



## Original Article

# The handaxe and the microscope: individual and social learning in a multidimensional model of adaptation <sup>☆</sup>

Laurent Lehmann <sup>a,\*</sup>, Joe Yuichiro Wakano <sup>b</sup>

<sup>a</sup> Department of Ecology and Evolution, University of Lausanne, Switzerland

<sup>b</sup> Meiji Institute for Advanced Study of Mathematical Sciences, Meiji University, Japan

## ARTICLE INFO

## Article history:

Initial receipt 21 May 2012

Final revision received 1 November 2012

## Keywords:

Cultural evolution  
Reinforcement learning  
Fictitious play  
Social learning  
Adaptation  
Demography  
Limits to learning

## ABSTRACT

When individuals learn by trial-and-error, they perform randomly chosen actions and then reinforce those actions that led to a high payoff. However, individuals do not always have to physically perform an action in order to evaluate its consequences. Rather, they may be able to mentally simulate actions and their consequences without actually performing them. Such fictitious learners can select actions with high payoffs without making long chains of trial-and-error learning. Here, we analyze the evolution of an  $n$ -dimensional cultural trait (or artifact) by learning, in a payoff landscape with a single optimum. We derive the stochastic learning dynamics of the distance to the optimum in trait space when choice between alternative artifacts follows the standard logit choice rule. We show that for both trial-and-error and fictitious learners, the learning dynamics stabilize at an approximate distance of  $\sqrt{n/(2\lambda_e)}$  away from the optimum, where  $\lambda_e$  is an effective learning performance parameter depending on the learning rule under scrutiny. Individual learners are thus unlikely to reach the optimum when traits are complex ( $n$  large), and so face a barrier to further improvement of the artifact. We show, however, that this barrier can be significantly reduced in a large population of learners performing payoff-biased social learning, in which case  $\lambda_e$  becomes proportional to population size. Overall, our results illustrate the effects of errors in learning, levels of cognition, and population size for the evolution of complex cultural traits.

© 2013 Elsevier Inc. All rights reserved.

## 1. Introduction

Making tools is a defining feature of *Homo sapiens*. From assembling molecules to building space-stations, no other species on earth has innovated as many technological artifacts as humans. This innovation process is leading to a demo-cultural explosion (Kremer, 1993; Ghirlanda & Enquist, 2007). Archaic humans, from *Homo habilis* to *Homo neanderthalensis*, on the other hand are characterized by small rates of cultural innovation. Stone tool artifacts have remained almost constant for about one million years and subsistence strategies did not change for long stretches of time (Klein, 2009). The enhanced learning ability and the increased population size of *Homo sapiens* are possible reasons for the acceleration of artifact evolution in modern humans (Klein, 2009).

Learning is the acquisition and/or change of a trait during an individual's lifespan through experience and social transmission (Dugatkin, 2004). Since most human artifacts are unlikely to be genetically encoded, they must be acquired by learning. The different rates of cultural evolution observed in the archeological record raise

the question as to what type of learning mechanism can increase the rate of innovation and the complexity of artifacts used by individuals in a population.

Artifacts, whether they are made of bone, stone, iron, or any other material, can be considered as multidimensional traits. The payoff (or utility) brought to its carrier by a specific kind of stone tool, such as a handaxe, depends on the adjustment and complementarity of several underlying quantitative factors, such as its weight, center of gravity, sharpness of the edge, shape of the grip, quality of the stone material, and so on. For example, small Neolithic arrowheads or spearheads, characteristic of *Homo sapiens*, might be almost useless if they are not combined with wooden arrows or spears. Only the correct combination of underlying factors may yield a large payoff to the carrier of an arrow or a spear. Adjustment and complementarity between factors are even more crucial for modern tools ranging from boomerangs and canoes to aircrafts.

In this paper, we extend Fisher's (Fisher, 1930) multidimensional model of adaptation in order to study rates of artifact evolution through individual and social learning. We envision an optimal artifact located at the origin of an  $n$ -dimensional quantitative factor space. The payoff brought by an artifact decreases as the Euclidean distance to the origin in this state space increases. We then compare mathematically the effect on the rate of artifact improvement of two individual learning rules.

<sup>☆</sup> This work is partly supported by Monbukagakusho grant 22101001 and Swiss NSF grant PP00P3-123344.

\* Corresponding author.

E-mail address: [laurent.lehmann@unil.ch](mailto:laurent.lehmann@unil.ch) (L. Lehmann).

The first learning rule is trial-and-error learning, which is standard in animal behavior, psychology, and economics (e.g., Bush & Mosteller, 1951; Fudenberg & Levine, 1998; Camerer & Ho, 1999; Dugatkin, 2004). This relies on the reinforcement of actions that lead to high payoffs, where the actions are generally obtained by random trials of behaviors. The second rule is fictitious learning, whose name and characteristics follow from the learning rule entitled “fictitious play” commonly used in game theory and economics (e.g., Fudenberg & Levine, 1998; Camerer & Ho, 1999; Hopkins, 2002). Here, an individual expresses actions that would lead to high payoffs if they were actually carried out. A fictitious learner (or a fictitious player in game theory) can evaluate through mental simulations the consequences of its actions on payoff, before expressing them. Instead of performing random trials of actions, it chooses the actions that are evaluated as best from a set of conceivable actions.

Here, we first analyze the stochastic learning dynamics of the distance to the optimum in an  $n$ -dimensional factor space, for both trial-and-error and fictitious learning. We then compare the results to the situation where a population of individuals can also use social transmission to learn the artifact.

## 2. Baseline modeling assumptions

### 2.1. Multidimensional artifacts

Consider an individual expressing a learned trait or artifact, which is made up of  $n$  underlying quantitative characters  $z_1, z_2, \dots, z_n$  that are all assumed to vary continuously. The artifact can be characterized by the vector  $\mathbf{z} = (z_1, z_2, \dots, z_n)$  of trait values in trait space ( $\mathbf{z} \in R^n$ ). For instance, a handaxe could be characterized by its tip length, base length, and width ( $n=3$ ).

The expression of trait  $\mathbf{z}$  is assumed to result in some payoff to its carrier. The payoff landscape is taken to be spherically symmetric and centered at the origin of the state space. The payoff  $u$  of expressing trait  $\mathbf{z}$  then depends only on the Euclidean distance  $\|\mathbf{z}\| = \sqrt{z_1^2 + z_2^2 + \dots + z_n^2}$  of the trait  $\mathbf{z}$  to the origin [ $u = u(\|\mathbf{z}\|)$ , Fig. 1].

In essence, we are describing Fisher’s geometrical model of adaptation (e.g., Fisher, 1930; Hartl & Taubes, 1996; Poon & Otto,

2000; Waxman, 2006; Martin & Lenormand, 2006). Fisher (1930) compared the requirement of conformity of the various factors underlying an adaptation with the mechanical adjustment of the various components of a microscope, which must be finely adjusted to produce distinct vision. His model was constructed to capture complementarity between traits. If two different traits  $\mathbf{z}$  and  $\mathbf{z}'$  are equally useful to its carrier [ $u(\|\mathbf{z}\|) = u(\|\mathbf{z}'\|)$ ], then  $\mathbf{z}/2 + \mathbf{z}'/2$ , which is the half mixture of both traits, results in a higher payoff than either  $\mathbf{z}$  and  $\mathbf{z}'$ . Hence, if traits conform to each other, they may markedly increase payoff.

### 2.2. Two individual learning rules

We depart from the standard formulation of Fisher’s model by assuming that the trait is learned instead of being genetically inherited. When an individual expresses a new trait  $\mathbf{z}' = \mathbf{z} + \delta$  (Fig. 2), where  $\delta \equiv (\delta_1, \delta_2, \dots, \delta_n)$  collects the changes in the  $n$  different factors, how likely is it that the new artifact  $\mathbf{z}'$  is closer to the optimum than the current one  $\mathbf{z}$ ? In other words, how likely is it that a learner improves the artifact? The answer will depend on the type of learning rule used by the individual, since this determines the distribution of the deviation  $\delta$  and the ability to detect an improvement if it happens. We compare here the performance of the following two individual learning rules:

#### 2.2.1. Trial-and-error leaning (TL)

Here, an individual is assumed in a first stage to try out a new trait  $\mathbf{z}'$ . We assume that each component,  $z'_i = z_i + \delta_i$ , of the new trait  $\mathbf{z}'$  follows an independent Normal distribution with mean  $(1 - \alpha)z_i$  and variance  $\sigma^2$ , where  $\alpha$  is a directionality parameter varying between zero and one [ $z'_i \sim N((1 - \alpha)z_i, \sigma^2)$ ]. When  $\alpha = 0$ , each component  $z'_i$  is normally distributed with variance  $\sigma^2$  around the current value  $z_i$ , while when  $\alpha = 1$ , each  $z_i$  is normally distributed around the optimum. The parameter  $\alpha$  thus allows one to tune how much “insight” an individual has on the location of the optimal trait, and the probability density function  $T(\mathbf{z}' | \mathbf{z})$  of going from trait  $\mathbf{z}$  to  $\mathbf{z}'$  follows a multivariate normal distribution. When  $\alpha = 0$ , a trial-and-error learner has no insight and the new trait is chosen fully at random.

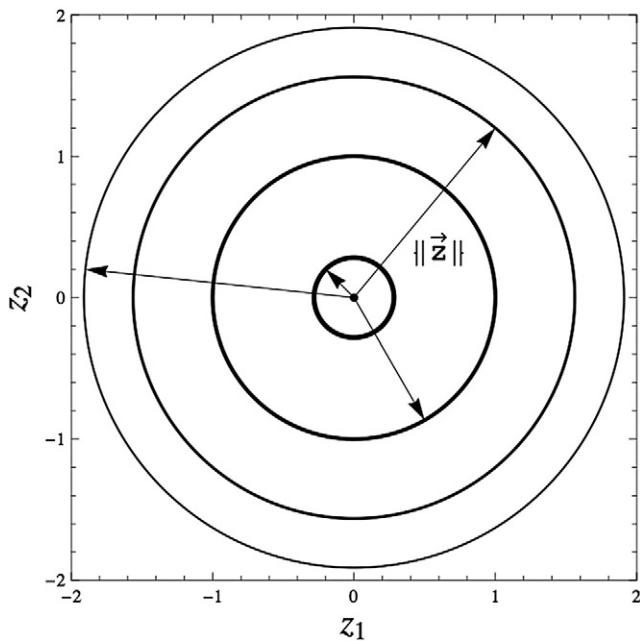


Fig. 1. Payoff isoclines in a two-dimensional artifact space (traits  $z_1$  and  $z_2$ ), where the payoff function  $u(\|\mathbf{z}\|)$  is spherically symmetric. All points on a solid circle result in similar payoff and are at the same distance  $\|\mathbf{z}\|$  from the optimum located at the origin (0,0) of the trait space. An arrow above a symbol denotes a vector.

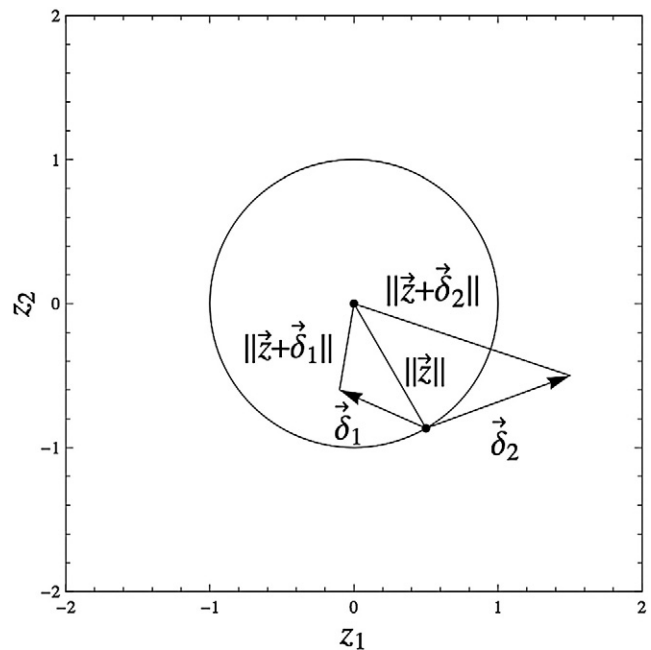


Fig. 2. Innovation vector in a two-dimensional artifact space. Innovation vector  $\delta_1$  results in a trait that is at closer distance  $\|\mathbf{z} + \delta_1\|$  to the optimum than the current trait  $\mathbf{z}$ , while innovation vector  $\delta_2$  results in a distance  $\|\mathbf{z} + \delta_2\|$  to the optimum that is further away from the optimum than the current trait  $\mathbf{z}$ .

