Evolutionarily stable learning schedules and cumulative culture in discrete generation models

Kenichi Aoki [a,*], Joe Yuichiro Wakano [b], Laurent Lehmann [c]

[a] Department of Biological Sciences, The University of Tokyo, Hongo 7-3-1, Bunkyo-ku, Tokyo 113-0033, Japan
[b] Meiji Institute for Advanced Study of Mathematical Sciences, Meiji University, 1-1-1 Higashi-Mita, Tama-ku, Kawasaki, Kanagawa 214-8571, Japan
[c] Department of Ecology and Evolution, UNIL Sorge, Le Biophore, CH-1015 Lausanne, Switzerland

ARTICLE INFO

ABSTRACT

Individual learning (e.g., trial-and-error) and social learning (e.g., imitation) are alternative ways of acquiring and expressing the appropriate phenotype in an environment. The optimal choice between using individual learning and/or social learning may be dictated by the life-stage or age of an organism. Of special interest is a learning schedule in which social learning precedes individual learning, because such a schedule is apparently a necessary condition for cumulative culture. Assuming two obligatory learning stages per discrete generation, we obtain the evolutionarily stable learning schedules for the three situations where the environment is constant, fluctuates between generations, or fluctuates within generations. During each learning stage, we assume that an organism may target the optimal phenotype in the current environment by individual learning, and/or the mature phenotype of the previous generation by oblique social learning. In the absence of exogenous costs to learning, the evolutionarily stable learning schedules are predicted to be either pure social learning followed by pure individual learning ("bang–bang" control) or pure individual learning at both stages ("flat" control). Moreover, we find for each situation that the evolutionarily stable learning schedule is also the one that optimizes the learned phenotype at equilibrium.

© 2012 Elsevier Inc. All rights reserved.

1. Introduction

Individual learning (IL; learning by oneself, e.g., by trial-and-error) and social learning (SL; learning from others, e.g., by imitation) are alternative ways of acquiring and expressing the appropriate phenotype in an environment (Galef, 1988; Whiten and Ham, 1992; Heyes, 1993). The two terms are roughly synonymous with producing and scrounging, respectively (Barnard and Sibly, 1981; Giraldeau et al., 1994; Laland, 2004; Arbilly et al., 2010, 2011). Both IL and SL are widely distributed among insects, fish, birds, and mammals (Dugatkin, 2004). For example, honeybee workers can search independently (use IL) or follow waggle dances (use SL) to find food (Leadbeater and Chittka, 2007). Similarly, primates can devise novel solutions to the problems they face during their lifetime (IL) or copy extant solutions from others in the population (SL; e.g., Dugatkin, 2004).

Since learning is costly in time or energy, organisms are often subject to a trade-off between IL and SL. This raises the question of when an organism should copy others (use SL) rather than directly sample the environment (use IL). Laland (2004) suggests

three relevant contexts (when an organism should use SL): (1) copy when established behavior (innate or otherwise) is unproductive, (2) copy when IL is costly, (3) copy when uncertain. To this list of "when" situations we wish to add a fourth possibility. Namely, the optimal choice between IL and SL may be dictated by the life-stage or age of an organism. Put another way, natural selection may favor certain broad developmental patterns in the differential use of IL and SL, with regard to phenotypes that may take a long time to be acquired.

There are few, if any, definitive empirical studies that bear on the ontogeny of IL versus SL with which to motivate such a proposal. First, in a comparative survey of the primate literature, Reader and Laland (2001) found that the reported incidence of innovations is higher in adults than non-adults, given the relative proportions of these groups. This observation suggests that as primates mature, they are more likely to rely on IL in an absolute sense and perhaps relative to SL as well. Second, in a study of honeybee workers throughout their foraging careers, Biesmeijer and Seeley (2005) noted that a novice forager is more likely than an experienced forager to use the information from waggle dances to guide her search (being recruited) than to search independently (scouting). (Clearly, factors such as availability of food sources also affect the division of labor between recruits and scouts, Beekman et al., 2007.) These studies imply that primates and honeybees may depend more on SL early in ontogeny and more on IL later.

* Corresponding author.
E-mail address: kenaoki@biol.s.u-tokyo.ac.jp (K. Aoki).

Third, Biondi et al. (2010a,b) compared the SL and IL proficiencies of juvenile and adult raptors (*Milvago chimango*), finding that juveniles were better than adults at learning from demonstrators (SL). However, they also discovered the same tendency for problem-solving (IL). Fourth, Cole et al. (2011) compared problem-solving (IL) abilities in juvenile and adult great tits, but did not observe a statistically significant main effect of age. Hence, these latter studies provide no evidence for a differential life-stage dependence on IL and SL.

SL by the waggle dance results in the (usually) horizontal transfer between honeybee workers of information regarding ephemeral food sources. As such, the information is only transiently useful and does not contribute to the formation of traditions (Leadbeater and Chittka, 2007). Arguably, it is the transmission of relatively stable behaviors across generations by vertical and/or oblique SL that is of greater interest. Moreover, in organisms where there is a transfer of information across generations, cumulative culture becomes a possibility (Tomasello, 1999; Tennie et al., 2009). Cumulative culture can only occur if organisms are equipped with the ability to accurately absorb the extant culture—complete accuracy is however not required (see Fig. 2)—and then to build on it. In other words, a necessary but not sufficient condition is that organisms use a composite learning strategy in which SL precedes IL (Boyd and Richerson, 1985; Enquist et al., 2007; Borenstein et al., 2008; Aoki, 2010). Although cumulative culture supported by this ordering of SL and IL may explain the ecological success of modern humans, it is not clear under what conditions such a learning schedule might have evolved (if indeed it did).

According to Vygotsky (1978, p. 86), (human) children can do some things under an adult's guidance (by SL) that they cannot do on their own (by IL) until a few years later ("zone of proximal development"). More recently, Striano et al. (2001) conducted experiments in which small children 2, 2.5, and 3 years old were tested for their ability to pretend play (e.g., making a carrot walk on the floor like a man). Observations were made on whether the children reproduced the actions shown to them by an adult experimenter or creatively produced novel actions not previously demonstrated during the experiment. Striano et al. (2001) showed that the proportion of novel actions as opposed to imitative actions increases with age, which suggests that reliance on IL versus SL also increases with age. An alternative interpretation is that older children have had more exposure to various kinds of pretend play outside the laboratory, and hence that actions coded as novel may in fact be "deferred imitation" (Striano et al., 2001).

In a review of hunter–gatherer diet and their possible consequences for human life history evolution, Kaplan et al. (2000) describe how hunter–gatherers utilize nutrient-dense, difficult-to-acquire foods such as roots, honey, and vertebrate meat. As a result, naive children produce little food on their own and require a long developmental period during which the necessary extraction and/or hunting skills can be learned. For example, in the case of vertebrate meat, which is procured by males, maximum productivity is attained only in middle age. Their data suggest that a long-term "apprenticeship" (i.e., SL) and much "practice" (i.e., trial-and-error IL), perhaps occurring in this order, are required before a member of hunter–gatherer society can be self-sufficient and moreover produce a surplus to support his/her offspring. The gradual improvement of foraging skills by learning is also seen in nonhuman primates (e.g., Gunst et al., 2010).

In this paper we obtain evolutionary predictions on the scheduling of IL and SL. We describe and analyze mathematical models for the scheduling of IL and SL, where an organism must learn a continuous phenotype, which has an optimal expression that depends on the current environment. An organism is assumed to experience two obligatory learning opportunities (which we call stage 0 and stage 1 learning) per discrete generation. During

each stage, it may apply IL which allows it to track (the changes in) the environment, oblique SL which allows it to access the information possessed by the previous generation, or a mixture of the two with complementary weights (see below). The phenotype resulting from the two-stage cumulative learning process during the lifespan of an organism determines its fitness.

Previous analytical models have made the following restrictive assumptions: (1) an organism may use a mixed strategy, but there is only one opportunity for learning (e.g., Boyd and Richerson, 1988; Feldman et al., 1996; Wakano and Aoki, 2006), or the learning schedule is the same across multiple learning stages (Lehmann et al., 2010); (2) a strategy with two learning stages is recognized, but the organism is compelled to use either pure IL or pure SL during each stage (e.g., Boyd and Richerson, 1985, Enquist et al., 2007, Borenstein et al., 2008 and Aoki, 2010); (3) time constraints on learning are ignored such that composite strategies in which IL and SL are applied sequentially compete directly with simple "one-shot" strategies using only pure IL or pure SL (e.g., Enquist et al., 2007, Borenstein et al., 2008 and Aoki, 2010).

