Behavioral Ecology

The official journal of the
ISBE
International Society for Behavioral Ecology

Behavioral Ecology (2016), 27(3), 842–850. doi:10.1093/beheco/arv184

Original Article

Environmental complexity favors the evolution of learning

Slimane Dridi[a,b] and Laurent Lehmann[a]
[a]Department of Ecology and Evolution, UNIL Sorge, Le Biophore, 1015, Lausanne, Switzerland and
[b]Department of Biology, University of Pennsylvania, 433 South University Avenue, Philadelphia, PA 19104, USA

Received 24 May 2015; revised 15 September 2015; accepted 25 September 2015; Advance Access publication 29 December 2015.

Learning is a fundamental biological adaptation that is widespread throughout the animal kingdom. According to previous research, 2 conditions are necessary for learning to be adaptive: between-generation environmental variability and within-generation environmental predictability. In this article, we show that between-generation variability is not necessary and that instrumental learning can provide a selective advantage in a complex environment, where an individual is exposed to a large number of different challenges during its lifespan. We construct an evolutionary model where individuals have a memory with limited storage capacity, and an evolving trait determines the fraction of that memory that should be allocated to innate responses to the environment versus learning these responses. The evolutionarily stable level of learning depends critically on the features of the environmental process, but generally increases with environmental complexity. We conclude by emphasizing that the specific advantages of learning should be distinguished from the general advantages of phenotypic plasticity, and we discuss possible routes to empirically test our claims.

Key words: **behavioral plasticity, environmental complexity, natural selection, short-term memory.**

INTRODUCTION

Learning allows an individual to use experience and thereby express payoff-relevant actions in novel environments. In particular, instrumental learning permits to build simple associations between newly encountered stimuli and appropriate actions (Pearce 2008). It is important to understand the adaptive value of this form of learning as it permeates the animal kingdom (Fawcett et al. 2013) and underlies the ecological success of the hominin lineage (Johnston 1982; Boyd and Richerson 1988; Shettleworth 2009). In previous work, the selective advantage of learning has been proposed to crucially rely on between-generation environmental variability (Boyd and Richerson 1988; Stephens 1991; Feldman et al. 1996; Kerr and Feldman 2003; Wakano et al. 2004; Dunlap and Stephens 2009; Aoki and Feldman 2014). The argument is that if offspring live in environments where the consequences of actions are totally different from that of parents, and were never experienced in the history of the population, then offspring can express novel appropriate actions only through learning. It has also been emphasized that the environment should not change too fast within an individual's lifespan for learning to evolve (Stephens 1991; Dunlap and Stephens 2009). In other words, the environment should be predictable enough for information to be useful, and where predictability will be understood throughout as temporal autocorrelation in the environment. This notion of predictability is thus distinct from the fact that cues from the environment can be good or bad indicators (i.e., predictors) of the true state of the environment (Burgess and Marshall 2014; Botero et al. 2015).

The general consensus in the literature on the evolution of learning is thus that 2 conditions are necessary for learning to be adaptive: between-generation environmental variability and within-generation predictability (Boyd and Richerson 1988; Stephens 1991; Feldman et al. 1996; Kerr and Feldman 2003; Wakano et al. 2004; Dunlap and Stephens 2009). The requirement of predictability seems unavoidable because learning can be effective only if there is a certain amount of temporal autocorrelation (Fawcett et al. 2014), that is, if information is reliable over time. The importance of between-generation variability is less clear, however, and we will show in this article that the occurrence of such variability is not a necessary condition for learning to evolve.

Though learning has indeed been shown to evolve under between-generation environmental variability (e.g., the infinite-environmental state model of Feldman et al. 1996; Wakano et al. 2004), the conditions under which general phenotypic plasticity evolves are very similar (i.e., variable environments, Gomulkiewicz and Kirkpatrick 1992; Pigliucci 2001). This blurs the specific advantages of learning over other forms of behavioral plasticity, such as innate behavioral plasticity (Mery and Burns 2010; Hollis and Guillette 2011; Snell-Rood 2013), which is exemplified by fearful reactions to predators, or preference for tasty food (Mery and Kawecki 2004; Riffell et al. 2008; Gong 2012). These forms of

---

Address correspondence to S. Dridi. E-mail: slimane.dridi@gmail.com.

plastic responses have also been called “activational” (Stamps and Groothuis 2010) or “contextual” plasticity (Snell-Rood 2013), and fall under the broad category of reversible plasticity (Botero et al. 2015). They may be defined as phenotypes that are fixed from birth (possibly modulo a developmental period) and are always expressed throughout the lifespan of an individual, whenever a particular stimulus is encountered. These phenotypes are thus reversible in the sense that they are not always expressed; they are only expressed when a particular stimulus is present and are therefore labile traits that can change multiple times during an individual’s lifetime.

Because innate behavioral plasticity and learning both refer to labile traits (in opposition to nonlabile traits or developmental plasticity, also called irreversible plasticity), they should in principle both provide an advantage in within-generation varying environments (Gomulkiewicz and Kirkpatrick 1992). Although recent research has made significant progress in this direction (Botero et al. 2015), more work is needed to disentangle the effects of environmental patterns on the evolution of the different forms of plasticity. In particular, another discriminating factor than environmental variability is required to understand the specific advantage of learning over innate behavioral plasticity. Such a distinction has not been made possible in previous theoretical work because in most models for the evolution of learning, learners are pitted against individuals that can only express one given genetically determined action (Boyd and Richerson 1988; Stephens 1991; Feldman et al. 1996; Kerr and Feldman 2003; Wakano et al. 2004; Dunlap and Stephens 2009). But this is not a very likely evolutionary transition to learning, as learning is more likely to evolve on top of innate behavioral plasticity (Kerr 2007), and it is indeed common to observe the coexistence within individuals of these 2 forms of plastic responses (Mery and Burns 2010; Snell-Rood 2013).

What distinguishes innate behavioral plasticity from learning is that, under learning, information about payoff-relevant actions is not genetically encoded, but acquired during lifetime. This suggests that factors that render difficult the storage of information in the genotype will favor learning over innate behavioral plasticity. Previous research has unraveled one way in which environmental information cannot persist in the genotype; namely, when offspring live in a completely novel environment, offspring carry genetic information that is not adapted to their current conditions; they will not reproduce and this will cause the loss of information inherited from previous generations (Boyd and Richerson 1988; Stephens 1991; Feldman et al. 1996; Kerr and Feldman 2003; Wakano et al. 2004; Dunlap and Stephens 2009).

But, even if parents and offspring face the same conditions, the environment may be too complex for all responses to it to be genetically encoded. By environmental complexity, we mean the number of distinct challenges or stimuli that an individual encounters within its lifespan. These may be, for instance, an encounter with a predator or with a food item of some nutritional value. Environmental complexity may then select for learning for the following reason. Because the range of challenges encountered during an individual’s lifespan can be extremely large, and each of these situations generates a particular combination of sensory perceptions in the animal’s brain, it seems unlikely that an animal is capable of storing the interactions with all these challenges and the associated responses (even with the abstract representation provided by neural networks, Enquist and Ghirlanda 2005).

