



Accumulation of independent cultural traits

Pontus Strimling^{*,1}, Jonas Sjöstrand¹, Magnus Enquist¹, Kimmo Eriksson¹

Centre for the Study of Cultural Evolution, Stockholm University, SE-106 91, Stockholm, Sweden

ARTICLE INFO

Article history:

Received 20 January 2009
Available online 8 May 2009

Keywords:

Cultural evolution
Cultural accumulation
Bird song
Moran model

ABSTRACT

In a species capable of (imperfect) social learning, how much culture can a population of a given size carry? And what is the relationship between the individual and the population? In the first study of these novel questions, here we develop a mathematical model of the accumulation of independent cultural traits in a finite population with overlapping generations.

© 2009 Elsevier Inc. All rights reserved.

1. Introduction

In a species capable of invention and social learning, populations will accumulate cultural traits. Examples include culturally acquired behaviors in chimpanzee populations and song elements in bird populations. The most prominent example is, of course, the massive accumulation of culture in historical and modern human populations. An essential feature of human culture in large populations is that every single individual carries a lot of cultural elements but still, only a fraction of the total amount in the population.

It is possible to measure the amount of culture (suitably defined) in a population. For instance, biologists have counted the number of bird song elements in different populations (Lynch, 1996), and linguists study the size of vocabularies in human languages. However, there seem to have been little attempts at developing a theory of cultural accumulation, and none that takes individuals' share of the total amount of culture into account. This is our aim in the present paper. The key assumption that makes our model amenable to analysis is that traits are *independent* of each other, i.e., the presence of one trait does not affect the probability of acquiring another trait. Although it is obvious that there exist many interesting dependencies within human culture, we believe that making an assumption of independence is a very reasonable first step in the development of a theory of how culture accumulates.

Cultural accumulation is likely to have been a crucial component of human ecological success, as a very large number of different inventions are needed for the survival and sustenance of billions of humans across a great variety of environments on Earth

(Richerson and Boyd, 2005). However, every single cultural element may typically have only marginal importance, and in this paper we are not interested in effects on reproductive fitness. In other words, we assume traits to be neutral in terms of biological evolution—but not neutral in terms of cultural evolution. Indeed, thanks to the independence assumption, our model can accommodate variation between cultural traits in how easily they are invented and socially transmitted.

To the best of our knowledge, the only previous models of cultural accumulation are by Enquist and Ghirlanda (2007) and Enquist et al. (2008). These models are macroscopic and deterministic, i.e., they work at the population level and ignore individuals and random processes. The basic findings in these models are:

- (1) *If the innovation rate is constant and a constant proportion of culture is lost, then the amount of culture carried by the population will reach an equilibrium.* Well-known examples of domains where the number of variants seems to have reached equilibria include words in use in languages, or baby names in use in a country.
- (2) *If the innovation rate grows with the amount of culture, then exponential growth is possible.* Several cultural domains, such as operas or mathematical theorems, seem to show exponential growth (Lehman, 1947; Enquist et al., 2008).

Our model offers the first microscopic approach to the first case, where the accumulation process reaches an equilibrium. Since we are interested in the effects of population size our model is based on a finite population, and hence the evolutionary process includes both innovation and drift. This is reminiscent of the neutral theory of genetic evolution (see e.g. Kimura and Crow, 1964; Futuyma, 1997) where genetic drift and mutations contribute to genetic diversity and genetic change. Similar random processes have been

* Corresponding author.

E-mail address: pontusstrimling@gmail.com (P. Strimling).

¹ All authors contributed equally to this work.

studied previously also in the context of cultural evolution, e.g. of bird song (Lynch et al., 1989; Lynch and Baker, 1993; Lynch, 1996). In many birds the young males learn their songs from their fathers or from neighboring males, and new songs are also regularly formed (Catchpole and Slater, 1995). Lynch and coworkers used a ‘haploid version’ of the genetic neutral model and found that the distribution of song types in wild bird populations is consistent with neutral cultural evolution. In another line of research, R.A. Bentley and coworkers have studied a number of human cultural traits that may be (roughly) neutral, like baby names, dog breeds and pottery designs (Hahn and Bentley, 2003; Herzog et al., 2004; Bentley et al., 2004). In these studies, empirical data of the frequency distribution of variants are shown to be consistent with a random copying model, where individuals obtain their cultural variants either from randomly drawn cultural parents or rare innovations. (This is mathematically equivalent to the infinite alleles version of the Wright–Fisher model in population genetics; in contrast, the model we present is a generalization of the infinite alleles version of the Moran model, cf. Ewens (2004).)

Both Bentley’s and Lynch’s studies illustrate that cultural neutral theory, in contrast with the genetic theory, is concerned with expressed traits or with the phenotype. Another difference that is not exploited in the above studies is that, while diploid organisms carry at most two alleles at an autosomal locus, a single individual may carry arbitrarily many cultural traits. For instance, in the case of bird song a more refined model would allow individuals to learn any number of song elements rather than one complete song. Learning of any number of elements can be described as individual accumulation of culture. The theory of cultural accumulation that we develop here would then predict both the total number of song elements in the populations and the average number of song elements per individual.

The paper is organized as follows. Section 2 describes our model in detail, and the main results of our mathematical analysis. In Section 3 we use these results to analyze the effects of various parameters on cultural accumulation, both on population level and individual level. We can also identify a shortcoming in the macroscopic accumulation models of Enquist and Ghirlanda, and we discuss its consequences. Section 4 deals with non-neutral evolution. Finally, in Section 5 we discuss what we have learned from our analysis. The Appendix contains the detailed mathematical analysis of the model and a comparison between simulations when generations are taken to be either discrete or overlapping.

