

Cumulative cultural dynamics and the coevolution of cultural innovation and transmission: an ESS model for panmictic and structured populations

L. LEHMANN*, M. W. FELDMAN† & R. KAEUFFER‡

*Institute of Biology, University of Neuchâtel, Neuchâtel, Switzerland

†Department of Biological Sciences, Stanford University, Stanford, CA, USA

‡Department of Biology, McGill University, Montreal, Canada

Keywords:

cultural accumulation;
cultural transmission;
individual and social learning;
innovation;
kin selection;
relatedness.

Abstract

When individuals in a population can acquire traits through learning, each individual may express a certain number of distinct cultural traits. These traits may have been either invented by the individual himself or acquired from others in the population. Here, we develop a game theoretic model for the accumulation of cultural traits through individual and social learning. We explore how the rates of innovation, decay, and transmission of cultural traits affect the evolutionary stable (ES) levels of individual and social learning and the number of cultural traits expressed by an individual when cultural dynamics are at a steady-state. We explore the evolution of these phenotypes in both panmictic and structured population settings. Our results suggest that in panmictic populations, the ES level of learning and number of traits tend to be independent of the social transmission rate of cultural traits and is mainly affected by the innovation and decay rates. By contrast, in structured populations, where interactions occur between relatives, the ES level of learning and the number of traits per individual can be increased (relative to the panmictic case) and may then markedly depend on the transmission rate of cultural traits. This suggests that kin selection may be one additional solution to Rogers's paradox of nonadaptive culture.

Introduction

Learned traits and in particular cultural traits are non-genetically determined phenotypes that are acquired during an individual's lifespan. They are not only characteristic of humans but are also expressed by many vertebrates (Laland & Janik, 2006). For instance, chimpanzees use sticks to catch prey and stones to crack nuts; and macaques wash potatoes and apples, unwrap and consume caramels, and can learn a whole spectrum of other feeding behaviours (Lefebvre, 1995; Whiten *et al.*, 1999; Dugatkin, 2004). Some birds are able to learn new songs but they can also acquire techniques to bait fish, batter or drop different types of prey on different substrates, use caps to carry water, use twigs to push

nuts, and pull fishing lines to get fish under water (Lefebvre *et al.*, 2002).

There are two basic ways by which an individual may learn a new trait (Rogers, 1988). First, the trait can be learned individually. Here, an individual interacts with its environment and learns the trait by trial-and-error, lucky accident, insight, or deduction. This can be viewed as cultural innovation, and this process may also depend on the number of traits already carried by the individuals in the population. Alternatively, a trait can be learned socially, in which case an individual obtains the trait by imitating or copying it from another individual in the population. This is cultural transmission. This second case is likely to involve social interactions between individuals in the population, and errors in transmission may further increase the rate of innovation of cultural traits.

Cultural innovation is to cultural evolution what mutation is to biological evolution: without innovation, cultural traits and therefore cultural transmission would not exist. In humans, these features may have led to the

Correspondence: Laurent Lehmann, Institute of Biology, University of Neuchâtel, Switzerland.
Tel.: 032 718 2234; fax: 032 718 3001;
e-mail: laurent.lehmann@unine.ch

relaxation of at least some of the environmental pressures faced by our early ancestors. Using basic objects found in their environment, such as wooden sticks and rocks, hominids developed new tools for foraging or hunting. Agricultural techniques were innovated later and led to new ways of sheltering and storing resources, which resulted in better control over environmental stochasticity and other hazards. The energy saved through increased ingenuity with which hominids put resources to use, allowed allocation of some energy to other physiological mechanisms such as brain development (Aiello & Wheeler, 1995), which, in turn, may have supported new cultural innovation. Eventually, these innovations led to the generation of technology that is the basis of economic growth (Kremer, 1993; Galor & Weil, 2000; Romer, 2006).

Similarly, in other vertebrate species such as primates and birds, innovation may be important in allowing individuals to adapt to changing environmental conditions (Sol *et al.*, 2005). Innovation allows individuals to adapt to new selective pressures by avoidance of new predators (Berger *et al.*, 2001) and by changing resource utilization (Estes *et al.*, 1998). Birds and mammals with a greater tendency to innovate also show a higher propensity to successfully migrate into a new environment (Sol *et al.*, 2005; 2008). There is also some evidence that the more innovative species are those that store less food (Lefebvre & Bolhuis, 2003). This observation suggests that a higher rate of innovation may result in a greater number of traits acquired during an individuals' lifespan and less sensitivity to environmental variations. The ability to innovate might thus have played a crucial role not only in the evolution of hominids but also in the evolution of other vertebrates.

What is the number of learned traits (or learned pieces of information) carried by an individual in humans or other species? One can speculate that the more traits an individual has, the more likely it is to cope with a variable, constantly changing environment. Hence, there might be a selection pressure for increasing the number of learned traits expressed by an individual during its lifetime. This number is likely to be constrained by at least two sets of factors. The first are physiological and environmental. To support individual and social learning, individuals need physiological mechanisms allowing them to invent, express, and remember traits, whose number is bound by the different number of objects (and possible combinations of them) encountered in the environment.

The second set of factors affecting the number of traits acquired by an individual during its lifespan depend on the type of social interactions it faces. Cultural traits are probably costlier to invent, in terms of time and energy, than they are to transmit from one individual to another in a population. Given the cost of producing a new trait that might be used freely by other individuals, the evolution of innovation and cultural transmission poses a social dilemma, which can be interpreted as a 'producer/

scrounger' game (Barnard & Sibly, 1981; Giraldeau *et al.*, 1994). It follows that the number of cultural traits carried by an individual depends on social interactions, the evolution of which depends considerably on life-history and demographic features (e.g. West *et al.*, 2007; Lion & van Baalen, 2007).

The accumulation of learned traits in a population thus depends on the interaction between many variables. But, surprisingly, the quantitative dynamics of the accumulation process remains largely unexplored, be it on a behavioural or evolutionary time scale. In this paper, we carry on the evolutionary approach to cumulative culture (Ghirlanda & Enquist, 2007; Enquist *et al.*, 2008; Strimling *et al.*, 2009; Lehmann & Feldman, 2009; Lehmann *et al.*, 2010; Ghirlanda *et al.*, 2010) and develop a game theoretic model to better understand how many traits an individual may learn during its lifetime as a function of several individual and social learning parameters. We evaluate the candidate evolutionarily stable level of learning, the fraction of time spent on individual versus social learning, and the associated steady-state number of traits expressed by an individual. We do so for two demographic scenarios: interactions occurring only between individuals in a panmictic population and interactions occurring between related individuals in a family or in a spatially structured population.

Model

Assumptions

Biological setting

We assume that haploid individuals live in a population made up of groups of finite size N that are connected by dispersal. The individuals may be iteroparous or semelparous, but we leave the exact details of the life history unspecified as it does not affect the argument presented in the following paragraphs. All that matters is that interactions between individuals can occur at a local scale, among group members, instead of occurring at random in the population. If dispersal is limited and group size is finite, then these interactions are likely to occur between relatives (Hamilton, 1971; Rousset, 2004).

We assume that the individuals in this population are endowed with physiological mechanisms allowing them to express and remember cultural traits. For each individual, the time interval during two reproductive events is assumed to be divided into a number of periods during which it forages to gain resources and when social interactions and cultural transmission can occur between neighbours. During each such time period, an individual may invent novel cultural traits (individual learning), acquire them by copying other individuals (social learning), or forget a certain number of traits acquired previously. In the case of social learning, we assume that the individual can acquire traits only from its group mates and not from individuals from other groups.

One time period can thus be thought of as a single period of cultural innovation and transmission. Each such learning period will affect the number of adaptive cultural traits carried by an individual (those traits that positively affect the organism's vital rates) and which we denote by A . Assuming a large number of rounds of cultural innovation and transmission, A may eventually converge to an equilibrium value if the time span between reproductive events is large. We assume that the number of offspring produced by an individual (or its survival) is an increasing function of the steady-state number of adaptive cultural traits it expresses and, for simplicity, we assume that individuals express only adaptive cultural traits.

Two-trait model

To analyse the coevolutionary dynamics of cultural innovation and transmission, we assume a two-locus (two-trait) model. The first locus controls the proportion of time l that an individual spends learning. The complementary fraction $1 - l$ of the individual's time is spent gathering resources, which can be converted into survival or fecundity. Because learning is either individual (producing cultural traits) or social (scrounging cultural traits from others), a second locus is assumed to control the proportion of time p that an individual that is learning spends producing novel cultural traits. A complementary fraction $1 - p$ of the time is thus allocated to socially learning cultural traits. Our aim is to study the evolutionary dynamics of learning, l , and producing, p .

