

Influence of plasticity and learning on evolution under directional selection

Ingo Paenke, Bernhard Sendhoff, Tadeusz Kawecki

2007

Preprint:

This is an accepted article published in American Naturalist. The final authenticated version is available online at: [https://doi.org/\[DOI not available\]](https://doi.org/[DOI not available])

Influence of Plasticity and Learning on Evolution under Directional Selection

Ingo Paenke,^{1,2,*} Bernhard Sendhoff,^{2,†} and Tadeusz J. Kawecki^{3‡}

1. Institute of Applied Informatics and Formal Description Methods, University of Karlsruhe, D-76128 Karlsruhe, Germany;

2. Honda Research Institute Europe, Carl-Legien-Strasse 30, D-63073 Offenbach/Main, Germany;

3. Department of Biology, University of Fribourg, Chemin du Musée 10, CH-1700 Fribourg, Switzerland

*Submitted November 15, 2006; Accepted March 8, 2007;
Electronically published June 5, 2007*

ABSTRACT: Phenotypic plasticity and related processes (learning, developmental noise) have been proposed to both accelerate and slow down genetically based evolutionary change. While both views have been supported by various mathematical models and simulations, no general predictions have been offered as to when these alternative outcomes should occur. Here we propose a general framework to study the effects of plasticity on the rate of evolution under directional selection. It is formulated in terms of the fitness gain gradient, which measures the effect of a marginal change in the degree of plasticity on the slope of the relationship between the genotypic value of the focal trait and log fitness. If the gain gradient has the same sign as the direction of selection, an increase in plasticity will magnify the response to selection; if the two signs are opposite, greater plasticity will lead to slower response. We use this general result to derive conditions for the acceleration/deceleration under several simple forms of plasticity, including developmental noise. We also show that our approach explains the results of several specific models from the literature and thus provides a unifying framework.

Keywords: Baldwin effect, genetic assimilation, learning, plasticity, developmental noise, fitness landscape.

The relationship between the genotype and the phenotype is molded by broadly defined phenotypic plasticity. This includes phenomena such as random events during development (developmental noise) and direct effects of the

environment (such as poor nutrition resulting in small size). It also includes fortuitous interference of environmental factors, with development resulting in qualitatively new phenotypes (e.g., morphological phenocopies produced by heat shock in the classic experiments of Waddington [1952]). Another category consists of adaptive responses that evolved to maximize fitness in a heterogeneous environment, for example, induced defenses against predators or parasites. Finally, phenomena such as learning, intelligence, or the “memory” of vertebrate adaptive immune systems allow an individual to develop, within its lifetime, an adaptive response to a novel challenge, even one never encountered in the evolutionary past of the species. Much theoretical and empirical research in the past two decades has focused on the evolution of those various mechanisms of plasticity.

Less attention has been paid to the consequences of those plastic processes for evolution. As plasticity changes the genotype-phenotype map, it will also usually change the relationship between genotype and fitness and thus affect the response to natural or artificial selection. In particular, adaptive plasticity or learning may allow genetically unfit individuals to compensate for the inadequacies of their genotypes and still attain high fitness. As a consequence, the genetic variation for fitness may become reduced and the response to selection may be slowed down (e.g., Johnston 1982; Gordon 1992; Papaj 1994; Mayley 1997). However, in some situations plasticity might change the genotype-phenotype relationship in a way that magnifies the genetic effects on fitness and thus may accelerate evolutionary change. This was first proposed in this journal more than 100 years ago by Baldwin (1896; see also Osborn 1896), who argued that learning may “guide” evolution. If an environmental change favors a new behavior, Baldwin argued, individuals capable of learning may develop this behavior within their lifetimes. Natural selection will subsequently favor genetic variants with a greater predisposition toward the optimal behavior (i.e., those that need less learning to develop it), so that eventually the new behavior will evolve to be a genetically determined “instinct.” Baldwin (1896) proposed that this process

* E-mail: ingo.paenke@aifb.uni-karlsruhe.de.

† E-mail: bernhard.sendhoff@honda-ri.de.