2. Cultural accumulation models

We will make the standard discreteness assumptions within cultural evolution modeling. To begin with, cultural traits are distinct entities, so social learners either obtain the trait or not. Further, we will assume traits in our models to be fitness-neutral, i.e., they incur no reproductive advantages for individuals who carry them. We will however allow for a cultural filter that weeds out traits that people do not like, for instance traits that do not fulfill their functions or are complicated to learn.

We assume that cultural transmission is never certain by introducing a , the *social learning efficiency*, which is simply the success rate of transmission. In other words, with probability $1 - a$ an individual fails in obtaining the cultural trait of the parent, yielding a loss of culture due to imperfect transmission.

In order to study cultural accumulation in individuals, we will assume traits that *coexist independently* of each other. For instance, whether an individual has a word A in her vocabulary is assumed not to interact with her capacity of learning word B. Thus, individuals may carry many traits, and therefore both society and the individual will accumulate culture. However, as we will discuss shortly, our model can also be given an interpretation in terms

of mutually exclusive variants. Of course, in reality many cultural traits will be neither perfectly independent nor perfectly mutually exclusive. The effects of linkage of traits form an important topic for future research, towards which the present model is a stepping stone.

Finally our model assumes overlapping generations, where only one individual leaves the population in each round. In the Appendix we relax this assumption to test the robustness of our model.

2.1. The random copying model with accumulation and overlapping generations

We will refer to our model as the *cultural accumulation model* (CAM). It assumes a population of N individuals, each of whom is carrying an individual set of traits. The following happens in each timestep.

- Phase 1: A random individual leaves the population.
- Phase 2: A new individual enters the population and chooses one random individual as her cultural parent. For each trait of the cultural parent, the new individual independently obtains it with probability a , the learning efficiency.
- Phase 3: The new individual invents a random number of traits with expected value μ . (This random number does not have to be independent of phase 2.)

2.2. Mathematical analysis

The outcome of our model of cultural evolution is that cultural traits are invented, possibly spread, finally to be replaced by new traits. At any given time, a certain trait v will be carried by a certain number of individuals, which number we call the *popularity* of v . In the Appendix, we study how the popularity distribution of traits develops over time, and we find its unique equilibrium.

Starting with no culture at time 0, we will denote the complete popularity distribution at time t by a column vector $\mathbf{f}^{(t)} = (f_1^{(t)}, \dots, f_N^{(t)})^T$, where $f_k^{(t)}$ is the expected number of traits of popularity k at time t . (It is impossible to have popularity larger than the size of the entire population, so for all $k > N$ we necessarily have $f_k^{(t)} = 0$.)

In the Appendix, we derive the recursion

$$\mathbf{f}^{(t+1)} = \mathbf{A}\mathbf{f}^{(t)} + (\mu, 0, 0, \dots, 0)^T,$$

where \mathbf{A} is a certain N -by- N tridiagonal matrix. Assuming $\mathbf{f}^{(0)} = \mathbf{0}$, we then obtain

$$\mathbf{f}^{(t)} = (\mathbf{I} + \mathbf{A} + \mathbf{A}^2 + \dots + \mathbf{A}^{t-1}) (\mu, 0, 0, \dots, 0)^T. \quad (1)$$

In the Appendix we also obtain the following formula for the expected number of traits of popularity k when the process has reached its stationary distribution:

$$f_k = \lim_{t \rightarrow \infty} f_k^{(t)} = \mu N \frac{a^{k-1}}{k} \prod_{i=1}^{k-1} \frac{N-i}{N-1-ai}. \quad (2)$$

We then derive our main result about cultural accumulation:

Proposition 1. *The expected number*

$$S = \sum_{k=1}^N f_k \quad (3)$$

of different cultural traits in the population at equilibrium is approximately

$$\tilde{S} := \frac{\mu N}{a} \ln \frac{1}{1-a} + \frac{\mu}{1-a}.$$

For any fixed $0 \leq a < 1$ the relative error $(S - \tilde{S})/S$ tends to zero as $N \rightarrow \infty$, and for any fixed $N \geq 1$ the relative error tends to zero as $a \rightarrow 1$.

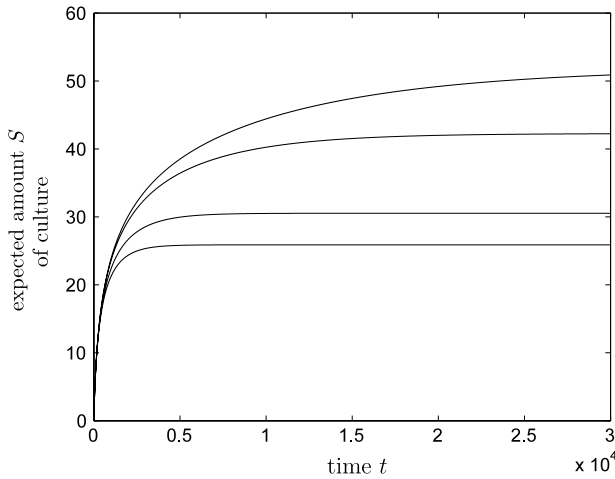


Fig. 1. The accumulation timeline for different values of a : 0.90, 0.94, 0.98 and 0.99; population size $N = 100$ and innovation rate $\mu = 0.1$.