ESS analysis

For analytical tractability, we investigate the evolutionary dynamics of the two traits under the assumptions of weak selection, additive gene action, and that only two alleles can segregate simultaneously at each locus in the population. These are standard assumptions, which allow us to use evolutionary game theory coupled with inclusive fitness theory (e.g. Taylor, 1996; Gandon, 1999; Rousset, 2004). Owing to the assumptions of weak selection, genetic associations do not significantly affect the evolutionary dynamics, and we can study the evolution of each trait by holding the other trait constant (Roze & Rousset, 2005; 2008). Hence, we do not consider genetic covariances between traits.

For each trait, we then focus on a mutant allele coding for a phenotype (fraction of time spent learning or producing) whose value deviates by a small amount from that expressed by an individual bearing a resident (wild-type) allele and ask whether the mutant allele will be selected for. By successive allelic replacement, each trait may eventually converge towards a candidate evolutionary stable state (ESS). Application of inclusive fitness theory to group structured populations of constant size without class structure (reviewed in Rousset, 2004) shows that a candidate ESS trait value z^* (where z refers

either to l or p) for a behaviour affecting fecundity satisfies the equation

$$\frac{\partial f}{\partial z_{\bullet}} + \kappa \frac{\partial f}{\partial z_0} = 0, \quad (1)$$

where $f \equiv f(z_{\bullet}, z_0)$ is the fecundity of a focal individual (number of offspring produced); $\partial f / \partial z_{\bullet}$ is the change in f stemming from the focal individual expressing a mutant allele (with phenotype denoted z_{\bullet}); $\partial f / \partial z_0$ is the change in the fecundity of the focal individual because of all of its patch neighbours expressing the mutant allele (with average phenotype z_0); and the partial derivatives are evaluated at $z_{\bullet} = z_0 = z^*$, the candidate ESS value of the trait under scrutiny. The $\partial f / \partial z_{\bullet}$ term can be thought of as the direct selective pressure on the trait z , whereas $\partial f / \partial z_0$ as the indirect selective pressure, and it is weighted by the coefficient κ , which can be thought of as a relatedness coefficient that has been rescaled to absorb any competitive effects because of limited dispersal and localized interactions that may decrease the selective pressure on the trait under study (Queller, 1994).

The coefficient κ is typically lower than the relatedness coefficient between group members. It has been calculated explicitly for many different life-cycle assumptions in patch-structured and isolation by distance models (e.g. Aoki, 1982; Rogers, 1990; Taylor, 1992; Taylor & Irwin, 2000; Gardner & West, 2006; Lehmann *et al.*, 2006; Lion & Gandon, 2009; Gardner, 2010) and is expressed in terms of the demographic parameters of the population (e.g. migration distribution, life-history features, local demographic conditions). Because our aim in this paper is to focus on the cumulative cultural dynamics occurring during an individual's lifetime, we treat κ as a model parameter but, importantly, our treatment can be embedded in the demographic assumptions of this previous work and many related models. For example, as a rough approximation, κ is often of order $1/N$ when migration is weak and patch size is large (Lehmann & Rousset, 2010, Table 2).

In the next section, we present expressions for the number of cultural traits carried by an individual as a function of its trait values and those of its patch mates. This allows us to evaluate f as a function of the number of cultural traits from which we can then evaluate the adaptive dynamics of both l and p .

Dynamics of cultural trait number

Our first goal is to derive a recurrence equation over one time period for the number $A_{\bullet,t}$ of adaptive cultural traits carried by a focal individual. This number will be affected by the set of phenotypes $\{l_{\bullet}, l_0, p_{\bullet}, p_0\}$, where l_{\bullet} (l_0) is the proportion of time that the focal individual (an average patch mate) expresses learning and p_{\bullet} (p_0) is the proportion of that time the focal individual (an average patch mate) expresses individual learning.

Because the focal individual may lose cultural traits from one time period to the next owing to failure of memory or obsolescence of usage of traits, may invent novel traits, or acquire them through social learning, we assume that the number of traits carried by the focal individual in period t is given by

$$A_{\bullet,t} = (1 - \epsilon)A_{\bullet,t-1} + I_{\bullet,t} + S_{\bullet,t} \quad (2)$$

where ϵ ($0 \leq \epsilon \leq 1$) is the rate of loss of a cultural trait (or obsolescence rate) over one time period, $I_{\bullet,t}$ is the number of new traits the focal individual acquires during period t by individual learning, and $S_{\bullet,t}$ is the number of such traits obtained through social learning.

The number of new traits invented during period t by the focal individual is assumed to be given by

$$I_{\bullet,t} = l_{\bullet} p_{\bullet} \mu, \quad (3)$$

where $l_{\bullet} p_{\bullet}$ is the fraction of time the focal individual spends on individual learning and μ is the rate of production of new traits per unit time spent learning, which are assumed to be independent of each other (e.g. Strimling *et al.*, 2009).

Independence of trait innovation may be a strong assumption. Nonindependence of trait innovation may occur, for instance, if the rate of innovation depends on the number of existing traits, $A_{\bullet,t-1}$, expressed by the focal individual. It may also occur if the innovation rate depends on the number of traits expressed by other individuals in the population, and errors occur in the transmission process, or if new traits are built on a combination of existing traits expressed by the focal and/or other individuals. But allowing for these realistic features would result in a much more complicated model, so they are not taken into account here, for simplicity.

When the focal individual acquires traits through social learning in period t , it may acquire two classes of traits. First, it may acquire novel traits invented by others during period t . Second, the focal individual may acquire traits invented by others in past time periods but which it did not acquire previously. To take these two cases into account, we assume that the number of traits acquired by the focal individual through social learning in period t is given by

$$S_{\bullet,t} = l_{\bullet}(1 - p_{\bullet})[\beta_n(N - 1)I_{0,t} + \beta_o(1 - \epsilon)O_{\bullet,t-1}], \quad (4)$$

where $l_{\bullet}(1 - p_{\bullet})$ is the fraction of time the focal individual spends on social learning, β_n is the probability that this individual learns from an average neighbour a random trait that was created by a neighbour in period t (β_n can be thought of as the contact rate between individuals multiplied by the rate of transmission of a random trait from the neighbour), $N - 1$ is the number of neighbours the focal individual may interact with, and

$$I_{0,t} = l_0 p_0 \mu \quad (5)$$

is the number of new traits invented during time period t by an average neighbour.

The second term in brackets in eqn 4 accounts for the traits acquired by social learning in period t but that were invented in previous time periods. Here, β_o is the probability that the focal individual learns a random trait invented previously that it has not yet acquired, $O_{\bullet,t-1}$ is the total number of different traits existing in the population at time $t - 1$ that the focal individual has not yet acquired (older traits), and $(1 - \epsilon)$ is the decay rate of the stock of adaptive traits not yet adopted by the focal individual. The number of cultural traits not yet acquired satisfies the recursion

$$O_{\bullet,t} = [1 - l_{\bullet}(1 - p_{\bullet})\beta_n](N - 1)I_{0,t} + [1 - l_{\bullet}(1 - p_{\bullet})\beta_o](1 - \epsilon)O_{\bullet,t-1} \quad (6)$$

because at time t , the focal individual fails to acquire a number $[1 - l_{\bullet}(1 - p_{\bullet})\beta_n](N - 1)I_{0,t}$ of novel cultural traits produced by neighbours, where $(N - 1)I_{0,t}$ is the total number of new traits produced by neighbours at t , and the focal individual also fails to acquire a number $[1 - l_{\bullet}(1 - p_{\bullet})\beta_o](1 - \epsilon)O_{\bullet,t-1}$ of traits it had not acquired previously.

The total number of distinct traits segregating in the focal group in period t is given by $T_t = A_{\bullet,t} + O_{\bullet,t}$. Because $\mu[l_{\bullet} p_{\bullet} + (N - 1)l_0 p_0]$ is the total number of traits innovated in the focal group per time period, T_t satisfies the recursion

$$T_{t+1} = (1 - \epsilon)T_t + \mu[l_{\bullet} p_{\bullet} + (N - 1)l_0 p_0], \quad (7)$$

which, for consistency, can be checked to hold by substituting eqns 2 and 6 in the left member.

Equilibrium trait number and effect on fecundity

If the number of learning periods occurring during an individual's lifespan is large, the numbers $A_{\bullet,t}$ and $O_{\bullet,t}$ may converge within that lifetime to their steady-state values, which we denote by A_{\bullet} and O_{\bullet} , respectively.

We assume that the number of offspring produced by an individual is an increasing linear function of the resources it obtains. The number of resources, in turn, is assumed to increase with the steady-state number of cultural traits an individual carries and to decrease with the fraction of time it spends learning. Hence, there are two factors allowing an individual to obtain resources: 'labor', $1 - l$, which is the time spent gathering resources and 'technology', $1 + A$, which is a baseline innate ability of an individual to extract resources augmented by the number of cultural traits it carries. It is a standard assumption of economics that the output per individual depends on 'labor' and 'technology', that is, on cultural traits (e.g. Galor & Weil, 2000; Romer, 2006), and it is

reasonable to postulate that this assumption applies not only to humans but also to other species. For simplicity, we also assume that these two factors of production combine multiplicatively to give the total amount of resources $(1 + A)(1 - l)$ available to an individual.