The discrepancy between the complexity of the environment and the capacity of an individual’s genetic memory to process information thus imposes a computational constraint on its decision system. Having a dynamic memory, which allows the individual to forget obsolete stimulus-response associations and learn new ones, may be useful for dealing with environmental complexity, as it makes feasible to react to an arbitrarily large number of situations. The contribution of forgetting to the adaptive value of learning has already been investigated (Kraemer and Golding 1997; Kerr and Feldman 2003), but only in situations where forgetting allows one to face the same challenge at distinct instants and if the optimal behavior for that challenge has changed (i.e., environmental variability). Forgetting may further contribute to the benefits provided by learning through the ability to encode different stimuli because different stimuli may be encountered at distinct instants of time. This is consistent with the functioning of short-term memory: Animate or inanimate features with which an animal interacts first enter the working memory and are transferred to the long-term memory only through a consolidation phase, which does not necessarily occur (Dudai 2004; Shettleworth 2009). When supplemented with forgetting, learning is thus likely to provide a powerful mean to cope with environmental complexity because it can scatter complexity over time; only a small portion of the environment’s complexity is dealt with per unit time.

In the rest of this article, our aim is to investigate the evolutionary transition from innate behavioral plasticity to learning and to formalize in an evolutionary model the above verbal argument that instrumental learning is adaptive under conditions of environmental complexity. In order to capture the limitations of an individual in terms of information processing, we assume that it is constrained by a maximum amount of memory. An evolving trait prescribes the allocation of this memory either to an innate memory or to a dynamic memory, which allows the individual to learn and forget associations between stimuli and actions. Because our goal is to find a minimal set of conditions where learning is selected for, the dynamic memory will only correspond to short-term memory, and we will ignore the possible gains provided by long-term memory. This is a conservative assumption because long-term memory is widespread in nature and is likely to further increase the selective advantage provided by learning (Shettleworth 2009). The environment consists of a finite (but possibly very large) number of challenges (or stimuli), each of which is characterized by its own optimal action(s). We will show that environmental complexity (operationalized as the number of potential stimuli in the environment) can generate a selection pressure in favor of a greater allocation of memory to learning.

MODEL

The individual and its environment

Consider an individual that interacts with its environment for successive discrete time steps. At each time step, the individual has to choose an action to respond to an environmentally determined challenge or stimulus, which is drawn from a set of N_s stimuli. Following previous formalizations (Feldman et al. 1996; Wakano et al. 2004), we assume that the chosen action is either the “correct” (or appropriate) response to the stimulus and gives payoff π_C or a “wrong” response and gives payoff π_W .

Because stimuli may depend on location, task to be performed, or time of the day, the individual is unlikely to meet all of them at once and we assume the following environmental process, where only a subset of the entire set of stimuli can be encountered per time step. Namely, any time step of an individual’s lifespan consists of 3 events. 1) With probability γ a block of stimuli of size $N_b \leq N_s$

is randomly drawn from the set of environmental stimuli, whereas with probability $1 - \gamma$, the individual faces the same block met at the previous time step. 2) A stimulus is uniformly drawn from the block (hence, a given stimulus in the block is sampled with probability $1/N_b$). 3) The individual chooses an action in response to this stimulus.

An important property of this environmental process is that the stationary distribution of stimuli is uniform, so that every stimulus has probability $1/N_s$ of being encountered in the stationary state (see the Appendix for a proof and a detailed mathematical description of the environmental process). We take N_s to be a measure of *environmental complexity*: Large values of N_s correspond to the case where an individual will encounter many different stimuli. However, the environment itself is not uniform and the parameter N_b ($1 \leq N_b \leq N_s$) captures *local complexity*: Low values of N_b correspond to a low number of stimuli possibly experienced in a given block of stimuli, whereas high N_b corresponds to a locally complex environment, where the individual is temporarily faced with a high number of stimuli. Finally, the parameter γ ($0 \leq \gamma \leq 1$) is the *environmental switching rate*: Low values of γ correspond to slow block turnover. In this environment, predictability, a feature that has been shown to critically affect the evolution of learning (Stephens 1991), is captured by the interaction between N_b and γ. When γ is close to 1, there is rapid block turnover, so the environment is not very predictable for any block size N_b. When γ is smaller, predictability depends on local complexity, N_b. A large value of N_b means that a lot of stimuli are being encountered in a given block so the probability to encounter the same stimulus repeatedly is low, and hence, the environment is less predictable. A small value of N_b corresponds to a more predictable environment where the individual only deals with a small number of stimuli during a period of interaction with a block.

In order to be able to store information about how to respond to stimuli, we assume that the individual has a memory that can store m associations between stimulus and action. These m memory "slots" could either be filled with fixed associations present at birth, which hold templates of stimuli together with the innate response to these stimuli, or with such associations that are learned during the individual's lifespan. Though the innate responses are not necessarily stored in the same neuronal memory as the dynamic responses (they may even be only physiological responses that are not neurally mediated), we use the term "memory" for both the innate and the dynamic storage of stimuli, in order to emphasize that the organism faces a trade-off in allocating physiological resources to these 2 different types of responses. We denote by g the number of associations that are innately determined. Hence, if $g < m$, a part of the memory, $m - g$ slots, is dynamic, and the individual can encode new stimuli encountered during its interactions with the environment. If $g = 0$, the individual is born with a "blank slate", with absolutely no innate tendency to respond to environmental stimuli. Our goal is to understand the selection pressure on the evolving trait g, given a fixed memory capacity m, and how this depends on environmental complexity (N_s), local complexity (N_b), and switching rate (γ). (see Table 1 for a list of all the symbols used in the model)

Fitness

In order to evaluate the selection pressure on g, we need a measure of expected payoff (or fitness) accruing to an individual expressing this trait value. To obtain this, we note that an encountered stimulus at a given time step can be in 3 possible states with regard to the individual's memory. First, the stimulus can be innately encoded, in which case we denote by π_I the average payoff obtained from the response to it. Second, the stimulus encountered can be present in the dynamic memory, in which case the response results in average payoff π_L. Third, the stimulus may not be present at all in the memory of the individual (it is "unknown"), in which case the individual's response results in average payoff π_U.