The factor μN can be interpreted as the expected number of innovations made in one generation. Thus, it follows that the expected lifetime T of a cultural trait at equilibrium is

$$T = \frac{S}{\mu N} \approx \frac{1}{a} \ln \frac{1}{1-a} + \frac{\mu}{1-a}.$$

3. Results and examples

The mathematical results of the previous section allow us to answer a number of questions about cultural accumulation given the assumptions made in our model. The rationale is, of course, that if our model captures the important aspects of cultural accumulation in the real world, then the results of our analysis can guide empirical studies and make predictions about what will be found.

First, we investigate how much time the accumulation process needs to reach close to equilibrium levels. We find that unless the social learning efficiency is very close to perfect, equilibrium will be reached fairly quickly. Then, assuming that we will usually find the world at states close to equilibrium, we study how the equilibrium amount of culture will depend on the exogenous parameters as well as how large proportion of the accumulated culture will be carried by an average individual. These entities are all empirically observable and measurable, at least in principle.

Finally, we will also study some theoretical aspects of the model: We will compare its results to the previous macroscopic model of cultural accumulation, and we will show how our model can be used also to study mutually exclusive traits.

3.1. How long does accumulation go on?

Assuming that social learning is not perfect (i.e., $a < 1$) the amount of culture in the population will always proceed to a single equilibrium. Let $x_t := \sum_{k=1}^N f_k^{(t)}$ be the number of cultural traits at timestep t . We do not have a closed expression for x_t , but using Eq. (1) we can compute x_t for any given values of the parameters N , μ and a . For a few different values of a , Fig. 1 illustrates how x_t grows with t —in other words, we see how the accumulation of cultural traits depends on time and on social learning efficiency.

It is clear from the figure that the amount of culture changes fast in the beginning when the amount of culture is still far from equilibrium, and slow when it is closer to equilibrium. In fact, from Eq. (16) in the Appendix, we know that the relative deviation from the equilibrium value will decrease with time approximately as $e^{-(1-a)t/N}$, so that x_t will behave approximately as $(1 - e^{-(1-a)t/N})S$. Thus, the equilibrium is $x_\infty = S$ and the number of timesteps

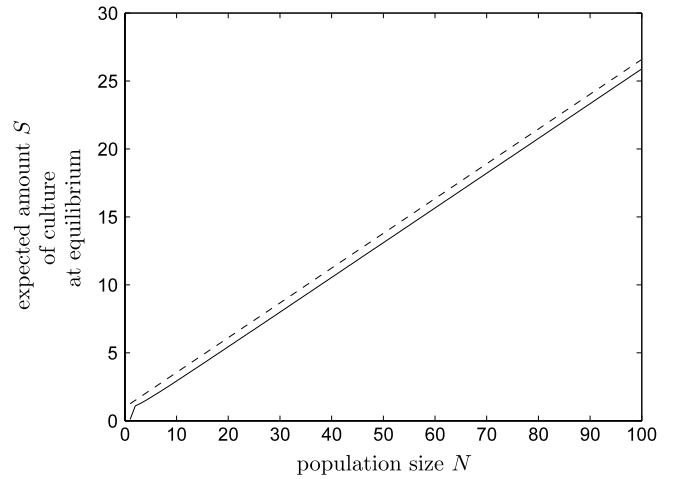


Fig. 2. The expected amount S of culture sustained at equilibrium as a function of population size N . The solid curve shows the exact value (given by Eq. (3)) and the dashed curve shows the approximation given by Proposition 1. Here, $a = 0.9$ and $\mu = 0.1$.

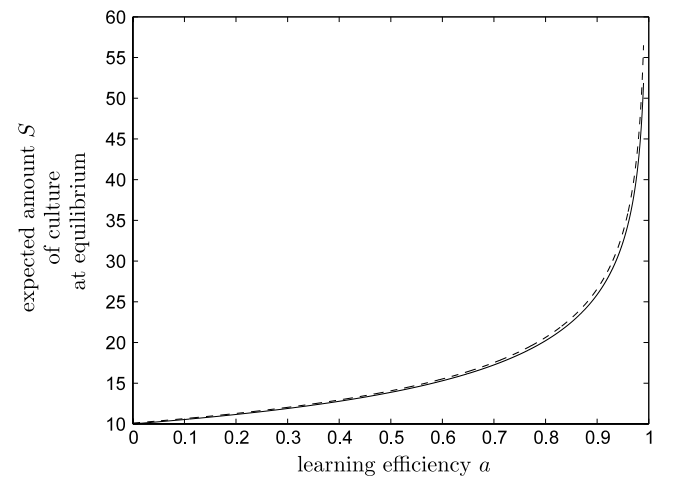


Fig. 3. The expected amount S of culture sustained at equilibrium as a function of social learning efficiency a . The solid curve shows the exact value (Eq. (3)) and the dashed curve shows the approximation given by Proposition 1. Here, $N = 100$ and $\mu = 0.1$.

needed to reach close to this equilibrium is of the order of $N/(1-a)$. Since one individual is replaced at each timestep, we need N steps to replace as many individuals as in one generation. Therefore, if we measure the time needed to approach equilibrium in generations, it is of the order of $1/(1-a)$.

3.2. Analysis of how the equilibrium amount of culture depends on the model parameters

As mentioned, assuming $a < 1$, the accumulation process will eventually reach an equilibrium with respect to the total amount of culture in the population. This equilibrium amount is proportional to the innovation rate μ , and, according to Proposition 1, grows approximately linearly with the population size N ; see Fig. 2.

The dependency on the social learning efficiency, a , is positive but non-linear. The amount of culture remains very low until a is close to 1, and then it accelerates quickly; see Fig. 3.