‡ Corresponding author; e-mail: tadeusz.kawecki@unifr.ch.

would lead to the evolution of a novel behavior more readily than natural selection acting directly on genetic variation of a given behavior in the absence of behavioral plasticity. Half a century later, Waddington (1952) experimentally demonstrated an analogous process for morphological traits by selecting for variants initially induced by environmental stress (e.g., heat shock or ether poisoning), which evolved to be constitutively expressed. He termed this effect genetic assimilation; however, the analogous process driven by learning is usually referred to as the Baldwin effect (Simpson 1953; Waddington 1953; Robinson and Dukas 1999; Hall 2001; Pigliucci and Murren 2003; Mery and Kawecki 2004). For an extensive discussion of similarities and differences between the genetic assimilation of morphological traits and the Baldwin effect, see Robinson and Dukas (1999). In a direct test of the Baldwin effect, Mery and Kawecki (2004) studied the influence of learning on the rate of experimental evolution. They subjected fruit flies (*Drosophila melanogaster*) to selection on oviposition substrate preference, either with or without giving them an opportunity to rely on learning when making their oviposition choices. They observed both acceleration and slowing of evolution by learning, depending on the identity of the medium that the flies were selected to prefer (Mery and Kawecki 2004).

A number of mathematical and simulation models have been developed to study the effect of plasticity and learning on the response to directional selection. Using a multilocus simulation model, Hinton and Nowlan (1987; see also Maynard Smith 1987) were the first to show that plasticity may accelerate evolutionary change. The analytical treatment by Fontanari and Meir (1990) provided additional support for the simulation results. However, subsequent models indicated that plasticity may also slow evolution (Anderson 1995; Ancel 2000) or that intermediate levels of plasticity may be most conducive to fast evolution (Keesing and Stork 1991; French and Messinger 1994; Mayley 1997). A recent article concluded that plasticity always slows the response to directional selection (Borenstein et al. 2006). (Several other models [e.g., Kirkpatrick 1982; Whitlock 1997; Price et al. 2003; Borenstein et al. 2006] predicted that plasticity helps the population to cross “adaptive valleys” on multipeaked adaptive landscapes. Because this article focuses on directional selection, we do not discuss them further.) Thus, model-based predictions concerning the effect of plasticity on the rate of evolution have often seemed contradictory. Those models rely on various approaches and make rather specific assumptions about the form of plasticity and the shape of the relationship between the phenotype and fitness (fitness function). To our knowledge, no general framework has been put forth to identify under what conditions plasticity speeds or slows the response to directional selection.

In this article, we propose such a framework. We derive general sufficient conditions for an increase in plasticity to accelerate or slow evolution under directional selection. We then apply these conditions to several specific forms of plasticity and show that they explain the results of several previous models. Thus, although our results follow from a simple mathematical model, they allow some general conclusions.

The Model

We consider a large population with discrete generations under directional selection on a continuous phenotypic trait z . In the absence of plasticity, the expected phenotype (the innate or default phenotype) is given by the genotypic value x . Both z and x are expressed as real one-dimensional numbers. Plasticity and similar processes (such as developmental noise or learning) may change the phenotype. Those plasticity processes are assumed to be characterized by a vector of parameters \mathbf{a} , so the phenotype can be written as $z = \phi(x, \mathbf{a})$; elements of \mathbf{a} are assumed to be real. We refer to ϕ as the reaction norm and note that it may be a random function (e.g., reflecting developmental noise). The plasticity parameters \mathbf{a} may reflect both the environment and the properties of the organism, such as developmental stability or learning ability. The model is thus formally similar to that of Rice (1998), but we treat parameters \mathbf{a} as given and concentrate on the evolution of x . The question we ask is, how different models of plasticity, that is, different forms of $\phi(x, \mathbf{a})$, affect the evolutionary change in the mean genotypic value of the population \bar{x} .