Because our main interest is in understanding the environmental conditions that favor learning, we assume (conservatively) that expressing a genetically determined action always leads to the "correct" payoff: $\pi_I = \pi_C$. When an individual encounters a stimulus that is not in its memory, we assume it samples an action at random. The expected payoff obtained by choosing an action randomly is denoted π_U, and is assumed to satisfy $\pi_W \leq \pi_U \leq \pi_C$. Finally, we assume that π_L, the payoff for learned responses to stimuli, is a constant satisfying $\pi_U \leq \pi_L \leq \pi_C$, so that an action for a stimulus present in dynamic memory leads to a higher payoff than if it was tried out randomly, because learning allows to sample the environment. A distinctive simplifying feature of our model, which gives analytical traction, is that we do not model explicitly the learning dynamics of such association between actions and stimuli (for instance by way of reinforcement learning). But by enforcing $\pi_U \leq \pi_L \leq \pi_C$, we implicitly capture any learning mechanism, from a very crude one where essentially no information is gathered if $\pi_L \approx \pi_U$ to a very sophisticated one if $\pi_L \approx \pi_C$.

Owing to the assumption that the stimuli are met by the individual in a stationary uniform distribution, the probability $P_I(g)$ that a currently encountered stimulus is innately encoded is independent of time; namely $P_I(g) = g/N_s$. Assuming that the individual interacts a very long time with its environment, we have that the asymptotic probability, $P_L(g)$ that an encountered stimulus is in the dynamic memory is also independent of time (this assumption can indeed be justified for the environmental process we consider in this article, see the Appendix). With this, we can then write the average payoff to an individual with genetic memory of size g as

$$f(g) = P_I(g)\pi_C + [1 - P_I(g)](P_L(g)\pi_L + [1 - P_L(g)]\pi_U). \qquad (1)$$

This equation captures the trade-off faced by the individual: Should it allocate memory to innate responses and respond optimally to only a limited number of stimuli (first term of Equation 1) or should it allocate memory to learning, and potentially learn to respond to many stimuli (second term of Equation 1)? Importantly, if the stimulus recall probability, $P_L(g)$, is a constant, independent of g, then the optimal number of innate memory slots, g, which maximizes payoff, is just $g^* = m$. Hence, learning does not evolve in this case (the same holds if $P_L(g)$ is increasing in g). This function thus requires that $P_L(g)$ is decreasing in g, at least on some subset of $[0, m]$, for learning to evolve. But the exact form of $P_L(g)$ will depend on how memory works, that is, for how long a stimulus is stored in memory before it is forgotten, which is detailed in the section "Memory" below.

We emphasize that Equation 1 holds for all generations of individuals and does not comprise random variables whose realizations might differ from one generation to the next. In particular, our model does not require that individuals encounter only a fraction of the total number N_s of stimuli, and in terms of the set of stimuli encountered by an individual, the environment is totally predictable between generations but not necessarily within generations. For this reason, and in contrast with previous approaches based on between-generation environmental fluctuations (Boyd and Richerson 1988; Stephens 1991; Feldman et al. 1996; Wakano et al. 2004; Aoki and Feldman 2014), the analysis of our model below will not require to explicitly capture evolutionary dynamics, but only within-generation dynamics, coupled with static optimization of Equation 1.

Memory

We endorse a simplified implementation of memory that is based on the functioning of the short-term memory in humans and animals (Baddeley 2003). Namely, we assume that when the individual has a dynamic memory ($g < m$) and meets an unknown stimulus, it always wants to store it. Because the dynamic memory is initially empty, the first $m - g$ encounters with non-innately encoded stimuli will simply result in the stimuli taking free slots until the dynamic memory is full. For subsequent decision steps, new stimuli will have to replace other ones in the dynamic memory. This is done via a replacement rule. We use the following replacement rule, which is taken from Kerr and Feldman (2003). A stimulus has a lifespan in memory of $m - g$ time steps (i.e., the size of the dynamic memory), starting from the last encounter with the stimulus. This means that if a stimulus is not met more than once in $m - g$ steps, it is forgotten. Otherwise, the stimulus stays in memory. With this rule, the dynamic memory will never contain more than $m - g$ stimuli.

Importantly, we assume that when a stimulus is replaced in memory, all the associated information is lost (note that this is another very conservative assumption because this means that we ignore the potential benefits of long-term memory). If this stimulus is encountered later, then the individual will have to relearn to interact with it (if the individual has the capacity to do so, i.e., if $g < m$). With this replacement rule, we can now evaluate $P_L(g)$ explicitly. We are then able to ascertain the evolutionarily stable value of g by taking the expected payoff (Equation 1) as our measure of fitness (Parker and Maynard-Smith 1990).

RESULTS

Stimulus recall probability

In the Appendix, we derive an expression for the stimulus recall probability, $P_L(g)$ (Equation A7). It turns out that this expression is cumbersome, but in Figure 1, we plot $P_L(g)$, which shows that it is decreasing in g, and when $g = m$, we have $P_L(m) = 0$. This decreasing pattern obtains because having a higher g means having fewer slots for the dynamic memory, which in turn implies that an individual will recall fewer steps of interaction with a given stimulus. The stimulus recall probability depends not only on the memory characteristics of the individual (m and g) but also on the 3 key environmental parameters of the model: environmental complexity (N_s), local complexity (N_b), and switching rate (γ).

As can be seen from Figure 1, $P_L(g)$ is decreasing in γ, which stems from the fact that when the block of stimuli changes more frequently, the probability to encounter the same given stimulus multiple times decreases (lower predictability). We also have that $P_L(g)$ is smaller in locally complex than in locally simple environments, because greater local complexity corresponds to more stimuli in a block. Finally, $P_L(g)$ is slowly decreasing with increasing N_s and eventually stabilizes for large N_s. This is mainly due to the fact that when the environment is complex, there is a very small probability that a given stimulus is in 2 different blocks of stimuli, so an individual will recall mainly interactions with stimuli within blocks, not across blocks.

The stimulus recall probability thus has 2 main properties, one related to the memory of the individual and the other one related to the environment: it is an increasing function of the allocation of memory to learning and also generally increases in the predictability of the environment.

Invasion of learning

When will learning be initially favored by selection? To answer this question, we consider a monomorphic population of "innates" with $g = m$ and ask when they will be invaded by a mutant learner strategy with only one memory slot allocated to the dynamic memory ($g = m - 1$), that is, when $f(m) < f(m - 1)$ is satisfied. This occurs when

$$P_L(m-1)[N_s - (m-1)](\pi_L - \pi_U) > \pi_C - \pi_U, \tag{2}$$

where

$$P_L(m-1) = \gamma \frac{1}{N_s} + (1-\gamma)\frac{1}{N_b} \tag{3}$$

is the stimulus recall probability for an individual with $m - 1$ innate responses. This is just the probability to encounter the same stimulus in 2 successive time steps (Equation A3) because an individual with one dynamic memory slot can only recall the last time step.

Equation 2 says that for a strategy with one dynamic memory slot to invade, the expected gain of learning summed over all non-innately encoded stimuli must exceed the gain to optimally respond to one stimulus. Indeed, the left-hand side of Equation 2 comprises the stimulus recall probability, $P_L(m - 1)$, the total number of non-innately encoded stimuli (and thus that can be learned about), $[N_s - (m - 1)]$, and the gain of learning to interact with one stimulus, $(\pi_L - \pi_U)$; the product of these 3 elements is the expected gain of learning for an individual with $g = m - 1$. The right-hand side of Equation 2 is the gain of optimally responding to one stimulus.