3.3. How much culture is carried by each individual?

Even if the population carries a large amount of culture, it is not necessary that any individual carries anywhere as many cultural

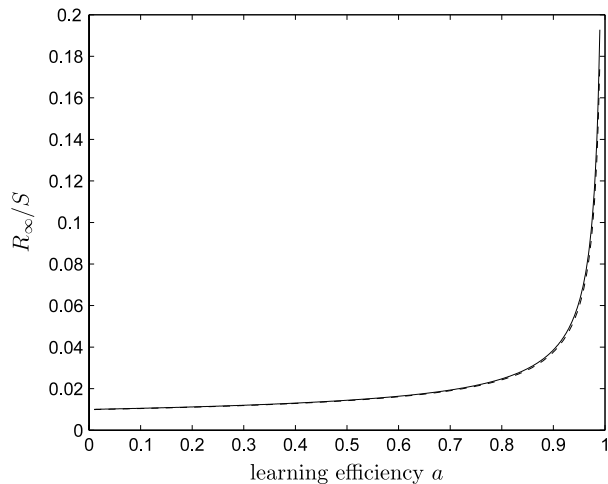


Fig. 4. Expected number of traits per individual divided by the expected number of traits in the population. Here, $N = 100$ and $\mu = 0.1$. The solid curve shows the exact value R_∞/S and the dashed curve shows the approximation R_∞/\bar{S} .

traits. To calculate the average amount of culture carried by individuals, consider how an individual acquires cultural traits in our model. If the average number of traits per individual in the last generation is R , then a new individual will on average have $aR + \mu$ traits. In equilibrium these amounts are equal, yielding the equilibrium amount

$$R_\infty = \frac{\mu}{1-a}. \quad (4)$$

Interestingly, the amount of culture per individual does not depend on the population size.

Now let us determine the proportion of traits in the population that is carried by an average individual (at equilibrium). Dividing the number of traits per individual, from (4), by the number of traits in the population, from Proposition 1, we obtain

$$\frac{\frac{\mu}{1-a}}{\frac{\mu N}{a} \ln \frac{1}{1-a} + \frac{\mu}{1-a}} = \frac{1}{1 - (1-1/a)N \ln \frac{1}{1-a}}.$$

From this expression we see that the proportion is independent of the innovation rate μ and ranges from $\frac{1}{N+1}$ (when $a = 0$) to 1 (when $a = 1$); see Fig. 4. (From the exact expression of \bar{S} , Eq. (3), the proportion is evidently $1/N$ when $a = 0$, so the use of \bar{S} underestimates the proportion a little when N is small.) Thus, individuals become more similar when a increases. As an individual carries a higher proportion of the total amount of culture in the population, she will also share a higher proportion of culture with other individuals.

3.4. Comparison between microscopic and macroscopic models of cultural accumulation

The accumulation of culture can be seen as the outcome of two opposing processes: invention and loss. An equilibrium is reached when these rates become equal. Cultural traits are lost due to two reasons. First, individuals can die without ever acting as “cultural parents”; second, social learning is not perfect ($a < 1$).

Previous macroscopic models of cultural accumulation (Enquist and Ghirlanda, 2007; Enquist et al., 2008) combine a constant innovation rate γ with a loss rate that is a constant proportion λ of the existing amount of culture. This amounts to a simple differential equation:

$$\dot{x} = -\lambda x + \gamma. \quad (5)$$

However, as illustrated in Fig. 5, the proportion of culture lost is not constant in our microscopic model. Thus, we have exposed a weakness in the macroscopic model. The loss rate is not constant

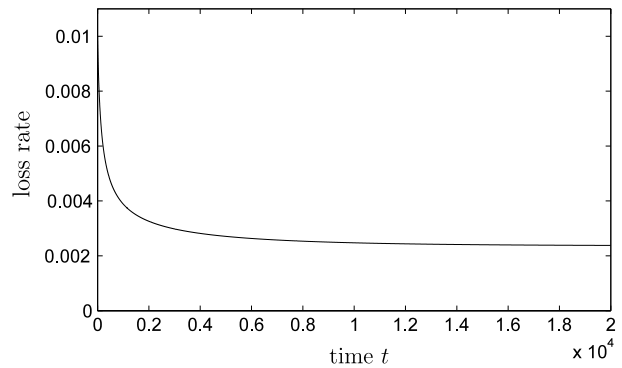


Fig. 5. The proportion of traits that is lost between timesteps. Here, $N = 100$, $a = 0.98$, and $\mu = 0.1$.

in the beginning of the accumulation process. This is due to the fact that, as the cultural accumulation approaches equilibrium, *trait profiles of different individuals become increasingly alike*. In the beginning, when traits exist only if they were recently invented, people’s trait profiles will be completely distinct from each other. As the proportion of cultural traits that is inherited rather than innovated increases, so does the expected number of traits that any two individuals share. This will affect the loss rate, because *when people share traits to a larger extent, fewer traits will be lost*.

3.5. Mutually exclusive traits and the infinite alleles version of the Moran model

The same mathematical model can be used also to describe evolution of mutually exclusive cultural traits. Assume that every individual is a bearer of at most one trait. If $\mu \leq 1 - a$, we can make two assumptions within the model:

- Innovation in phase 3 is dependent on phase 2 such that innovations can only happen in case social learning failed, and
- at most one innovation is made by the new individual.

The consequence is that no individual will ever carry more than one trait, and hence the traits can be interpreted as mutually exclusive variants. All our results are valid also for this situation, like e.g. Eqs. (1) and (2), and Proposition 1.