With these assumptions, the steady-state fertility of the focal individual is given by

$$f = (1 + A_{\bullet})(1 - l_{\bullet}), \quad (8)$$

where l_{\bullet} is its time spent learning and A_{\bullet} its steady-state number of cultural traits, which is a function of the phenotypes of interacting individuals ($p_{\bullet}, p_0, l_{\bullet}, l_0$) and of the model's parameter values ($\epsilon, \mu, \beta_o,$ and β_n , see eqns 26 and 27). Using eqns 8 and 26 in eqn 1, we can now study the adaptive dynamics of l and p . Before doing so and to gain intuition about the values that the number of traits carried by an individual can take, we examine a monomorphic population.

Results

Trait number in a monomorphic population

From eqn 26, the number of traits carried by a focal individual in a monomorphic population ($l_{\bullet} = l_0 = l, p_{\bullet} = p_0 = p$) is given by

$$A = \frac{l(p\mu + (1 - p)[n_s + \beta_o(1 - \epsilon)O])}{\epsilon}, \quad (9)$$

where $n_s = \beta_n(N - 1)lp\mu$. Equation 9 is the ratio of the total number of new traits acquired per unit time by an individual to the number of traits lost per unit time. The number of new traits depends on the number of traits, μ , acquired per unit time of individual learning and the number of traits acquired per unit time by social learning, which depends on those traits generated in the present time period by others (n_s) and those generated in past time periods and not yet copied ($\beta_o(1 - \epsilon)O$). The steady-state number of traits generated in past time periods and not yet acquired is given by

$$O = \frac{[1 - l(1 - p)\beta_n](N - 1)lp\mu}{1 - [1 - l(1 - p)\beta_o](1 - \epsilon)}, \quad (10)$$

which when combined with eqn 9 allows us to evaluate the total number of different traits in the population as

$$T = A + O = \frac{lpN\mu}{\epsilon}, \quad (11)$$

which is an increasing function of the population-wide number of traits $lpN\mu$ invented per unit time.

The main qualitative features of eqns 9–11 are that A tends to increase with the intensity of learning, l , but depending on the values of the transmission rates (β_n and β_o) and of ϵ , it can be a dome-shaped function of the

proportion p of time spent producing traits. The number of different traits not yet acquired by an individual, O , tends to increase with p but can be a dome-shaped curve of the proportion l of time spent learning under a certain range of parameter values. However, the total number of different traits in the population, T , is increasing in both l and p . Finally, we note that from eqns 9 to 11, we can evaluate the probability that two randomly sampled traits from two distinct individuals from the same group are identical as

$$\Theta = \frac{A - O / (N - 1)}{A} = \frac{Nl(1 - p)[\beta_n\epsilon + \beta_o(1 - \epsilon)]}{\epsilon + l(1 - p)(N[\beta_n\epsilon + \beta_o(1 - \epsilon)] - \beta_n\epsilon)}, \quad (12)$$

which is the number of distinct traits shared by two randomly sampled individuals divided by the total number of traits carried by a single individual. The similarity index Θ is an increasing function of the cultural transmission rates.

In the next sections, we investigate the evolutionary dynamics of l and p , which then allows us to evaluate the evolutionary stable (ES) trait number expressed at steady state. The evolutionary dynamics of the full model (with all parameters taking positive values) is complicated, and the ES levels of learning and producing cannot always be evaluated analytically. We studied the evolution of l and p using analytical expressions when we were able to derive them (or when they were not too complicated) and used numerical analysis for the more complicated cases.

Panmictic population

ES level of learning: baseline case

We first assume in this section that there are no effects of relatives on the two evolving traits ($\kappa = 0$ in eqn 1) and that individuals only acquire through social learning traits that were generated in the current time period and not those that were invented in past time periods ($\beta_n > 0$ and $\beta_o = 0$).

Inserting eqns 8 and 26 into eqn 1, letting $\kappa = 0$ and $\beta_o = 0$, taking learning as the focal trait ($z = l$), and holding producing constant ($p_{\bullet} = p_0 = p$), we find that the selective pressure on learning is given by

$$\begin{aligned} \frac{\partial f}{\partial l_{\bullet}} &= \frac{\partial A_{\bullet}}{\partial l_{\bullet}}(1 - l) - (1 + A) \\ &= \frac{\{p\mu + (1 - p)n_s\}(1 - l)}{\epsilon} - \left(1 + \frac{l\{p\mu + (1 - p)n_s\}}{\epsilon}\right), \end{aligned} \quad (13)$$

which reflects a trade-off between the increase in resources owing to the additional trait number accruing to the focal individual (first term of eqn 13) and the loss of resources from spending time learning instead of gathering resources (second term of eqn 13).

At an evolutionary equilibrium, gains and losses balance each other out (i.e. $\partial f/\partial l_{\bullet} = 0$), and from eqn 13, the candidate ES level of learning is

$$l^* = \frac{1}{2} \left(1 - \frac{\epsilon}{\{p\mu + (1-p)n_s\}} \right), \quad (14)$$

which has a maximum value of one half (when μ becomes infinitely large) and is a decreasing function of the decay rate ϵ and of the number n_s of traits acquired socially.

ES level of producing: baseline case

Inserting eqns 8 and 26 into eqn 1, letting $\kappa = 0$ and $\beta_o = 0$, but with producing as the focal trait in eqn 1 ($z = p$) and holding learning at its monomorphic value ($l_{\bullet} = l_o = l$), we find that the selective pressure on producing is

$$\begin{aligned} \frac{\partial f}{\partial p_{\bullet}} &= \frac{\partial A_{\bullet}}{\partial p_{\bullet}} (1-l) \\ &= \frac{l(\mu - n_s)(1-l)}{\epsilon}, \end{aligned} \quad (15)$$

where the term in the second parentheses is the number of traits that accrue to the focal individual when it spends one additional time unit producing instead of scrounging, in which case it gains μ additional traits and loses n_s traits.

At an evolutionary equilibrium, gains and losses balances each other out, that is $\partial f/\partial p_{\bullet} = 0$, which, from eqn 15, gives $\mu = n_s$. Using $n_s = \beta_n(N-1)lp\mu$, the candidate ES level of producing is then given by

$$p^* = \frac{1}{l\beta_n(N-1)}. \quad (16)$$

Hence, the proportion of time spent producing is equal to the number of traits acquired per unit time by producing relative to the number acquired per unit time by social learning. This is qualitatively similar to the polymorphic equilibrium of producing found in a two-allele model with similar baseline structure but with pure strategies and intergenerational effects of cultural transmission (Lehmann & Feldman, 2009, eqn 3.4).

Joint ESS

Equation 14 is a function of p , and eqn 16 is a function of l . Setting the trait values in both equations at their ES values and solving for l^* and p^* , we find that the candidate optimal levels of learning and producing are

$$l^* = \frac{\mu - \epsilon}{2\mu}, \quad (17)$$

which is increasing in μ and

$$p^* = \frac{2\mu}{(\mu - \epsilon)\beta_n(N-1)}, \quad (18)$$

which is decreasing in μ .

With these two candidate ESS levels, we can now determine the associated candidate optimal number of cultural traits. By inserting eqn 18 into eqn 9, we find that

$$A^* = \frac{\mu}{\epsilon} l^* \quad (19)$$

and substituting from eqn 17, we finally have

$$A^* = \frac{\mu - \epsilon}{2\epsilon}. \quad (20)$$

The candidate optimal number of traits A^* is independent of the transmission rate β_n (see Fig. 1). Hence, social learning does not affect the number of adaptive traits expressed by an individual at steady state and, therefore, does not increase the average fitness of individuals in the population. This is a qualitative result that has been observed in several earlier models for the evolution of social learning (Rogers, 1988; Boyd & Richerson, 1995; Wakano *et al.*, 2004; Enquist & Ghirlanda, 2007; Lehmann & Feldman, 2009) and has been called Rogers's paradox of nonadaptive culture (Enquist *et al.*, 2007; Rendell *et al.*, 2010).