By trying out  $\mathbf{z}'$  (e.g. constructing a handaxe and using it to scrape skins, or modifying a spear and throwing it), the individual can estimate the payoff  $u(\|\mathbf{z}'\|)$  of the new trait. In a second stage, the individual then chooses to accept or reject the new trait according to a payoff comparison. We assume that the acceptance probability of  $\mathbf{z}'$  is given by the standard discrete logit choice rule, as used in psychology and the social sciences (Luce, 1959; McFadden, 1974; Fudenberg & Levine, 1998; Camerer & Ho, 1999). The probability that an individual with current trait  $\mathbf{z}$  (distance  $\|\mathbf{z}\|$  from the optimum) accepts trait  $\mathbf{z}'$  is then given by

$$A(\mathbf{z}', \mathbf{z}) = \frac{e^{\lambda u(\|\mathbf{z}'\|)}}{e^{\lambda u(\|\mathbf{z}'\|)} + e^{\lambda u(\|\mathbf{z}\|)}} \quad (1)$$

where  $\lambda$  describes the sensitivity to payoff. When  $\lambda \rightarrow 0$ , choice is random, while when  $\lambda \rightarrow \infty$  an individual chooses with probability one the trait resulting in the highest payoff (Fig. 3).

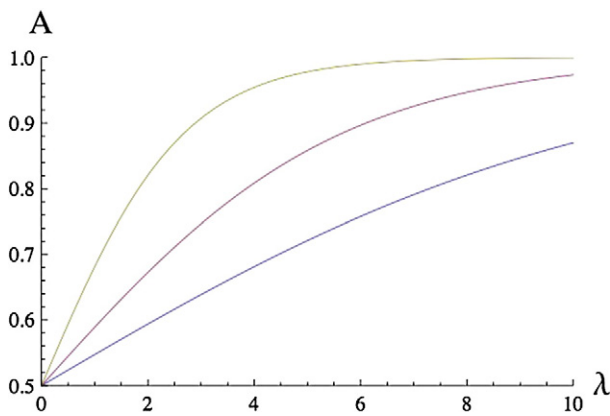
### 2.2.2. Fictitious learning (FL)

Here, an individual with trait  $\mathbf{z}$  can estimate without any trial the payoff of all traits located at distance equal or less than  $k$  around the current trait  $\mathbf{z}$  (Fig. 4). Hence, in order to associate a payoff to each trait, the individual makes fictitious trials for all trait values in the disk of radius  $k$  for a two-dimensional artifact ( $n=2$ ), in the sphere of radius  $k$  for a three-dimensional artifact ( $n=3$ ), and more generally in the so-called hyper-sphere of radius  $k$  for a higher dimensional artifact. The volumes of these spheres define imagination sets, which are the sets of traits an individual can psychologically conceive and estimate the payoff of.

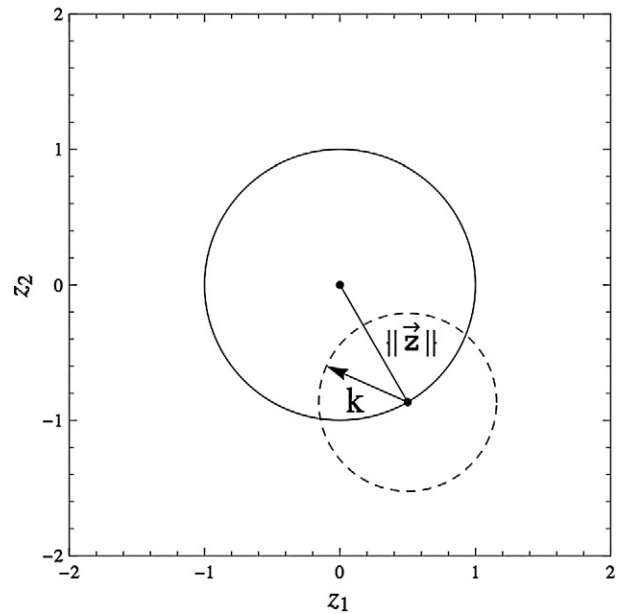
For fictitious learning, the trial of a new trait  $\mathbf{z}'$  is given by the continuous logit choice function. This means that the probability of trying trait  $\mathbf{z}'$  is proportional to  $e^{\gamma u(\|\mathbf{z}'\|)}$ , where  $\gamma$  is a parameter tuning the accuracy by which an individual is able to conceive the various alternative artifacts and associate a payoff to them. This gives the trial density function of going from trait  $\mathbf{z}$  to  $\mathbf{z}'$  as

$$T(\mathbf{z}'|\mathbf{z}) = \frac{e^{\gamma u(\|\mathbf{z}'\|)}}{\int_B e^{\gamma u(\|\mathbf{v}\|)} d\mathbf{v}_1 d\mathbf{v}_2 \dots d\mathbf{v}_n} \quad (2)$$

if  $\|\mathbf{z}' - \mathbf{z}\| \leq k$ , zero otherwise. Here,  $B = \{\mathbf{v} \in \mathbb{R}^n : \|\mathbf{v} - \mathbf{z}\| \leq k\}$  is the region of integration (the imagination set), which is an  $n$ -ball of radius  $k$  centered at  $\mathbf{z}$  (see Fig. 4). This formulation entails that the



**Fig. 3.** Acceptance probability  $A$  (Eq. 1) for a quadratic payoff function  $u(z) = -z^2$ , graphed as a function of  $\lambda$ , where component  $i$  of  $\mathbf{z}$  is varied from  $z_i$  to  $z_i = z_i + \delta$  to produce  $\mathbf{z}'$ . This gives the acceptance probability  $A = 1/(1 + \exp[-\lambda\delta(2z_i - \delta)])$  of the innovation, which is graphed for  $z_i = 1$  and  $\delta = 0.1$  (lower curve);  $z_i = 1$  and  $\delta = 0.2$  (middle curve); and  $z_i = 2$  and  $\delta = 0.2$  (upper curve). The innovation thus results in a larger payoff than the current artifact. When  $\lambda = 0$ , choice is random and  $A = 0.5$ . The acceptance probability of the innovation then always increases ( $A > 0.5$ ) as  $\lambda$  increases, but the more so if the payoff differences between the two alternatives are larger ( $\delta$  increases).



**Fig. 4.** Imagination set with radius  $k$  of a fictitious player. Such an individual can attach a payoff to each trait inside the circle of radius  $k$  centered around the current trait  $\mathbf{z}$ .

imagination set is also the feasible set, but this could be the whole state space so that if  $\gamma$  is very large an individual may jump to the optimum in one trial. A fictitious learner thus tries out a new trait  $\mathbf{z}'$  sampled from the skewed trial distribution (Eq. 2), which relies on an estimation (by mental simulations) of the payoffs associated to each artifact in its imagination set. By trying out  $\mathbf{z}'$  (e.g., constructing the handaxe and scraping skins), the individual gains information about the true payoff  $u(\|\mathbf{z}'\|)$  of the new trait, and accepts or rejects it according to the acceptance probability (Eq. 1), like a trial-and-error learner.

Note that the parameter  $\lambda$  (sensitivity to payoff) used in the acceptance probability (Eq. 1) is likely to depend on different underlying processes than the parameter  $\gamma$ , which describes the accuracy by which different artifacts are conceived mentally. For instance,  $\lambda$  may depend on how well outcomes are perceived and remembered, but  $\gamma$  may also depend on how well an individual understands the causal structure of the payoff landscape, or on how well intractable computations are replaced by simple heuristics. Further,  $\lambda$  may itself take different values under the different learning rules, and this variation itself may depend on the trait under focus.

Finally, we emphasize that a trial-and-error learner with insight and a fictitious learner both have larger cognitive abilities than a trial-and-error learner without insight, which chooses actions at random. In effect, trial-and-error learning with insight involves a (partial) conceptualization of the optimal artifact so that the learner can sense (to some extent at least) the direction of change towards the optimum. This requires certain cognitive abilities. Further, a fictitious learner is potentially able to conceive all artifacts he/she is able to construct, which requires very high computational abilities. As such, the cognitive abilities of a trial-and-error learner with insight and a fictitious learner are probably closer to each other than to a trial-and-error learner without insight, but we keep the distinction between the two former learning rules because they involve different underlying learning mechanisms.

### 3. Learning dynamics

We envision the innovation process (trial and acceptance) as occurring during an individual's lifespan. Hence, an individual constantly tries out a new trait. For instance, a hunter may invent

one day an alternative spear by sampling in the trial distribution,  $T(\mathbf{z}'|\mathbf{z})$ . The individual then uses the spear for a while to compare outcomes and adopts the variant perceived as more efficient according to the acceptance probability,  $A(\mathbf{z}',\mathbf{z})$ . After some time period, the hunter may innovate a new spear, which may occur from a chance event (i.e., trial-and-error learning without insight) or on purpose, in which simulations of payoff may be used to modify the spear (i.e., fictitious learning). By the repeated occurrence of trial and acceptance (or rejection), the dynamics of the  $n$ -dimensional artifact evolve in the state space and may or may not converge to the optimum.

In order to determine under what conditions the direction of the learning dynamics points towards the optimum, we use in the rest of this paper the quadratic payoff function  $u(\|\mathbf{z}\|) = -\|\mathbf{z}\|^2$ , which describes partial complementarity between traits. Due to the complexity of the process, in particular the interaction between the multidimensional trial distribution and the acceptance function, an exact analysis of the process is very difficult. Nevertheless, we can show that an individual learner can be entirely characterized by the probability density function  $p(z,t)$  of being at distance  $z = \|\mathbf{z}\|$  from the optimum at time  $t$ , which depends on a transition function whose state space is the distance to the optimum [Supplementary online material, appendix A (abbreviated SO Appendix A), Eqs. A-1–A-8 for trial-and-error learning, and SO Appendix B, Eqs. B-1–B-22 for fictitious learning].

We can thus summarize the learning dynamics as a one-dimensional stochastic process regardless of the dimensionality  $n$  of the artifact space, which can itself be thought of as a measure of artifact complexity. But the exact transition functions of the learning dynamics remain complicated (e.g., Eq. A-4 for trial-and-error learning and Eq. B-21 for fictitious learning). In order to understand the conditions under which artifact evolution will approach the optimum, we thus approximate the stochastic learning dynamics by a diffusion process, which entails assuming small parameter values (e.g., (Grimmett & Stirzaker, 2001; Karlin & Taylor, 1981)). This leads to the results described in the following sections.

## 4. Results for trial-and-error learning

### 4.1. Mean change to the optimum

Assuming small parameter values ( $\alpha$ ,  $\sigma^2$ , and  $\lambda$ ), the stochastic dynamics of the distance to the optimum for a trial-and-error learner are given by

$$a = \frac{\sigma^2}{4z} \left[ (n-1) - 2(\lambda + \alpha/\sigma^2)z^2 \right]$$

$$b = \frac{\sigma^2}{2}, \quad (3)$$

which are, respectively, the mean and variance of the change of the distance to the optimum given current distance  $z$  ( $a \equiv E[\Delta z|z]$ , and  $b \equiv E[(\Delta z)^2|z]$ , SO Appendix A, Eqs. A-9–A-19).

When  $a < 0$ , the mean change is negative, which means that the change occurs in the direction of the optimum and the individual improves its artifact, on average. When  $a > 0$ , the artifact deteriorates on average. A steady-state obtains if  $a = 0$ . Eq. (3) shows that for a one-dimensional artifact ( $n = 1$ ), there is always improvement since the mean change  $a$  is always negative. Hence the learning process converges, on average, to the optimum. This accords well with intuition.

Eq. (3) reveals that as the complexity of the trait increases ( $n$  increases), it becomes less likely that the change in trait points toward the optimum. In this case, the expected direction of change depends on a balance between the artifact complexity value  $n - 1$ , and the

learning rate  $2(\lambda + \alpha/\sigma^2)$  weighted by the current position squared,  $z^2$ . It is now possible that the trait deteriorates on average as a result of experimenting with new traits.

For a trial-and-error learner without “insight” ( $\alpha = 0$ ), which amounts to standard reinforcement learning, the trait complexity cannot exceed the threshold value  $1 + 2\lambda z^2$ , otherwise the learning dynamics move away from the optimum, on average. This shows that the closer one is to the optimum, the more difficult it becomes to further approach it. This threshold value can be thought of as a barrier to improvement of the artifact. Close to the optimum, a sample of the trial distribution can easily overshoot it, and there are many more ways to try (at random) a trait that leads to a decrease in payoff rather than to an increase. Hence unless sensitivity to payoff,  $\lambda$ , is infinitely large, such sub-optimal traits can sometimes be accepted, and because many of them are tried, the dynamics can move away from the optimum on average.

For a trial-and-error learner with “insight” ( $\alpha > 0$ ), small  $\sigma$  and large  $\alpha$  improve the performance because “insight” provides a reliable step toward the optimum, while the variance  $\sigma^2$  in trials tends to drive trait  $z$  further away from the optimum whenever  $n > 1$ . Note that even if the guess of the location of the optimum is perfect ( $\alpha = 1$ ), but there are errors in trials ( $\sigma^2 > 0$ ), the artifact dynamics may be driven away from the optimum. This illustrates the very important role played by errors in the learning dynamics of complex traits.