Making the environment more complex (increasing N_s) widens the range of parameters where learning invades because this makes larger the left-hand side of Equation 2. However, it is not enough that environmental complexity is higher than memory size; the difference $[N_s - (m - 1)]$ has to be substantial for learning to evolve. It is important to emphasize that even when the payoff due to learning, π_L, is only slightly higher than the random payoff, π_U, environmental complexity still has a positive effect on the evolution of learning; one just needs to make the environment complex enough for Equation 2 to be satisfied. Hence, the model captures well the evolutionary transition from optimal innate behavioral plasticity to an imperfect learning system, which is in theory the possible first state of a learning ability.

Equation 2 also shows that the environment should be predictable enough for learning to be favored by selection. In terms of our model parameters, this translates as a sufficiently low environmental switching rate (γ small enough) and as a sufficiently low local complexity (N_b smaller enough than N_s). This is because $P_L(m - 1)$ is decreasing in both N_b and γ, and so is the left-hand side of Equation 2. Intuitively, if $N_b = N_s$, there is only one block of stimuli of size N_s, and the individual must cope with the total amount of complexity at once; a learner with a limited memory size cannot cope with such a task. Also, γ should be small enough so that the average period of interaction with a block, $1/\gamma$ is large enough compared with the local complexity, N_b. This will allow a learner to interact many times with the same stimulus and learn the best response to it.

Optimal memory allocation

We now turn to investigate numerically the optimal value g^* that maximizes fitness (Equation 1), that is, the evolutionarily stable allocation of memory to innate behavioral plasticity. First and foremost, we find that increasing the complexity of the environment increases the optimal size of the dynamic memory: g^* is decreasing with increasing N_s. This is because in complex environments, where N_s is much larger than the memory size m, allocating one more slot to the innate memory has only a small effect on fitness (first term of Equation 1); in contrast, allocating this slot to the dynamic memory always results in an increase of the stimulus recall probability, even for large N_s (second term of Equation 1 and Figure 1). Likewise, increasing the efficiency of the learning system, π_L, also decreases the value of g^* (Figure 2a) because increasing π_L means having a higher benefit of learning (i.e., this increases the value of the second term of Equation 1) for an individual with a given g.

The environmental switching rate, γ, has a simple effect on optimal memory allocation. In agreement with the invasion condition (Equation 2) and the above fact that the probability to recall a stimulus is decreasing in γ, the optimal number of innate responses, g^*, is increasing in γ. This makes sense since in an environment changing more frequently, it is less beneficial to recall events farther in the past. However, there is a threshold effect when we have no environmental change at all ($\gamma = 0$), and the individual is faced with only a random sample from the environment for its entire lifespan. This leads to a null model where the environment is totally random and learning is useless.

Local complexity, N_b, has a non-monotonic effect on g^* (see in particular Figure 2b–d). For locally complex environments, g^* is increasing in N_b, whereas for locally simple environments, it is decreasing in N_b. Hence, the maximum allocation of memory to learning occurs at moderate levels of local complexity. This pattern is explained in terms of the marginal gains of allocating memory slots to either part of the memory. When N_b is high, the gains from allocating a memory slot to the innate memory are higher than the gains from allocating it to the dynamic memory because there is only a small probability to recall interactions with stimuli (low predictability). When N_b is low, the converse is true: It is more beneficial to allocate a slot to the dynamic memory because the stimulus recall probability is high (high predictability).

We also looked at various 2-way interactions between parameters. First, it is interesting to investigate the interaction between the switching rate and local complexity (Figure 2b) because together they determine the predictability of the environment. We find that the environmental switching rate sharpens the non-monotonic effect of N_b described above. This is because in fast-changing environments, there is very low predictability irrespective of local complexity because there is only a small probability to encounter twice the same stimulus (in this case, $P_L(g)$ is flat, Figure 1). In slowly changing environments where interaction periods are longer, the stimulus recall probability is now highly dependent on N_b and g. Hence, for slowly changing environments (low γ), high N_b corresponds to low predictability and low N_b to high predictability, and we recover the non-monotonic effect of N_b found above.

Strikingly, there is no interaction effect between local complexity and memory size on g^*, provided $N_s - m$ is kept constant (Figure 2c). This is intriguing because one would think that if the environment is locally complex (N_b large), then there is a very small probability to interact many times with a given stimulus in a short period of time, thereby rendering learning more difficult for an individual with a given memory m. However, making N_b large means making N_s at least as large, so the interaction between N_b and m is already captured by the main effect of $N_s - m$, which explains why there is no interaction effect between N_b and $N_s - m$ on g^*. The payoff for learned responses to stimuli, π_L, finally, has the role of making the effect of the other parameters more abrupt. For instance, when π_L increases, we observe that below a threshold value of block size N_b, all memory slots are dedicated to learning ($g^* = 0$, Figure 2d).

DISCUSSION

In this study, we investigated the evolutionary transition from innate behavioral plasticity to learning and showed with an analytical model that environmental complexity (operationalized as the number of stimuli in the environment, N_s) favors the evolution of learning. Because we considered an environment that is constant across generations, yet where learning can invade, our results demonstrate that between-generation environmental variability is not necessary for learning to evolve. Moreover, we found broad conditions where learning coexists with innate behavioral plasticity (i.e., $0 < g^* < m$).

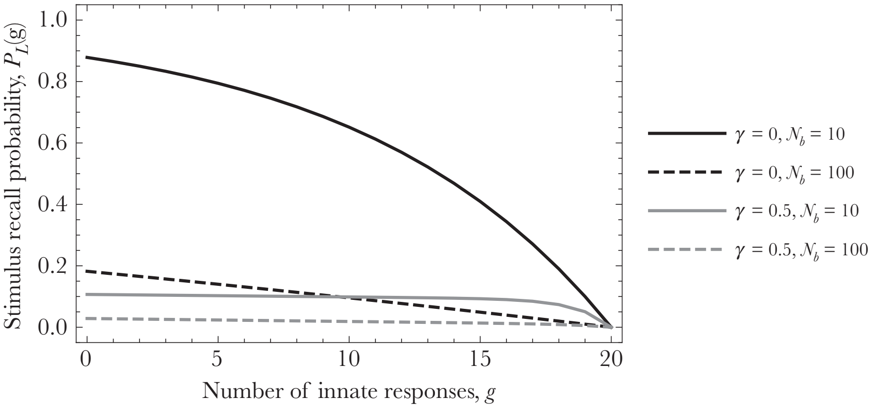

Figure 1

Stimulus recall probability, $P_L(g)$, as a function of number of innate responses (g) environmental switching rate (γ), and local complexity (N_b). Parameter values: $m = 20$, $N_s = 1000$.