Using the theory of evolutionarily stable (ES) strategies (Maynard Smith, 1982), we identify the ES learning schedules for the three situations (models) where the environment is constant, fluctuates between generations, or fluctuates within generations. Environmental variability is known theoretically to be an important factor in the evolution of learning (Boyd and Richerson, 1985; Rogers, 1988; Feldman et al., 1996). In the absence of exogenous costs to either IL or SL, we find that the ES learning schedule is either pure SL followed by pure IL ("bang–bang" control, Sydsæter et al., 2008), or pure IL at both stages (we use the descriptive neologism "flat" control).

In evolutionary games between unrelated individuals, each organism "tries" to maximize its personal fitness, with the consequence that the mean fitness or "efficiency" of the population is not necessarily maximized. In our case, if all organisms in a panmictic population adopt an ES learning schedule in which SL precedes IL, cumulative cultural evolution may or may not result in a phenotype that most closely approaches the optimal phenotype in the given environment, and population fitness will be maximized only if it does. Hence, for the three models, we consider whether the ES learning schedule is also the one that "optimizes" the learned phenotype at equilibrium.

2. Models and mathematical analyses

2.1. Overview

We consider a panmictic population of infinite size with asexual organisms that are semelparous. Generations overlap only insofar as oblique SL occurs. We assume that organisms can obtain information about the environment, or possible responses to that environment, directly by individual learning (IL) and/or indirectly by social learning (SL). This information is then expressed as the phenotype, which is assumed to vary continuously.

The phenotypes that we have in mind are those that are usually regarded as being acquired and expressed by learning, such as the choice of food by rats (e.g., Aisner and Terkel, 1992), the quantity of food foraged by some primates (e.g., Kaplan et al., 2000 and Gunst et al., 2010), the length of a stone artifact (e.g., Lycett and Gowlett, 2008), or canoe design features (e.g., Rogers and Ehrlich, 2008). Of these, the second, third, and (some of the) fourth are clearly continuous phenotypes. We specifically wish to understand the evolution of learning schedules in the (many) species where IL and SL are already well-established. As such, we ignore the possibility that the phenotype may be innately determined.

The phenotype of a naive newborn before any learning occurs will be called the "initial phenotype" and is set to 0. This requires

that the continuous scale on which the phenotype is measured be suitably transformed. We use the terms information and phenotype synonymously. There are two learning opportunities per generation (stage 0 and stage 1 learning), during each of which an organism can use either IL or SL or a mixture of the two to update its phenotype. Learning strategies, which are genetically determined, assign weights (between 0 and 1) to the information gathered by IL and by SL during each stage. We refer to the phenotype after stage 0 learning as the "intermediate phenotype" and the phenotype after stage 1 learning as the "mature phenotype". With each learning strategy are associated one intermediate phenotype and one mature phenotype that are expressed sequentially. Weights should not be confused with probabilities in a mixed strategy that determine which phenotype among multiple alternatives is expressed at any one time. Only the mature phenotype contributes to fitness. Post-learning organisms reproduce in proportion to their fitness (fertility selection) and survive to serve as exemplars (cultural parents) for the next generation.

The target for IL is the optimal phenotype in the current state of the environment, which may fluctuate within or between generations. The target for SL is always the mature phenotype of the previous generation if the population is genetically monomorphic, or the population mean of the mature phenotypes if more than one learning strategy is segregating. The efficiencies of IL and SL during stage i learning are α_i and β_i ($i = 0, 1$), respectively, where efficiency is defined as the proportional reduction in the deviation from the target and assumed to be the same for all learning strategies. We assume $0 < \alpha_i$, $\beta_i < 1$ ($i = 0, 1$), which entails that the mature phenotype approaches but never converges to the optimal phenotype (see Eqs. (5), (13) and (19)).

Our model is a generalization, in some respects, of the "guided variation" model proposed by Boyd and Richerson (1985). The recursions detailed below are in standard use in the modeling of the social transmission of continuous phenotypes (Cavalli-Sforza and Feldman, 1981), and hence in the evolutionary theory of learning.

2.2. Constant environment

Denote the optimal phenotype by z^*, which is a positive constant, and the mature phenotype of the resident (wild-type) strategy (schedule) in generation t by z_t. This strategy assigns weight u_i (the control) to IL and complementary weight $1 - u_i$ to SL during stage i learning, where $0 \le u_i \le 1$ ($i = 0, 1$). We assume that, during each learning stage, the incremental change in the phenotype due to learning is proportional to the distance between the current phenotype and the target phenotype, which is z^* for IL and the mature phenotype of the previous generation for SL. The constants of proportionality are the efficiencies of learning, i.e., α_i and β_i ($i = 0, 1$).

Since the initial phenotype is set to 0, pure IL would produce the phenotype $\alpha_0 z^*$ during stage 0 of generation $t + 1$, whereas pure SL would yield the phenotype $\beta_0 z_t$. Hence, taking the weights in consideration, we have the intermediate phenotype

$$z_{t+1}^{\text{int}} \equiv u_0 \alpha_0 z^* + (1 - u_0) \beta_0 z_t. \tag{1}$$

On further applying pure IL or pure SL during stage 1, the incremental changes in the phenotype would be $\alpha_1(z^* - z_{t+1}^{\text{int}})$ and $\beta_1(z_t - z_{t+1}^{\text{int}})$, respectively. Thus, the mature phenotype is

$$z_{t+1} = z_{t+1}^{\text{int}} + u_1 \alpha_1(z^* - z_{t+1}^{\text{int}}) + (1 - u_1)\beta_1(z_t - z_{t+1}^{\text{int}}). \tag{2}$$

See Fig. 1 for a schematic diagram illustrating the process described here.
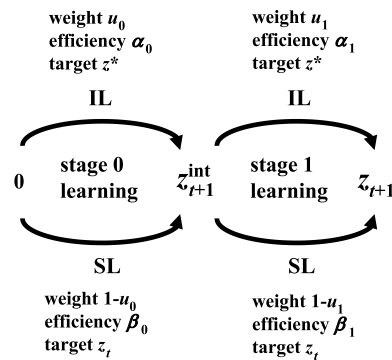

Fig. 1. A schematic diagram illustrating the two-stage learning process before reproduction in generation $t + 1$. The initial phenotype is 0 by assumption. The intermediate phenotype is z_{t+1}^{int}, and the mature phenotype is z_{t+1}. See text for details.

Substituting Eq. (1) into Eq. (2), the recursion in the mature phenotype across generations can be written as

$$z_{t+1} = A z_t + B z^*, \tag{3}$$

where

$$A = (1 - u_0)\beta_0 [1 - u_1 \alpha_1 - (1 - u_1)\beta_1] + (1 - u_1)\beta_1, \tag{4a}$$

$$B = u_0 \alpha_0 [1 - u_1 \alpha_1 - (1 - u_1)\beta_1] + u_1 \alpha_1. \tag{4b}$$

At equilibrium,

$$\hat{z} = B z^* / (1 - A), \tag{5}$$

where $0 < A < 1$ and $0 < B < 1$. The two panels in Fig. 2 illustrate the approach to equilibrium during the first five generations for two different sets of weights.

Next introduce a rare mutant strategy of small effect into the population of resident strategists at the equilibrium defined by Eq. (5). We assume that only two learning strategies, the resident and the mutant, can simultaneously segregate in the population. Denote the control for this mutant strategy by u_i^m ($i = 0, 1$), and let A_m and B_m be the functions of u_i^m ($i = 0, 1$) obtained by substituting $u_i = u_i^m$ ($i = 0, 1$) in Eqs. (4a) and (4b), respectively. Then, as long as the mutant strategy remains rare, the mature phenotype of the mutant strategy is given approximately by

$$z^m \equiv A_m \hat{z} + B_m z^* \tag{6}$$

(Appendix A).

We assume that the mature phenotype determines the fitness of an organism, where deviations from the optimal phenotype, z^*, are penalized. Explicitly, we write the objective function as
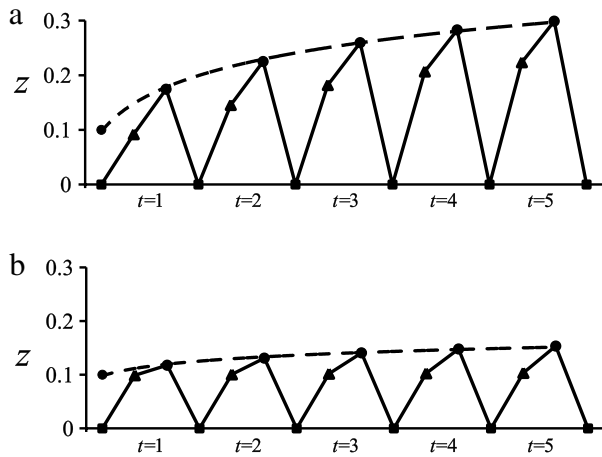
$$F\left(u_0^m, u_1^m, u_0, u_1\right) = \exp\left\{-\frac{(z^m - z^*)^2}{w}\right\}, \tag{7}$$

which is Gaussian with intensity of selection parameter w.