### 4.2. Stationary distribution

The mean and the variance (Eq. 3) fully characterize the stochastic learning dynamic of the distance to the optimum,  $p(z,t)$ , which can be evaluated analytically (Eq. A-21). This eventually settles into the stationary probability density function  $p(z)$  of being at distance  $z$  from the optimum [ $p(z) = \lim_{t \rightarrow \infty} p(z,t)$ ]. The stationary distribution is

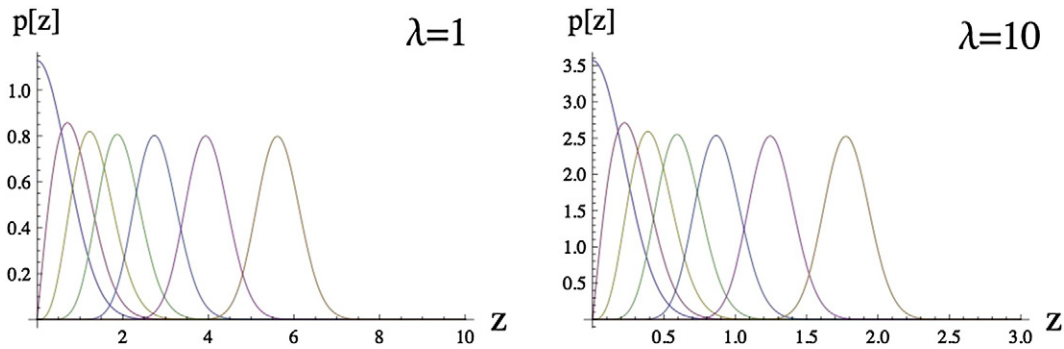
$$p(z) = \frac{2 \exp(-\lambda_e z^2) z^{n-1} \lambda_e^{n/2}}{\Gamma(n/2)}, \quad (4)$$

which is a generalized gamma distribution with scale parameter  $n$  and shape parameter  $\lambda_e = \lambda + \alpha/\sigma^2$  (SO, Eqs. A-20–A-22). The parameter  $\lambda_e$  can be thought of as an effective learning rate, and this notation is convenient to make comparisons between models. The resulting mean and variance of the distance to the optimum are, respectively,  $\sqrt{n/(2\lambda_e)}$  and  $1/(4\lambda_e)$ , when  $n$  is large, and thus decrease as  $\lambda_e$  increases. Hence, at steady-state the mean distance to the optimum is likely to be non-zero ( $\sqrt{n/(2\lambda_e)} > 0$ ); this mean can also be obtained by setting  $a = 0$  in Eq. (3) and solving for  $z$ .

For a trial-and-error learner without “insight” ( $\alpha = 0$ ), the stationary distribution can stabilize far away from the optimum as trait complexity increases (Fig. 5). Extensive numerical simulations show that the analytic expressions for the stationary distribution given by Eq. (4) fit very well with exact results obtained from stochastic simulations even when  $\lambda$  becomes very large (Fig. 6), in which case deleterious trials ( $z' - z > 0$ ) are almost always rejected. This suggests that the barrier to improvement encapsulated in the mean change described by Eq. (3) is a robust phenomenon, not sensitive to the assumption of small parameter values used in its derivation.

For a trial-and-error learner with “insight” ( $\alpha > 0$ ), the stationary distribution  $p(z)$  is graphed in Fig. 7, which shows that the equilibrium trait values stabilize closer to the optimum as  $\alpha$  increases. Numerical simulations again show good agreements even for large  $\lambda$  value provided  $\alpha$  is not too large (Fig. 7). Since Eq. (4) tends to





**Fig. 5.** Stationary phenotypic distribution  $p(z)$  given by Eq. (4) for a trial-and-error learner without insight ( $\lambda_e = \lambda$ ). From left to right in each panel we have  $n = 1, 2, 4, 16, 32, 64,$  and  $128,$  while  $\lambda = 1$  in the first panel and  $\lambda = 10$  in the second panel.

underestimate the barrier to learning (Fig. 7), our analytical results are conservative.

It may be felt that the barrier to improvement we identified is the result of an individual modifying several traits at the same time. If an individual can modify only a single trait at a time, isn't convergence to the optimum more likely? Our diffusion approximations to the learning process (Eq. (3) and Eqs. (5 and 6)) were actually constructed precisely along these lines (see SO Appendix A). Here, each trait can be thought of as evolving independently of each other and convergence to the optimum, on average, for each trait is indeed guaranteed (Eq. A-14).

But importantly, errors are still possible in one dimension so there is always a variance around the optimum. It is the superposition of this variance per trait over the  $n$  dimensions of the artifact, which creates departure from the distance to the optimum of a multi-dimensional trait. Thus, only sub-optimal traits are likely to be expressed at steady-state and this is delineated by the stationary distribution (Eq. 4). The mean  $\sqrt{n/(2\lambda_e)}$  (for large  $n$ ) of this distribution summarizes this out-of-optimality result, and is graphed as a function of the model's parameters in Fig. 8.

### 5. Results for fictitious learning

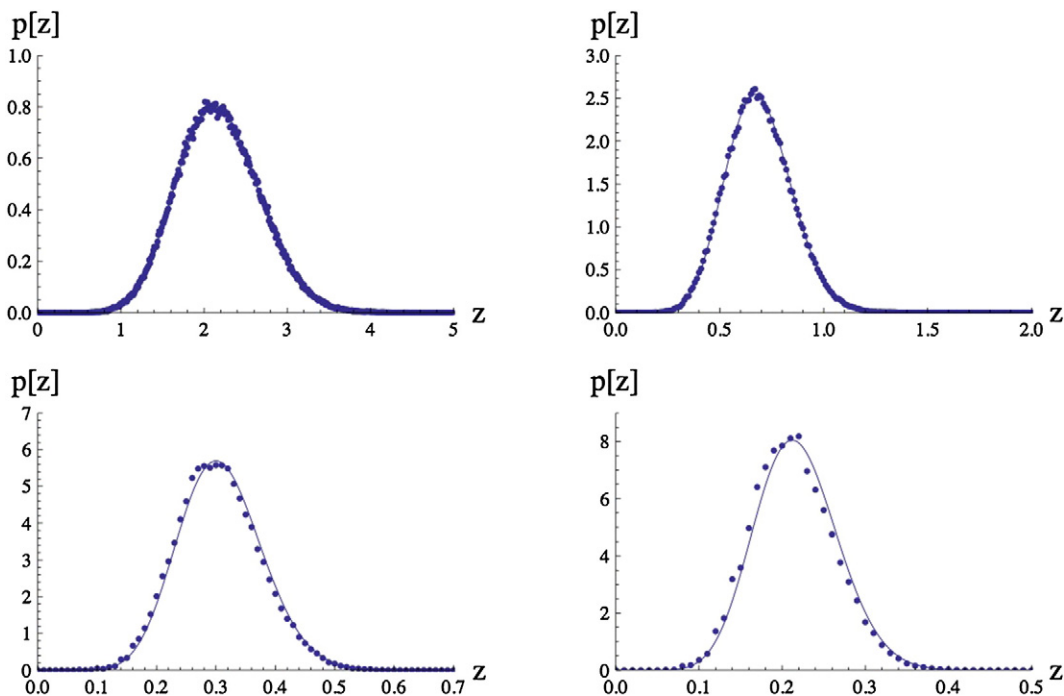
Assuming small parameter values ( $k, \gamma,$  and  $\lambda$ ), the stochastic dynamics of the distance to the optimum for a fictitious learner satisfy a diffusion process with mean and variance given by

$$a = \frac{k^2}{12z} [(n-1) - 2(\lambda + 2\gamma)z^2]$$

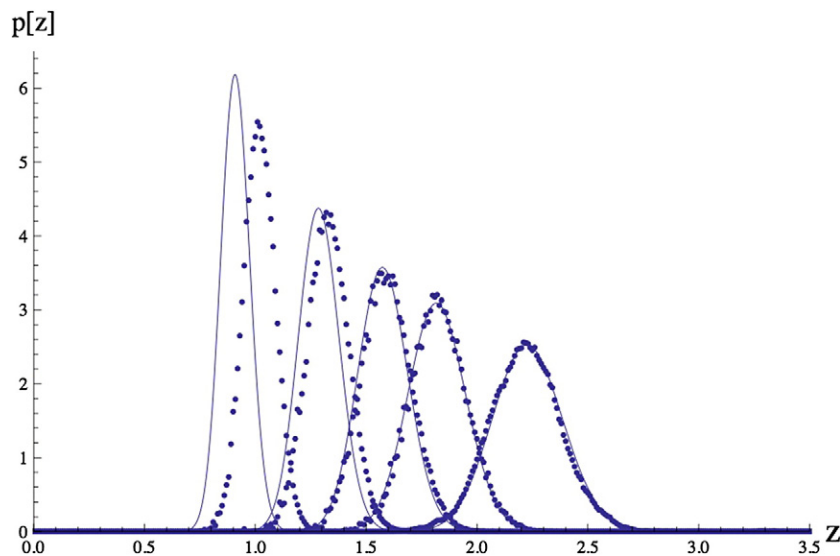
$$b = \frac{k^2}{6} \tag{5}$$

(SO, Eqs. B-23–B-27). This is qualitatively of the same form as Eq. (3). The change in the direction to the optimum occurs where the artifact complexity value  $n-1$  is balanced by the learning efficiency  $2(\lambda + 2\gamma)z^2$ . We can thus think in terms of an effective learning performance parameter  $\lambda_e$  affecting the mean change towards the optimum, which is given here by  $\lambda_e = \lambda + 2\gamma$ .

Eq. (5) shows that when the dimensionality of the trait exceeds a threshold value  $1 + 2(\lambda + 2\gamma)z^2$ , a fictitious learner cannot improve



**Fig. 6.** Comparison of the stationary phenotypic distribution  $p(z)$  for a trial-and-error learner without insight ( $\lambda_e = \lambda$ ) with that obtained from stochastic simulations using the exact transition density of the process (SO, Eq. A-8). The solid lines show the analytical solutions, while the circles denote the simulation results. In all panels  $n = 10$ . First row of panels, from left to right  $\lambda = 1$  and  $\lambda = 10$ . Second row of panels, from left to right  $\lambda = 50$  and  $\lambda = 100$ .



**Fig. 7.** Comparison of the stationary phenotypic distribution  $p(z)$  for a trial-and-error learner with insight ( $\lambda_e = \lambda + \alpha/\sigma^2$ ) with that obtained from stochastic simulations of the exact transition density of the process (SO, Eq. A-8). The solid lines show the analytical solutions, while the circles denote the simulation results. In all curves,  $n = 100$ ,  $\lambda = 10$ , and  $\sigma = 0.1$ , whilst  $\alpha$  is varied. From right to left, the curves are for  $\alpha = 0, 0.05, 0.1, 0.2$ , and  $0.5$ . As  $\alpha$  becomes large ( $\sim 0.5$ ), the diffusion approximation (Eq. 4) assuming small  $\alpha$  tends to underestimate the stationary distance to the optimum. Hence, it is a conservative approximation.

its artifact on average and this does not depend on the size of the imagination set  $k$ . This is a counterintuitive result because a fictitious learner can ascertain the payoff of each artifact in its imagination set, which could actually be the whole state space. Eq. (5) also implies that at steady-state the distance to the optimum is distributed according to Eq. (4) with  $\lambda_e$  given by  $\lambda + 2\gamma$ , and that the mean distance to the optimum is approximatively given by  $\sqrt{n/[2(\lambda + 2\gamma)]}$ . Hence, as was the case for a trial-and-error learner, a fictitious learner may express sub-optimal artifacts at steady-state.

Note that  $(1/3)k^2$  is equivalent to the variance of a uniform distribution on  $[-k, k]$ , so that the variance in change in the trait ( $b$  in Eq. (5)) is half that of a uniform distribution. Hence, when the accuracy of mental simulation is zero ( $\gamma = 0$ ), the dynamics of a fictitious learner are exactly the same as those of a trial-and-error learner without insight (Eq. (3) with  $\alpha = 0$ ), but where the trait distribution follows a uniform instead of a normal distribution.

In order to check the accuracy of our approximation, we compare in Fig. 9 the stationary distribution  $p(z)$  for a fictitious learner (Eq. (4) with  $\lambda_e = \lambda + 2\gamma$ ) to that obtained from the exact model (SO, Eq. B-20). As can be seen from Fig. 9, the agreement between the two models is good even for large  $\gamma$  values. For large  $k$  values, Eq. (4) tends to underestimate the barrier to learning (Fig. 9), which was observed under extensive numerical analysis. Our analytical results are thus again conservative.

## 6. Population process with individual and social learning

### 6.1. Assumptions

So far, we have considered that the distance to the optimum characterizes the state of a single individual. Alternatively, one can look at the process at the population level and define  $p(z, t)$  as being the probability density function that a population with  $N$  learners is at distance  $z$  from the optimum at time  $t$ . This allows us to introduce social learning into the model (Cavalli-Sforza & Feldman, 1981; Lumsden & Wilson, 1981; Boyd & Richerson, 1985), and we apply this to a population of trial-and-error learners using a Moran cultural updating process (Lumsden & Wilson, 1981; Lehmann, Aoki, & Feldman, 2011; Aoki, Lehmann, & Feldman, 2011).