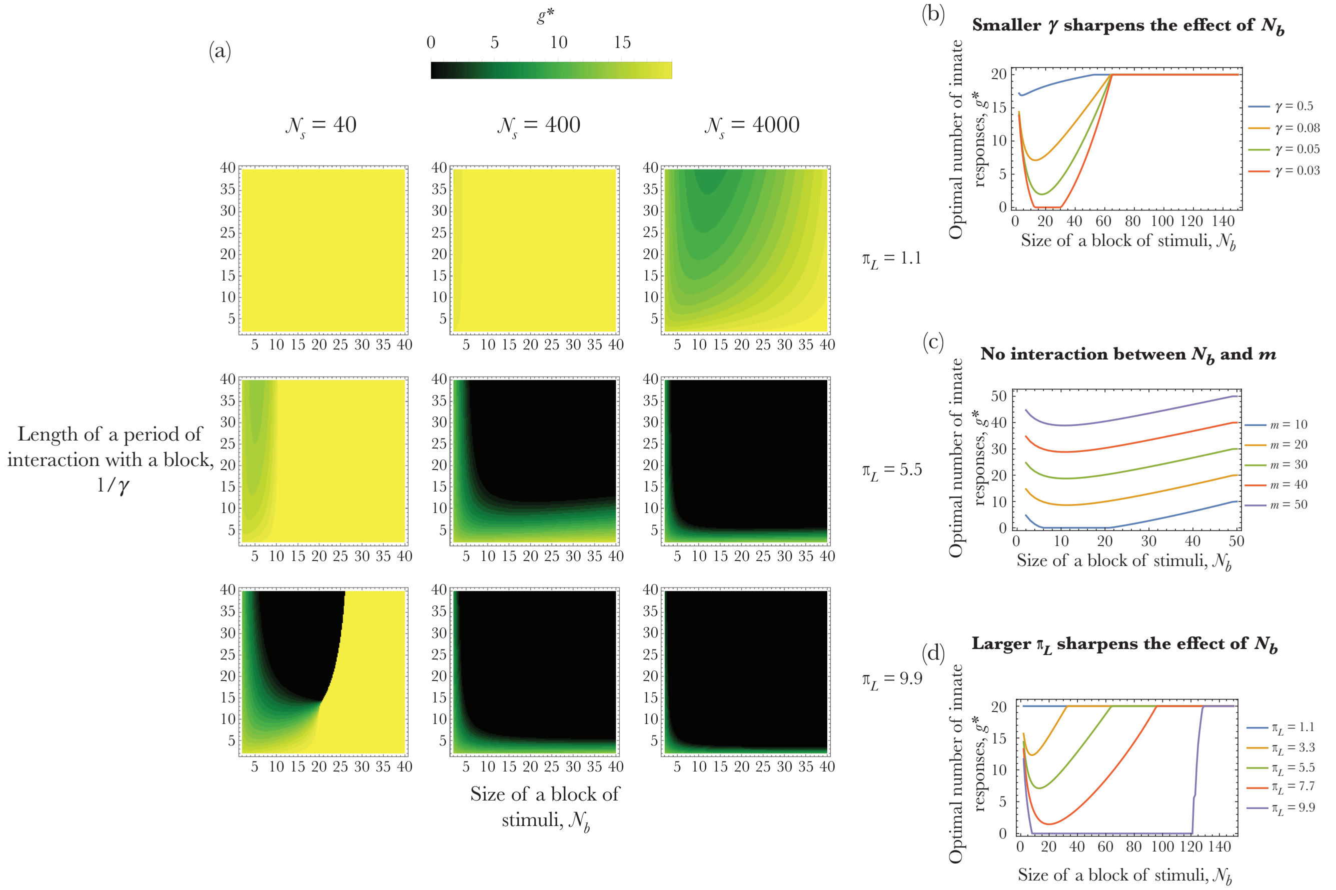

Figure 2

Optimal number of innate responses, g^*, that maximizes Equation 1, as a function of environmental parameters and efficiency of learning mechanism. (a) Concomitant effects of environmental complexity (N_s, columns), efficiency of learning mechanism (π_L, rows), local complexity (N_b, x axis), and environmental switching rate (γ, y axis) (parameter values: $m = 20$, $\pi_C = 10$, $\pi_U = 1$). (b) Interaction between N_b and γ (parameter values: same as in (a) and $N_s = 150$, $\pi_L = 5.5$). (c) Interaction between N_b and m (parameter values: same as in (a) and $N_s = m + 100$, $\pi_L = 5.5$, $\gamma = 1/12$). (d) Interaction between N_b and π_L (parameter values: same as in (a) and $m = 20$, $N_s = 150$, $\gamma = 1/12$).

Summary of results

Our results are 2-fold. First, we provide conditions for learners to invade a population of individuals relying on innate behavioral plasticity. We find that increasing the complexity of the environment widens the range of conditions under which learners can invade. This is due to the fact that with innate behavioral plasticity in a complex environment, an individual cannot form new stimulus-response associations during its lifespan and thus can respond to only a limited number of stimuli. We also find that environmental complexity is not a sufficient condition for learning to evolve. Namely, we confirm previous results showing that the environment needs to display predictability within an individual's lifespan (Stephens 1991; Dunlap and Stephens 2009, 2014). In locally complex environments, where an individual interacts with blocks consisting of many stimuli at the same time, predictability is very low, that is, there is a small probability to encounter the same stimulus in a short period of time. In this case, learning cannot invade because the limited memory of the individual is unable to deal with the entire environmental complexity at once.

Our second type of result is related to the optimal allocation of memory between innateness and learning. In our model, this optimal allocation is determined by the trade-off between using memory to respond optimally and innately to only a few stimuli, versus using this memory to learn to respond suboptimally to potentially many stimuli. As could be anticipated from the invasion results, we find that the optimal allocation of memory to learning increases with the complexity and the predictability of the environment. Moreover, we find that the maximum allocation to learning occurs in moderately predictable environments. These are the environments where block turnover is small (i.e., low switching rate) but where local complexity (or block size) is intermediate. When these 2 last conditions are met, it can even be optimal that individuals are born with a "blank slate" (all memory slots allocated to the dynamic memory). Our numerical results indicate that regions of the parameter space where the evolution of a blank slate is possible are relatively large, especially when the learning mechanism is very efficient ($\pi_L \approx \pi_C$, bottom row of Figure 2). To reconcile this result with the observation that blank slate organisms arguably do not exist in nature (Mameli and Bateson 2006), this suggests that there should be constraints impeding the evolution and maintenance of close-to-optimal learning efficiency (i.e., in our parameter space, realistic regions are such that $\pi_L \ll \pi_C$). Overall, the results show that learning supplemented with forgetting represents an efficient way to deal with environments that are complex on the global scale but are relatively simple on a local

Table 1

Symbols used in the model

Symbol	Signification
N_s	Number of stimuli in the environment
N_b	Size of a block of stimuli
γ	Environmental switching rate (probability to draw a new block of stimuli)
π_C	Payoff associated to a "correct" or optimal response to a stimulus
π_W	Payoff associated to a "wrong" or suboptimal response to a stimulus
π_U	Payoff obtained by choosing responses randomly
π_L	Payoff associated to a learned response to a stimulus
m	Memory size or number of stimuli that can be stored in memory
g	Number of stimuli to which the individual has an innate response
$P_I(g)$	Probability that a stimulus is in the innate memory of an individual with g innate responses
$P_L(g)$	Probability that a stimulus is in the dynamic memory of an individual with g innate responses

scale. But, how can we measure complexity in the real world? Also, what phenotypes are affected by environmental complexity?