An ES control u_i^* ($i = 0, 1$) is one such that $F\left(u_0^m, u_1^m, u_0, u_1\right)$ takes a (local) maximum when $u_0^m = u_0 = u_0^*$ and $u_1^m = u_1 = u_1^*$ (Maynard Smith, 1982). Set

$$s_0 = \left.\frac{\partial F}{\partial u_0^m}\right|_{u_0^m = u_0, u_1^m = u_1}, \qquad s_1 = \left.\frac{\partial F}{\partial u_1^m}\right|_{u_0^m = u_0, u_1^m = u_1}, \tag{8}$$

which represent the selection gradients on the mutant control for stage 0 and stage 1 learning, respectively. The partial derivatives in Eq. (8) follow from our assumption that each of the two controls, u_0 and u_1, evolves by holding the other control constant (i.e., no pleiotropy or double mutants). In other words, only one mutant control, u_0^m or u_1^m, deviates from the resident control at any one time. Nevertheless, we use the notation of Eq. (7)—which gives the

Fig. 2. Saw-tooth diagrams illustrating the cultural dynamics of the phenotype (here denoted by z) in a constant environment over five generations for a cohort adopting the same learning schedule (u_0, u_1). Panel (a) $u_0 = 0.2, u_1 = 0.9$: primarily SL during stage 0 and primarily IL during stage 1. Panel (b) $u_0 = 0.9, u_1 = 0.2$: primarily IL during stage 0 and primarily SL during stage 1. Solid lines depict the change in the phenotype in each generation. The initial phenotype (squares) can be set to 0 without loss of generality. The intermediate phenotype (triangles) is achieved by a mixture of individual learning (IL, weight u_0) and social learning (SL, weight $1 - u_0$) during stage 0 (Eq. (1)). This intermediate phenotype is then upgraded during stage 1 (Eq. (2)), again by a mixture of IL (weight u_1) and SL (weight $1 - u_1$), to yield the mature phenotype (circles). The broken curve represents the cultural dynamics of the mature phenotype (Eq. (3)), which can be seen to increase from its initial value of 0.1 toward its equilibrium value (Eq. (4)). The optimal phenotype is $z^* = 1$. The increase over generations in the mature phenotype can be interpreted as the accumulation of cultural information. The other parameters are $\alpha_0 = \alpha_1 = 0.1$ and $\beta_0 = \beta_1 = 0.9$, which entails that the evolutionarily stable (ES) learning schedule is $u_0^* = 0$, $u_1^* = 1$ (inequality (11c)). Note that in panel (a) the weights more closely approximate the ES values, so that the equilibrium value of the mature phenotype and the rate of approach to equilibrium are larger.

impression that u_0^m and u_1^m can vary simultaneously—for simplicity of presentation.

From Eqs. (7) and (8) we then have

$$s_0 = -\frac{2(\hat{z} - z^*)}{w} \exp\left\{-\frac{(\hat{z} - z^*)^2}{w}\right\} \left.\frac{\partial z^m}{\partial u_0^m}\right|_{u_0^m = u_0, u_1^m = u_1}, \qquad (9a)$$

$$s_1 = -\frac{2(\hat{z} - z^*)}{w} \exp\left\{-\frac{(\hat{z} - z^*)^2}{w}\right\} \left.\frac{\partial z^m}{\partial u_1^m}\right|_{u_0^m = u_0, u_1^m = u_1}, \qquad (9b)$$

where we have used $z^m|_{u_0^m = u_0, u_1^m = u_1} = \hat{z}$.

It can readily be shown that

$$z^* - z^m|_{u_0^m = u_0, u_1^m = u_1} = z^*(1 - A - B)/(1 - A) > 0. \qquad (10)$$

Inequality 10 entails that the sign of s_i will be identical to the sign of $\left.\frac{\partial z^m}{\partial u_i^m}\right| \equiv \left.\frac{\partial z^m}{\partial u_i^m}\right|_{u_0^m = u_0, u_1^m = u_1}$ (we drop the subscripts for brevity). Computing the latter, we obtain
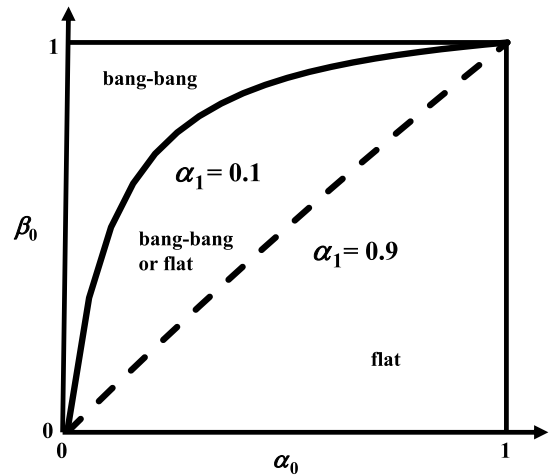
$$\left(z^*\right)^{-1} [1 - u_1\alpha_1 - (1 - u_1)\beta_1]^{-1} (1 - A) \left.\frac{\partial z^m}{\partial u_0^m}\right|$$
$$= \alpha_0(1 - \beta_0)(1 - \beta_1) + u_1 [\alpha_0(1 - \beta_0)\beta_1 - \alpha_1\beta_0(1 - \alpha_0)] \qquad (11a)$$

and

$$\left(z^*\right)^{-1} (1 - A) \left.\frac{\partial z^m}{\partial u_1^m}\right| = [1 - u_0\alpha_0 - (1 - u_0)\beta_0]\alpha_1(1 - \beta_1). \qquad (11b)$$

Clearly the right hand side and the coefficient of $\left.\frac{\partial z^m}{\partial u_1^m}\right|$ in Eq. (11b) are positive. Hence $\left.\frac{\partial z^m}{\partial u_1^m}\right| > 0$ for all u_0 and u_1 ($0 \le u_0$, $u_1 \le 1$),

Fig. 3. Dependence of the evolutionarily stable (ES) learning schedule in a constant environment on the parameters α_0, α_1 and β_0. The region in which the bang–bang control (pure SL followed by pure IL) is ES (i.e., inequality (11c) is satisfied) lies above the continuous line when $\alpha_1 = 0.1$ and above the dashed line when $\alpha_1 = 0.9$. The area of this region is monotone increasing in α_1. The ES learning schedule coincides with the optimal learning schedule.

from which $u_1^* = 1$ is the ES control for stage 1 learning. That is, stage 1 learning should be pure IL. Substituting $u_1 = 1$ in Eq. (11a), we find that $\left.\frac{\partial z^m}{\partial u_0^m}\right| < 0$ if and only if

$$\alpha_0(1 - \beta_0) - \alpha_1(1 - \alpha_0)\beta_0 < 0. \qquad (11c)$$

Thus, the ES control for stage 0 learning is $u_0^* = 0$ when inequality (11c) is satisfied (α_0 small, β_0 large, α_1 large) and $u_0^* = 1$ when this inequality is reversed (α_0 large, β_0 small, α_1 small) (Fig. 3 and Table 1).

We have shown that either $u_0^* = 0$, $u_1^* = 1$ (bang–bang control), or $u_0^* = u_1^* = 1$ (flat control) is the ES strategy. Since the selection gradients do not vanish at these points, we need not calculate the second order derivatives. Eq. (6) entails an approximation, which might seem to neglect the dynamic interaction between cultural (mature phenotype) and genetic (frequency of mutant strategy) evolution. However, a more detailed analysis given in Appendix A shows that the selection gradients (defined by Eqs. (6)–(8)) provide the first order effect on mutant frequency change at all frequencies and therefore that, by repeated invasions of mutant strategies of small effect, the evolutionary dynamics of the controls will converge to either $u_0^* = 0$, $u_1^* = 1$ (when inequality (11c) holds) or $u_0^* = u_1^* = 1$ (otherwise).