Biological interpretation of this model is thus most straightforward for types of plasticity in which an innate phenotype can be objectively defined, particularly where the plasticity is at least in part based on mechanisms not involved in expression of the innate phenotype. One example is learning, whereby an innate behavioral phenotype shown by naive individuals can be modified as a result of experience. Another is immune defense, which typically has a constitutive and an inducible component, often based on different mechanisms (e.g., antimicrobial peptides vs. antibodies). Still another category comprises behavioral responses that can compensate for physiological deficiencies, such as behavioral thermoregulation in reptiles (Huey et al. 2003). We also apply the model to developmental noise, where the genotypic value is simply the mean phenotype expressed by a given genotype. The model may also be formally applied to plastic responses such as the dependence of body size on developmental temperature (seen in many insects), with x defined as the expected phenotype at some particular reference temperature. In such a case, however, the distinction between the innate

phenotype and the plasticity would be arbitrary, dependent on the choice of reference temperature. It is thus mostly the other forms of plasticity described above that we focus on in this article.

We assume that selection is acting on the phenotype z , with individual fitness given by the fitness function (fitness landscape) $f(z)$. When the fitness of a genotype x , $w(x)$, is averaged over all phenotypes, it produces

$$w(x) = E[f(\phi(x, \mathbf{a}))]. \quad (1)$$

The response of the mean genotypic value \bar{x} to selection depends on the selection differential on x :

$$S_x = \frac{\int_{x_{\min}}^{x_{\max}} xp(x)w(x, \mathbf{a})dx}{\int_{x_{\min}}^{x_{\max}} p(x)w(x, \mathbf{a})dx} - \bar{x} \quad (2)$$

(Lande 1979), where \bar{x} is the mean x , (x_{\min}, x_{\max}) is the range of x in the population, $p(x)$ is the probability density function of the distribution of x , and the denominator is the mean fitness. In the following analysis, we assume that within the range (x_{\min}, x_{\max}) , $w(x, \mathbf{a})$ is a monotonic function of x and is differentiable in both variables. If all genetic variation in x is additive (or if reproduction is asexual), the response to selection simply equals the selection differential. This does not hold with dominance or epistasis, but it is reasonable to assume that in most cases the response to selection will be positively correlated with S_x . Thus, the impact of plasticity on the evolutionary change in \bar{x} can be predicted by analyzing its effect on S_x . We analyze this effect by focusing on one plasticity parameter a_k at a time and studying $\partial S_x / \partial a_k$. If $\partial S_x / \partial a_k$ has the same sign as S_x , a marginal increase in a_k will make the selection differential greater in magnitude, thus resulting in stronger response to selection. The reverse holds if $\partial S_x / \partial a_k$ and S_x have opposite signs.

As can be seen from equation (2), the sign of $\partial S_x / \partial a_k$ will in general depend on the shape of $w(x, \mathbf{a})$ and on the distribution of x in the population. To make further progress, we assume that the sign of $\partial^2 \ln w(x, \mathbf{a}) / \partial x \partial a_k$ is constant in the interval (x_{\min}, x_{\max}) . In other words, we assume that the proportional effect of the plasticity parameter a_k on the genotypic fitness $w(x, \mathbf{a})$ is either monotonically increasing or monotonically decreasing within the range of genotypic variation. In the appendix we show that, in this case,

$$\text{sign}\left(\frac{\partial S_x}{\partial a_k}\right) = \text{sign}\left(\frac{\partial^2}{\partial x \partial a_k} \ln w(x, \mathbf{a})\right). \quad (3)$$

This result has a relatively straightforward intuitive interpretation. Consider first the case $\partial^2 \ln w(x, \mathbf{a}) / \partial x \partial a_k = 0$.