Empirical predictions

The expression "environmental complexity" used in this article refers to the number of fitness-relevant stimuli a given individual is likely to encounter and distinguish in the course of its lifespan. One could have defined environmental complexity in a number of other ways—for instance complexity could refer to the "difficulty" of responding to each particular stimulus in the environment—but our use of complexity is related to previous discussions of this notion in the context of the evolution of cognition (Godfrey-Smith 2002; Jones and Blackwell 2011; Fawcett et al. 2014), and has links to the concept of "space complexity" in computer science (Adami et al. 2000). Loosely speaking, space complexity is the amount of memory (as a function of input size) required by an algorithm to solve a given problem. In our setting, an algorithm (i.e., genotype) that would encode an optimal response to every possible stimulus would have a high "space complexity."

Another advantage of our definition is that it may be easier to measure empirically. There are at least 3 measurable ecological and psychological factors that directly or indirectly influence complexity. First, the complexity of an organism's environment should be positively correlated with the organism's habitat range (appropriately rescaled to account for body size). Individuals from species exploring vast areas in order to forage, migrate, or reproduce should encounter more types of biotic and abiotic stimuli than individuals from other species. Second, the level of detail that an organism's sensory system can perceive is also likely to allow an individual to distinguish between many stimuli (typically, species with a visual ability that are color-blind "miss" one dimension of the world's complexity). Third, lifespan is a factor that will affect the number of distinct stimuli or challenges encountered by a given individual: Species with longer lifespan should be exposed to a greater variety of stimuli.

Our results thus predict that these factors should be positively correlated with learning ability. Mainly, we expect that species scoring very low on the 3 dimensions of complexity highlighted above should be those species that have a scant ability to learn, and rely on simpler forms of plasticity (that we termed "innate behavioral plasticity"). Although measuring learning ability across species may not be easy, classical studies in comparative cognition have shown that instrumental learning occurs in many species (Shettleworth 2009), and parallel instrumental learning of several stimuli (Banai et al. 2010) may be a good way to evaluate the kind of learning that is required in our model. Recent empirical advances have also demonstrated that comparing specific cognitive abilities across a number of different species is possible provided species' specialists coordinate their efforts (MacLean et al. 2014). Finally, our results suggest that the capacity (or size) of individuals' working memory is likely to increase with environmental complexity. The size of working memory corresponds to the number of stimuli (or chunks of information) an individual can hold in short-term memory for further processing and use (Carruthers 2013).

Model realism

Our model is obviously a simplification, but we argue that the environment we considered is representative of those faced by many animals. To give a concrete example of the range of settings where our model applies, we can take daily routines (Houston and McNamara 1999). In each part of its routine, an animal interacts with a given subset of stimuli that are present at a given location and time because of statistical regularities in the environment. For example, when an individual visits a particular food patch in the course of foraging, it may encounter different types of food items, but also individuals from other species that have overlapping diet, as well as predators awaiting their preys. All of these constitute the block of stimuli met by the individual on this particular food patch; when visiting other food patches and performing other tasks, the individual will encounter other blocks of stimuli (that may or may not contain the stimuli previously encountered). In nature an animal can further decide which stimulus (or block of stimuli) it interacts with, thus the complexity of the environment is under partial control of the animal. If animals are stimulus specialists (i.e., they tend to prefer and interact with a small number of stimuli) rather than generalists, this may reduce the effective complexity of the environment and potentially reduce the range of conditions where learning can evolve. However, whether animals are specialists or generalist may depend on many factors. This points toward a more advanced modeling of such problems, where stimulus preference (or stimulus choice) is under selection and can also be learned (Kerr 2007).

Although our model is general enough to capture features of real environments, our implementation of memory is relatively specific. In order to ground our results on the most possible conservative assumptions, we made 2 notable simplifications. First, we focused on instrumental rather than associative learning. In artificial selection experiments where the evolution of learning was shown to be favored by between-generation environmental variability (Mery and Kawecki 2004; Dunlap and Stephens 2009, 2014), it was associative rather than instrumental learning that was considered. It is likely that the conditions favoring associative learning are different than the ones favoring instrumental learning. However, in most models of the evolution of learning, the modeling approach is so abstract that it may encompass both forms of learning (Boyd and Richerson 1988; Stephens 1991; Feldman et al. 1996; Wakano et al. 2004). These mechanistic considerations need further evolutionary investigation and clarification.

Our second simplification is that we considered only features of the short-term (or working) memory (Shettleworth 2009; Banai et al. 2010). In contrast with long-term memory, the events stored in working memory follow a dynamic process such that they enter memory when an individual starts interacting with a given stimulus, but such interactions are replaced by other ones as the individual interacts with different stimuli. Indeed, animals and humans can hold only a given, small number of items or stimuli in working memory (Miller 1956; Dudai, 2004). This is captured by the dynamic part of the memory in our model, where stimulus-action associations are totally removed from memory once they are replaced by other ones.

The "innate" part of the memory, on the other hand, captures many of the examples in nature showing that animals tend to have innate, hard-wired responses to stimuli (Mery and Kawecki 2004; Riffell et al. 2008; Gong 2012). For instance, certain ants have innate templates of enemies in memory (Dorosheva et al. 2011) and human infants innately distinguish between face-like stimuli and other stimuli, indicating that the neuronal networks responsible for visual perception have a particular innate wiring structure (Slater and Kirby 1998; see also Perin et al. 2011 on a generalization of this idea to the innate structure of the whole neocortex of mice).

In conclusion, we have shown in this article that environmental complexity can give rise to the evolutionary emergence of a learning ability, when environmental information cannot be genetically encoded. This finding was possible because we treated learning as one form of behavioral plasticity, and searched for the specific advantage of learning over innate behavioral plasticity. When memory is finite, learning can invade a population where individuals have optimal innate responses because, when coupled with forgetting, it enables individuals to respond to an overall larger spectrum of stimuli. We finally produced a number of predictions relating environmental complexity to measurable ecological and psychological parameters that invite to empirical tests of our model.

FUNDING

This work was supported by Swiss NSF grant PP00P3-123344.

We thank J.-B. André for stimulating us to investigate the role of environmental complexity in the evolution of learning. We also thank T. Kawecki, S. Powers, M. Dos Santos, and M. González-Forero for comments and suggestions about the model.