2.3. Environment fluctuates between generations (with period 2)

Let the environment fluctuate between generations with period 2, such that the optimal phenotype is $z^* > 0$ in generation t, $-z^* < 0$ in generation $t + 1$, z^* again in generation $t + 2$, etc. As before, we set the initial phenotype to 0. Unlike the case of the constant environment, however, this choice of the initial phenotype relative to the optimal phenotypes entails a special assumption regarding the phenotypic scale. For example, we must exclude the case where the optimal phenotypes in the two environments deviate from the initial phenotype in the same direction.

For this model, the recursion in the mature phenotype of the resident strategy is

$$z_{t+1} = Az_t - Bz^*, \qquad (12a)$$

$$z_{t+2} = Az_{t+1} + Bz^*, \qquad (12b)$$

$$z_{t+3} = Az_{t+2} - Bz^*, \qquad (12c)$$

Table 1
Evolutionarily stable (ES) learning schedules for the three models of environmental change.

Environmental change	ES learning schedule		Parameter dependence
	Stage 0	Stage 1	
Constant	Pure SL	Pure IL	$\alpha_0(1-\beta_0)-\alpha_1(1-\alpha_0)\beta_0 < 0$
Constant	Pure IL	Pure IL	$\alpha_0(1-\beta_0)-\alpha_1(1-\alpha_0)\beta_0 > 0$
Between generations	Pure IL	Pure IL	None
Within generations	Pure SL	Pure IL	None

α_0:efficiency of IL at stage 0, β_0: efficiency of SL at stage 0, α_1: efficiency of IL at stage 1.

etc. Hence, there is a periodic equilibrium where

$$\hat{z}^+ = Bz^*/(1+A) \tag{13a}$$

and

$$\hat{z}^- = -Bz^*/(1+A) \tag{13b}$$

alternate.

For a rare mutant strategy in generation t, when the optimal phenotype is z^* and the mature phenotype of the resident strategy in the previous generation $(t-1)$ is $\hat{z}^-$ (Eq. (13b)), we have approximately

$$z_t^m \approx A_m\hat{z}^- + B_mz^* = -z^*\left(\frac{A_mB}{1+A} - B_m\right). \tag{14a}$$

Similarly, in generation $t+1$ we have approximately

$$z_{t+1}^m \approx A_m\hat{z}^+ - B_mz^* = z^*\left(\frac{A_mB}{1+A} - B_m\right). \tag{14b}$$

This periodic pattern repeats for as long as the mutant remains rare. Define

$$z^m \equiv A_m\hat{z}^- + B_mz^*. \tag{14c}$$

Then clearly $z_t^m \approx z^m$ and $z_{t+1}^m \approx -z^m$.

In the fluctuating environment of period two, the objective function can be defined as the geometric mean of Eq. (7) at two consecutive generations. Hence, recalling that the optimal phenotype is z^* and $-z^*$ in generations t and $t+1$, respectively, the objective function is approximately

$$F\left(u_0^m, u_1^m, u_0, u_1\right) = \exp\left\{-\frac{(z^m - z^*)^2}{w}\right\},$$

which is identical to Eq. (7). The selection gradients on the mutant control for stage 0 and stage 1 learning are given by Eqs. (8) and (9), where we substitute $\hat{z}^+$ for $\hat{z}$ in the latter. As in the constant environment case, these selection gradients provide (up to a constant of proportionality) the first order effects of selection at all frequencies of the mutant strategy (Appendix B).

Also as in the constant environment case, we find

$$z^* - z^m\big|_{u_0^m=u_0, u_1^m=u_1} = z^*(1+A-B)/(1+A) > 0. \tag{15}$$

Moreover,

$$\left(z^*\right)^{-1}[1-u_1\alpha_1-(1-u_1)\beta_1]^{-1}(1+A)\left.\frac{\partial z^m}{\partial u_0^m}\right|$$
$$= \beta_0B + \alpha_0(1+A) \tag{16a}$$

and

$$\left(z^*\right)^{-1}(1+A)\left.\frac{\partial z^m}{\partial u_1^m}\right| = \{(1-u_0)\alpha_1\beta_0 + \beta_1[1-(1-u_0)\beta_0]\}B$$
$$+ [\alpha_1(1-u_0\alpha_0) + u_0\alpha_0\beta_1](1+A). \tag{16b}$$

The right hand sides and coefficients in Eqs. (16a) and (16b) are both clearly positive. Hence, noting inequality (15), we conclude that the ES control is $u_0^* = u_1^* = 1$. In this case, learning should be entirely by IL (Table 1).

2.4. Environment fluctuates within generations (with period 2)

Assume that the optimal phenotype is $-z^* < 0$ during stage 0 learning but shifts to $z^* > 0$ during stage 1 learning. The initial phenotype is assumed to be 0. Then, the recursion in the mature phenotype of the resident strategy is

$$z_{t+1} = Az_t + Cz^*, \tag{17}$$

where A is given by Eq. (4a) and

$$C = u_1\alpha_1 - u_0\alpha_0[1 - u_1\alpha_1 - (1-u_1)\beta_1]. \tag{18}$$

At equilibrium

$$\hat{z} = Cz^*/(1-A). \tag{19}$$

For the rare mutant strategy, we have approximately

$$z^m \equiv A_m\hat{z} + C_mz^* \tag{20}$$

(Appendix C), where $\hat{z}$ is given by Eq. (19), and A_m and C_m are the functions of u_i^m $(i = 0, 1)$ obtained by substituting $u_i = u_i^m$ $(i = 0, 1)$ in Eqs. (4a) and (18), respectively. The selection gradients on the mutant control are again given by Eqs. (8) and (9) (where $\hat{z}$ satisfies Eq. (19)). As before, we can show

$$z^* - z^m\big|_{u_0^m=u_0, u_1^m=u_1} = z^*(1-A-C)/(1-A) > 0. \tag{21}$$

Moreover,

$$\left(z^*\right)^{-1}[1-u_1\alpha_1-(1-u_1)\beta_1]^{-1}(1-A)\left.\frac{\partial z^m}{\partial u_0^m}\right|$$
$$= -\{u_1\alpha_1\beta_0(1+\alpha_0) + \alpha_0(1-\beta_0)[1-(1-u_1)\beta_1]\} \tag{22a}$$

and

$$\left(z^*\right)^{-1}(1-A)\left.\frac{\partial z^m}{\partial u_1^m}\right| = \alpha_1(1-\beta_1)[1+u_0\alpha_0-(1-u_0)\beta_0]. \tag{22b}$$

Clearly, Eq. (22a) entails $\left.\frac{\partial z^m}{\partial u_0^m}\right| < 0$ whereas Eq. (22b) entails $\left.\frac{\partial z^m}{\partial u_1^m}\right| > 0$. Hence, the ES control is $u_0^* = 0$, $u_1^* = 1$ (Table 1).

3. Evolutionarily stability and optimality

We show here that the ES learning schedule in each of the three models results in the equilibrium value of the mature phenotype that most closely approaches the optimal phenotype in the given environment.

For the model of the constant environment (Section 2.2), the optimal phenotype is $z^* > 0$, and hence we need to show that $\hat{z}$ (which is $<z^*$ from Eq. (10)) is maximized on the ES learning schedule. To do so, we regard $\hat{z}$ in Eq. (5) as a function of u_0 and u_1, and take partial derivatives. We can show that $\frac{\partial\hat{z}}{\partial u_1} > 0$ always holds and that

$$\left.\frac{\partial\hat{z}}{\partial u_0}\right|_{u_1=1} = \frac{z^*(1-\alpha_1)}{[1-(1-u_0)\beta_0(1-\alpha_1)]^2}$$
$$\times [\alpha_0(1-\beta_0) - \alpha_1(1-\alpha_0)\beta_0], \tag{23}$$

by the same calculations used to obtain Eqs. (11a) and (11b). Note that the constant in square brackets in the numerator of Eq. (23), which determines the sign of this equation, is identical to the diagnostic quantity in inequality (11c). Hence, when inequality (11c) is satisfied (i.e., when Eq. (23) is negative), we conclude that smaller values of u_0 (i.e., a greater reliance on SL) result in values of $\hat{z}$ that are closer to the optimal phenotype, and more specifically that $u_0 = 0$, $u_1 = 1$ maximizes $\hat{z}$. In this case, we have from Eq. (5) that

$$\hat{z} = z^* \alpha_1 / [1 - (1 - \alpha_1)\beta_0] . \tag{24a}$$

Similarly, $u_0 = 1$, $u_1 = 1$ maximizes $\hat{z}$ when this inequality is reversed, in which case

$$\hat{z} = z^* [\alpha_1 + \alpha_0(1 - \alpha_1)] . \tag{24b}$$

Thus, the model under consideration has the property that the ES learning schedule optimizes the mature phenotype at equilibrium.