This implies $\partial \ln w(x, \mathbf{a}) / \partial a_k$ is a constant; that is, a marginal increase in a_k has the same proportional effect on the fitness of all genotypes. This is equivalent to multiplying $w(x, \mathbf{a})$ by a constant, which has no effect on the selection differential S_x . If, in turn, $\partial^2 \ln w(x, \mathbf{a}) / \partial x \partial a_k > 0$, the fitness of genotypes with greater x increases proportionally more (or decreases proportionally less) due to an increase in a_k than does the fitness of genotypes with smaller x . If at the same time genotypic fitness increases with x (which implies $S_x > 0$), an increase in a_k will magnify the relative difference in fitness among genotypes, resulting in greater S_x . If, in contrast, genotypic fitness decreases with x (i.e., $S_x < 0$), an increase in a_k will reduce variation in relative fitness among genotypes, resulting in S_x smaller in absolute value (with $S_x < 0$, this implies $\partial S_x / \partial a_k > 0$). Similar reasoning can be carried out for the case $\partial^2 \ln w(x, \mathbf{a}) / \partial x \partial a_k < 0$. Because $\partial^2 \ln w(x, \mathbf{a}) / \partial x \partial a_k$ quantifies how marginal gain in relative fitness from increased plasticity changes with the genotypic value, we refer to it as the fitness gain gradient.

To summarize, result (3) implies that an increase in parameter a_k will make the selection differential greater in magnitude and so accelerate evolution if the fitness gain gradient has the same sign as the selection differential; it will slow evolution if the two signs are opposite. Below we apply this result to several forms of plasticity to derive the conditions under which they are expected to facilitate or hinder evolutionary change. We also use it to explain the results, sometimes seemingly disparate, of several specific models from the literature.

Developmental Noise

In the classical quantitative genetic model, developmental noise and microenvironmental variation add a normally distributed random variate ε with zero mean and variance V_E to the phenotypic value. Here V_E is the only plasticity parameter (i.e., $\mathbf{a} = [V_E]$), so the reaction norm can be written as

$$\phi(x, V_E) = x + \varepsilon. \quad (4)$$

Based on the third-order Taylor approximation, the genotypic fitness is then

$$w(x, V_E) \cong f(x) + \frac{V_E}{2} f''(x), \quad (5)$$

where $f''(x)$ denotes the second derivative of the phenotypic fitness function $f(z)$ evaluated at x . This approximation applies to any distribution of ε that is symmetric around zero, not just normal. Thus, greater developmental noise will increase the genotypic fitness if the phenotypic

fitness function $f(z)$ is convex and reduce the genotypic fitness if $f(z)$ is concave. The magnitude of this effect increases with the curvature of $f(z)$. Thus, if the curvature of a convex $f(z)$ increases in the direction of increasing fitness, the fitness of already fitter genotypes will increase more; developmental noise will magnify the differences in absolute fitness between genotypes. Similarly, if the curvature of a concave $f(z)$ decreases in the direction of higher fitness, the fitness of already fitter genotypes will be reduced less by the developmental noise, and so the absolute differences in genotypic fitness will again be magnified. Both these cases correspond to f''' having the same sign as f' . Conversely, if the curvature of a convex $f(z)$ decreases or if the curvature of a concave $f(z)$ increases in the direction of increasing fitness (i.e., f''' is the opposite sign to f'), the absolute differences in genotypic fitness will be reduced. However, the rate of evolution depends on the relative rather than absolute differences in fitness. To find out the conditions under which greater developmental noise will magnify or reduce the relative differences in genotypic fitness, we apply result (3) to derive the expression for the fitness gain gradient:

$$\frac{\partial^2}{\partial x \partial V_E} \ln w(x, V_E) \cong 2 \frac{f(x)f'''(x) - f'(x)f''(x)}{(2f(x) + V_E f''(x))^2}. \quad (6)$$

The denominator of the right-hand side is positive, so the sign will depend on the numerator. Note that the sign of S_x is the same as the sign of $f'(x)$. Hence, based on result (3), an increase in the noise variance V_E will magnify relative differences in fitness and thus result in a stronger response to selection if the expression in the numerator of equation (6) has the same sign as $f'(x)$. This result holds for all symmetric noise distributions as long as the fitness function $f(z)$ can be sufficiently well approximated by the third-order Taylor series. For a linear fitness function the fitness gain gradient is zero; that is, V_E does not affect evolution.