In a population with  $N$  individuals, each of them may express a different artifact. Tracking the joint dynamics of artifact change in

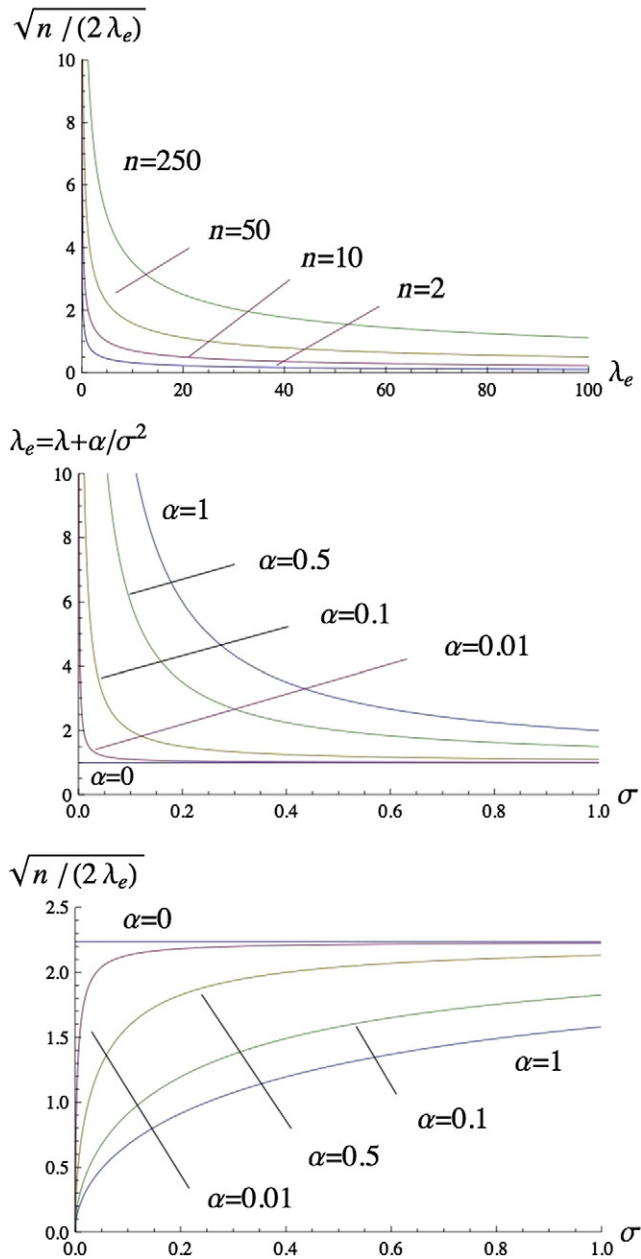
each individual is very involved. So in order to circumvent this complication we assume a separation of time scales between innovation of new artifacts (slow process) and social learning (fast process). This allows us to postulate that only two variants can simultaneously coexist in the population, and a novel variant appears through innovation only once the existing polymorphism has been lost through social learning (SO Appendix C, Eq. C-1).

The separation of time scales between individual and social learning works as follows. Suppose that every member of the population currently expresses a trait at distance  $z$  from the optimum. With a very small probability  $\mu$ , one randomly sampled individual applies individual learning and tries out a new artifact (at new distance  $z'$  from the optimum) according to the trial distribution and accepts or rejects the innovation. With complementary probability  $1 - \mu$ , the individual applies social learning. In this case, the learner chooses an exemplar individual at random and adopts the exemplar's cultural trait according to the logit choice rule, involving comparing the payoff of its own trait with that of the exemplar (SO Appendix C, Eqs. C-2–C-3).

When an innovation occurs and the innovator accepts the new trait, there is polymorphism in the population. Because  $\mu$  is very small, change in variant frequency in the population is subsequently likely to occur only through social learning until the new artifact is either adopted by all individuals or eliminated. Eventually, after polymorphism is lost, a new innovation will occur. This process results in a trait substitution sequence in state space as was the case for a single learner. This separation of time scales assumption, which is in standard use in evolutionary biology (e.g., Kimura, 1971; Gillespie, 1991; Orr, 1998; Poon & Otto, 2000; Metz, Geritz, Mesz ena, & van Heerwaarden, 1996), allows us to approximate the dynamics of artifact evolution under a population process when individuals spend markedly more time on social than individual learning. By contrast, when the rate of individual learning is much larger than social learning, we can approximate the learning dynamics with the model described above, which focuses on a single individual (Eq. 3). Hence, these two cases (individual and population process) can be thought of as the two extremes of a continuum of situations involving various rates of individual and social learning.

### 6.2. Results

With the above assumptions and considering small parameter values ( $\alpha$ ,  $\sigma^2$ , and  $\lambda$ ), the stochastic dynamics of the distance to the



**Fig. 8.** Top panel: average distance  $\sqrt{n/(2\lambda_e)}$  to the optimum at steady-state graphed as a function of trait complexity  $n$  and effective learning parameter  $\lambda_e$ . Middle panel: effective learning parameter  $\lambda_e = \lambda + \alpha/\sigma^2$  for a trial-and-error learner with insight graphed for  $\lambda = 1$  as a function of the standard deviation of the trial distribution  $\sigma$  and level of insight  $\alpha$ . Lower panel: average distance  $\sqrt{n/(2\lambda_e)}$  to the optimum for a trial-and-error learner with insight ( $\lambda_e = \lambda + \alpha/\sigma^2$ ) graphed for the same parameters as in the previous panel for  $n = 10$ . The straight lines in the middle and lower panels are for an experienced-based learner without insight. Hence, the level of insight,  $\alpha$ , provides a reliable guide to the optimum only if the variance in the trial distribution is not too large.

optimum for a population of trial-and-error learners satisfy a diffusion process with mean and variance given by

$$a = \frac{\mu\sigma^2}{4z} \left[ (n-1) - 2(N\lambda + \alpha/\sigma^2)z^2 \right]$$

$$b = \frac{\mu\sigma^2}{2} \quad (6)$$

(SO Appendix C, Eqs. C-4–C-8). Comparing this to Eq. (3), shows that the effective learning performance parameter can now be written

$\lambda_e = N\lambda + \alpha/\sigma^2$ . Holding everything else constant, large  $N$  values increase learning performance because it reduces the probability that a trait with lower performance fixes at the level of the whole population. Some individuals may accept a trait that decreases payoff, but it is less likely that a whole population accepts this trait. The population thus acts as a corrector mechanism, which tends to eliminate innovations with lower payoff than the current trait.

For the population process, the stationary distribution  $p(z)$  of the distance to the optimum is still given by Eq. (4), but with  $\lambda_e = N\lambda + \alpha/\sigma^2$ . Hence, the mean distance to the optimum at steady-state is given for large  $n$  by  $\sqrt{n/(2\lambda_e)}$  and for a population of trial-and-error learners without “insight” one has  $\sqrt{n/(2\lambda N)}$ . This goes to zero as population size becomes very large. Hence, convergence to the optimum is very likely to occur in large populations.

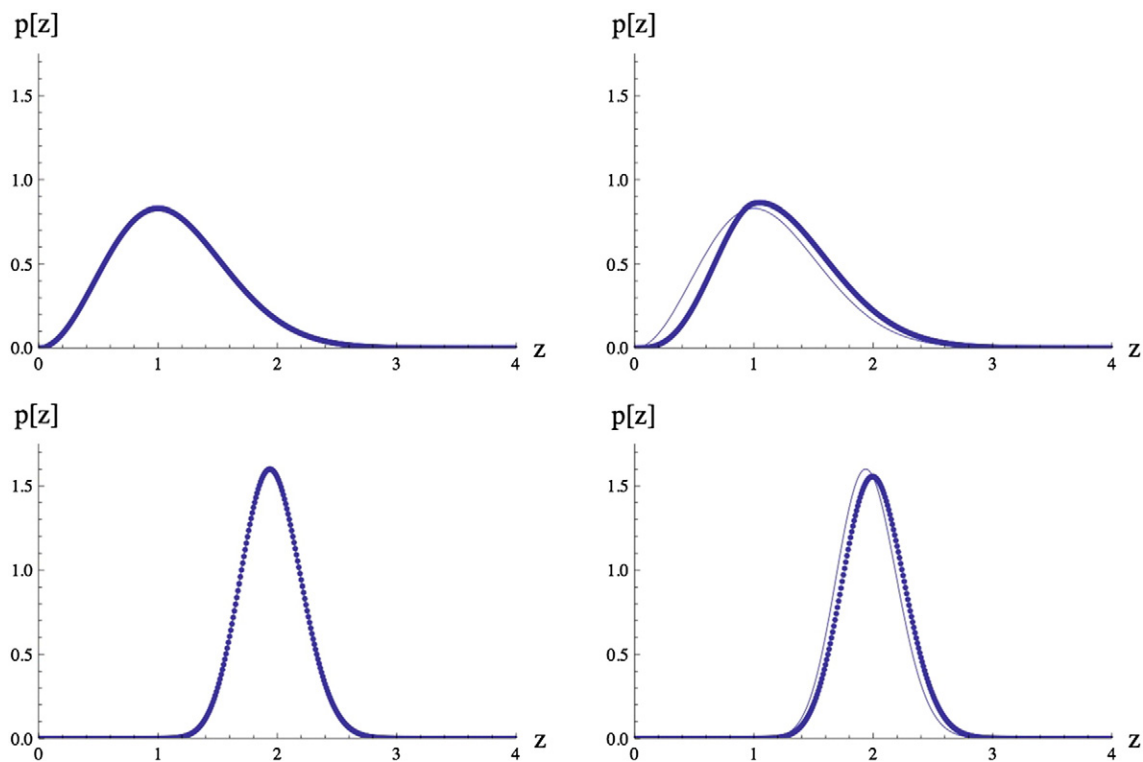
Importantly, the population correction mechanism occurs because all individuals were assumed to update their trait according to the payoff comparison rule encapsulated in the acceptance function (SO Appendix C, Eqs. C-2–C-3). Alternatively, one may postulate that social learning occurs by random copying of an individual in the population. In this case, estimates are not used during social learning, so learning is cognitively less demanding. The dynamics of the distance to the optimum are then given by Eq. (6), setting  $N = 1$  regardless of the true population size (SO Appendix C, Eq. C-9). The approach to the steady-state is then qualitatively similar to that of a single isolated learner (Eq. 3).

For the population to act as a corrector mechanism, several individuals must thus be sensitive to payoff. Further, the salient result given by Eq. (6) rests on the assumption that errors in estimating alternative variants are uncorrelated between individuals. This is likely to be the case when the sensitivity to payoff,  $\lambda$ , depends on observational errors. However, certain errors may be correlated, since individuals are likely to express systematic cognitive biases (Tversky & Kahneman, 1974). In this case, the error of accepting an artifact with lower payoff by one individual is likely to be repeated by another. Hence, Eq. (6) is better thought of as representing one extreme of a continuum of situations where a population of social learners may or may not (as exemplified by random copying) result in a better approach to the optimum than an individual taken in isolation.

## 7. Discussion

We have shown that for different learning rules, the stochastic learning dynamics of an  $n$ -dimensional artifact stabilize at approximate distance  $\sqrt{n/(2\lambda_e)}$  away from the optimum (Eq. (4), Fig. 8), where  $n$  is artifact complexity and  $\lambda_e$  an effective learning performance parameter, which depends on the learning rule under focus. Because the learning dynamic does not converge to the optimum, our model identifies a complexity barrier to artifact improvement. This results in individuals expressing on average sub-optimal artifacts, even if the payoff landscape is single peaked. This barrier to learning is non-existent for one-dimensional artifacts (Eqs. 3–5). For a multidimensional artifact ( $n > 1$ ), the barrier appears as a consequence of errors in the learning dynamics, and occurs for related but slightly different reasons under trial-and-error and fictitious learning.

A trial-and-error learner without “insight” ( $\alpha = 0$ ) is a standard reinforcement learner, as it tries out at random a new artifact by sampling in the trial distribution. For a multidimensional trait, such a learner is very likely to fall outside the region of improvement of the artifact. This stems from the fact that when conformity among factors is required to produce high payoff (Fig. 1), there are many more ways to decrease payoff than there are to increase it, especially when the current trait is close to the optimum. This is qualitatively similar to Fisher’s original point that mutations occurring at random in an organism are likely to be deleterious (e.g., (Fisher, 1930; Hartl & Taubes, 1996; Poon & Otto, 2000; Waxman, 2006; Martin & Lenormand, 2006)). If an individual then does not have a perfect



**Fig. 9.** Comparison of the stationary phenotypic distribution  $p(z)$  for a fictitious learner ( $\lambda_e = \lambda + 2\gamma$  in Eq. 4) with that obtained from the exact probability transition function (SO, Eq. B-20). The solid lines show the analytical solutions, while the circles denote the simulation results. In all graphs  $\lambda = 0$ . In the first row of panels  $n = 3$  and  $\gamma = 0.5$ , and from left to right  $k = 0.1$  and  $k = 1$ . In the second row of panels  $n = 31$  and  $\gamma = 2$ , and from left to right  $k = 0.1$  and  $k = 1$ .

acceptance criteria, such that errors occur in choice ( $\lambda < \infty$ ), it may adopt a trait with lower payoff and on average may not converge to the optimum.