Next, for the model of environmental fluctuations between generations (Section 2.3), we have $\frac{\partial \hat{z}^+}{\partial u_0} > 0$ and $\frac{\partial \hat{z}^+}{\partial u_1} > 0$ in a generation in which the optimal phenotype is $z^* > 0$. The calculations are the same as involved in the derivation of Eqs. (16a) and (16b). Similarly, $\frac{\partial \hat{z}^-}{\partial u_0} < 0$ and $\frac{\partial \hat{z}^-}{\partial u_1} < 0$ in a generation in which the optimal phenotype is $-z^* < 0$. Hence, the mature phenotype at equilibrium is optimized in both generations when $u_0 = u_1 = 1$, which coincides with the ES learning schedule.

Finally, for the model of environmental fluctuations within generations (Section 2.4), we find $\frac{\partial \hat{z}}{\partial u_0} < 0$ and $\frac{\partial \hat{z}}{\partial u_1} > 0$ so that $u_0 = 0$, $u_1 = 1$, corresponding to the ES learning schedule, results in the mature phenotype at equilibrium that is closest to the optimal phenotype $z^* > 0$.

4. Discussion

This paper addresses the following question: when an organism has several opportunities for IL and SL during its lifetime, what is the evolutionarily stable (ES) schedule of IL and SL as a function of life-stage (or age)? More specifically, we have analyzed a model where an organism has two obligatory learning opportunities (stage 0 and stage 1), at each of which it can apply IL or oblique SL or a mixture of the two to acquire information about, or the appropriate response to, the environment. Fitness is determined by the mature phenotype resulting from this two-stage cumulative learning process (Section 2.1 for details), but since the mature phenotype to be learned is assumed to vary continuously (e.g., the quantity of food foraged, the length of a stone artifact) and an organism can learn from the parental generation (oblique SL), our model incorporates the possibility that the mature phenotype can change cumulatively over several generations (Fig. 2; Section 2.2 for details).

We then asked what the ES weights (controls) are for IL and SL at each of the two learning stages under this model. In the absence of exogenous costs to either IL or SL, we have found that the ES controls can only take the values 0 or 1, and that the ES learning schedules maximize the mature phenotype at equilibrium (see Section 3). Specifically, in a constant environment the ES learning schedule is either pure SL followed by pure IL (bang–bang control) or pure IL at both stages (flat control), depending on the efficiencies of IL and SL at both stages (except the efficiency of SL at stage 1). With environmental fluctuations between or within generations, the ES learning schedules are unaffected by these efficiencies. They are flat control and bang–bang control, respectively. As summarized in Table 1, pure SL followed by pure IL – the learning schedule supportive of cumulative culture – is ES when the environment is constant (and the efficiencies

of learning satisfy inequality (11c)) or when the environment fluctuates within generations.

Previous studies incorporating two opportunities for learning per generation have arbitrarily assumed that an organism can apply pure IL or pure SL during each stage, but not a mixture of the two (Boyd and Richerson, 1985; Enquist et al., 2007; Borenstein et al., 2008; Aoki, 2010). It is interesting that our analyses, proceeding from fairly general assumptions regarding the controls, predict that the ES weights, in the absence of exogenous costs to learning, will take only the values 0 (pure SL) or 1 (pure IL). Hence, our results provide some justification for the arbitrary assumption made in these earlier studies. On the other hand, our results rule out a learning schedule in which pure IL is followed by pure SL, with once again the caveat that we have ignored the exogenous costs. This timing of pure IL followed by pure SL, which is rejected by our analyses, is found in "conditional social learning" (Enquist et al., 2007). However, Enquist et al. (2007) consider conditional social learning to be an inferior strategy compared to "critical social learning", which posits pure SL followed by pure IL, when the relative magnitudes of the exogenous costs of IL and SL are taken into account. Boyd and Richerson (1988) have also described a strategy in which IL can be interpreted as occurring before SL (i.e., resort to SL when IL proves indecisive). However, this strategy is mathematically equivalent to a mixed strategy with only one opportunity for learning (Wakano and Aoki, 2007).

We note another difference between the composite strategies envisioned by Enquist et al. (2007) and by us (see also Borenstein et al., 2008 and Aoki, 2010). Conditional social learning and critical social learning are both conditional strategies in that an organism proceeds to stage 1 learning only when the solution obtained by stage 0 learning is unsatisfactory. By contrast, the two-stage learning process in our models is obligatory.

Recall that the efficiency of learning is defined as the proportional reduction in the deviation from the target. In a constant environment, bang–bang control (pure SL followed by pure IL) is more likely to be favored when the efficiencies of SL in stage 0 and IL in stage 1 are both high (and the efficiency of IL in stage 0 is low) (Fig. 3). This result may be felt to be intuitively reasonable, but entails that flat control (pure IL at both stages) is more likely to be favored when the efficiency of IL in stage 1 is low. We believe the reason for this latter, perhaps counterintuitive, result is that the mature phenotype at equilibrium will be far from optimal in this case, so that it is better for the next generation to avoid using SL to target the mature phenotype of the previous generation. In either case, the ES control in stage 1 is pure IL, which apparently contradicts the consensus view – derived from models allowing for only one learning opportunity during the lifespan of an organism – that IL is disadvantaged in a constant environment (Aoki et al., 2005 and Wakano et al., 2004; but see Enquist et al., 2007).

That the ES learning schedule should be flat control when the environment fluctuates between generations is consistent with previous work by Feldman et al. (1996), which examined the effect of environmental periodicity on the probabilities of using IL or SL in a mixed strategy model. In a model with one learning opportunity per generation, Feldman et al. (1996) showed that in the extreme case of period 2 (corresponding to our model), the pure IL strategy would be evolutionarily stable (see also Boyd and Richerson, 1985 and Rogers, 1988). On the other hand, a mixed strategy of both IL and SL was predicted when environmental changes occurred less frequently. We expect that, if we extend our model by incorporating longer environmental periodicities, we will see a transition in the ES learning schedule between flat control and bang–bang control as the period of environmental change increases (for values of the efficiencies of learning satisfying inequality (11c)). In addition, we believe that the generality of our

results are unlikely to be compromised by the assumption of a periodic environment with two states. Feldman et al. (1996) and Aoki et al. (2005) have shown that environmental change, whether periodic or random, and whether occurring between two or an infinite number of states, have qualitatively similar consequences for the evolution of learning strategies.

When the environment fluctuates within generations, we find that the ES learning schedule is bang–bang control (pure SL followed by pure IL). That IL should be avoided during stage 0 may also be felt to be intuitively reasonable, since the optimal phenotype changes between stage 0 and stage 1, and an organism using IL to track the environment during stage 0 would be "on a wild goose chase". A related result has been obtained by Aoki (2010), who showed in a model with spatial environmental heterogeneity that the learning strategy of SL followed by IL can evolve if migration occurs, at a relatively high rate, after SL but before IL. In this case, a large fraction – but less than half – of the organisms migrate and thus experience different environments during SL and IL. Thus, bang–bang control is found to be the ES schedule when there are two learning stages per generation and the environment changes either temporally or spatially between the two stages.

However, the generality of this prediction – SL is used early in ontogeny and IL later in life – may be questionable, possibly depending on our assumption of two learning opportunities per generation. The conditions for the evolution of bang–bang control, when the environment fluctuates within generations, may in fact be more restrictive. If there are three learning opportunities, for example, such that the optimal phenotype fluctuates as $-z^*$, z^*, $-z^*$ in one generation and z^*, $-z^*$, z^* in the next, it is not clear what the ES learning schedule would be. Clearly, this model needs to be pursued in greater detail and generality.

The model of environmental fluctuations within generations has points in common with foraging models that posit spatial and temporal variation in food availability within the lifetime of an organism. Building on the concept of encounter stochasticity (the number of food items encountered by an organism is a random variable; Eliassen et al., 2009), Arbilly et al. (2011) (see also Arbilly et al., 2010) argue that (horizontal) SL may be selected for over IL when the best foods are also the least likely to find. Our analysis agrees in showing that the ES learning schedule includes SL when environmental changes occur within generations. This is in contrast to the effect of frequent environmental changes between generations, which has been shown in numerous studies to favor IL.

