we see that for this one-to-many (teacher) transmission model the dependence on  $N$  disappears and

$$N_e \rightarrow \exp(2 - \sqrt{2}) / [\exp(2 - \sqrt{2}) - 1] \approx 2.26, \quad (\text{G.4})$$

which we can interpret as extreme cultural inbreeding—the cultural variation is dictated by the equivalent of only 2.26 individuals, a small number indeed.

## References

- Bentley, R.A., Hahn, M.W., Shennan, S.J., 2004. Random drift and culture change. *Proc. R. Soc. Lond. Ser. B* 271, 1443–1450.
- Bentley, R.A., Lipo, C.P., Herzog, H.A., Hahn, M.W., 2007. Regular rates of popular culture change reflect random copying. *Evol. Hum. Behav.* 28, 151–158.
- Boyd, R., Richerson, P.J., 1985. *Culture and the Evolutionary Process*. Univ. Chicago Press, Chicago, IL.
- Brown, M.J., Feldman, M.W., 2009. Sociocultural epistasis and cultural exaptation in footbinding, marriage form, and religious practices in early 20th century Taiwan. *Proc. Natl. Acad. Sci. USA* 106, 22139–22144.
- Cavalli-Sforza, L.L., Feldman, M.W., 1981. *Cultural Transmission and Evolution*. Princeton Univ. Press, Princeton, NJ.
- Dunbar, R.I.M., 1992. Neocortex size as a constraint on group size in primates. *J. Hum. Evol.* 20, 469–493.
- Eerkens, J.W., Lipo, C.P., 2007. Cultural transmission theory and the archaeological record: providing context to understanding variation and temporal changes in material culture. *J. Archaeol. Res.* 15, 239–274.
- Enquist, M., Strimling, P., Eriksson, K., Laland, K., Sjostrand, J., 2010. One cultural parent makes no culture. *Anim. Behav.* 79, 1353–1362.
- Ewens, W.J., 2004. *Mathematical Population Genetics*, 2nd ed. Springer, New York.
- Felsenstein, J., 1971. Inbreeding and variance effective numbers in populations with overlapping generations. *Genetics* 68, 581–597.
- Fisher, R.A., 1930. *The Genetic Theory of Natural Selection*. Clarendon, Oxford, UK.
- Gale, J.S., 1990. *Theoretical Population Genetics*. Unwin Hyman, London, UK.
- Guglielmino, C.R., Viganotti, C., Hewlett, B., Cavalli-Sforza, L.L., 1995. Cultural variation in Africa: role of mechanisms of transmission and adaptation. *Proc. Natl. Acad. Sci. USA* 92, 7585–7589.
- Henrich, J., 2004. Demography and cultural evolution: how adaptive cultural processes can produce maladaptive losses—the Tasmanian case. *Am. Antiq.* 69, 197–214.

- Hewlett, B.S., Cavalli-Sforza, L.L., 1986. Cultural transmission among Aka pygmies. *Am. Anthropol.* 88, 922–934.
- Ichikawa, M., 1978. The residential groups of the Mbuti pygmies. *Senri Ethnol. Stud.* 1, 131–188.
- Kelly, R.L., 1995. *The Foraging Spectrum*. Smithsonian Institution Press, Washington, DC.
- Kimura, M., 1969. The number of heterozygous nucleotide sites maintained in a finite population due to steady flux of mutations. *Genetics* 61, 893–903.
- Kimura, M., Ohta, T., 1969. The average number of generations until fixation of a mutant gene in a finite population. *Genetics* 61, 763–771.
- Lee, R.B., 1979. *The !Kung San*. Cambridge Univ Press, Cambridge, UK.
- Lehmann, L., Aoki, K., Feldman, M.W., 2011. On the number of independent cultural traits carried by individuals and populations. *Philos. Trans. R. Soc. Lond. Ser. B* 366, 424–435.
- Lumsden, C., Wilson, E.O., 1981. *Genes, Mind, and Culture*. Harvard Univ. Press, Cambridge, MA.
- Lycett, S.J., Gowlett, A.J., 2008. On questions surrounding the Acheulean tradition. *World Archaeol.* 40, 295–315.
- MacDonald, D.H., 1998. Subsistence, sex, and cultural transmission in Folsom culture. *J. Anthropol. Archaeol.* 17, 217–239.
- Maruyama, T., 1974. The age of an allele in a finite population. *Genet. Res. Camb.* 23, 137–143.
- Moran, P.A.P., 1958. Random processes in genetics. *Proc. Cambridge Philos. Soc.* 54, 60–71.
- Pigeot, N., 1990. Technical and social actors flintknapping specialists and apprentices at Magdalenian Etioilles. *Archaeol. Rev. Camb.* 9, 126–141.
- Rogers, E.M., 1995. *Diffusion of Innovations*, 4th ed. Free Press, New York.
- Rogers, D.S., Ehrlich, P.R., 2008. Natural selection and cultural rates of change. *Proc. Natl. Acad. Sci. USA* 105, 3416–3420.
- Rogers, D.S., Feldman, M.W., Ehrlich, P.R., 2009. Inferring population histories using cultural data. *Proc. R. Soc. Lond. Ser. B* 276, 3835–3843.
- Ryan, B., Gross, N.C., 1943. The diffusion of hybrid seed corn in two Iowa communities. *Rural Sociol.* 8, 15–24.
- Shennan, S., 2001. Demography and cultural innovation: a model and its implications for the emergence of modern human culture. *Camb. Archaeol. J.* 11, 5–16.
- Terashima, H., 1985. Variation and composition principles of the residence group (band) of the Mbuti pygmies—beyond a typical/atypical dichotomy. *African Study Monogr. (Suppl. 4)*, 103–120.
- Wright, S., 1931. Evolution in Mendelian populations. *Genetics* 16, 97–159.