A trial-and-error learner with “insight” ( $\alpha > 0$ ) is not a simple trial-and-error learner, as it can carry out additional computation to estimate the direction of improvement of the artifact, which can be achieved by using a guessing algorithm. But even if such a learner can perfectly locate the optimum on average ( $\alpha = 1$ ), the dynamics may stabilize away from the optimum if errors are made in choice ( $\lambda < \infty$ ). This illustrates the fundamental role played by errors in the learning dynamics of complex traits (those with large  $n$ ) even if innovations do not occur at random and the payoff landscape is single peaked.

A fictitious learner is able to mentally evaluate the payoff of each trait in a set of radius  $k$  around its current trait (the imagination set). This set could be made as large as the whole state space so that a fictitious learner may potentially “know” the payoff of each conceivable artifact, which, in practice, requires unreasonable powers of computation. But as the trait complexity  $n$  increases, introducing more traits in the imagination set may worsen rather than improve the learning process since for very complex artifacts, the imagination set may then consist entirely of non-improving traits. If errors occur in trials ( $\gamma < \infty$ ), a fictitious learner may then adopt a trait with lower payoff than the current one because there are more ways to make the wrong choice than the correct one. This leads to the result that, holding everything else the same, a fictitious learner can perform only twice as good as a trial-and-error learner when sensitivity to realized payoff is zero, i.e.,  $\lambda = 0$  in Eq. (5). This is one of the counterintuitive results of our analysis and delineates a possible limit to learning even for individuals having unrealistically large powers of computation.

We have also shown that when individuals can socially learn a new trait in a payoff-biased manner with uncorrelated errors in estimation, a population of individuals can drastically increase the performance of learning because the population serves as a corrector mechanism to

errors in choice (Eq. 6). In this case, if one individual designs a bad artifact and uses it, this artifact design is unlikely to spread through the whole population. It follows that a population of payoff-biased learners is likely to converge to the optimum even for arbitrarily complex artifacts, if the population can be made arbitrarily large (Eq. 6).

That population size can affect the dynamics of cultural traits is not new (Cavalli-Sforza & Feldman, 1981; Henrich, 2004; Bentley, & Shennan, 2004; Lehmann & Feldman, 2009; Powell, Shennan, & Thomas, 2009). The smaller the population size, the more likely are stochastic factors to overwhelm the filtering of alternative heritable designs by natural selection (Crow & Kimura, 1970; Cavalli-Sforza & Feldman, 1981). A single isolated learner can then be seen as a population of effective size one, thereby reducing the correction mechanisms occurring when several individuals must accept a new artifact for it to fix in the population.

For tractability, we used a quadratic payoff function to derive our results, which is spherically symmetric ( $u(z) = -z^2$ , Fig. 1). To what extent will a change in this assumption change our qualitative result that the learning dynamics stabilize away from the optimum? The assumption of spherical symmetry is not crucial to the observation that, as the number of dimensions increases, the number of ways of making payoff-improving innovations to the number of ways of making payoff-decreasing innovations decreases markedly (Waxman & Welch, 2005; Waxman, 2006). This ratio of “correct” to “wrong” innovations is what drives the dynamics away from the optimum in the presence of error in choice in small populations. Hence, the relaxation of the specific assumptions behind our payoff function is unlikely to qualitatively affect our results (in SO Appendix D, Eq. D-1, we show that the results are unchanged for  $n=2$  under various standard non-spherically symmetric payoff functions).

Our results suggest that a population of simple interconnected reinforcement learners trying out actions at random can, under certain conditions, perform better than a population of isolated



learners that have high computational abilities. This should select for social learning but also for filtering mechanisms at the individual level that increase the sensitivity and/or estimation of artifact payoff ( $\lambda$ ). Many cultural traits themselves, from modes of social organization to accounting systems, may have allowed for a reduction in errors of acceptance of new traits, and therefore paved the way for the evolution of higher dimensional artifacts.

While our analysis delineates a possible stasis of artifact evolution when learning efficiency and/or population size is low, we have only considered the dynamics of a single artifact in isolation (such as a handaxe or a spear). In this context, our measure of “cultural complexity” was the number of dimensions of the trait under focus. Many human artifacts are indeed multidimensional traits whose payoff to their carriers depends on the adjustment and complementarity of several underlying quantitative factors. Nevertheless, this is but one way of considering cultural complexity. Another natural and relevant measure of cultural complexity is the number of different artifacts used by an individual or a group of individuals, i.e., the sum total of technology, such as the number of different types of weapons and cooking techniques.

The number of cultural traits in a population can accumulate at very different rates depending on the modes of individual and social learning used by individuals (Ghirlanda & Enquist, 2007; Strimling, Sjöstrand, Enquist, & Eriksson, 2009; Lehmann et al., 2011; Kobayashi & Aoki, 2012). Indeed, a stasis of cultural evolution can also occur at this level. Hence, in order to obtain a better understanding of the stasis of artifact evolution in archaic humans and its acceleration in modern ones, the increase in the dimension of artifacts and their number (or diversity) should probably be considered simultaneously. This opens the door to further explorations of individual and social learning dynamics.

## Supplementary Materials

Supplementary data to this article can be found online at <http://dx.doi.org/10.1016/j.evolhumbehav.2012.11.001>.

## Acknowledgments

We thank Kenichi Aoki, Yutaka Kobayashi, Nobuyuki Takahashi, Yoshihiro Nishiaki, and Charles Mullan for various discussions. We are grateful to Simon Powers for improving the English and to two reviewers for extremely useful comments on the model.

## References

- Aoki, K., Lehmann, L., & Feldman, M. W. (2011). Rates of cultural change and patterns of cultural accumulation in stochastic models of social transmission. *Theoretical Population Biology*, 79, 192–202.
- Bentley, R. A., H.M.W., & Shennan, S. J. (2004). Random drift and culture change. *Proceedings of the Royal Society B-Biological Sciences*, 271, 1443–1450.
- Boyd, R., & Richerson, P. J. (1985). *Culture and the evolutionary process*. Chicago: University of Chicago Press.

- Bush, R., & Mosteller, F. (1951). A mathematical model for simple learning. *Psychological Review*, 58, 313–323.
- Camerer, C., & Ho, T. (1999). Experience-weighted attraction learning in normal form games. *Econometrica*, 67, 827–874.
- Cavalli-Sforza, L., & Feldman, M. W. (1981). *Cultural transmission and evolution*. NJ: Princeton University Press.
- Crow, J. F., & Kimura, M. (1970). *An introduction to population genetics theory*. New York: Harper and Row.
- Dugatkin, L. A. (2004). *Principles of animal behavior*. London: W. W. Norton and Company.
- Fisher, R. A. (1930). *The genetical theory of natural selection*. Oxford: Clarendon Press.
- Fudenberg, D., & Levine, D. K. (1998). *Theory of learning in games*. Cambridge, MA: MIT Press.
- Ghirlanda, S., & Enquist, M. (2007). Cumulative culture and explosive demographic transitions. *Quality and Quantity*, 41, 581–600.
- Gillespie, J. H. (1991). *The causes of molecular evolution*. Oxford: Oxford University Press.
- Grimmett, G., & Stirzaker, D. (2001). *Probability and random processes*. Oxford: Oxford University Press.
- Hartl, D., & Taubes, H. C. (1996). Compensatory nearly neutral mutations: Selection without adaptation. *Journal of Theoretical Biology*, 182, 303–309.
- Henrich, J. (2004). Demography and cultural evolution: How adaptive cultural processes can produce maladaptive losses: The Tasmanian case. *American Antiquity*, 69, 197–214.
- Hopkins, E. (2002). Two competing models of how people learn in games. *Econometrica*, 70, 2141–2166.
- Karlin, S., & Taylor, H. M. (1981). *A second course in stochastic processes*. San Diego: Academic Press.
- Kimura, M. (1971). Theoretical foundation of population genetics at the molecular level. *Theoretical Population Biology*, 2, 174–208.
- Klein, R. G. (2009). *The human career* (3th edn). Chicago: The University of Chicago Press.
- Kobayashi, Y., & Aoki, K. (2012). Innovativeness, population size and cumulative cultural evolution. *Theoretical Population Biology*, 82, 38–47.
- Kremer, M. (1993). Population growth and technological change: One million BC to 1990. *The Quarterly Review of Economics*, 108, 681–716.
- Lehmann, L., Aoki, K., & Feldman, M. W. (2011). On the number of independent cultural traits carried by individuals and populations. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 366, 424–435.
- Lehmann, L., & Feldman, M. W. (2009). Coevolution of adaptive technology, maladaptive culture, and population size in a producer–scrounger game. *Proceedings of the Royal Society of London, Series B: Biological Sciences*, 276, 3853–3862.
- Luce, R. D. (1959). *Individual choice behavior*. New York: Wiley.
- Lumsden, C. J., & Wilson, E. O. (1981). *Genes, mind and culture*. MA: Harvard University Press.
- Martin, G., & Lenormand, T. (2006). A general multivariate extension of Fisher's geometrical model and the distribution of mutation fitness effects across species. *Evolution*, 60, 893–907.
- McFadden, D. (1974). Conditional logit analysis of qualitative choice behavior. In P. Zarembka (Ed.), *Frontiers in econometrics* (pp. 105–142). New York: Academic Press.
- Metz, J. J., Geritz, S. A. H., Meszéna, G., Jacobs, F. J. A., & van Heerwaarden, J. (1996). Adaptive dynamics: A geometrical study of the consequences of nearly faithful reproduction. In S. J. van Strien, & S. M. Verduyn Lunel (Eds.), *Stochastic and spatial structures of dynamical systems* (pp. 183–231). Amsterdam: North-Holland.
- Orr, H. A. (1998). The population genetics of adaptation: The distribution of factors fixed during adaptive evolution. *Evolution*, 52, 935–949.
- Poon, A., & Otto, S. P. (2000). Compensating for our load of mutations: Freezing the meltdown of small populations. *Evolution*, 54, 1467–1479.
- Powell, A., Shennan, S., & Thomas, M. G. (2009). Late Pleistocene demography and the appearance of modern human behavior. *Science*, 324, 1298–1301.
- Strimling, P., Sjöstrand, J., Enquist, M., & Eriksson, K. (2009). Accumulation of independent cultural traits. *Theoretical Population Biology*, 76, 77–83.
- Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, 185, 1124–1131.
- Waxman, D. (2006). Fisher's geometrical model of evolutionary adaptation—beyond spherical geometry. *Journal of Theoretical Biology*, 241, 887–895.
- Waxman, D., & Welch, J. J. (2005). Fisher's microscope and Haldane's ellipse. *The American Naturalist*, 166, 447–457.

## Online Supplementary material for “The handaxe and the microscope—Individual and social learning in a multidimensional model of artifacts evolution”

### Appendix A: trial-and-error learning

#### Markov process for the distance to the origin

Here, we derive a one dimensional transition probability density  $p(z' | z)$  that a trial-and-error learner goes from state  $z$  to  $z'$  from the knowledge of the trial and acceptance distribution.

#### Trial distribution

In order to obtain  $p(z' | z)$ , we need to evaluate the probability density function  $T(z' | z)$  that a trait at distance  $z' = \sqrt{\sum_{i=1}^n (z'_i)^2}$  from the origin is tried out when the current distance is  $z = \sqrt{\sum_{i=1}^n z_i^2}$ . Here,  $z'_i$  follows a Normal distribution with mean  $(1 - \alpha)z_i$  and variance  $\sigma^2$ .

We then write  $z' = \sigma x$ , where  $x = \sqrt{\sum_{i=1}^n (z'_i)^2 / \sigma^2}$  follows the noncentral Chi distribution:

$$h(x) = \eta^{1-\frac{n}{2}} x^{n/2} e^{-\frac{1}{2}(\eta^2+x^2)} I_{n/2-1}(x\eta), \quad (\text{A-1})$$

where  $I_\nu(\cdot)$  is the modified Bessel function of the first kind. The parameters of this distribution are  $n$  and

$$\eta = \sqrt{\sum_{i=1}^n \frac{[(1-\alpha)z_i]^2}{\sigma^2}}, \quad (\text{A-2})$$

which can be rearranged as

$$\begin{aligned} \eta &= \frac{(1-\alpha)}{\sigma} \sqrt{\sum_{i=1}^n z_i^2} \\ &= \frac{(1-\alpha)z}{\sigma}. \end{aligned} \quad (\text{A-3})$$

Since  $z' = \sigma x$  and the distribution of  $x$  is given by eq. A-1, the distribution of  $z'$  is obtained by a change of variable and is given by  $T(z' | z) = h(z'/\sigma)/\sigma$ . This gives the probability density function

$$T(z' | z) = \frac{(z')^{n/2} [(1-\alpha)z]^{1-\frac{n}{2}}}{\sigma} \exp\left(-\frac{(z')^2 + [(1-\alpha)z]^2}{2\sigma^2}\right) I_{n/2-1}\left(\frac{z'(1-\alpha)z}{\sigma^2}\right), \quad (\text{A-4})$$

which depends only on the norm after ( $z'$ ) and before ( $z$ ) innovation.