However, SL in previous models is often synonymous with information "scrounging", as organisms adopting this strategy do not contribute to the production of information and exploit producers (IL). By contrast, a learning schedule in which oblique SL precedes IL allows for the possibility that information possessed by past generations can be acquired by SL, which can then be further improved by IL. In particular, when pure SL is the ES control for stage 0 learning ($u_0^* = 0$, $u_1^* = 1$), even the partial use of SL allows for an improvement in the quality of information gathered by an organism over the two learning stages, and effectively leads to cumulative cultural evolution (the ratchet effect; Fig. 2). Here, SL is no longer simply synonymous with information scrounging, but connotes information "enhancing" as well (Section 3).

So far we have ignored the exogenous costs that may be associated with IL and/or SL. For example, the resources required for the development and maintenance of the neural substrate for learning, and the attention diverted from predators during learning may generate costs, which are to be distinguished from the endogenous costs of missing the target. Exogenous costs can be incorporated into our models by multiplying the objective function Eq. (7) by the factor $(1 - c_0 u_0^m)(1 - c_1 u_1^m)$, where c_i ($i = 0, 1$) is

positive or negative depending on whether a larger net exogenous cost accrues to IL than SL during stage i ($i = 0, 1$). When these exogenous costs are introduced, we find as expected that the ES controls, u_0^* and u_1^*, can take intermediate values between 0 and 1 (numerical work not shown). In other words, the ES learning schedule may entail a partial reliance on SL and IL at either or both stages, which is perhaps more consonant with actual learning behavior.

Another type of exogenous cost may derive from the requirement of additional memory or processing power for learning schedules with intermediate weights ($0 < u_0 < 1$ and/or $0 < u_1 < 1$) implied by our model. Hence, we expect a priori that learning schedules with intermediate weights are less likely to be ES. Interestingly, such schedules were found not to be ES even when such an exogenous cost was ignored.

The learning schedule in which pure SL and pure IL occur in this order has variously been called guided variation (Boyd and Richerson, 1985), critical social learning (Enquist et al., 2007), and "social-learner-explorer" (Borenstein et al., 2008; Aoki, 2010). It is arguably one of the few plausible learning strategies supportive of cumulative culture. We find that this learning schedule can be ES when the environment is constant, provided the efficiency of SL is high during stage 0 learning, as inequality (11c) is more likely to be satisfied for large β_0 (Fig. 3). In addition, we observe a "non-monotone" dependence on the efficiencies of IL, where α_0 and α_1 have opposite effects in inequality (11c). These predictions are entirely consistent with the conditions for critical social learning to be ES, which Enquist et al. (2007) state as "less environmental changes and more faithful cultural transmission". Moreover, "increasing the probability of finding an OK solution through individual learning … first favors and then slightly disfavors critical social learning". Enquist et al. (2007) do not address the case of environmental fluctuations within generations, so that a comparison is not possible.

In closing, we emphasize the restrictive nature of some of our simplifying assumptions, which may affect the generality of our conclusions. First, the learning strategies considered here can each express only one mature phenotype, whereas it may be more pertinent to allow for mixed strategies. Second, we assumed two learning opportunities per generation, with possible repercussions as already mentioned particularly in the case where the environment fluctuates within generations. Third, we set the initial phenotype at halfway between the two optimal phenotypes when the environment fluctuates between or within generations, which entails a special assumption regarding the phenotypic scale. Fourth, future work should incorporate the possibility that the ontogeny of learning may by contingent on the learning experience itself.

While acknowledging these limitations, we feel that our paper contributes to the evolutionary theory of learning. It does so by providing a framework for investigating the broad developmental patterns in the differential use of individual learning and social learning that may be favored by natural selection under various regimes of environmental change.

Acknowledgments

Appendix A. Justification of the use of Eq. (6)

Let the frequency of the mutant strategy among the newborns of generation t be q_t. Moreover, let the mature phenotypes of the

resident and mutant strategies in the same generation be z_t and z_t^m, respectively. Then

$$z_t = A\bar{z}_{t-1} + Bz^*, \tag{A.1a}$$

$$z_t^m = A_m\bar{z}_{t-1} + B_mz^*, \tag{A.1b}$$

where

$$\bar{z}_{t-1} = (1 - q_{t-1})z_{t-1} + q_{t-1}z_{t-1}^m \tag{A.1c}$$

is the population mean of the mature phenotypes in generation $t - 1$, and A, B, A_m, B_m are as defined in Eqs. (4a) and (4b), and immediately above Eq. (6). Our system is completely described by Eq. (A.1) for the cultural dynamics together with Eq. (A.8) (see below) for the mutant frequency dynamics.

Assume for now that a separation of time scales is possible, such that changes in the mature phenotypes, z_t and z_t^m, occur much more quickly than changes in the mutant frequency, q_t. Then, setting $q_t = \varepsilon$ (constant), and noting Eqs. (5) and (6), we have

$$z_t - \hat{z} = A(1 - \varepsilon)(z_{t-1} - \hat{z}) + A\varepsilon(z_{t-1}^m - z^m)$$
$$+ A\varepsilon(z^m - \hat{z}), \tag{A.2a}$$

$$z_t^m - z^m = A_m(1 - \varepsilon)(z_{t-1} - \hat{z}) + A_m\varepsilon(z_{t-1}^m - z^m)$$
$$+ A_m\varepsilon(z^m - \hat{z}). \tag{A.2b}$$

In order to justify our use of the approximate Eq. (6), we look at one control at a time. We focus in this Appendix on the control u_0, but exactly the same argument applies to u_1. Thus, we set $u_0^m = u_0 + \delta$ and $u_1^m = u_1$, from which we can write $A_m = A + A_1\delta$, $B_m = B + B_1\delta$. Substituting in the linear equations Eq. (A.2) and solving, we obtain

$$\begin{pmatrix} z_t - \hat{z} \\ z_t^m - z^m \end{pmatrix} = \frac{1 - (A + A_1\delta\varepsilon)^t}{1 - (A + A_1\delta\varepsilon)} \begin{pmatrix} A(1 - \varepsilon) & A\varepsilon \\ A_m(1 - \varepsilon) & A_m\varepsilon \end{pmatrix}$$
$$\times \begin{pmatrix} A \\ A_m \end{pmatrix} \varepsilon\delta(A_1\hat{z} + B_1z^*). \tag{A.3}$$

The solution Eq. (A.3) of Eq. (A.2) assumes that the mutant arises in generation 0, so that the initial conditions are $z_0 = \hat{z}$ and $z_0^m = z^m$.

Eq. (A.3) shows that the approximate value of the mature phenotype of the mutant strategy, z^m, deviates from the exact value, z_t^m, to order ε. Since ε can be taken arbitrarily small in the analysis of evolutionary stability, we conclude that the approximation is valid.

We will now give a more general proof that assumes neither a separation of time scales nor a low frequency of the mutant strategy. In fact, we will show that the explicit expression for the change in the frequency of the mutant strategy, $\Delta q_t = q_{t+1} - q_t$, can be written, to first order in the small quantity δ, in terms of the approximate Eq. (6) for any t.

First, on subtracting Eq. (A.1a) from Eq. (A.1b), we have

$$z_{t+1}^m - z_{t+1} = \delta(A_1\bar{z}_t + B_1z^*). \tag{A.4}$$

Assuming $z_0 = \hat{z}$ and $z_0^m = z^m$, it is possible to show that $z_t = \hat{z} + O(\delta)$ and $z_t^m = \hat{z} + O(\delta)$, and hence $\bar{z}_t = \hat{z} + O(\delta)$. Substituting the last expression and the explicit forms of A_1 and B_1 in Eq. (A.4), we obtain

$$z_{t+1}^m - z_{t+1} = \delta(\alpha_0z^* - \beta_0\hat{z})[1 - u_1\alpha_1 - (1 - u_1)\beta_1] + O(\delta^2). \tag{A.5}$$

Eq. (A.5) holds for all time t and any frequency q_t of the mutant strategy. Moreover, the coefficient of δ on the right hand side of Eq. (A.5) is independent of t, which means that there exists a constant c, which does not depend on t, such that

$$\left| z_{t+1}^m - z_{t+1} \right| < c\delta$$

for sufficiently small δ. We conclude that the mature phenotypes of the mutant and resident strategies remain close to each other for all time and that their difference does not accumulate to the first order in δ. Using Eq. (6), we can rewrite Eq. (A.5) as

$$z_t^m - z_t = \delta\left.\frac{\partial z^m}{\partial u_0^m}\right|_{u_0^m = u_0} + O(\delta^2). \tag{A.6}$$