In this special case, with $\mu = 1 - a$, our model is almost equivalent to the infinite alleles version of the Moran model in mathematical population genetics (cf. Ewens, 2004). The only difference is that, we have assumed that the death phase precedes the birth phase, whereas the Moran model has this the other way around. Indeed, from our analysis it is straightforward to obtain corresponding results under this alternative assumption. Simply add one individual to the population size (i.e. replace N by $N + 1$) and then perform phase 1, the death of one individual, as described by Eq. (10):

$$f_{k,N}^{\text{death-last}} = f_{k,N+1} \left(1 - \frac{k}{N+1} \right) + \frac{k+1}{N+1} f_{k+1,N+1}. \quad (6)$$

From expression (2), where we replace N by $N + 1$, we then obtain a sum that simplifies to

$$f_{k,N}^{\text{death-last}} = \mu N \frac{a^{k-1}}{k} \prod_{i=1}^k \frac{N+1-i}{N-ai}, \quad (7)$$

so the difference between the death-first and death-last models is very slight.

The special case corresponding to the Moran model is when $\mu = 1 - a$. A classic result for the Moran model (cf. Ewens, 2004) is an exact counterpart of our approximative result in

Proposition 1: Under the stationary distribution, the expected number of variants/alleles is exactly

$$S_{\mu=1-a}^{\text{death-last}} = \sum_{k=1}^N \frac{\theta}{\theta + k - 1},$$

where $\theta := N(1-a)/a$. For any fixed N one can check in a symbolic algebra system (like Maple) that for the special case of $\mu = 1 - a$ the identity

$$\sum_{k=1}^N \mu N \frac{a^{k-1}}{k} \prod_{i=1}^k \frac{N+1-i}{N-ai} = \sum_{k=1}^N \frac{\theta}{\theta + k - 1}$$

holds (as it must do, by combination of our analysis and the classic result). We can now use it to derive an exact expression for $S^{\text{death-last}}$ also in the general case when we do not have the constraint on μ . First observe that in the above identity, μ only appears as a factor in the left-hand side; this means that multiplication by $(1-a)/\mu$ has the same effect as setting $\mu = 1 - a$. Consequently, we obtain the following general identity:

$$S^{\text{death-last}} = \frac{\mu}{1-a} \sum_{k=1}^N \frac{\theta}{\theta + k - 1} = N\mu \sum_{k=1}^N \frac{1}{N - a(N + 1 - k)}, \tag{8}$$

where the last equality follows from the definition of θ .

For most practical purposes, it is of course preferable to use the approximation in Proposition 1, which holds equally well for the death-last version.

4. Cultural non-neutrality

By assumption, each cultural element is inherited independently of other traits in our model. Thus if cultural traits are different in how easy they are to invent or learn, we can track the different categories of traits independently of each other. As a straightforward example, let us assume a distinction between two categories of elements: functional and nonfunctional. A proportion q of all innovations is functional, and they are socially learnt with efficiency a . Thanks to a “filter” (a tendency to prefer functional culture), non-functional elements are socially transmitted with lower efficiency, say ϕa where ϕ is the proportion of nonfunctional elements that passes the filter.

From Proposition 1, we can now set up expressions for the amounts of functional and nonfunctional culture at equilibrium, respectively:

$$q\mu \left(\frac{N}{a} \ln \frac{1}{1-a} + \frac{1}{1-a} \right) \quad \text{and} \\ (1-q)\mu \left(\frac{N}{\phi a} \ln \frac{1}{1-\phi a} + \frac{1}{1-\phi a} \right).$$

It follows that the *proportion of functional elements* at equilibrium does not depend on the innovation rate μ . However, the expressions show that the relative effect of the filter on the proportion of functional culture will depend on all other parameters, including the population size. For instance, take social learning efficiency $a = 0.9$, filter $\phi = 0.1$, and proportion of functional innovations $q = 0.1$. Then the proportion of functional culture in equilibrium is approximately 0.21 for large populations. (So the filter has slightly more than doubled the share of functional culture compared to its share of innovations.) The effect of the filter is somewhat even greater when the population is small; e.g., for $N = 10$ the proportion of functional culture is approximately 0.25.

5. Discussion

In this paper we have demonstrated that it is possible to create an analytically tractable individual-based model of cultural accumulation. The key is the possibility to keep track of the

Table 1
Model variables and their influence on accumulation.

	Social learning efficiency (a)	Population size (N)	Innovation rate (μ)
Time to equilibrium (in generations)	+		
Culture at equilibrium	+	+	+
Lifetime of traits (in generations)	+	-	
Culture per individual at equilibrium	+		+
Proportion of all culture per individual	+	-	

popularity distribution, which is our fundamental object of analysis. How the different variables influence accumulation are summarized in Table 1.

At this point, the model relies on three fundamental simplifying assumptions, each of which deserves to be discussed:

First, we assume that culture is transmitted in a *single learning event* per individual, which is a very common assumption in cultural evolution models. Extending the model to investigate the robustness of the results with respect to the number of learning events will be an important next step.

Second, our model is purely cultural, i.e., it accounts for *no interaction with biological reproduction*. As we argued in the introduction, it seems highly plausible that the accumulation of culture can play an important role for ecological success. Our model can potentially be used as a stepping stone for the theoretical study of such issues.

Third, we assume that *cultural elements are independent* of each other, such that having one does not increase or decrease the probability of learning another. It is evident, though, that various forms of dependence constitute a salient feature of many collections of cultural elements. Incorporating dependency in the model to analyze its impact on accumulation is an important direction of future research.

We will now discuss how this model can be used to inform empirical studies of cultural accumulation. There seem to be several promising areas of empirical study, such as the accumulation of bird song elements that we discussed in the introduction. Other examples of possibilities include collections of culturally transmitted behaviors in primates, and the number of food dishes in use in human populations. The models make the following predictions.