The steady-state number of cultural traits A^* is not affected by social learning, that is, by β_n , because social learning results in the aggregation of traits from several different individuals in the population. Although an increase in the transmission rate β_n decreases the selective pressure on producing and thus on the time spent innovating (eqn 18), this loss is compensated by an increase in trait number acquired through social learning. At an evolutionary equilibrium, the same amount of cultural traits can then be maintained with fewer individuals producing it (Lehmann & Feldman, 2009), and social learning is selected against at the point where it results in a decrease in trait number relative to that when only individual learning occurs. Nevertheless, it is

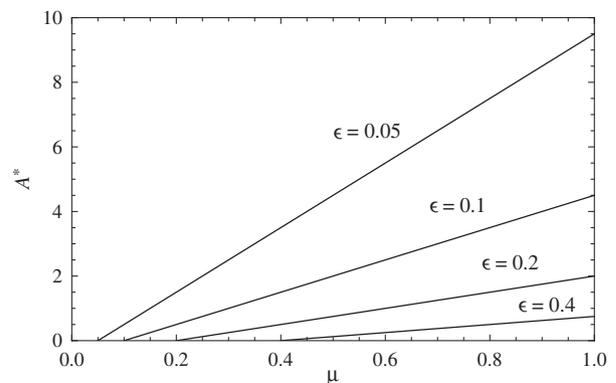


Fig. 1 Candidate ES expected number of traits A^* , eqn 20, carried by an individual at an evolutionary equilibrium as functions of the innovation rate μ . From the top to the bottom curve: $\epsilon = 0.05, 0.1, 0.2$, and 0.4 . Individuals cannot acquire traits from past time periods ($\beta_o = 0$).

important to note that the result that β_n does not affect A^* is not general and is likely to depend on the functional relationship between trait values and cultural dynamics as will be illustrated in the next section.

Effect of acquiring traits generated in past time periods

If we assume that individuals can acquire traits from past learning periods that they have not yet acquired ($\beta_o > 0$), the model becomes more complicated and we have analysed it numerically (the selective pressure on learning and producing is given by eqns 28 and 30, respectively). In the top panel of Fig. 2, the ES levels of learning and producing are graphed as functions of β_o . The main effect of increasing this parameter is that the level of producing decreases, while l tends to decrease or remains approximately constant. This follows from the fact

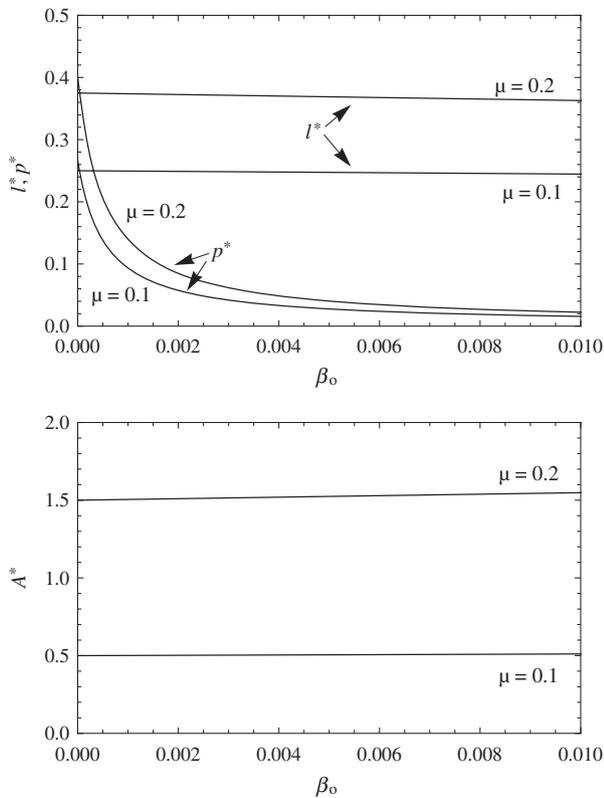


Fig. 2 Upper panel: candidate ES level of learning, l^* , and producing, p^* , as functions of β_o with parameter values $N = 1000$, $\epsilon = 0.1$, $\beta_n = 0.01$, and two values of the mutation rate (the upper decreasing curve is for $\mu = 0.2$, whereas the lower is for $\mu = 0.1$). When $\beta_o = 0$, the values of learning and producing are those given by eqns 17 and 18. But when the value of β_o increases and approaches that of β_n , the level of producing decreases. By contrast, as β_o varies, l^* remains approximately constant (the upper flat curve is for $\mu = 0.2$, whereas the lower is for $\mu = 0.1$). Lower panel: expected number of traits A^* carried by an individual at an evolutionary equilibrium as a function of β_o for the same parameter values as those in the first panel of the figure (the upper curve is for $\mu = 0.2$ and the lower is for $\mu = 0.1$).

that when individuals can also acquire traits generated in past time periods, the number of traits they can acquire per unit time increases, which increases the selective pressure in favour of scrounging and decreases the equilibrium level of producing. At the same time, the selective pressure on learning decreases because the increase in the number of traits acquired socially makes the benefit of learning balance the cost at lower levels of learning.

If we assume that $\beta_o = \beta_n = \beta$, which seems to be a natural assumption because then individuals do not discriminate between traits generated in present and past learning periods, and that the parameter β is small (i.e. taking into account only first-order effects in β near $\beta = 0$), we find that the candidate ES level of producing is given by

$$p^* = \frac{\epsilon}{l\beta(N-1)}, \quad (21)$$

(eqns 28–31). When $\epsilon = 1$, traits from past time periods are no longer adaptive and eqn 21 then agrees with eqn 16; otherwise, the proportion of time spent producing is lower than that in eqn 16 because the selective pressure for scrounging is stronger. Under the assumption that β is small, the candidate ES level of learning takes the same value as that found previously (eqn 17), namely $l^* = (\mu - \epsilon)/(2\mu)$ (see eqns 28–31). Substituting this equation and eqn 21 into eqn 32, yields the candidate ES number of traits as $A^* = (\mu - \epsilon)/(2\epsilon)$, which is equivalent to that found previously (eqn 20).

The aforementioned analytic approximations for p^* and l^* rely on assuming that β is small. This can be justified by noting that this parameter involves the product of the contact rate between individuals and the transmission probability of a trait. If the contact rate between individuals is approximately equal to the inverse of population size, namely, each individual interacts with each other according to its frequency in the population, then β will be small unless population size is very small. This seems to be a natural assumption, and it suggests that l^* might be well approximated by eqn 17 in large populations, whether or not individuals can acquire traits from past generations.

When the proportion of time spent learning remains approximately the same, regardless of the magnitude of β , we expect that the equilibrium number of traits A^* carried by an individual will also not vary with β . This is indeed the case and for small β , A^* is given by eqn 20, whether or not individuals can acquire traits from past learning periods (see Fig. 2 and eqn 32). For large values of β , A^* may increase as a function of β (Fig. 3). Here, social learning increases the average fitness of individuals in the population. This may be explained by noting that even if $\beta_n = 1$ and $\beta_o = 0$, there are some traits produced in a given time period that cannot be acquired by social learning because of the trade-off between producing and

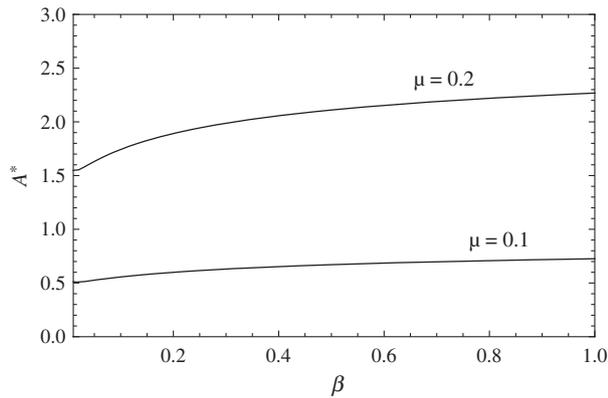


Fig. 3 Expected number of traits A^* carried by an individual at an evolutionary equilibrium as functions of $\beta = \beta_o = \beta_n$ with parameters values $N = 1000$, and $\epsilon = 0.1$, and two values of the mutation rate (the upper curve is for $\mu = 0.2$, and the lower is for $\mu = 0.1$).

scrounging (at an ESS an individual spends at least some fraction of its time producing, otherwise there would be no cultural traits). Allowing an individual to acquire traits generated in past time periods by social learning in that case ($\beta_o > 0$) and holding the level of producing constant, an individual can then acquire more adaptive traits. But this effect, where social learning now tends to increase the mean fitness of individuals in the population because it allows to acquire traits that would otherwise not be possible to acquire, is not strong (lower panel of Fig. 3), but it illustrates that whether social learners affect A^* may depend on the functional relationship between trait values expressed by individuals and cultural dynamics.

In summary, if each individual has a small probability of interacting with each other individual in the population, the ES trait number expressed by an individual is independent of the transmission rate β so that social learning does not increase the average fitness of individuals in the population and affects only the frequency of producers (eqn 21). This result is only weakly affected by allowing for large β values but we will see that it is strongly affected by allowing cultural transmission among relatives.