Now set

$$G(x) = \exp\left\{ -\frac{(x - z^*)^2}{w} \right\}. \tag{A.7}$$

Then the fitnesses (relative fertilities) of the resident and mutant strategies, when the dependence on δ is made explicit, are given by $G(z_t(\delta))$ and $G(z_t^m(\delta))$, respectively. Hence, the exact equation describing the change per generation of the frequency of the mutant strategy is

$$\Delta q_t = q_t(1 - q_t)\left(\frac{G(z_t^m(\delta)) - G(z_t(\delta))}{q_tG(z_t^m(\delta)) + (1 - q_t)G(z_t(\delta))} \right). \tag{A.8}$$

By way of a first order Taylor expansion around $\delta = 0$ of this equation, we obtain

$$\Delta q_t = \delta q_t(1 - q_t)S_{0,t} + O(\delta^2), \tag{A.9a}$$

where

$$S_{0,t} = \frac{1}{G(\hat{z})} \times \left.\frac{dG(z_t^m)}{dz_t^m}\right|_{z_t^m = \hat{z}} \times \left.\frac{d}{d\delta}(z_t^m - z_t)\right|_{\delta = 0}. \tag{A.9b}$$

If the selection gradient, $S_{0,t}$, is a constant independent of time, t, then we can set $S_{0,t} = S_0$, where $S_0 > 0$ implies both allele invasion and substitution. To show this, we note that $z_t^m|_{\delta=0} = \hat{z}$, and hence by Eq. (A.7) the derivative $\left.\frac{dG(z_t^m)}{dz_t^m}\right|_{z_t^m = \hat{z}}$ does not depend on time. Further, from Eq. (A.6) we see that $\frac{d}{d\delta}(z_t^m - z_t)|_{\delta=0}$ does not depend on time either and is equal to $\left.\frac{\partial z^m}{\partial u_0^m}\right|_{u_0^m = u_0}$.

To complete the proof, we rewrite the objective function defined in Eq. (7) as

$$F(u_0^m, u_1, u_0, u_1) = \exp\left\{ -\frac{(z^m - z^*)^2}{w} \right\}$$

since we are here considering only variation in the control u_0^m. Then, $\left.\frac{dG(z_t^m)}{dz_t^m}\right|_{z_t^m = \hat{z}} = \left.\frac{\partial F}{\partial z^m}\right|_{z^m = \hat{z}}$ and $G(\hat{z}) = F(u_0, u_1, u_0, u_1)$. Hence, the selection gradient is a constant independent of time, which reduces to

$$S_{0,t} = S_0 = \frac{1}{F(u_0, u_1)} \times \left.\frac{\partial F}{\partial z^m}\right|_{z^m = \hat{z}} \times \left.\frac{\partial z^m}{\partial u_0^m}\right|_{u_0^m = u_0}$$

$$= \frac{1}{F(u_0, u_1)}\left.\frac{\partial F}{\partial u_0^m}\right|_{u_0^m = u_0}. \tag{A.10}$$

This selection gradient can be expressed as $S_0 = s_0/F(u_0, u_1, u_0, u_1)$, where s_0 is as defined in Eq. (8). Hence, the optimal control found from solving the heuristic equation $s_0 = 0$ agrees with that found by solving $S_0 = 0$ derived from the population genetic model.

Appendix B. Justification of the use of Eq. (14c)

We give a population genetic justification for the approximation used in the model of environmental fluctuations between

generations. Following the convention of Section 2.3, the optimal phenotype is $z^* > 0$ in generation t, $-z^* < 0$ in generation $t + 1$, etc. We denote by

$$G^+(x) = \exp\left\{-\frac{(x - z^*)^2}{w}\right\}$$
(B.1a)

the objective function when the optimal phenotype is z^*, and by

$$G^-(x) = \exp\left\{-\frac{(x + z^*)^2}{w}\right\}$$
(B.1b)

the objective function when the optimal phenotype is $-z^*$.

Then, using Eq. (A.8) with $G^+(x)$ instead of $G(x)$, the change in the frequency of the mutant strategy from generation t to $t + 1$ can be written to first order in δ as

$$q_{t+1} = q_t + \delta q_t(1 - q_t)S_{0,t}^+ + O(\delta^2),$$
(B.2a)

where

$$S_{0,t}^+ = \frac{1}{G^+(\hat{z}^+)} \times \left.\frac{dG^+(z_t^m)}{dz_t^m}\right|_{z_t^m = \hat{z}^+} \times \left.\frac{d}{d\delta}(z_t^m - z_t)\right|_{\delta=0}.$$
(B.2b)

Likewise, to first order in δ, the change in the mutant frequency from generation $t + 1$ (when the optimal phenotype is $-z^*$) to $t + 2$ can be written as

$$q_{t+2} = q_{t+1} + \delta q_{t+1}(1 - q_{t+1})S_{0,t+1}^- + O(\delta^2),$$
(B.3a)

where

$$S_{0,t+1}^- = \frac{1}{G^-(\hat{z}^-)} \times \left.\frac{dG^-(z_{t+1}^m)}{dz_{t+1}^m}\right|_{z_{t+1}^m = \hat{z}^-} \times \left.\frac{d}{d\delta}(z_{t+1}^m - z_{t+1})\right|_{\delta=0}.$$
(B.3b)

Substituting Eq. (B.2a) into (B.3a) and simplifying yields

$$q_{t+2} - q_t = \delta q_t(1 - q_t)\left(S_{0,t}^+ + S_{0,t+1}^-\right) + O(\delta^2).$$
(B.4)

If the average selection gradient, $\left(S_{0,t}^+ + S_{0,t+1}^-\right)/2$, is a constant independent of time, t, then its sign predicts the fate of the mutant strategy at all frequencies. Following the same line of argument as in Appendix A, we will show that $\left(S_{0,t}^+ + S_{0,t+1}^-\right)/2$ is indeed independent of time and that it has the same sign as the selection gradient, $\left.\frac{\partial F}{\partial u_0^m}\right|_{u_0^m = u_0}$, defined in Section 2.3.

First, we write the exact recursions for the mature phenotypes of the resident and mutant strategies in generation t as

$$z_t = A\bar{z}_{t-1} + Bz^*,$$
(B.5a)
$$z_t^m = A_m\bar{z}_{t-1} + B_mz^*,$$
(B.5b)

which by using $\bar{z}_{t-1} = \hat{z}^- + O(\delta)$ gives

$$z_t^m - z_t = \delta(\alpha_0 z^* - \beta_0\hat{z}^-)\left[1 - u_1\alpha_1 - (1 - u_1)\beta_1\right] + O(\delta^2).$$
(B.6)

Hence, from Eq. (14c) we have

$$\left.\frac{d}{d\delta}(z_t^m - z_t)\right|_{\delta=0} = \left.\frac{\partial z^m}{\partial u_0^m}\right|_{u_0^m = u_0}.$$
(B.7)

Next, the exact recursions for the mature phenotypes of the resident and mutant strategies in generation $t + 1$ are

$$z_{t+1} = A\bar{z}_t - Bz^*,$$
(B.8a)
$$z_{t+1}^m = A_m\bar{z}_t - B_mz^*,$$
(B.8b)

which by using $\bar{z}_t = \hat{z}^+ + O(\delta)$ gives

$$z_{t+1}^m - z_{t+1}$$
$$= -\delta(\alpha_0 z^* + \beta_0\hat{z}^+)\left[1 - u_1\alpha_1 - (1 - u_1)\beta_1\right] + O(\delta^2).$$
(B.9)

Using Eq. (14c) again and noting $\hat{z}^+ = -\hat{z}^-$, we have

$$\left.\frac{d}{d\delta}(z_{t+1}^m - z_{t+1})\right|_{\delta=0} = -\left.\frac{\partial z^m}{\partial u_0^m}\right|_{u_0^m = u_0}.$$
(B.10)

The equality $\hat{z}^+ = -\hat{z}^-$ also entails that

$$\frac{1}{G^+(\hat{z}^+)} \times \left.\frac{dG^+(z_t^m)}{dz_t^m}\right|_{z_t^m = \hat{z}^+}$$
$$= \frac{2(z^* - \hat{z}^+)}{w} = -\frac{1}{G^-(\hat{z}^-)} \times \left.\frac{dG^-(z_{t+1}^m)}{dz_{t+1}^m}\right|_{z_{t+1}^m = \hat{z}^-}.$$
(B.11)

Hence, using Eqs. (B.7), (B.10), (B.11), (B.2b) and (B.3b), the average selection gradient can be written as

$$\left(S_{0,t}^+ + S_{0,t+1}^-\right)/2 = \frac{2(z^* - \hat{z}^+)}{w}\left.\frac{\partial z^m}{\partial u_0^m}\right|_{u_0^m = u_0}.$$
(B.12)

Finally, since $\hat{z}^+ = z^m|_{u_0^m = u_0, u_1^m = u_1} < z^*$, the optimal control found from Eq. (16a) agrees with that predicted from $\left(S_{0,t}^+ + S_{0,t+1}^-\right)/2$, which is derived from the population genetic model.