This general result can be used to explain the results of several previous models (Cavalli-Sforza and Feldman 1976; Anderson 1995; Ancel 2000), which predicted that normally distributed noise should slow the response to selection. Those authors assumed a Gaussian fitness function $f(z) = c \exp[-s(z - \theta)^2]$, where θ is the optimal phenotype and s determines the strength of selection; the parameter c has only a scaling effect (both s and c are assumed positive). Setting, without loss of generality, $\theta = 0$ and substituting this fitness function in the numerator of equation (6), one arrives at

$$f(x)f'''(x) - f'(x)f''(x) = 8c^2s^2x \exp(-2sx^2). \quad (7)$$

This expression is negative for $x < 0$ and positive for $x >$

0; that is, its sign is always opposite to that of $f'(x)$. Thus, our model predicts that, with a Gaussian fitness landscape, developmental noise of any symmetric distribution (not just normal) is expected to slow evolution, in agreement with the results of the models cited above.

Consider, however, an alternative fitness landscape $f(z) = c \exp[-s(z - \theta)^4]$. In this case, again setting $\theta = 0$,

$$f(x)f'''(x) - f'(x)f''(x) = 24c^2sx(4sx^4 - 1) \exp(-2sx^4), \quad (8)$$

which for $|x| < 2^{-1/2}s^{-1/4}$ has the same sign as $f'(x)$. Thus, for this fitness function, our model predicts that increased noise should magnify the selection differential when most of the population distribution is within $0.71s^{-1/4}$ of the optimum. This effect is illustrated by a numerical example in figure 1, which also shows how it affects the evolutionary trajectory of the population mean. Thus, the conclusion of Cavalli-Sforza and Feldman (1976), Anderson (1995), and Ancel (2000)—that in a quantitative genetic framework, developmental noise slows the response to directional selection—resulted from their choice of a Gaussian fitness function.

Directional Plasticity or Learning

Learning and many forms of plasticity result in a directional change of the phenotype, often but not always in the direction of increasing fitness. As the simplest model of such a directional plasticity, we consider a shift of the phenotype by a constant value u , with a negligible noise component; that is,

$$\phi(x, u) = x + u, \quad (9)$$

so the genotypic fitness $w(x, u) = f(x + u)$. In this case, the gain gradient becomes

$$\frac{\partial^2}{\partial x \partial u} \ln w(x, u) = \frac{f(z)f''(z) - (f'(z))^2}{(f(z))^2} = \frac{d^2}{dz^2} \ln f(z), \quad (10)$$

where f and its derivatives are evaluated at $z = \phi(x, u) = x + u$. To interpret result (10) in the light of equation (3), note that greater plasticity implies greater $|u|$; that is, for $u < 0$, greater plasticity means smaller u . Consider first adaptive plasticity, that is, a shift of the phenotype toward higher fitness (u has the same sign as $f'(z)$ and S_x). Inspection of equations (3) and (10) reveals that such plasticity will accelerate evolution if the phenotypic fitness landscape $f(z)$ is logarithmically convex and will slow it down if it is logarithmically concave. Con-