## Jump stochastic process

From eq. 1 of the main text, we can write

$$A(z', z) = \frac{e^{\lambda u(z')}}{e^{\lambda u(z')} + e^{\lambda u(z)}}, \quad (\text{A-5})$$

which is also expressed only in terms of the norm after ( $z'$ ) and before ( $z$ ) innovation. Hence, an individual can be characterized by its distance  $z$  from the optimum and we can write the probability density function that it goes from state  $z$  to state  $z'$  as

$$R(z' | z) = A(z', z)T(z' | z) \quad (\text{A-6})$$

for  $z' \neq z$ . By contrast, for  $z' = z$ , we need to take into account that when a trial is rejected, the individual always reverts to  $z'$ . Hence, the probability density of remaining in state  $z'$  is

$$R(z' | z') = A(z', z')T(z' | z') + \int_0^\infty [1 - A(y, z')]T(y | z') dy. \quad (\text{A-7})$$

Combining both eq. A-6 and eq. A-7, the transition density function  $p(z' | z)$  that the individual goes from state  $z$  to  $z'$  is

$$p(z' | z) = R(z' | z) + \delta(z' - z) \left[ 1 - \int_0^\infty R(y | z) dy \right], \quad (\text{A-8})$$

where  $\delta(z' - z)$  is the Dirac's delta function. This probability density determines a Markov jump process for the probability density function  $p(z, t)$  of being at distance  $z$  from the optimum at time  $t$  with substitution rate  $R(z' | z)$  (Gardiner, 2009, pp 51-52). This rate is expressed in terms of eq. A-4 and eq. A-5, and was used to numerically simulate the stochastic dynamics of the distance to the origin.

## Diffusion process

Here, we derive a heuristic diffusion approximation to the Markov process determined by eq. A-8. This approximation is based on the derivation of an independent one dimensional diffusion equation for each trait value  $z_i$  of the trait vector  $\mathbf{z}$ . From these  $n$  independent diffusion processes, we then construct a diffusion process for  $z$  by using Itô's transformation formula (Gardiner, 2009; Karlin and Taylor, 1981). Although this derivation does not start directly by using the original transition probability density function for the norm (eq. A-8), it leads to a diffusion process for the distance to the optimum, which accurately describes the original process under a large spectrum of parameter values (see section "Trial-and-error learning" of the main text).

### One dimensional diffusion

In order to obtain a diffusion equation for  $z_i$ , we consider that only a change in the  $i$ -th component of  $\mathbf{z}$  can occur. From the assumptions that  $z'_i$  is normally distributed with mean

$(1 - \alpha)z_i$  and variance  $\sigma^2$ , the trial distribution for the  $i$ -th component is

$$T(z'_i | z_i) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(z'_i - (1 - \alpha)z_i)^2}{2\sigma^2}\right). \quad (\text{A-9})$$

Because only the  $i$ -th component of  $\mathbf{z}$  varies, it follows from the acceptance probability (eq. 1 of the main text) and  $u(z) = -z^2$  that the acceptance probability  $A(\|\mathbf{z}'\|, \|\mathbf{z}\|)$  of the new trait  $\mathbf{z}'$  depends only on the values of  $z'_i$  and  $z_i$ . This acceptance probability is

$$A(z'_i, z_i) = \frac{e^{\lambda(z'_i)^2}}{e^{\lambda(z'_i)^2} + e^{\lambda z_i^2}}. \quad (\text{A-10})$$

With this, eq. A-9, and following the same argument that lead to eq. A-8, we can write the transition probability density function  $p(z'_i | z_i)$  that an individual having currently trait  $z_i$  will switch to  $z'_i$  as

$$p(z'_i | z_i) = A(z'_i, z_i)T(z'_i | z_i) + \delta(z'_i - z_i) \int_{-\infty}^{\infty} [1 - A(y, z_i)] T(y | z_i) dy. \quad (\text{A-11})$$

From this density function, we can express the  $k$ -th conditional moment of change in trait value as

$$\mathbb{E}[(z'_i - z_i)^k | z_i] = \int_{-\infty}^{\infty} (z'_i - z_i)^k A(z'_i, z_i) T(z'_i | z_i) dz'_i \quad (\text{A-12})$$

since the integrand takes value zero when  $z'_i = z_i$ . By way of a Taylor expansion around  $\lambda = 0$  and  $\alpha = 0$ , we obtain

$$\begin{aligned} \mathbb{E}[z'_i - z_i | z_i] &= -\frac{z_i}{2} (\lambda\sigma^2 + \alpha) + O(\lambda^2) + O(\alpha^2) + O(\lambda\sigma^4) + O(\lambda\sigma^2\alpha) \\ \mathbb{E}[(z'_i - z_i)^2 | z_i] &= \frac{\sigma^2}{2} + O(\lambda^2) + O(\alpha^2) + O(\lambda\sigma^4) + O(\lambda\sigma^2\alpha) \\ \mathbb{E}[(z'_i - z_i)^3 | z_i] &= O(\lambda^2) + O(\alpha^2) + O(\lambda\sigma^4) + O(\lambda\sigma^2\alpha). \end{aligned} \quad (\text{A-13})$$

Neglecting higher order terms in eq. A-13, and assuming that  $\mathbb{E}[(z'_i - z_i)^k]$  for  $k > 3$  are also negligible, we take the first two moments in eq. A-13 as defining a diffusion process with, respectively, infinitesimal mean and variance

$$\begin{aligned} a_i &= -\frac{(\lambda\sigma^2 + \alpha)}{2} z_i \\ b_i &= \frac{\sigma^2}{2}. \end{aligned} \quad (\text{A-14})$$

This can be recognized as an Ornstein-Uhlenbeck process (Gardiner, 2009, p. 72; Karlin and Taylor, 1981, p. 172) so that given initial value  $z_i(0)$ , the density function  $p(z_i, t)$  is Normal with mean  $z_i(0) \exp(-\frac{1}{2}(\lambda\sigma^2 + \alpha)t)$  and variance  $(\sigma^2/2) \left(1 - e^{-(\lambda\sigma^2 + \alpha)t}\right) / (\lambda\sigma^2 + \alpha)$ . Asymptotically,  $t \rightarrow \infty$ ,  $z_i$  is thus normally distributed with zero mean and variance  $1/[2(\lambda + \alpha/\sigma^2)]$ .



### Multidimensional diffusion

Having a diffusion process for each phenotypic component  $z_i$  of the vector  $\mathbf{z}$  (eq. A-14), we now use Itô's formula to obtain a diffusion equation for the norm  $z = \sqrt{z_1^2 + z_2^2 + \dots + z_i^2}$ . Because the covariances between the  $n$  diffusion processes are equal to zero (each diffusion is independent of each other), the multi-variate Itô transformation formula (Gardiner, 2009, eq. 4.3.26) informs us that the time dynamics  $z(t)$  of the norm (a random variable) satisfies the stochastic differential equation

$$dz = \left[ \sum_i a_i \frac{\partial z}{\partial z_i} + \frac{1}{2} \sum_i b_i \frac{\partial^2 z}{\partial z_i^2} \right] dt + \sum_i \sqrt{b_i} \frac{\partial z}{\partial z_i} dW_i, \quad (\text{A-15})$$

where  $dW_i$  is the white noise process from the Brownian motion (Gardiner, 2009; Karlin and Taylor, 1981). From the definition of the norm, we have

$$\begin{aligned} \frac{\partial z}{\partial z_i} &= \frac{z_i}{z} \\ \frac{\partial^2 z}{\partial z_i^2} &= \frac{1}{z} - \frac{z_i^2}{z^3}, \end{aligned} \quad (\text{A-16})$$

which on substitution into eq. A-18 along with eq. A-14 produces

$$dz = \frac{1}{2} \left[ -(\lambda\sigma^2 + \alpha) \sum_i \frac{z_i^2}{z} + \frac{\sigma^2}{2} \sum_i \left( \frac{1}{z} - \frac{z_i^2}{z^3} \right) \right] dt + \frac{\sigma}{\sqrt{2}} \sum_i \frac{z_i}{z} dW_i. \quad (\text{A-17})$$

From our notations  $\sum_i z_i^2 = z^2$  and since  $\sum_i \frac{z_i}{z} dW_i$  follows the one-dimensional Brownian motion  $dW$  (Øksendal, 2007, p. 149, Theorem 8.4.2), we obtain after simplifications a one-dimensional stochastic differential equation

$$dz = a(z) dt + \sqrt{b(z)} dW, \quad (\text{A-18})$$

where

$$\begin{aligned} a(z) &= \frac{\sigma^2}{4z} [(n-1) - 2\lambda_e z^2] \\ b(z) &= \frac{\sigma^2}{2} \end{aligned} \quad (\text{A-19})$$

are the infinitesimal mean and variance of the change in the norm and  $\lambda_e = \lambda + \alpha/\sigma^2$ .

Eq. A-19 also provides the infinitesimal mean and variance of the diffusion process that arises as a solution of eq. A-18 (Karlin and Taylor, 1981, p. 376), and which describes the dynamics of the probability density function  $p(z, t)$  that realized value  $z$  of the norm is observed at time  $t$  as

$$\frac{\partial p(z, t)}{\partial t} = -\frac{\partial}{\partial z} [a(z)p(z, t)] + \frac{1}{2} \frac{\partial^2}{\partial z^2} [b(z)p(z, t)]. \quad (\text{A-20})$$

Using eq. A-19, the explicit solution to this radial Ornstein-Uhlenbeck process can be written as

$$p(z, t) = \frac{2z^{\nu+1}e^{t\nu\lambda_e\sigma^2/2}}{z_0^\nu\sigma^2\kappa(t)} \exp\left[-\frac{z^2 + z_0^2e^{-t\lambda_e\sigma^2}}{\sigma^2\kappa(t)}\right] I_\nu\left(\frac{2zz_0e^{-t\lambda_e\sigma^2/2}}{\sigma^2\kappa(t)}\right), \quad (\text{A-21})$$

where  $\nu = n/2 - 1$ ,  $\kappa(t) = (1 - e^{-t\lambda_e\sigma^2})/(\lambda_e\sigma^2)$ , and  $z_0$  is the initial value (Barczy and Pap, 2005).

The radial Ornstein-Uhlenbeck process eventually reaches a stationary state determined by  $p(z) = \lim_{t \rightarrow \infty} p(z, t)$ . From eq. A-21, this is

$$p(z) = \frac{2 \exp(-\lambda_e z^2) z^{n-1} \lambda_e^{n/2}}{\Gamma(n/2)}, \quad (\text{A-22})$$

which is a generalized gamma distribution with scale parameter  $n$  and shape parameter  $\lambda_e$ .

## Appendix B: fictitious learning

### Markov process for the distance to the origin

Here, we derive a one dimensional transition probability density  $p(z' | z)$  for the Markov process that a fictitious player goes from state  $z$  to  $z'$  from the knowledge of the acceptance probability and the trial probability density function  $T(\mathbf{z}' | \mathbf{z})$  of choosing trait  $\mathbf{z}'$  when the current trait is  $\mathbf{z}$  (eq. 2 of the main text). This probability density is

$$T(\mathbf{z}' | \mathbf{z}) = \frac{e^{\gamma u(\|\mathbf{z}'\|)}}{\int_B e^{\gamma u(\|\mathbf{v}\|)} dv_1 dv_2 \dots dv_n} \quad (\text{B-1})$$

if  $\|\mathbf{z}' - \mathbf{z}\| \leq k$ , zero otherwise. The denominator is the normalization constant, which depends on the volume of integration  $B = \{\mathbf{v} \in \mathcal{R}^n : \|\mathbf{z}' - \mathbf{z}\| \leq k\}$ . This is an  $n$ -ball of radius  $k$  centered at  $\mathbf{z}$ .