To begin with, according to our results accumulation will reach equilibrium after a moderate number of generations; so we will typically expect to observe equilibrium levels.

About these equilibrium levels, our results offer two quite surprising predictions. One is that *population structure will not be important* for the total amount of culture in a population; according to our main result the equilibrium amount of culture is basically proportional to population size, which means that two isolated halves of a population will together carry as much culture as if they were one large interconnected population. The other surprising prediction is that the *cultural accumulation within individuals will be independent of population size*. In other words, we will expect, say, the same number of song elements carried by individual birds in both larger and smaller populations. (Of course, both these results rely on populations not being too small, for our main result to apply.)

Finally, from an empirical study that covers a not too small sample of the population one can, at least in theory, obtain a good estimate of the popularity distribution. From this distribution it is possible to use our analysis to indirectly estimate the values of the innovation rate μ and the social learning efficiency a . For instance, we have seen that shared culture requires efficient social learning, so to the extent that culture is shared within the population we will obtain a large value of a compared to μ .

Acknowledgment

This research was supported by the CULTAPTATION project (European Commission contract FP6-2004-NEST-PATH-043434). The authors are grateful to two anonymous referees and Marc Feldman for pointing out the connection between their original manuscript and the model of Moran as well as giving several other helpful comments.

Appendix. Mathematical treatment

Let $\text{pop}(v)$ denote the current popularity of a trait v .

A.1. The popularity distribution development over time

Our first task is to derive $\mathbf{f}^{(t+1)}$ from $\mathbf{f}^{(t)}$. Let us perform phase 1, i.e. we remove one random individual from the population. For each trait v , the leaving individual has it with probability $\text{pop}(v)/N$. Thus, the expected number $E_{k \leftarrow k+1}^{\text{phase 1}}$ of traits of popularity $k + 1$ that lose a supporter during phase 1 is

$$E_{k \leftarrow k+1}^{\text{phase 1}} = (k + 1)f_{k+1}^{(t)}/N. \tag{9}$$

Next, we compute the expected number $f_k^{\text{phase 1}}$ of traits of popularity k after phase 1:

$$f_k^{\text{phase 1}} = f_k^{(t)} + E_{k \leftarrow k+1}^{\text{phase 1}} - E_{k-1 \leftarrow k}^{\text{phase 1}} = f_k^{(t)} \left(1 - \frac{k}{N} \right) + \frac{k + 1}{N} f_{k+1}^{(t)}. \tag{10}$$

During phase 2, for each trait v , the new individual obtains it with probability $a \text{pop}(v)/(N - 1)$. Thus, the expected number $E_{k \rightarrow k+1}^{\text{phase 2}}$ of traits that have popularity k after phase 1 and popularity $k + 1$ after phase 2 is

$$E_{k \rightarrow k+1}^{\text{phase 2}} = akf_k^{\text{phase 1}}/(N - 1). \tag{11}$$

Finally, by definition, during phase 3 the expected number of new traits created is μ . Nothing else is affected; so we let

$$E_{k \rightarrow k+1}^{\text{phase 3}} = \begin{cases} \mu & \text{if } k = 0, \\ 0 & \text{if } k \geq 1. \end{cases} \tag{12}$$

For $k = 1, \dots, N$ we have

$$f_k^{(t+1)} = f_k^{(t)} + E_{k \leftarrow k+1}^{\text{phase 1}} - E_{k-1 \leftarrow k}^{\text{phase 1}} + E_{k-1 \rightarrow k}^{\text{phase 2}} - E_{k \rightarrow k+1}^{\text{phase 2}} + E_{k-1 \rightarrow k}^{\text{phase 3}} \\ = A_{k,k-1}f_{k-1}^{(t)} + A_{k,k}f_k^{(t)} + A_{k,k+1}f_{k+1}^{(t)} + E_{k-1 \rightarrow k}^{\text{phase 3}}, \tag{13}$$

where

$$A_{k,k-1} = \frac{a(k-1)(N-k+1)}{N(N-1)}, \\ A_{k,k} = 1 - \frac{k}{N} - \frac{ak(N-2k+1)}{N(N-1)}, \\ A_{k,k+1} = \frac{(N-1-ak)(k+1)}{N(N-1)}.$$

By letting \mathbf{A} be an N -by- N tridiagonal matrix whose entries are given by the equations above, we can write (13) as

$$\mathbf{f}^{(t+1)} = \mathbf{A}\mathbf{f}^{(t)} + (\mu, 0, 0, \dots, 0)^T.$$

(Note that $f_{N+1}^{(t)} = A_{1,0} = 0$.) If $\mathbf{f}^{(0)} = \mathbf{0}$, we obtain

$$\mathbf{f}^{(t)} = (\mathbf{I} + \mathbf{A} + \mathbf{A}^2 + \dots + \mathbf{A}^{t-1})(\mu, 0, 0, \dots, 0)^T. \tag{14}$$

A.2. The unique equilibrium

The following proposition guarantees that the series $\mathbf{I} + \mathbf{A} + \mathbf{A}^2 + \dots$ converges so that we reach an equilibrium, which is then given by

$$\mathbf{f} := \lim_{t \rightarrow \infty} \mathbf{f}^{(t)} = (\mathbf{I} - \mathbf{A})^{-1}(\mu, 0, 0, \dots, 0)^T.$$