Population structure: effect of relatives

We now investigate the effect of introducing interactions between relatives on the coevolutionary dynamics of producing and learning. This may be important in groups of small size under limited dispersal (e.g. spatially structured populations) or when family members interact. These two cases are taken into account by letting the parameter κ be positive, so that we have also to take into account the effect of neighbours on the fitness of a focal individual when computing the selection gradients on producing and learning (see eqn 1). That is, we need to

evaluate not only the direct selective pressure ($\partial f/\partial z_o$) on a focal trait z (either l or p) but also the indirect selective pressure ($\partial f/\partial z_o$). We evaluated the direct selective pressure on learning and producing in the last section (e.g. eqns 13 and 15); it now remains to evaluate the indirect selective pressure on these two traits. To that end, we first assume that individuals only acquire traits through social learning that were generated in the current period ($\beta_n > 0$ and $\beta_o = 0$).

Joint ES level of learning and producing

The change in the fecundity of the focal individual because of its patch mates spending one additional unit of time learning instead of producing resources (indirect selective pressure on learning) is obtained by inserting eqns 8 and 26 into $\partial f/\partial l_o$ and holding producing constant ($p_o = p_o = p$), whereby

$$\frac{\partial f}{\partial l_o} = \frac{(1-l)(1-p)n_s}{\epsilon}. \quad (22)$$

This selective pressure is positive for all parameter values so that the net selective pressure on learning, which is obtained from $\partial f/\partial l_o + \kappa \partial f/\partial l_o$ by combining eqns 13 and 22, increases as a result of interactions taking place between relatives.

The change in the fecundity of the focal individual because of its patch mates spending one additional unit of time producing instead of scrounging (indirect selective pressure on producing) is obtained by inserting eqns 8 and 26 into $\partial f/\partial p_o$ and holding learning at its monomorphic values ($l_o = l_o = l$) to give

$$\frac{\partial f}{\partial p_o} = \frac{(1-l)l(1-p)n_s}{\epsilon p}. \quad (23)$$

This term is also always positive because by investing more into producing, the patch mates of the focal individual increase the number of traits it receives through social learning. Hence, the net selective pressure on producing, which is obtained from $\partial f/\partial p_o + \kappa \partial f/\partial p_o$ with eqns 15 and 23, will increase as a result of the positive indirect effect.

Solving $\partial f/\partial l_o + \kappa \partial f/\partial l_o = 0$ (with eqns 13 and 22) and $\partial f/\partial p_o + \kappa \partial f/\partial p_o = 0$ (with eqns 15 and 23) for l and p gives the candidate ES value of producing as

$$p^* = \frac{1}{1 + \kappa} \left[\kappa + \frac{1}{l^* \beta_n (N-1)} \right], \quad (24)$$

where l^* is a somewhat complicated functions of the parameters (β_n , μ , ϵ , N , and κ ; see eqn 34). By comparing eqn 24 with eqn 16, we see that the level of producing increases with κ , approximately by the constant factor κ when this parameter is small, and it was observed numerically that l^* tends to be an increasing function of κ for the whole range of parameter values we investigated. Hence, individuals spend more time learning and

producing novel traits when interactions occur between relatives.

ES trait number

The previous analyses show that both the ES levels of learning (eqn 34) and producing (eqn 24) will be greater in the presence of interactions taking place between relatives ($\kappa > 0$), which suggests that the associated stable number A^* of cultural traits carried by an individual may also increase. This is indeed the case and on substitution of eqn 24 into eqn 9 (and setting $\beta_o = 0$), we have

$$A^* = \frac{\mu}{\epsilon} \left[l^* + \kappa \frac{(l^* \beta_n (N-1) - 1)^2}{\beta_n (N-1)(1+\kappa)^2} \right], \quad (25)$$

which is graphed in Fig. 4 as a function of κ . The lower panel of Fig. 4 also suggests that A^* now varies more strongly as a function of the transmission rate β_n than

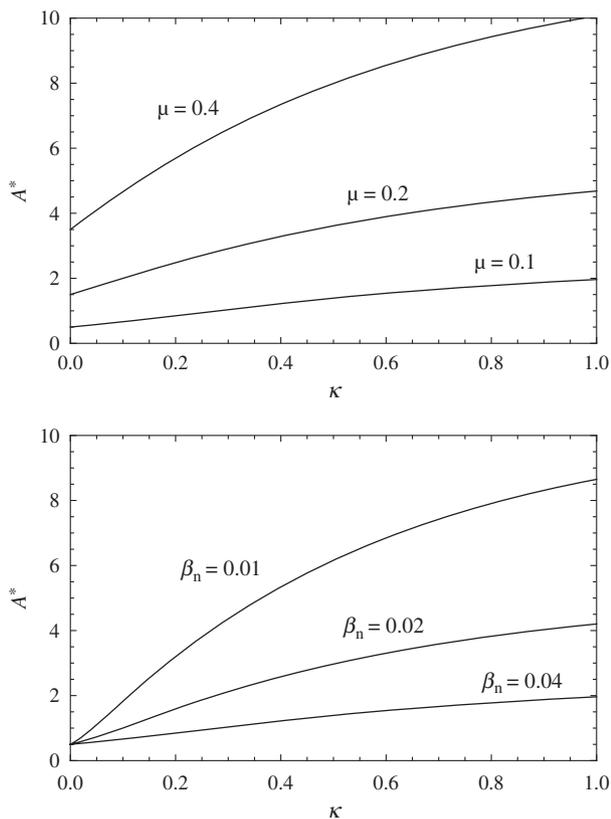


Fig. 4 Expected number A^* of traits carried by an individual at an evolutionary equilibrium as a function of the relatedness coefficient κ when individuals cannot acquire traits from past time periods ($\beta_o = 0$). Upper panel: from the top to the bottom curve $\mu = 0.4, 0.2$, and 0.1 ; the other parameter values are $N = 1000$, $\epsilon = 0.05$, and $\beta_n = 0.01$. Lower panel: from the top to the bottom curve $\beta_n = 0.01, 0.02$, and 0.04 ; the other parameter values are $N = 1000$, $\epsilon = 0.05$, $\mu = 0.1$.

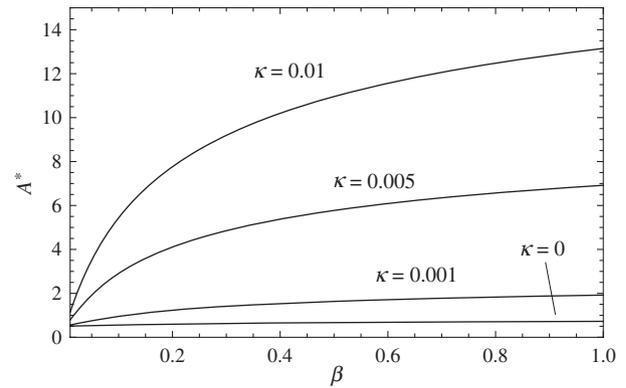


Fig. 5 Expected number A^* of traits carried by an individual at an evolutionary equilibrium as a function of $\beta = \beta_o = \beta_n$ for various values of κ . From the top to the bottom curve $\kappa = 0.01, 0.005, 0.001$, and 0 ; the other parameter values are $N = 1000$, $\epsilon = 0.05$, and $\mu = 0.1$.

was the case when there were no interactions between relatives (Fig. 2).

In Fig. 5, we graph A^* as a function of the transmission rate when $\beta_o = \beta_n = \beta$ (no distinction between traits generated in present and past time periods), and the figure shows that A^* increases greatly with β as κ itself increases. Here, κ and β interact to determine the number of cultural traits carried by an individual. In contrast to the panmictic population case, where the equilibrium number of traits carried by an individual at steady state was approximately independent of β , this number may be strongly affected by β when interactions occur between relatives (Fig. 5), so that social learning may markedly increase the average fitness of individuals in the population.

Discussion

Justification of the model

Cultural transmission, the exchange between individuals of nongenetically determined behaviours, would not be possible if individuals did not imitate or communicate, and it would not exist if individuals did not invent new traits. In this paper, we analysed a stylized model for the coevolution of cultural innovation and transmission in a situation where these features determine the accumulation of learned traits during an individual's lifespan and where the total number of accumulated traits affects an individual's fitness.

The structure of our model is slightly different from classical models of cultural evolution. Instead of postulating the existence of a given cultural trait and focus on the dynamics of different variants of this trait (e.g. Cavalli-Sforza & Feldman, 1981; Boyd & Richerson, 1985), we assumed that each trait is either present or absent and evaluated the dynamics of the number of

different traits present. By contrast to genetically determined traits, the number of different cultural traits carried by an individual may change even during its lifespan, especially in hominids. It is thus interesting to try to understand how many novel cultural traits an individual acquires during its lifespan, how does the accumulation process operate, and what is the selection pressure on it.

Although we did not consider this here, there is nothing that prevents from each trait having different variants, so that complexity may also accumulate at each trait, as in standard models of cultural evolution. Accumulation of distinct traits and gradual evolution within traits are not mutually exclusive processes. On the contrary, models of cultural evolution should probably include these two processes simultaneously to be more realistic, but this is beyond the scope of this paper.