### Changing the coordinates basis

In eq. B-1, the coordinates in the integration factor belong to the standard basis so that  $\mathbf{v} = \sum_{i=1}^n v_i \mathbf{e}_i$ , where  $\mathbf{e}_i$  is the base vector of the coordinate system, which is centered at the optimum with elements equal to zero except element  $i$  equal to one. We now introduce a linear transformation of this coordinate system and write  $\mathbf{v} = \mathbf{z} + x\mathbf{x} + \sum_{i=1}^{n-1} y_i \mathbf{y}_i$ , where  $\mathbf{x}, \mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_{n-1}$  are normalized vectors (new basis), which are orthogonal to each other, and

$$\mathbf{x} = \frac{\mathbf{z}}{\|\mathbf{z}\|} = \frac{1}{z} \sum_{i=1}^n z_i \mathbf{e}_i. \quad (\text{B-2})$$

This vector is expressed in terms of the coordinates of the current trait so that  $\mathbf{z} = z\mathbf{x}$ , which gives

$$\mathbf{v} = (z+x)\mathbf{x} + \sum_{i=1}^{n-1} y_i \mathbf{y}_i. \quad (\text{B-3})$$

Geometrically speaking, the new basis is taken so that the  $x$ -axis of the transformed system contains the origin  $\mathbf{0} = (0,0,0..)$  of the trait space and  $\mathbf{z}$ ; namely, the  $x$ -axis of the transformed system is parallel to  $\mathbf{z}$ . Expressed in the new basis, the coordinates of vector  $\mathbf{v}$  are  $(z+x, y_1, y_2, \dots, y_{n-1})$  and the payoff of expressing this trait is given by  $u\left(\sqrt{(z+x)^2 + (y_1)^2 + (y_2)^2 + \dots + (y_{n-1})^2}\right)$ .

With the change of basis, we can write the transition density function to the new trait  $\mathbf{z}'$  with coordinates  $(z+x', y'_1, y'_2, \dots, y'_{n-1})$  as

$$T(\mathbf{z}' | \mathbf{z}) = \frac{1}{K} e^{\gamma u(\sqrt{(z+x')^2 + (y'_1)^2 + (y'_2)^2 + \dots + (y'_{n-1})^2})}, \quad (\text{B-4})$$

where

$$K = \int_{-k}^k \int_{\Omega(x)} e^{\gamma u(\sqrt{(z+x)^2 + (y_1)^2 + (y_2)^2 + \dots + (y_{n-1})^2})} dx dy_1 dy_2 \dots dy_{n-1}, \quad (\text{B-5})$$

is the normalization constant. Here,  $\Omega(x) = \{(y_1, y_2, \dots, y_{n-1}) \in \mathcal{R}^{n-1} : \sum_{i=1}^{n-1} y_i^2 \leq k^2 - x^2\}$  is the  $n-1$ -ball of radius  $\sqrt{k^2 - x^2}$  centered at  $\mathbf{z}$ .

Note that  $x$  ranges from  $-k$  to  $k$  if and only if  $z \geq k$ . When the current position is too close to the origin ( $z < k$ ), the range of  $x$  becomes  $[-z, k]$ . In the following analysis, however, we only consider the case where  $z \geq k$ .

### Using spherical coordinates

We use polar coordinates to carry out the integration in eq. B-5. To that aim, let  $y^2 = (y_1)^2 + (y_2)^2 + \dots + (y_{n-1})^2$  be the hypersphere with radius  $y$  and write the infinitesimal volume  $dy_1 dy_2 \dots dy_{n-1}$  in terms of spherical polar coordinates as  $J dy d\phi_1 d\phi_2 \dots d\phi_{n-2}$ . Here, the Jacobian of the transformation is  $J = y^{n-2} \cos(\phi_1)^{n-3} \cos(\phi_2)^{n-4} \dots \cos(\phi_{n-3})$  and  $\phi_i$  varies between  $-\pi/2$  and  $\pi/2$  except  $\phi_{n-2}$ , which varies between 0 and  $2\pi$  (Kendall, 2004, p. 16–17). This change of variables allows us to write

$$\begin{aligned} & \int_{\Omega(x)} e^{\gamma u(\sqrt{(z+x)^2 + (y_1)^2 + (y_2)^2 + \dots + (y_{n-1})^2})} dy_1 dy_2 \dots dy_{n-1} \\ &= \int_0^{\sqrt{k^2 - x^2}} \int_{\phi_1 = -\pi/2}^{\pi/2} \dots \int_{\phi_{n-3} = -\pi/2}^{\pi/2} \int_{\phi_{n-2} = 0}^{2\pi} e^{\gamma u(\sqrt{(z+x)^2 + y^2})} \\ & \quad \times y^{n-2} \cos(\phi_1)^{n-3} \cos(\phi_2)^{n-4} \dots \cos(\phi_{n-3}) dy d\phi_1 d\phi_2 \dots d\phi_{n-2}. \quad (\text{B-6}) \end{aligned}$$

We have

$$\int_{\phi_1=-\pi/2}^{\pi/2} \cdots \int_{\phi_{n-3}=-\pi/2}^{\pi/2} \int_{\phi_{n-2}=0}^{2\pi} \cos(\phi_1)^{n-3} \cos(\phi_2)^{n-4} \cdots \cos(\phi_{n-3}) d\phi_1 d\phi_2 \cdots d\phi_{n-2} = \frac{2\pi^{\frac{n-1}{2}}}{\Gamma\left(\frac{n-1}{2}\right)}, \quad (\text{B-7})$$

which is the surface area of an  $n - 1$  ball of unit radius (Kendall, 2004, p. 36). Hence,

$$\int_{\Omega(x)} e^{\gamma u(\sqrt{(z+x)^2+(y_1)^2+(y_2)^2+\cdots+(y_{n-1})^2})} dy_1 dy_2 \cdots dy_{n-1} = \int_0^{\sqrt{k^2-x^2}} e^{\gamma u(\sqrt{(z+x)^2+y^2})} \frac{2\pi^{\frac{n-1}{2}}}{\Gamma\left(\frac{n-1}{2}\right)} y^{n-2} dy, \quad (\text{B-8})$$

whereby the constant of integration given by eq. B-5 becomes

$$K = \int_{-k}^k \int_0^{\sqrt{k^2-x^2}} e^{\gamma u(\sqrt{(z+x)^2+y^2})} \frac{2\pi^{\frac{n-1}{2}}}{\Gamma\left(\frac{n-1}{2}\right)} y^{n-2} dy dx. \quad (\text{B-9})$$

### Retrieving the distribution to the optimum

From the constant of integration (eq. B-9), we see that the probability  $T(x, y | \mathbf{z}) dy dx$  that the new trait is within the intervals  $[x, x+dx)$  and  $[y, y+dy)$  is given by the integrand, which depends only on the current distance  $z$  to the optimum. We can then write the probability density function  $T(x, y | z)$  of going from a current distance  $z$  to point  $(x, y)$  as

$$T(x, y | z) = \frac{e^{\gamma u(\sqrt{(z+x)^2+y^2})}}{K} \frac{2\pi^{\frac{n-1}{2}}}{\Gamma\left(\frac{n-1}{2}\right)} y^{n-2}. \quad (\text{B-10})$$

The new distance to the optimum,  $z' = \sqrt{(z+x)^2+y^2}$ , depends on both  $x$  and  $y$ . Because we have the joint distribution of  $x$  and  $y$ , we can obtain the probability density function of  $z'$  by applying the rules of transformations of random variables.

We consider the transformation

$$\begin{aligned} z' &= \sqrt{(z+x)^2+y^2} \\ w &= y \end{aligned} \quad (\text{B-11})$$

Solving eq. B-11 for  $x$  and  $y$ , we obtain two roots:

$$\begin{aligned} x &= -z \pm \sqrt{(z')^2 - w^2} \\ y &= w. \end{aligned} \quad (\text{B-12})$$

Owing to the limitation that the range of  $x$  is  $[-k, k]$  and that of  $y$  is  $[0, \sqrt{k^2-x^2}]$ , there is a unique relevant root to the transformation:

$$\begin{aligned} x &= -z + \sqrt{(z')^2 - w^2} \\ y &= w. \end{aligned} \quad (\text{B-13})$$



Hence, the transformation has an inverse and the probability density function  $T(z', y | z)$  of going from current distance  $z$  to the point  $(z', w)$  is

$$T(z', w | z) = T_{XY}(x(z, z', w), y(w) | z) \times |\det(J(x, y))| \Big|_{x=-z+\sqrt{(z')^2+w^2}, y=w} \quad (\text{B-14})$$

where  $T_{XY}(x, y | z)$  is eq. B-10 and  $|\det(J(x, y))|$  is the absolute value of the determinant of the Jacobian matrix of the transformation. The Jacobian matrix is

$$J(x, y) = \begin{pmatrix} \frac{\partial x}{\partial z'} & \frac{\partial x}{\partial w} \\ \frac{\partial y}{\partial z'} & \frac{\partial y}{\partial w} \end{pmatrix} = \begin{pmatrix} \frac{z'}{\sqrt{(z')^2-w^2}} & -\frac{w}{\sqrt{(z')^2-w^2}} \\ 0 & 1 \end{pmatrix}, \quad (\text{B-15})$$

which yields the determinant  $|\det(J(x, y))| = z'/\sqrt{(z')^2-w^2}$ . On substitution of eq. B-13 into  $\sqrt{(z+x)^2+y^2}$ , one gets  $z'$ . Thereby

$$T(z', w | z) = \frac{1}{K} \frac{z' e^{\gamma u(z')}}{\sqrt{(z')^2-w^2}} \frac{2\pi^{\frac{n-1}{2}}}{\Gamma(\frac{n-1}{2})} w^{n-2}. \quad (\text{B-16})$$

By integrating over  $w$ , we can now obtain the probability density function of  $z'$  as

$$T(z' | z) = \int_0^{w_{\max}} \frac{1}{K} \frac{z' e^{\gamma u(z')}}{\sqrt{(z')^2-w^2}} \frac{2\pi^{\frac{n-1}{2}}}{\Gamma(\frac{n-1}{2})} w^{n-2} dw, \quad (\text{B-17})$$

where  $w_{\max}$  is the upper limit of integration. This has to be expressed in terms of  $z'$  and  $k$  (instead of  $x$  and  $k$  as in eq. B-9). In order to find the upper limit  $w_{\max}$  of  $w \in [0, w_{\max}]$ , we use the constraints for given values of  $z'$  and  $k$ :

$$\begin{aligned} z' &= \sqrt{(z+x_{\max})^2+w_{\max}^2} \\ k &= \sqrt{x_{\max}^2+w_{\max}^2}, \end{aligned} \quad (\text{B-18})$$

which leads to the relevant root

$$\begin{aligned} x_{\max} &= \frac{(z')^2 - k^2 - z^2}{2z} \\ w_{\max} &= \frac{\sqrt{(k+z-z')(k-z+z')(-k+z+z')(k+z+z')}}{2z}. \end{aligned} \quad (\text{B-19})$$

Using the explicit expression for the integration limit in eq. B-17, the symbolic algebra system Mathematica (Wolfram, 2003) produces the explicit trial distribution

$$T(z' | z) = (2z)^{1-n} \pi^{(n-1)/2} e^{\gamma u(z')} X(z', z)^{(n-1)/2} F_2 \left( \frac{1}{2}, \frac{n-1}{2}, \frac{n+1}{2}, \frac{X(z', z)}{4z^2(z')^2} \right), \quad (\text{B-20})$$

where

$$F_2(a, b, c, z) \equiv \sum_{n=0}^{\infty} \frac{\Gamma(a+n)\Gamma(b+n)\Gamma(c)z^n}{\Gamma(a)\Gamma(b)\Gamma(c+n)n!} \quad (\text{B-21})$$

is the Gauss hypergeometric function (Abramowitz and Stegun, 1964) and

$$X(z', z) \equiv (k + z + z')(k + z - z')(k - z + z')(-k + z + z'). \quad (\text{B-22})$$

Inserting  $T(z' | z)$  (eq. B-20) into eq. A-6 defines a one-dimensional Markov process for arbitrary  $n$ , which we solved numerically to compare to the results obtained by the diffusion approximation derived in the next section (see Fig. 9 of the main text).

## Diffusion process

Here, we derive a heuristic diffusion approximation to the Markov process determined by eq. A-8 with eq. B-20. As we did for trial-and-error learning, the diffusion for the distance to the origin is based on the derivation of an independent one-dimensional diffusion equation for each trait value  $z_i$  of the trait vector  $\mathbf{z}$ .

For a single axis  $z_i$ , we assume that the trial density distribution is given by a one-dimensional version of eq. B-1. With the quadratic utility function  $u(z) = -z^2$ , this becomes

$$T(z'_i | z_i) = \frac{e^{-\gamma(z'_i)^2}}{\int_{z_i-k}^{z_i+k} e^{-\gamma y^2} dy} \quad (\text{B-23})$$

for  $\|z'_i - z_i\| \leq k$ , zero otherwise. With this equation, we can obtain the transition probability density function  $p(z'_i | z_i)$  that an individual having currently trait  $z_i$  will switch to  $z'_i$ . This is given by eq. A-11 from which we can evaluate the  $k$ -th conditional moment of change in trait value as  $E[(z'_i - z_i)^k | z_i] = \int_{-\infty}^{\infty} (z'_i - z_i)^k A(z'_i, z_i) T(z'_i | z_i) dz'_i$  (eq. A-12). Substituting eq. A-10 and eq. B-23, and by way of a Taylor expansion around  $\lambda = 0$ ,  $\gamma = 0$ , and  $k^2 = 0$ , we then obtain the three moments of the change in trait value:

$$\begin{aligned} E[z'_i - z_i | z_i] &= -\frac{2}{6} (\lambda + 2\gamma) k^2 z_i + O(\lambda^2) + O(k^4) \\ E[(z'_i - z_i)^2 | z_i] &= \frac{k^2}{6} + O(\lambda^2) + O(k^4) \\ E[(z'_i - z_i)^3 | z_i] &= O(\lambda^2) + O(k^4). \end{aligned} \quad (\text{B-24})$$

Neglecting higher order terms in this equation, and assuming that  $E[(z'_i - z_i)^k]$  for  $k > 3$  are also negligible, we assume that the two moments

$$\begin{aligned} a_i &= -\frac{2}{6} (\lambda + 2\gamma) k^2 z_i \\ b_i &= \frac{k^2}{6} \end{aligned} \quad (\text{B-25})$$

determine a diffusion process for  $z_i$ .