Proposition 2. The eigenvalues of \mathbf{A} are $1 > \lambda_1 > \dots > \lambda_N = 0$, where

$$\lambda_m = \left(1 - \frac{m}{N} \right) \left(1 + \frac{am}{N-1} \right).$$

An (right) eigenvector $\mathbf{x} = (x_1, \dots, x_N)^T$ with an eigenvalue λ_1 is given by

$$x_k = a^{k-1} \prod_{i=1}^{k-1} \frac{N-i}{N-1-ai}.$$

Proof. Fix m . We will show that to the eigenvalue λ_m defined above, there corresponds a left eigenvector $\mathbf{y} = (y_1, \dots, y_N)$ to \mathbf{A} of the form $y_k = c_1k + c_2k^2 + \dots + c_mk^m$, where the coefficients c_i are independent of k . To this end, we have to check that \mathbf{y} satisfies the relations

$$A_{k-1,k}y_{k-1} + A_{k,k}y_k + A_{k+1,k}y_{k+1} = \lambda_my_k$$

for $k = 1, \dots, N$, where we have put $y_0 = A_{N+1,N} = 0$ for convenience. This can be written as

$$P(k) := ((N-1+a)k - ak^2)y_{k-1} \\ + (N(N-1) - (N-m)(N-1+am)) \\ - (N-1+aN+a)k + 2ak^2)y_k + (aNk - ak^2)y_{k+1} = 0$$

for $k = 1, \dots, N$. The left-hand side $P(k)$ is a polynomial in k , and we will try to choose c_1, \dots, c_m so that the coefficients $[k^j]P(k)$ all vanish. We have $[k^j]P(k) = 0$ for all $j > m + 2$ and a simple computation reveals that $[k^{m+2}]P(k) = [k^{m+1}]P(k) = [k^m]P(k) = [k^0]P(k) = 0$ for any values of c_i . What remains is a homogeneous linear equation system with $m - 1$ equations and m unknowns, and choosing c_1, \dots, c_m as a non-trivial solution to this system makes \mathbf{y} a left eigenvector to \mathbf{A} with an eigenvalue λ_m .

To see that $\mathbf{x} = (x_1, \dots, x_N)^T$ is an (right) eigenvector to \mathbf{A} with an eigenvalue $\lambda_1 = 1 - \frac{1-a}{N}$, one only has to check the relations

$$A_{k,k-1}x_{k-1} + A_{k,k}x_k + A_{k,k+1}x_{k+1} = \lambda_1x_k$$

for $k = 1, \dots, N$, where we have put $x_{N+1} = A_{1,0} = 0$ for convenience. \square

Our next task is to find an explicit formula for the equilibrium $\mathbf{f} = (f_1, \dots, f_N)^T$. At equilibrium we must have

$$E_{k \leftarrow k+1}^{\text{phase 1}} - E_{k \rightarrow k+1}^{\text{phase 2}} - E_{k \rightarrow k+1}^{\text{phase 3}} = 0 \tag{15}$$

for $k = 0, \dots, N - 1$. Plug (10) into (11) and then plug (9), (11) and (12) into (15). This yields the recurrence relation

$$\begin{cases} f_1 = \mu N \\ f_{k+1} = \frac{a(N-k)k}{(N-1-ak)(k+1)} f_k \quad \text{if } k \geq 1 \end{cases}$$

and Eq. (2) follows.

Finally, we will examine how fast the equilibrium is approached. Since

$$\mathbf{f}^{(t)} = (\mathbf{I} + \mathbf{A} + \mathbf{A}^2 + \dots + \mathbf{A}^{t-1})(\mu, 0, 0, \dots, 0)^T = (\mathbf{I} - \mathbf{A}^t)\mathbf{f},$$

the relative deviation $|\mathbf{f}^{(t)} - \mathbf{f}|/|\mathbf{f}|$ from equilibrium is $O(\lambda_1^t)$, where λ_1 is the largest eigenvalue of \mathbf{A} . (Here, $|\mathbf{f}|$ denotes the standard ℓ^2 -norm.) By Proposition 2, we have

$$\lambda_1^t = \left(1 - \frac{1-a}{N} \right)^t \approx e^{-(1-a)t/N} \tag{16}$$

if N is large.

A.3. The number of different cultural traits

We are now ready to prove Proposition 1, the main result about cultural accumulation in our model.

Proof. Keeping a fixed and letting N tend to infinity, we obtain

$$\begin{aligned} \lim_{N \rightarrow \infty} S/N &= \lim_{N \rightarrow \infty} \sum_{k=1}^N \mu \frac{a^{k-1}}{k} \prod_{i=1}^{k-1} \frac{N-i}{N-1-ai} = \mu \sum_{k=1}^{\infty} \frac{a^{k-1}}{k} \\ &= \frac{\mu}{a} \ln \frac{1}{1-a} = \lim_{N \rightarrow \infty} \tilde{S}/N. \end{aligned}$$

When we keep N fixed and let a tend to 1, the last term in the sum $S = \sum_{k=1}^N f_k$ dominates:

$$\begin{aligned} \lim_{a \rightarrow 1} (1-a)S &= \lim_{a \rightarrow 1} \mu N \frac{(1-a)a^{N-1}}{N} \prod_{i=1}^{N-1} \frac{N-i}{N-1-ai} \\ &= \lim_{a \rightarrow 1} \mu (1-a) \frac{(N-1) \prod_{i=2}^{N-1} (N-i)}{(N-1)(1-a) \prod_{i=1}^{N-2} (N-1-ai)} \\ &= \lim_{a \rightarrow 1} \mu \prod_{i=1}^{N-2} \frac{N-1-i}{N-1-ai} = \mu = \lim_{a \rightarrow 1} (1-a)\tilde{S}. \quad \square \end{aligned}$$

A.4. Comparison between cultural accumulation models with discrete and overlapping generations

One of our assumptions is that of overlapping generations. Since it is common in the cultural evolution literature that the models assume discrete generations (Cavalli-Sforza and Feldman, 1981; Boyd and Richerson, 1985; Bentley et al., 2004) we want to test whether that assumption would change our results essentially. Intuitively, we do not expect any great difference in behavior from this slight difference in model specification. However, it is conceivable that parameter values do not directly translate between the two models. For instance, since there is no common time scale between the two models it is not clear that the innovation rate in one model should correspond to the same rate in the other model.