Handling editor: David Stephens

APPENDIX

Here, we derive an expression for the probability to recall a stimulus $P_{\mathrm{L}}(g)$. At each time step t ($t = 1, 2, \dots$) of the environmental process, a stimulus s_t is drawn from the set of environmental stimuli, S, according to the procedure described in the main text (i.e., s_t is a random variable). Now, given our implementation of memory, where a given stimulus stays in dynamic memory $m - g$ time steps, $P_{\mathrm{L}}(g)$ is the probability that, conditional on being met once, this stimulus is met a second time in a period of length less or equal than $m - g$ time steps. Let us denote by $R_{\mathrm{s}} \in \{1,2,\dots\}$ the number of time steps occurring between 2 encounters with a given stimulus (this is a random variable that is independent of time at stationarity). Then

$$P_{\mathrm{L}}(g) = P\{R_{\mathrm{s}} \leq m - g\}. \tag{A1}$$

We will compute this expression by letting R_{s} be the return (or recurrence) time of a backward Markov chain (Grimmett and Stirzaker 2001) that can be constructed from our assumptions about the environment.

To that end, let $B_t \subset S$ be the block of stimuli ($|B_t| = N_{\mathrm{b}}$) encountered at time t, $s_t \in B_t$ be the stimulus encountered at t, and $s_{\mathrm{f}} \in S$ be a given focal stimulus. We can then define the 3 mutually exclusive events

$$\begin{aligned} s &= \{s_t = s_{\mathrm{f}}\}, \\ i &= \{s_t \neq s_{\mathrm{f}}, s_{\mathrm{f}} \in B_t\}, \\ o &= \{s_{\mathrm{f}} \notin B_t\}, \end{aligned} \tag{A2}$$

where s is the event that the stimulus encountered at time t is the focal one, i is the event that the focal stimulus is in the current block but is not the currently encountered stimulus, and o is the event that the focal stimulus is not in the current block.

We can now define a Markov chain on these 3 states: s, i, and o and compute from it the recurrence time to s. From our assumptions, the forward transition probabilities ρ_{jk} of this chain are

$$\begin{aligned}
\rho_{ss} &= \gamma \frac{N_{\mathrm{b}}}{N_{\mathrm{s}}} \frac{1}{N_{\mathrm{b}}} + (1-\gamma)\frac{1}{N_{\mathrm{b}}}, \\
\rho_{si} &= \gamma \frac{N_{\mathrm{b}}}{N_{\mathrm{s}}} \frac{N_{\mathrm{b}}-1}{N_{\mathrm{b}}} + (1-\gamma)\frac{N_{\mathrm{b}}-1}{N_{\mathrm{b}}}, \\
\rho_{so} &= \gamma \frac{N_{\mathrm{s}}-N_{\mathrm{b}}}{N_{\mathrm{s}}}, \\
\rho_{is} &= \gamma \frac{N_{\mathrm{b}}}{N_{\mathrm{s}}} \frac{1}{N_{\mathrm{b}}} + (1-\gamma)\frac{1}{N_{\mathrm{b}}}, \\
\rho_{ii} &= \gamma \frac{N_{\mathrm{b}}}{N_{\mathrm{s}}} \frac{N_{\mathrm{b}}-1}{N_{\mathrm{b}}} + (1-\gamma)\frac{N_{\mathrm{b}}-1}{N_{\mathrm{b}}}, \\
\rho_{io} &= \gamma \frac{N_{\mathrm{s}}-N_{\mathrm{b}}}{N_{\mathrm{s}}}, \\
\rho_{os} &= \gamma \frac{N_{\mathrm{b}}}{N_{\mathrm{s}}} \frac{1}{N_{\mathrm{b}}}, \\
\rho_{oi} &= \gamma \frac{N_{\mathrm{b}}}{N_{\mathrm{s}}} \frac{N_{\mathrm{b}}-1}{N_{\mathrm{b}}}, \\
\rho_{oo} &= \gamma \frac{N_{\mathrm{s}}-N_{\mathrm{b}}}{N_{\mathrm{s}}} + 1 - \gamma.
\end{aligned} \tag{A3}$$

For instance, the probability ρ_{ss} to move from state s to itself takes this form because a new block is drawn with probability γ, in which case the focal stimulus s_{f} makes part of the new block with probability $N_{\mathrm{b}}/N_{\mathrm{s}}$, and is drawn from within the block with probability $1/N_{\mathrm{b}}$. If one does not change block, which happens with probability $1 - \gamma$, the probability to draw s_{f} from the current block is $1/N_{\mathrm{b}}$ because s_{f} already makes part of the current block.

From the above transition probabilities, we can define the backward transition probabilities $\hat{\rho}_{kj}$. and owing to our stationarity assumption, this is given by

$$\hat{\rho}_{kj} = \frac{\mu_j}{\mu_k} \rho_{jk} \tag{A4}$$

(see, e.g., Theorem 1.9.1 of Noris 1998), where μ_j and μ_k are, respectively, the probabilities of state j and k under the stationary distribution of the forward Markov chain defined by the transition probabilities ρ_{jk}. Equation A4 defines a backward Markov chain. Because the stationary probabilities are given by $\mu_s = 1/N_{\mathrm{s}}$, $\mu_i = (N_{\mathrm{b}} - 1)/N_{\mathrm{s}}$, and $\mu_o = (N_{\mathrm{s}} - N_{\mathrm{b}})/N_{\mathrm{s}}$, we find using Equation A3 that

$$\hat{\rho}_{jk} = \rho_{jk}, \quad \forall j,k \in \{s,i,o\}. \tag{A5}$$

Now, we can compute the distribution of return times R_s. To do so, denote v^h_{jk} the probability that, starting from state j, the first visit to state k occurs h steps in the past. With this, we have that $P\{R_s = h\} = v^h_{ss}$. In order to find v^h_{ss}, we note that the probabilities of first visit obey the recursions

$$\begin{aligned} v^h_{ss} &= \hat{\rho}_{so} v^{h-1}_{os} + \hat{\rho}_{si} v^{h-1}_{is}, \\ v^h_{is} &= \hat{\rho}_{io} v^{h-1}_{os} + \hat{\rho}_{ii} v^{h-1}_{is}, \\ v^h_{os} &= \hat{\rho}_{oo} v^{h-1}_{os} + \hat{\rho}_{oi} v^{h-1}_{is}. \end{aligned} \tag{A6}$$

Solving this linear system of difference equations provides the probabilities of first visit on the left-hand side, including $v^h_{ss} = P\{R_s = h\}$ (this expression is too long to be displayed here, but a Mathematica notebook containing it is available on request). The probability that $R_s \le m - g$ $(m - g > 0)$ can then be computed by summing all the possible cases up to $m - g$, namely,

$$P_L(g) = P\{R_s \le m - g\} = \sum_{h=1}^{m-g} P\{R_s = h\}. \tag{A7}$$

It turns out that, substituting the explicit expression of $P\{R_s = h\} = v^h_{ss}$, we obtain a closed-form expression, but that is unfortunately too long to provide direct insight (again, a Mathematica notebook containing this expression is available on request).