We can now use the Itô transformation formula (eq. A-15) in order to amalgamate  $n$  independent diffusion processes determined by eq. B-25. By the same line of arguments

as we used for the EL model, we obtain after simplifications a one-dimensional stochastic differential equation for the distance to the origin

$$dz = a(z) dt + \sqrt{b(z)} dW, \quad (\text{B-26})$$

where

$$\begin{aligned} a(z) &= \frac{k^2}{12z} [(n-1) - 2(\lambda + 2\gamma)z^2] \\ b(z) &= \frac{k^2}{6} \end{aligned} \quad (\text{B-27})$$

are the infinitesimal mean and variance of the change in the norm.

We mention that the derivation of eq. B-26 is not rigorous and that the transition function eq. B-23 is not continuous thereby breaking the requirements for obtaining stochastic differential equations (Karlin and Taylor, 1981). However, our aim is essentially to obtain an approximate analytical expression for the mean change  $a(z)$  for the norm, which allows comparison between analytical learning models and whose accuracy can then be checked by comparison to the exact results obtained from eq. B-20. Extensive numerical simulations shows that for small  $k$  values the stationary distribution obtained by numerically solving the Markov process described by eq. B-20 agrees almost perfectly with the stationary distribution obtained by using eq. B-27 in eq. A-21, which leads to eq. A-22 with  $\lambda_e = (\lambda + 2\gamma)$  (see main text).

## Appendix C: population process

### Substitution rate

Here, we derive the dynamics of the distance to the origin for a population of  $N$  trial-and-error learners under the separation of time scales assumption described in the main text. Under this separation of time scales between individual and social learning, the substitution rate of state  $z$  by new state  $z'$  can be written as

$$R(z', z) = \mu NT(z' | z)A(z', z)\Pi(z', z), \quad (\text{C-1})$$

where  $\mu N$  is the number of trials carried out by the  $N$  individuals in the population per time step,  $T(z' | z)$  is the probability that a trait at distance  $z'$  from the optimum is tried out by an individual (given by eq. A-4), and  $A(z', z)$  is the probability that this individual accepts the new trait (given by eq. A-5). Thus,  $\mu NT(z' | z)A(z', z)$  can be interpreted as the probability that an individual with a new trait  $z'$  appears in the population per unit time. Then,  $\Pi(z', z)$  is the probability that this trait is adopted by all individuals in the population when there is initially a single individual with trait  $z'$  and  $N-1$  other individuals having trait  $z$ . This is the fixation probability of the innovation and Eq. C-1 is an instance of a long-term cultural evolutionary rate (Aoki et al., 2011, eq. 1.2).

## Fixation probability

The fixation probability  $\Pi(z', z)$  of innovation  $z'$  in eq. C-1 depends on the type of social learning in use in the population. We now derive this expression under a Moran cultural process, where only one individual updates its cultural per unit time (Aoki et al., 2011; Lehmann et al., 2011). In doing so, the updating individual randomly samples an exemplar individual from the population and adopts its cultural trait according to a logit choice rule by comparing the payoff of the trait it carries and that from the exemplar individual. This leads to a payoff-biased social learning rule.

Under these assumptions, the birth ( $\lambda_i$ ) and death ( $\mu_i$ ) rates of an individual of type  $z'$  in a population where  $i$  individuals have trait  $z'$  and  $N - i$  individuals have trait  $z$  are

$$\begin{aligned}\lambda_i &= \frac{(N-i)i}{N(N-1)}A(z', z) \\ \mu_i &= \frac{(N-i)i}{N(N-1)}A(z, z').\end{aligned}\tag{C-2}$$

Using the standard expression for the fixation probability under a birth and death process (Ewens, 2004, eq. 2.158), the fixation probability of a single innovation is

$$\Pi(z', z) = \frac{1 - e^{\lambda[(z')^2 - z^2]}}{1 - \left(e^{\lambda[(z')^2 - z^2]}\right)^N}.\tag{C-3}$$

## Diffusion for the norm

### Payoff-based social learning

As under the individual level process for trial-and-error learning described above, we will derive a diffusion equation for the norm  $z$  by deriving first a diffusion equation for each  $z_i$ . Because only the  $i$ -th component of  $\mathbf{z}$  is assumed to vary, we have from eq. C-1 that the substitution rate of a  $z_i$  population by a  $z'_i$  population is

$$R(z'_i, z_i) = \mu NT(z'_i | z_i)A(z'_i, z_i)\Pi(z'_i, z_i),\tag{C-4}$$

where  $T(z'_i | z_i)$  is given by eq. A-9,  $A(z'_i, z_i)$  by eq. A-10, and  $\Pi(z'_i, z_i)$  is obtained from eq. C-3 and reduce to

$$\Pi(z'_i, z_i) = \frac{1 - e^{\lambda[(z'_i)^2 - z_i^2]}}{1 - \left(e^{\lambda[(z'_i)^2 - z_i^2]}\right)^N}\tag{C-5}$$

when only the trait value on the  $i$ 's axis is varied. The transition probability density function  $p(z'_i | z_i)$  that a population with current trait  $z_i$  will switch to  $z'_i$  is then given by eq. A-11 with eq. C-4.



From the transition probability density function, we can obtain by way of a Taylor expansion around  $\lambda = 0$  and  $\alpha = 0$ , the three moments of the change in trait value

$$\begin{aligned} \mathbb{E}[z'_i - z_i | z_i] &= -\frac{1}{2}\mu z_i (N\lambda\sigma^2 + \alpha) + O(\lambda^2) + O(\alpha^2) + O(\lambda\sigma^4) + O(\lambda\sigma^2\alpha) \\ \mathbb{E}[(z'_i - z_i)^2 | z_i] &= \frac{\mu\sigma^2}{2} + O(\lambda^2) + O(\alpha^2) + O(\lambda\sigma^4) + O(\lambda\sigma^2\alpha) \\ \mathbb{E}[(z'_i - z_i)^3 | z_i] &= O(\lambda^2) + O(\alpha^2) + O(\lambda\sigma^4) + O(\lambda\sigma^2\alpha). \end{aligned} \quad (\text{C-6})$$

Neglecting higher order terms in eq. A-13, and assuming that  $\mathbb{E}[(z'_i - z_i)^k] \sim 0$  for  $k > 3$ , we use the two first moments to define a diffusion process with, respectively, infinitesimal mean and variance

$$\begin{aligned} a_i &= -\frac{1}{2}\mu z_i (N\lambda\sigma^2 + \alpha) \\ b_i &= \frac{\mu\sigma^2}{2}. \end{aligned} \quad (\text{C-7})$$

This defines again an Ornstein-Uhlenbeck process (Gardiner, 2009, p. 72; Karlin and Taylor, 1981, p. 172).

Substituting the moments into the Itô transformation formula (eq. A-15) and following the same line of arguments as for the individual process, we obtain after simplifications that the dynamics of the norm follows the stochastic differential equation eq. A-18 with infinitesimal parameters

$$\begin{aligned} a(z) &= \frac{\mu\sigma^2}{4z} [(n-1) - 2(\lambda N + \alpha/\sigma^2) z^2] \\ b(z) &= \frac{\mu\sigma^2}{2}. \end{aligned} \quad (\text{C-8})$$

This is similar in form to eq. A-19 with  $\lambda_e = \lambda N + \alpha/\sigma^2$ .

### Random copying

Instead of assuming that social learning occurs according to a payoff based comparison rule, one may assume that social learning occurs through random copying of an exemplar individual. In this case, the fixation probability  $\Pi(z'_i, z_i)$  of the innovation is equal to  $1/N$  (Aoki et al., 2011, eq. 2.5). Then, performing exactly the same calculations as in the last section with the right member of eq. C-5 replaced by  $1/N$ , we find that the dynamics of the norm follows the stochastic differential equation eq. A-18 with infinitesimal parameters

$$\begin{aligned} a(z) &= \frac{\mu\sigma^2}{4z} [(n-1) - 2(\lambda + \alpha/\sigma^2) z^2] \\ b(z) &= \frac{\mu\sigma^2}{2}. \end{aligned} \quad (\text{C-9})$$

The term in brackets is the same as in eq. A-19, so that the dynamics is qualitatively the same as if the population consisted of a single isolated learner.

## Appendix D: two-dimensional trait without spherical symmetry

Here, we relax the assumption of spherically symmetric payoff function for the case where  $n = 2$  and analyze the mean change of the distance to the optimum  $z$  ( $a \equiv \mathbb{E}[\Delta z|z]$ ) for a trial-and-error learner. The trait is thus  $\mathbf{z} = (z_1, z_2)$  and the payoff function can be generically written  $u(z_1, z_2)$ . Then, given that the current traits is at distance  $z = \sqrt{z_1^2 + z_2^2}$  from the optimum, the expected change in the distance to the optimum is

$$\mathbb{E}[z' - z|z] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left( \sqrt{(z'_1)^2 + (z'_2)^2} - \sqrt{z_1^2 + z_2^2} \right) \times \frac{e^{\lambda u(z'_1, z'_2)}}{e^{\lambda u(z'_1, z'_2)} + e^{\lambda u(z_1, z_2)}} T(z'_1 | z_1) T(z'_2 | z_2) dz'_1 dz'_2, \quad (\text{D-1})$$

where the trial distributions are given by eq. A-9.

We computed this mean change for various standard payoff functions relaxing the assumption of spherical symmetry. In particular, we considered the payoff function  $u(z_1, z_2) = -(z_1^2 + 2\beta z_1 z_2 + z_2^2)$ , which is a form used in the evolutionary literature (Waxman and Welch, 2005). Here,  $\beta$  tunes the departure from spherical symmetry, and for  $\beta > 0$  it results in constant payoff curves that are ellipses with major axes equidistant from both the  $z_1$  and  $z_2$  axis. We also considered the function  $u(z_1, z_2) = -(\beta z_1^2 + (1 - \beta)z_2^2)$ . Now,  $\beta$  tunes the importance of trait  $z_1$  over  $z_2$ . When  $\beta > 0.5$ , constant payoff curves are also ellipses, but with the major axes parallel to the  $z_1$  axes. We also considered the payoff function  $u(z_1, z_2) = -(z_1^\beta + z_2^\beta)^\beta$ . This is the  $\beta$ -norm, and here  $\beta$  tunes complementarity between traits, which increases above that obtained under spherical symmetry when  $\beta > 2$ . If  $\beta$  becomes very large, then varying only one traits does not increase payoff at all. Finally, we also considered the case where all these payoff functions are exponentiated (payoff given by  $\exp(u(\mathbf{z}))$  in the above functions), which results in payoff functions that are steeper around the optimum.

For all the payoff function described above, the qualitative results that the learning dynamics stabilize on average away from the optimum is preserved and we did not observe important quantitative changes from the case of spherical symmetry by varying  $\beta$ . Hence, our qualitative results are robust to changes in the assumption of spherically symmetric payoff function.

## References

Abramowitz, M. and I. A. Stegun. 1964. Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables. Dover Publications, New York.

- Aoki, K., L. Lehmann, and M. W. Feldman. 2011. Rates of cultural change and patterns of cultural accumulation in stochastic models of social transmission. *Theoretical Population Biology* 79:192–202.
- Barczy, M. and G. Pap. 2005. Connection between deriving bridges and radial parts from multidimensional Ornstein-Uhlenbeck processes. *Periodica Mathematica Hungarica* 50:47–60.
- Ewens, W. J. 2004. *Mathematical Population Genetics*. Springer-Verlag, New York.
- Gardiner, C. W. 2009. *Stochastic Methods*. Springer-Verlag, Berlin, 4th edn.
- Karlin, S. and H. M. Taylor. 1981. *A Second Course in Stochastic Processes*. Academic Press, San Diego.
- Kendall, M. G. 2004. *A Course in the Geometry of n Dimensions*. Dover, New York.
- Lehmann, L., K. Aoki, and M. W. Feldman. 2011. On the number of independent cultural traits carried by individuals and populations. *Philosophical transactions of the Royal Society of London Series B-Biological Sciences* 366:424–35.
- Øksendal, B. 2007. *Stochastic Differential Equation*. Springer-Verlag, Berlin, 6th edn.
- Waxman, D. and J. J. Welch. 2005. Fisher’s microscope and Haldane’s ellipse. *American Naturalist* 166:447–57.
- Wolfram, S. 2003. *Mathematica*, 5th ed. Cambridge University Press, Cambridge.