Here, our aim was to investigate the accumulation side of cultural evolution with and without interactions occurring between relatives. To do this, we analysed the coevolution between the proportion of time, l , that an individual spends learning instead of producing resources, and the proportion of learning time, p , spent producing novel cultural traits instead of socially learning them. These two traits determine the number of cultural traits, A , expressed by an individual at the steady state of the learning dynamics, which determines fitness.

We wish to emphasize that the separation of individual learning, p , vs. social learning, $1 - p$, is a limiting assumption, which may be problematic for situations where social interactions between individuals do not result in true cultural inheritance but rather facilitate the process of individual learning. This may be the case in many vertebrates. Hence, our model assumes that social learning results in inheritance of cultural traits.

Coevolution of learning and producing

Learning and producing with unrelated individuals

Under our assumptions about the functional relationship between fitness and steady-state number of cultural traits (eqn 8) and in the absence of interactions between relatives ($\kappa = 0$), evolution leads to a proportion of time spent learning and a trait number per individual that are both increasing functions of the innovation rate, μ , and decreasing functions of the obsolescence rate, ϵ (eqns 17, 20, and Fig. 1). Because the time spent producing is inversely related to the time spent learning, the candidate ES level of producing decreases with μ and increases with ϵ (eqn 18).

These results also apply when in each learning period an individual can acquire traits that it has not acquired previously, provided the transmission rate, β , of cultural traits is low (Fig. 2). A low β value can generally be justified if the population is large, in which case the contact rate between individuals is likely to be low as it is inversely related to population size. This suggests that at

an evolutionary equilibrium, the number of cultural traits carried by an individual in a panmictic population of large size is likely to depend only weakly on the social learning rate β .

Learning and producing with relatives

Adding interactions between relatives increases both the proportions of time spent learning and producing (eqns 22 and 23). This can be understood as follows. When a focal individual increases its net investment into producing, a related neighbour that is scrounging gains additional cultural traits that increase its fitness, thereby increasing the inclusive fitness of the focal individual. The indirect selective pressure on learning and producing thus depends on having social learners in the population (factor $1 - p$ in both eqns 22 and 23), that is, it depends strongly on β , without which the selective pressure on learning and producing would not be increased.

Increase in learning and producing leads to an increase in the number of learned traits carried by an individual at steady state (eqn 34, Fig. 4). In contrast to the panmictic case, the effect of the transmission rate β on the number of cultural traits expressed at steady state by an individual is now important (Fig. 5). There is thus an interaction between the social learning rate β and the relatedness coefficient κ , which suggests that kin selection may, among other factors (e.g. Enquist *et al.*, 2007; Lehmann & Feldman, 2009; Rendell *et al.*, 2010), mediate Rogers's paradox of nonadaptive culture.

Variations in parameter values

Individual variation in innovation and obsolescence rates

We assumed that the innovation rate of cultural traits, μ , is exogenously determined, but this parameter may actually depend on several endogenous factors. For instance, observations in birds suggest that behaviour plays a central role in innovation as some individuals are more attracted by novel objects in their environment (neophilic individuals, for a description of this behaviour in birds see Reader, 2003). These individuals are then more likely to associate objects with each other and then have a higher innovation rate than others.

It has been suggested that responses to novel objects and propensities to innovate are heritable (Reader, 2003). If there are genetic variations in the phenotypes underlying μ , then this parameter may itself be under selection. It has also been shown that the allocation of energy into innovation and learning vs. its allocation into other life-history components might change during an individual's lifespan and differ between the sexes (Laland & Reader, 1999a,b), which suggests that models could be constructed that take into account class-specific (sex, age, stage) innovation rates that may result in cultural structure among age or stages classes.

Individuals may not only vary in their ability to innovate but also in the rate of loss of cultural trait, ϵ ,

which can be interpreted as forgetfulness. For instance, it has been demonstrated that individual's memories evolve across the lifespan. Memory is also known to decrease with age in humans, but poor memory is also found at very young ages (Neisser, 2004) and may be adaptive (Dunlap *et al.*, 2009). Because the mechanisms that affect memory across the lifespan are not well understood, we kept this parameter constant in our model but it could also evolve with an age-specific expression schedule, and one could include density-dependent effects on memory, which were neglected in this present study.

Environmental variation in innovation and obsolescence rates

The innovation rate of cultural traits may also depend on the complexity of the environment, that is, the number of objects encountered by an individual in its environment. The more objects there are in the environment, the more combinations are possible between these objects, and the higher may be the rate of innovation, if everything else remains the same.

Changes in the environment can also force individuals to innovate and therefore increase μ , when the existing traits are not well adapted to a new environment. For instance, new conditions might diminish trophic resources; in guppies, it has been shown that the more food-deprived fish show the highest level of innovation (Laland & Reader, 1999a). Environmental variation may also affect the rate of obsolescence ϵ of the traits, because traits might become maladaptive in a new environment (Rogers, 1988; Galef, 2009). The more rapidly the environment changes, the higher ϵ is likely to be. For instance, one can interpret the value $\epsilon = 1$ as a situation when the environment completely changes from one learning period to the next.

Group size (or 'social environment') may also influence the rate of innovation. For instance, when individuals live in groups, cooperation between them may facilitate the ability to solve complicated tasks (Seed *et al.*, 2008). Also in large groups, the probability of solving a new problem is increased because the number of attempts to solve it is increased. Furthermore, in large groups, variation in individual neophobia, but also in personal experiences, increases the chance of finding solutions in face of environmental changes (Liker & Bokony, 2009).

Transmission rate

The transmission rate parameters (β_n and β_o , or simply β if they are equal) describe the ease with which cultural traits are transmitted from one individual to the other in the population. The parameter β tunes the extent to which it may be profitable to rely on social information instead of private information (e.g. Danchin *et al.*, 2004). It can be thought of as the contact rate between individuals times the rate of transmission of a random

trait expressed by another individual in the population. The higher the transmission rate, the greater the benefits of scrounging. On the other hand, if individual learning is too costly in terms of time and energy, individuals may perform more social learning (Webster & Laland, 2008), holding β constant. In addition, the transmission rate can also be interpreted as capturing the quality of social information and thus can be taken to be low for less adaptive traits copied from others. It has also been shown empirically that social learning may be maladaptive (Laland & Williams, 1998; Galef, 2009), but maladaptations are not directly captured by our model and this deserves further formalization.

The transmission rate β may also capture two different types of transmission processes. The first is the process by which the focal individual simply imitates an exemplar individual, in which case the latter individual plays only a passive role. This may be the case when a new behaviour created by a producer is easily accessible to a social learner (scrounger), who may then imitate the trait that the producer has inadvertently expressed (unintentional communication, Danchin *et al.*, 2004). Second, β , may involve active communication between the focal individual and the individual it interacts with, that is, an iterative process of message exchange, which may end up with the focal individual learning a novel trait. If behaviours are complicated, they might only be acquired by communication, which may also involve teaching, a process that has been demonstrated to occur in several species (Hoppitt *et al.*, 2008; Thornton & Raihani, 2008).

Because communicating is probably costlier (at least in time for the producer) than imitating, the evolution of the ability to communicate (including teaching) cultural traits poses a social dilemma as well (Thornton & Raihani, 2008). We investigated a direct extension of our model where we let communication evolve, but besides adding complexity it did not produce additional qualitative results: the ES levels of learning and producing in the absence of interactions between relatives are weakly affected by β , whereas in the presence of interactions between relatives, communication increases the value of β relative to imitation or copying.

Intergenerational versus intragenerational effects

We analysed our model only in the presence of intragenerational effects (only horizontal transmission). But cultural traits can also be passed on from parents to offspring (vertical transmission). If the number of periods of cultural transmission is small, then such intergenerational effects will affect the number of traits A carried by an individual as it will determine the initial conditions of the system of equations (eqns 2 and 6) describing cumulative cultural dynamics. Our model can be extended to take both inter- and intragenerational effects into account. We carried out an analysis with intergenerational

effects when evolution occurs in a panmictic population and found that such effects do not affect the main qualitative results reported here. However, taking into account intergenerational effects in a subdivided population is more involved and deserves further formalization. It would also be interesting in this situation to include the possibility that individuals copy traits from individuals in other groups.

A case where intergenerational transmission of cultural traits might have some important consequences is when maternal care is important (Estes *et al.*, 1998). This strong social link may increase the rate of transmission between individuals from different generations, which may also occur in the case of grand mothering. The grandmother hypothesis usually invokes the importance of grandmothers taking care of grandchildren, when mothers are not able to look after them (O'Connell *et al.*, 1999). However, we might also hypothesize the importance of their role in the transmission of adaptive cultural traits and even the transmission of obsolete traits. Thus, given the relatively large amount of time between the innovation of a trait and its use in subsequent generations, there is an increase in the probability that an environmental change occurs and, therefore, that traits inherited from grandmothers become obsolete. Hence, under different regimes of environmental fluctuation, the selective pressure on grand mothering might be different as a result of different effects on cultural transmission.