Appendix C. Justification of the use of Eq. (20)

When the environment fluctuates within generations, Eqs. (A.8), (A.9a) and (A.9b) still apply. In order to justify the use of z^m as defined by Eq. (20) we need to show that

$$\left.\frac{d}{d\delta}(z_t^m - z_t)\right|_{\delta=0} = \left.\frac{\partial z^m}{\partial u_0^m}\right|_{u_0^m = u_0},$$
(C.1)

whereby the selection gradient on the mutant control derived from the population genetic model is given by Eq. (A.10).

To prove Eq. (C.1), we first note from Eqs. (17)–(20) that the exact recursions for the mature phenotypes of the resident and mutant strategies are

$$z_t = A\bar{z}_{t-1} + Cz^*,$$
(C.2a)
$$z_t^m = A_m\bar{z}_{t-1} + C_mz^*.$$
(C.2b)

Following the same line of argument as in Appendix A, we have

$$z_t^m - z_t = -\delta(\alpha_0 z^* + \beta_0\hat{z})\left[1 - u_1\alpha_1 - (1 - u_1)\beta_1\right] + O(\delta^2),$$
(C.3)

and by substituting Eqs. (C.3) and (20) into Eq. (C.1), we find that the equality indeed holds.

References

Aisner, R., Terkel, J., 1992. Ontogeny of pine cone opening behavior in the black rat, Rattus rattus. Anim. Behav. 44, 327–336.

Aoki, K., 2010. Evolution of the social-learner-explorer strategy in an environmentally heterogeneous two-island model. Evolution 64, 2575–2586.

Aoki, K., Wakano, J.Y., Feldman, M.W., 2005. The emergence of social learning in a temporally changing environment: a theoretical model. Curr. Anthropol. 46, 334–340.

Arbilly, M., Motro, U., Feldman, M.W., Lotem, A., 2010. Co-evolution of learning complexity and social foraging strategies. J. Theoret. Biol. 267, 573–581.

Arbilly, M., Motro, U., Feldman, M.W., Lotem, A., 2011. Evolution of social learning when high expected payoffs are associated with high risk failure. J. R. Soc. Interface 8, 1604–1615.

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.

Beekman, M., Gilchrist, A.L., Duncan, M., Sumpter, D.J.T., 2007. What makes a honeybee scout? Behav. Ecol. Sociobiol. 61, 985–995.

Biesmeijer, J.C., Seeley, T.D., 2005. The use of waggle dance information by honey bees throughout their foraging careers. Behav. Ecol. Sociobiol. 59, 133–142.

Biondi, L.M., Bó, M.S., Vassallo, A.I., 2010a. Inter-individual and age differences in exploration, neophobia and problem-solving ability in a Neotropical raptor (Milvago chimango). Anim. Cogn. 13, 701–710.

Biondi, L.M., Garcia, G.O., Bó, M.S., Vassallo, A.I., 2010b. Social learning in the Caracara Chimango, Milvago Chimango (Aves: Falconiformes): an age comparison. Ethology 116, 722–735.

Borenstein, E., Feldman, M.W., Aoki, K., 2008. Evolution of learning in fluctuating environments: when selection favors both social and exploratory individual learning. Evolution 62, 586–602.

Boyd, R., Richerson, P.J., 1985. Culture and the Evolutionary Process. Univ. of Chicago Press, Chicago, IL.

Boyd, R., Richerson, P.J., 1988. An evolutionary model of social learning: the effect of spatial and temporal variation. In: Zentall, T., Galef Jr., B.G. (Eds.), Social Learning. Erlbaum, Hillsdale, NJ, pp. 29–48.

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

Cole, E.F., Cram, D.L., Quinn, J.L., 2011. Individual variation in spontaneous problem-solving performance among wild great tits. Anim. Behav. 81, 491–498.

Dugatkin, L.A., 2004. Principles of Animal Behavior. W.W. Norton and Company.

Eliassen, S., Jørgensen, C., Mangel, M., Giske, J., 2009. Quantifying the adaptive value of learning in foraging behavior. Am. Nat. 174, 478–489.

Enquist, M., Eriksson, K., Ghirlanda, S., 2007. Critical social learning: a solution to Roger's paradox of nonadaptive culture. Am. Anthropol. 109, 727–734.

Feldman, M.W., Aoki, K., Kumm, J., 1996. Individual versus social learning: evolutionary analysis in a fluctuating environment. Anthropol. Sci. 104, 209–232.

Galef Jr., B.G., 1988. Imitation in animals: history, definitions, and interpretations of data from the psychological laboratory. In: Zentall, T., Galef Jr., B.G. (Eds.), Social Learning. Erlbaum, Hillsdale, NJ, pp. 2–28.

Giraldeau, L., Caraco, T., Valone, T., 1994. Social foraging: individual learning and cultural transmission of innovations. Behav. Ecol. 5, 35–43.

Gunst, N., Leca, J.-B., Boinski, S., Fragazy, D., 2010. The ontogeny of handling hard-to-process food in wild brown capuchins (Cebus apella apella): evidence from foraging on the fruit of Maximiliana maripa. Am. J. Primatol. 72, 960–973.

Heyes, C.M., 1993. Imitation, culture, and cognition. Anim. Behav. 46, 999–1010.

Kaplan, H., Hill, K., Lancaster, J., Hurtado, A.M., 2000. A theory of human life history evolution: diet, intelligence, and longevity. Evol. Anthropol. 9, 156–183.

Laland, K.N., 2004. Social learning strategies. Learn. Behav. 32, 4–14.

Leadbeater, E., Chittka, L., 2007. Social learning in insects—from miniature brains to consensus building. Curr. Biol. 17, R703–R713.

Lehmann, L., Feldman, M.W., Kaeuffer, R., 2010. Cumulative cultural dynamics and the coevolution of cultural innovation and transmission: an ESS model for panmictic and structured populations. J. Evol. Biol. 23, 2356–2369.

Lycett, S.J., Gowlett, A.J., 2008. On questions surrounding the Acheulean tradition. World Archaeol. 40, 295–315.

Maynard Smith, J., 1982. Evolution and the Theory of Games. Cambridge Univ. Press.

Reader, S.M., Laland, K.N., 2001. Primate innovation: sex, age and social rank differences. Int. J. Primatol. 22, 787–805.

Rogers, A.R., 1988. Does biology constrain culture? Am. Anthropol. 90, 819–831.

Rogers, D.S., Ehrlich, P.R., 2008. Natural selection and cultural rates of change. Proc. Natl. Acad. Sci. USA 105, 3416–3420.

Striano, T., Tomasello, M., Rochat, P., 2001. Social and object support for early symbolic play. Dev. Sci. 4, 442–455.

Sydsæter, K., Hammond, P., Seierstad, A., Strøm, A., 2008. Further Mathematics for Economic Analysis, second ed. Pearson Education.

Tennie, C., Call, J., Tomasello, M., 2009. Ratcheting up the ratchet: on the evolution of cumulative culture. Phil. Trans. R. Soc. B 364, 2405–2415.

Tomasello, M., 1999. The human adaptation for culture. Annu. Rev. Anthropol. 28, 509–529.

Vygotsky, L.S., 1978. Mind in Society: The Development of Higher Psychological Processes. Harvard Univ. Press.

Wakano, J.Y., Aoki, K., 2006. A mixed strategy model for the emergence and intensification of social learning in a periodically changing natural environment. Theor. Popul. Biol. 70, 486–497.

Wakano, J.Y., Aoki, K., 2007. Do social learning and conformist bias coevolve? Henrich and Boyd revisited. Theor. Popul. Biol. 72, 504–512.

Wakano, J.Y., Aoki, K., Feldman, M.W., 2004. Evolution of social learning: a mathematical analysis. Theor. Popul. Biol. 66, 249–258.

Whiten, A., Ham, R., 1992. On the nature and evolution of imitation in the animal kingdom: reappraisal of a century of research. In: Rosenblatt, J.S., et al. (Eds.), Advances in the Study of Behavior, vol. 21. Academic Press, New York, pp. 239–283.