REFERENCES

Adami C, Ofria C, Collier TC. 2000. Evolution of biological complexity. Proc Natl Acad Sci USA. 97:4463–4468.

Aoki K, Feldman MW. 2014. Evolution of learning strategies in temporally and spatially variable environments: a review of theory. Theor Popul Biol. 91:3–19.

Baddeley A. 2003. Working memory: looking back and looking forward. Nat Rev Neurosci. 4:829–839.

Banai K, Ortiz JA, Oppenheimer JD, Wright BA. 2010. Learning two things at once: differential constraints on the acquisition and consolidation of perceptual learning. Neuroscience. 165:436–444.

Botero CA, Weissing FJ, Wright J, Rubenstein DR. 2015. Evolutionary tipping points in the capacity to adapt to environmental change. Proc Natl Acad Sci USA. 112:184–189.

Boyd R, Richerson PJ. 1988. Culture and the evolutionary process. Chicago (IL): University of Chicago Press.

Burgess SC, Marshall DJ. 2014. Adaptive parental effects: the importance of estimating environmental predictability and offspring fitness appropriately. Oikos. 123:769–776.

Carruthers P. 2013. Evolution of working memory. Proc Natl Acad Sci USA. 110(Suppl 2):10371–10378.

Dorosheva EA, Yakovlev IK, Reznikova ZI. 2011. An innate template for enemy recognition in red wood ants. Entomol Rev. 91:274–280.

Dudai Y. 2004. The neurobiology of consolidations, or, how stable is the engram? Annu Rev Psychol. 55:51–86.

Dunlap AS, Stephens DW. 2009. Components of change in the evolution of learning and unlearned preference. Proc Biol Sci. 276:3201–3208.

Dunlap AS, Stephens DW. 2014. Experimental evolution of prepared learning. Proc Natl Acad Sci USA. 111:11750–11755.

Enquist ME, Ghirlanda S. 2005. Neural networks and animal behavior. Princeton (NJ): Princeton University Press.

Fawcett TW, Hamblin S, Giraldeau L-A. 2013. Exposing the behavioral gambit: the evolution of learning and decision rules. Behav Ecol. 24:2–11.

Fawcett TW, Fallenstein B, Higginson AD, Houston AI, Mallpress DEW, Trimmer PC, McNamara JM. 2014. The evolution of decision rules in complex environments. Trends Cogn Sci. 18:153–161.

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

Godfrey-Smith P. 2002. Environmental complexity and the evolution of cognition. In: Sternberg RJ, Kaufman JA, editors. The evolution of intelligence. Mahwah (NJ): Psychology Press. p. 223–250.

Gomulkiewicz R, Kirkpatrick M. 1992. Quantitative genetics and the evolution of reaction norms. Evolution. 46:390–411.

Gong Z. 2012. Innate preference in *Drosophila melanogaster*. Sci China Life Sci. 55:8–14.

Grimmett GR, Stirzaker DR. 2001. Probability and random processes. 3rd edn. Oxford: Oxford University Press.

Hollis K, Guillette L. 2011. Associative learning in insects: evolutionary models, mushroom bodies, and a neuroscientific conundrum. Comp Cogn Behav Rev. 6:25–46.

Houston A, McNamara JM. 1999. Models of adaptive behaviour. Cambridge (UK): Cambridge University Press.

Johnston TD. 1982. Selective costs and benefits in the evolution of learning. In: Jay RAH, Rosenblatt S, editors. Advances in the study of behavior. Vol. 12. New York: Academic Press. p. 65–106.

Jones D, Blackwell T. 2011. Social learning and evolution in a structured environment. Proceedings of the Eleventh European Conference on the Synthesis and Simulation of Living Systems; August 2008; Southampton, UK. Cambridge (MA): MIT Press. 380–387 p.

Kerr B. 2007. Niche construction and cognitive evolution. Biol Theory. 2:250–262.

Kerr B, Feldman MW. 2003. Carving the cognitive niche: optimal learning strategies in homogeneous and heterogeneous environments. J Theor Biol. 220:169–188.

Kraemer PJ, Golding JM. 1997. Adaptive forgetting in animals. Psychon Bull Rev. 4:480–491.

MacLean EL, Hare B, Nunn CL, Addessi E, Amici F, Anderson RC, Aureli F, Baker JM, Bania AE, Barnard AM et al. 2014. The evolution of self-control. Proc Natl Acad Sci USA. 111:E2140–E2148.

Mameli M, Bateson P. 2006. Innateness and the sciences. Biol Philos. 21:155–188.

Mery F, Burns JG. 2010. Behavioural plasticity: an interaction between evolution and experience. Evol Ecol. 24:571–583.

Mery F, Kawecki TJ. 2004. The effect of learning on experimental evolution of resource preference in *Drosophila melanogaster*. Evolution. 58:757–767.

Miller GA. 1956. The magical number seven, plus or minus two: some limits on our capacity for processing information. Psychol Rev. 63:81–97.

Norris J. 1998. Markov chains. Cambridge (UK): Cambridge University Press.

Parker GA, Maynard-Smith J. 1990. Optimality theory in evolutionary biology. Nature. 348:27–33.

Pearce JM. 2008. Animal learning and cognition: an introduction. 3rd ed. Hove (UK): Psychology Press.

Perin R, Berger TK, Markram H. 2011. A synaptic organizing principle for cortical neuronal groups. Proc Natl Acad Sci USA. 108:5419–5424.

Pigliucci M. 2001. Phenotypic plasticity: beyond nature and nurture. Baltimore (MD): Johns Hopkins University Press.

Riffell JA, Alarcón R, Abrell L, Davidowitz G, Bronstein JL, Hildebrand JG. 2008. Behavioral consequences of innate preferences and olfactory learning in hawkmoth–flower interactions. Proc Natl Acad Sci USA. 105:3404–3409.

Shettleworth SJ. 2009. Cognition, evolution, and behavior. Oxford: Oxford University Press.

Slater A, Kirby R. 1998. Innate and learned perceptual abilities in the newborn infant. Exp Brain Res. 123:90–94.

Snell-Rood EC. 2013. An overview of the evolutionary causes and consequences of behavioural plasticity. Anim Behav. 85:1004–1011.

Stamps J, Groothuis TG. 2010. The development of animal personality: relevance, concepts and perspectives. Biol Rev Camb Philos Soc. 85:301–325.

Stephens DW. 1991. Change, regularity, and value in the evolution of animal learning. Behav Ecol. 2:77–89.

Wakano JY, Aoki K, Feldman MW. 2004. Evolution of social learning: a mathematical analysis. Theor Popul Biol. 66:249–258.