Conclusion

Our theoretical analysis suggests that when interactions occur between individuals in a panmictic population, the number of cultural traits carried by an individual at an evolutionary steady state varies only weakly with variation in the social transmission rate β and depends mainly on the innovation rate, μ , and obsolescence rate, ϵ , of cultural traits [although introducing other features, like critical social learning (Enquist *et al.*, 2007), may change this result]. By contrast, when interactions occur between relatives, the number of cultural traits carried by an individual is markedly affected by the social transmission rate β , which determines the extent to which the relatives of an actor may benefit from the latter producing novel traits. This interaction between transmission and relatedness may be relevant for understanding the evolution of communication, whose role in the evolution of cumulative cultural dynamics remains to be further investigated.

Acknowledgments

We thank R. Bshary and E. van de Waal for comments. This work is supported by grant PP00P3-123344 from the Swiss NSF to LL, by NIH grant GM28016 to MWF, and by a grant FQRNT to RK.

References

- Aiello, L.C. & Wheeler, P. 1995. The expensive-tissue hypothesis: the brain and the digestive-system in human and primate evolution. *Curr. Anthropol.* **36**: 199–221.
- Aoki, K. 1982. A condition for group selection to prevail over counteracting individual selection. *Evolution* **36**: 832–842.
- Barnard, C.J. & Sibly, R.M. 1981. Producers and scroungers: a general model and its application to captive flocks of house sparrows. *Anim. Behav.* **29**: 543–550.
- Berger, J., Swenson, J. & Persson, I. 2001. Recolonizing carnivores and naive prey: Conservation lessons from Pleistocene extinctions. *Science* **291**: 1036–1039.
- Boyd, R. & Richerson, P.J. 1985. *Culture and the Evolutionary Process*. University of Chicago Press, Chicago.
- Boyd, R. & Richerson, P.J. 1995. Why does culture increase human adaptability? *Ethol. Sociobiol.* **16**: 125–143.
- Cavalli-Sforza, L. & Feldman, M.W. 1981. *Cultural Transmission and Evolution*. Princeton University Press, NJ.
- Danchin, E., Giraldeau, L.-A., Valone, T.J. & Wagner, R.H. 2004. Public information: from nosy neighbors to cultural evolution. *Science* **305**: 487–491.
- Dugatkin, L.A. 2004. *Principles of Animal Behavior*. W. W. Norton and Company, London.
- Dunlap, A.S., Mclinn, C.M., Maccormick, H.A., Scott, M.E., & Kerr, B. 2009. Why some memories do not last a lifetime: dynamic long-term retrieval in changing environments. *Behav. Ecol.* **20**: 1096–1105.
- Enquist, M., Eriksson, K. & Ghirlanda, S. 2007. Critical social learning: a solution to Rogers's paradox of nonadaptive culture. *Am. Anthropol.* **109**: 727–734.
- Enquist, M. & Ghirlanda, S. 2007. Evolution of social learning does not explain the origin of human cumulative culture. *J. Theor. Biol.* **246**: 129–135.
- Enquist, M., Ghirlanda, S., Jarrick, A. & Wachtmeister, C.A. 2008. Why does human culture increase exponentially? *Theor. Popul. Biol.* **74**: 46–55.
- Estes, J., Tinker, M., Williams, T. & Doak, D. 1998. Killer whale predation on sea otters linking oceanic and nearshore ecosystems. *Science* **282**: 473–476.
- Galef, B. 2009. Strategies for social learning: testing predictions from formal theory. In *Advances in the Study of Behavior*, Vol. 39 (J. Brockmann, C.T. Snowdon, T.J. Roper, M. Naguib & K.E. Wynne-Edwards, eds), pp. 117–151. Elsevier Academic Press Inc., San Diego.
- Galor, O. & Weil, D.N. 2000. Population, technology, and growth: from Malthusian stagnation to the demographic transition and beyond. *Am. Econ. Assoc.* **90**: 806–828.
- Gandon, S. 1999. Kin competition, the cost of inbreeding and the evolution of dispersal. *J. Theor. Biol.* **200**: 345–364.
- Gardner, A. 2010. Sex-biased dispersal of adults mediates the evolution of altruism among juveniles. *J. Theor. Biol.* **262**: 339–345.
- Gardner, A. & West, S.A. 2006. Demography, altruism, and the benefits of budding. *J. Evol. Biol.* **19**: 1707–1716.
- Ghirlanda, S. & Enquist, M. 2007. Cumulative culture and explosive demographic transitions. *Qual. Quant.* **41**: 581–600.
- Ghirlanda, S., Enquist, M. & Perc, M. 2010. Sustainability of culture-driven population dynamics. *Theor. Popul. Biol.* **77**: 181–188.

- Giraldeau, L., Caraco, T. & Valone, T. 1994. Social foraging: individual learning and cultural transmission of innovations. *Behav. Ecol.* **5**: 35–43.
- Hamilton, W.D. 1971. Selection of selfish and altruistic behaviour in some extreme models. In: *Man and Beast: Comparative Social Behavior* (J. Eisenberg & W. Dillon, eds), pp. 59–91. Smithsonian Institutions Press, Washington, DC.
- Hoppitt, W.J. E., Brown, G., Kendal, R., Rendell, L., Thornton, A., Webster, M. & Laland, K. 2008. Lessons from animal teaching. *Trends Ecol. Evol.* **23**: 486–493.
- Kremer, M. 1993. Population growth and technological change: one Million BC to 1990. *Q. Rev. Econ.* **108**: 681–716.
- Laland, K. & Williams, K. 1998. Social transmission of maladaptive information in the guppy. *Behav. Ecol.* **9**: 493–499.
- Laland, K.N. & Janik, V.M. 2006. The animal cultures debate. *Trends Ecol. Evol.* **21**: 542–547.
- Laland, K.N. & Reader, S. 1999a. Foraging innovation in the guppy. *Anim. Behav.* **57**: 331–340.
- Laland, K.N. & Reader, S. 1999b. Foraging innovation is inversely related to competitive ability in male but not in female guppies. *Behav. Ecol.* **10**: 270–274.
- Lefebvre, L. 1995. Culturally-transmitted feeding behaviour in primates: Evidence for accelerating learning rates. *Primates* **36**: 227–239.
- Lefebvre, L. & Bolhuis, J.J. 2003. Positive and negative correlates of feeding innovations in birds: evidence for limited modularity. In: *Animal Innovation* (K.N. Laland & S.M. Reader, eds), pp. 39–61. Oxford University Press, Oxford.
- Lefebvre, L., Nicolakakis, N. & Boire, D. 2002. Tools and brains in birds. *Behaviour* **139**: 939–973.
- Lehmann, L., Aoki, K. & Feldman, M.W. 2010. On the number of independent cultural traits carried by individuals and populations. *Philos. Trans. R. Soc. Lond. B Biol. Sci.*, in press.
- Lehmann, L. & Feldman, M.W. 2009. Coevolution of adaptive technology, maladaptive culture, and population size in a producer-scrounger game. *Proc. R. Soc. Lond. B Biol. Sci.* **276**: 3853–3862.
- Lehmann, L., Perrin, N. & Rousset, F. 2006. Population demography and the evolution of helping behaviors. *Evolution* **60**: 1137–1151.
- Lehmann, L. & Rousset, F. 2010. How life-history and demography promote or inhibit the evolution of helping behaviors. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **365**: 2599–2617.
- Liker, A. & Bokony, V. 2009. Larger groups are more successful in innovative problem solving in house sparrows. *Proc. Natl. Acad. Sci. USA* **106**: 7893–7898.
- Lion, S. & Gandon, S. 2009. Habitat saturation and the spatial evolutionary ecology of altruism. *J. Evol. Biol.* **22**: 1487–1502.
- Lion, S. & van Baalen, M. 2007. Self-structuring in spatial evolutionary ecology. *Ecol. Lett.* **11**: 277–295.
- Neisser, U. 2004. Memory development: new questions and old. *Dev. Rev.* **24**: 154–158.
- O'Connell, J.F., Hawkes, K. & Jones, N.G.B. 1999. Grandmothering and the evolution of *Homo erectus*. *J. Hum. Evol.* **36**: 461–485.
- Queller, D.C. 1994. Genetic relatedness in viscous populations. *Evol. Ecol.* **8**: 70–73.
- Reader, S. 2003. *Innovation and Social Learning: Individual Variation and Brain Evolution*. Brill Academic Publishers, Leiden.
- Rendell, L., Fogarty, L. & Laland, K.N. 2010. Rogers's paradox recast and resolved: population structure and the evolution of social learning strategies. *Evolution* **64**: 534–548.
- Rogers, A.R. 1988. Does biology constrain culture? *Am. Anthropol.* **90**: 819–831.
- Rogers, A.R. 1990. Group selection by selective emigration: the effects of migration and kin structure. *Am. Nat.* **135**: 398–413.
- Romer, D. 2006. *Advanced Macroeconomics*, 3rd edn. McGraw-Hill, Boston.
- Rousset, F. 2004. *Genetic Structure and Selection in Subdivided Populations*. Princeton University Press, Princeton, NJ.
- Roze, D. & Rousset, F. 2005. Inbreeding depression and the evolution of dispersal rates: a multilocus model. *Am. Nat.* **166**: 708–721.
- Roze, D. & Rousset, F. 2008. Multilocus models in the infinite island model of population structure. *Theor. Popul. Biol.* **73**: 529–542.
- Seed, A.M., Clayton, N.S. & Emery, N.J. 2008. Cooperative problem solving in rooks (*Corvus frugilegus*). *Proc. R. Soc. Lond. B Biol. Sci.* **275**: 1421–1429.
- Sol, D., Bacher, S., Reader, S.M. & Lefebvre, L. 2008. Brain size predicts the success of mammal species introduced into novel environments. *Am. Nat.* **172**(Suppl. 1) S63–S71.
- Sol, D., Duncan, R., Blackburn, T., Cassey, P. & Lefebvre, L. 2005. Big brains, enhanced cognition, and response of birds to novel environments. *Proc. Natl. Acad. Sci. USA* **102**: 5460–5465.
- Strimling, P., Sjöstrand, J., Enquist, M. & Eriksson, K. 2009. Accumulation of independent cultural traits. *Theor. Popul. Biol.* **76**: 77–83.
- Taylor, P.D. 1992. Altruism in viscous populations - an inclusive fitness model. *Evol. Ecol.* **6**: 352–356.
- Taylor, P.D. 1996. Inclusive fitness arguments in genetic models of behaviour. *J. Math. Biol.* **34**: 654–674.
- Taylor, P.D. & Irwin, A.J. 2000. Overlapping generations can promote altruistic behavior. *Evolution* **54**: 1135–1141.
- Thornton, A. & Raihani, N.J. 2008. The evolution of teaching. *Anim. Behav.* **75**: 1823–1836.
- Wakano, J.Y., Aoki, K. & Feldman, M.W. 2004. Evolution of social learning: a mathematical analysis. *Theor. Popul. Biol.* **66**: 249–258.
- Webster, M.M. & Laland, K.N. 2008. Social learning strategies and predation risk: minnows copy only when using private information would be costly. *Proc. R. Soc. Lond. B Biol. Sci.* **271**: 957–962.
- West, S.A., Griffin, A.S. & Gardner, A. 2007. Evolutionary explanations for cooperation. *Curr. Biol.* **17**: 661–672.
- Whiten, A., Goodall, J., McGrew, W.C., Nishida, T., Reynolds, V., Sugiyama Y., Tutin, C.E.G., Wrangham, R.W. & Boesch, C. 1999. Cultures in chimpanzees. *Nature* **399**: 682–685.