In order to compare the models, fix a population size N . Let $f_k^{\mu,a}$ denote the expected number of traits of popularity k at equilibrium for CAM with innovation rate μ and social learning efficiency a . Similarly, let $\tilde{f}_k^{\tilde{\mu},\tilde{a}}$ denote the expected number of traits of popularity k at equilibrium for a model with discrete generations and with innovation rate $\tilde{\mu}$ and social learning efficiency \tilde{a} .

Computer simulations indicate that given any particular parameter values for the discrete generations model, we can find a CAM with approximately the same equilibrium behavior. In other words, given any $\tilde{\mu} > 0$ and $0 < \tilde{a} < 1$ it seems that we are always able to choose μ and a such that $\tilde{f}_k^{\tilde{\mu},\tilde{a}} \approx f_k^{\mu,a}$ for all $k = 1, \dots, N$.

For instance, if $N = 100$, $\tilde{a} = 0.98$, and $\tilde{\mu} = 0.1$, then we should choose $a = 0.9597$ and $\mu = 0.2040$, as shown in Fig. 6. The

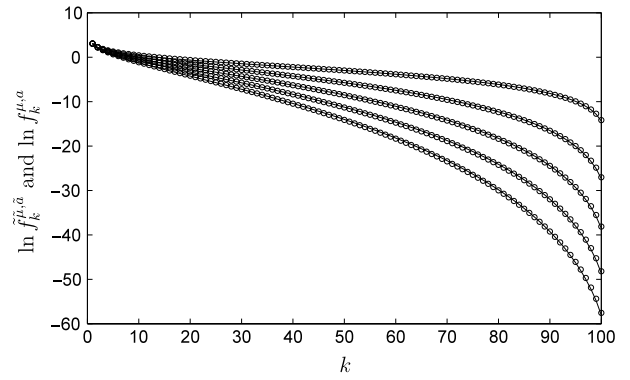


Fig. 6. The natural logarithm of $\tilde{f}_k^{\tilde{\mu},\tilde{a}}$ (curves) and of $f_k^{\mu,a}$ (circles). Here, $N = 100$ and $\tilde{\mu} = 0.1$ everywhere, while the other parameter values vary from the bottom curve to the top curve like this: $\tilde{a} = 0.90, 0.92, 0.94, 0.96, 0.98, a = 0.8074, 0.8439, 0.8815, 0.9202, 0.9597$, and $\mu = 0.2211, 0.2171, 0.2129, 0.2082, 0.2040$.

parameters a and μ are chosen to minimize the square sum

$$\sum_{k=1}^N (\ln f_k^{\mu,a} - \ln \tilde{f}_k^{\tilde{\mu},\tilde{a}})^2.$$

Though this intimate correspondence between models with overlapping and discrete generations is intuitively not very surprising, we have not found a formal proof.

References

Bentley, R.A., Hahn, M.W., Shennan, S.J., 2004. Random drift and culture change. *Proceedings of the Royal Society B* 271, 1443–1450.

Boyd, R., Richerson, P.J., 1985. *Culture and The Evolutionary Process*. University of Chicago Press.

Catchpole, C.K., Slater, P.J.B., 1995. *Bird Song: Biological Themes and Variations*. Cambridge University Press.

Cavalli-Sforza, L.L., Feldman, M.W., 1981. *Cultural Transmission and Evolution*. Princeton University Press.

Enquist, M., Ghirlanda, S., 2007. Evolution of social learning does not explain the origin of human cumulative culture. *Journal of Theoretical Biology* 246, 129–135.

Enquist, M., Ghirlanda, S., Jarrick, A., Wachtmeister, C.-A., 2008. Why does human culture increase exponentially? *Theoretical Population Biology* 74 (1), 46–55.

Ewens, W.J., 2004. *Mathematical Population Genetics*, 2nd ed. Springer.

Futuyma, D.J., 1997. *Evolutionary Biology*, 3rd ed. Sinauer Associates.

Hahn, M.W., Bentley, R.A., 2003. Drift as a mechanism for cultural change: An example from baby names. *Proceedings of the Royal Society B* 270, S120–S123.

Herzog, H.A., Bentley, R.A., Hahn, M.W., 2004. Random drift and large shifts in popularity of dog breeds. *Proceedings of the Royal Society B* 271, S353–S356.

Kimura, M., Crow, J.F., 1964. The number of alleles that can be maintained in a finite population. *Genetics* 49, 725–738.

Lehman, H.C., 1947. The exponential increase in man’s cultural output. *Social Forces* 25, 281–290.

Lynch, A., 1996. The population memetics of bird-song. In: Kroodsma, D.E., Kroodsma, E.H.M. (Eds.), *Ecology and Evolution of Acoustic Communication in Birds*. Comstock Books, pp. 181–198.

Lynch, A., Baker, A.J., 1993. A population memetics approach to cultural evolution in chaffinch song: Meme diversity within populations. *The American Naturalist* 141, 597–620.

Lynch, A., Plunkett, G.P., Baker, A.J., Jenkins, P.F., 1989. A model of cultural evolution of chaffinch song derived with the meme concept. *The American Naturalist* 133 (5), 634–653.

Richerson, P.J., Boyd, R., 2005. *Not by genes alone*. University of Chicago Press.