Appendix

Steady-state number of cultural traits

The system of recurrence equations presented in the main text for $A_{\bullet,t}$ and $O_{\bullet,t}$ (eqns 2 and 6) is linear and can be solved analytically. At steady state, when $t \rightarrow \infty$, $A_{\bullet,t} = A_{\bullet,t-1}$ and $O_{\bullet,t} = O_{\bullet,t-1}$ and from eqns 2 and 6, we find that

$$A_{\bullet} = \frac{l_{\bullet}(p_{\bullet}\mu + (1 - p_{\bullet})[\beta_n(N - 1)l_0p_0\mu + \beta_o(1 - \epsilon)O_{\bullet}])}{\epsilon}, \quad (26)$$

where

$$O_{\bullet} = \frac{[1 - l_{\bullet}(1 - p_{\bullet})\beta_n](N - 1)l_0p_0\mu}{1 - [1 - l_{\bullet}(1 - p_{\bullet})\beta_o](1 - \epsilon)}. \quad (27)$$

In a monomorphic population $l_0 = l_{\bullet} = l$ and $p_0 = p_{\bullet} = p$, then eqns 26 and 27 reduce to eqns 9 and 10 of the main text.

Low β approximation

Inserting eqn 8 and eqns 26 and 27 into $\partial f/\partial l_{\bullet}$ and holding producing constant give the gradient of selection on learning as

$$\frac{\partial f}{\partial l_{\bullet}} = \frac{p\mu + (1 - p)[n_s + \beta_o(1 - \epsilon)O](1 - l)}{\epsilon} + \frac{l(1 - p)\beta_o(1 - \epsilon)[\partial O_{\bullet}/\partial l_{\bullet}](1 - l)}{\epsilon} - (1 + A), \quad (28)$$

where A is given by eqn 9, and it can be seen from eqn 27 that $\partial O_{\bullet}/\partial l_{\bullet} \leq 0$. If $\beta_o = \beta_n = \beta$, a first-order Taylor expansion of the selective pressure around $\beta = 0$ gives

$$\frac{\partial f}{\partial l_{\bullet}} = \frac{\{p\mu + (1 - p)[n_s + \beta(1 - \epsilon)(N - 1)lp\mu/\epsilon]\}(1 - l)}{\epsilon} - \left(1 + \frac{l\{p\mu + (1 - p)[n_s + \beta(1 - \epsilon)(N - 1)lp\mu/\epsilon]\}}{\epsilon}\right) + \mathcal{O}(\beta^2), \quad (29)$$

where we used $O = [(N - 1)lp\mu]/\epsilon + \mathcal{O}(\beta)$ and $\partial O_{\bullet}/\partial l_{\bullet} = \mathcal{O}(\beta)$.

Inserting eqn 8 and eqns 26 and 27 into $\partial f/\partial p_{\bullet}$ and holding learning constant give the gradient of selection on producing as

$$\frac{\partial f}{\partial p_{\bullet}} = \frac{l(1 - l)}{\epsilon} \left(\mu - [n_s + \beta_o(1 - \epsilon)O] + (1 - p)\beta_o(1 - \epsilon) \frac{\partial O_{\bullet}}{\partial p_{\bullet}} \right), \quad (30)$$

and it can be seen from eqn 27 that $\partial O_{\bullet}/\partial p_{\bullet} \geq 0$. When $\beta_o = \beta_n = \beta$, a first-order Taylor expansion of eqn 30 around $\beta = 0$ gives

$$\frac{\partial f}{\partial p_{\bullet}} = \frac{(1 - l)l\mu}{\epsilon} \left(1 - \frac{\beta(N - 1)lp}{\epsilon} \right) + \mathcal{O}(\beta^2). \quad (31)$$

Setting the partial derivatives in eqns 29 and 31 to zero, neglecting terms $\mathcal{O}(\beta^2)$, and solving for l and p give $p = \epsilon/[l\beta(N - 1)]$ and $l = (\mu - \epsilon)/(2\mu)$.

If $\beta_o = \beta_n = \beta$, a first-order Taylor expansion of eqn 9 around $\beta = 0$ gives

$$A = \frac{lp\mu}{\epsilon} \left(1 + \frac{\beta(N - 1)l(1 - p)}{\epsilon} \right) + \mathcal{O}(\beta^2). \quad (32)$$

Neglecting $\mathcal{O}(\beta^2)$ and inserting $p = \epsilon/[l\beta(N - 1)]$ and $l = (\mu - \epsilon)/(2\mu)$ into this equation produce $A = (\mu - \epsilon)/(2\epsilon)$, which is eqn 20 of the main text.

ESS learning and producing with relatives

Solving $\partial f/\partial l_{\bullet} + \kappa \partial f/\partial l_0 = 0$ (with eqns 13 and 22) and $\partial f/\partial p_{\bullet} + \kappa \partial f/\partial p_0 = 0$ (with eqns 15 and 23) for l and p gives the candidate ES value of producing

$$p^* = \frac{1}{1 + \kappa} \left[\kappa + \frac{1}{l^* \beta_n (N - 1)} \right], \quad (33)$$

where the ES candidate level of learning is given by a somewhat complicated formula:

$$l^* = \frac{\{\kappa(\kappa + 1)[(N - 1)\beta - 1] - 2\}\sqrt{\mu}}{2(N - 1)\kappa(\kappa + 2)\beta\sqrt{\mu}} + \frac{(\kappa + 1)\sqrt{(\kappa[(N - 1)\beta - 1] + 2)^2\mu - 4(N - 1)\kappa(\kappa + 2)\beta\epsilon}}{2(N - 1)\kappa(\kappa + 2)\beta\sqrt{\mu}}. \quad (34)$$

Received 25 May 2010; revised 20 July 2010; accepted 28 July 2010