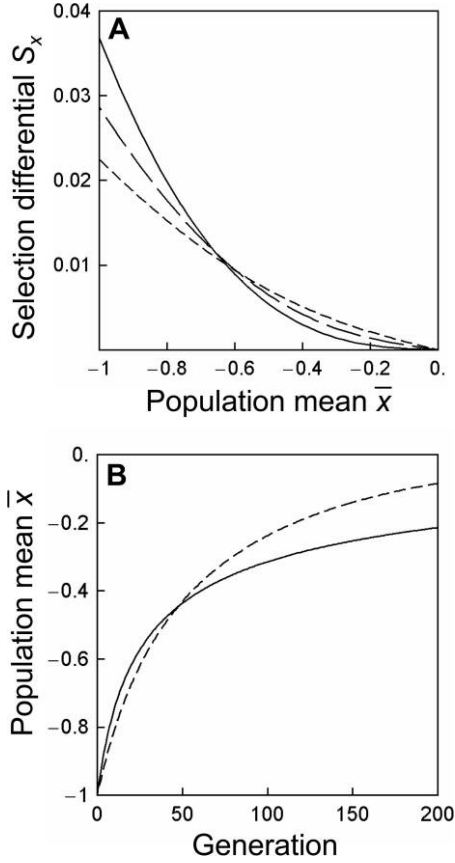


Figure 1: Effect of normally distributed noise with variance V_E on the rate of evolution under fitness function given by $f(z) = \exp[-s(z - \theta)^2]$, with $s = 1$ and $\theta = 0$. A, Selection differential S_x as a function of the population mean, assuming $V_E = 0$ (solid line), $V_E = 0.04$ (long-dashed line), and $V_E = 0.1$ (short-dashed line). B, Predicted evolutionary trajectory of the population mean, assuming $\bar{x} = -1$ at generation 0. Distribution of x was assumed to be normal with additive genetic variance $V_G = 0.01$; the response to selection was modeled as $\Delta \bar{x} = V_G S_x$. $V_E = 0$ (solid line) or $V_E = 0.1$ (short-dashed line). Genotypic fitness and selection differential were calculated by numerical integration of equations (1) and (2) using Mathematica, version 5.0 (i.e., the approximation in eq. [5] was not used).

versely, maladaptive plasticity (u of the opposite sign from $f'(z)$) will slow evolution for logarithmically convex $f(z)$ and accelerate evolution for logarithmically concave $f(z)$. Note that a Gaussian function is always logarithmically concave, and so is a linear function; hence, adaptive plasticity with a constant shift in the phenotype will slow the response to selection described by one of these functions.

Obviously, the form of directional plasticity assumed above, with a constant shift in the phenotype, is only the simplest one. A slightly more general form of the reaction norm allows for the phenotypic effect to depend on the genotypic value:

$$\phi(x, u) = x + uy(x), \quad (11)$$

where u is a parameter quantifying the degree of plasticity and $y(x)$ is a monotonic function. For this reaction norm, the gain gradient is

$$\begin{aligned} \frac{\partial^2}{\partial x \partial u} \ln w(x, u) &= y(x)(1 + uy'(x)) \frac{f(z)f''(z) - (f'(z))^2}{(f(z))^2} \\ &+ y'(x) \frac{f'(z)}{f(z)}, \end{aligned} \quad (12)$$

where f and its derivatives are evaluated at $z = \phi(x, u)$. The first term on the right-hand side is analogous to the entire right-hand side of equation (10); it reflects the fact that a plastic shift in the phenotype exposes the population to a different portion of the fitness function, which may have a different logarithmic slope. The second term reflects the fact that the shift may magnify or reduce the phenotypic differences between genotypes. Note that this latter effect is different from the effect of developmental noise, which increases phenotypic variance without magnifying differences among genotypes. For the sake of simplicity, we restrict the following discussion to $f'(z) > 0$ (selection for larger phenotypic values), as well as $u \geq 0$ and $y(x) \geq 0$ (adaptive plasticity).

Consider four cases, illustrated in figure 2. If the fitness landscape is logarithmically convex ($\partial^2 \ln f(z) / \partial z^2 > 0$) and the shift of the phenotype is greater for greater genotypic values ($y'(x) > 0$), plasticity will move the phenotypes toward a steeper portion of the fitness curve and simultaneously magnify phenotypic variation due to differences in x (i.e., stretch the distribution; fig. 2A). Both effects magnify the variation in log fitness among genotypes, so evolution is expected to be accelerated. Conversely, if the fitness function is logarithmically concave ($\partial^2 \ln f(z) / \partial z^2 < 0$) and the shift in the phenotype is greater for smaller x ($y'(x) > 0$), plasticity moves the distribution of phenotypes toward a logarithmically less steep part of the fitness landscape and additionally reduces the phenotypic differences among genotypes (fig. 2D). Both reduce the variation in log fitness among the genotypes, and thus the response to selection is predicted to be slowed. If $\partial^2 \ln f(z) / \partial z^2 < 0$ but $y'(x) > 0$, plasticity will stretch the phenotypic distribution but expose the population to a less steep part of the fitness function (fig. 2B). The overall effect on the fitness differences among genotypes and thus on the selection differential will depend on which of those two antagonistic effects is greater. The same holds for the case in figure 2C, where $\partial^2 \ln f(z) / \partial z^2 > 0$ but $y'(x) < 0$;

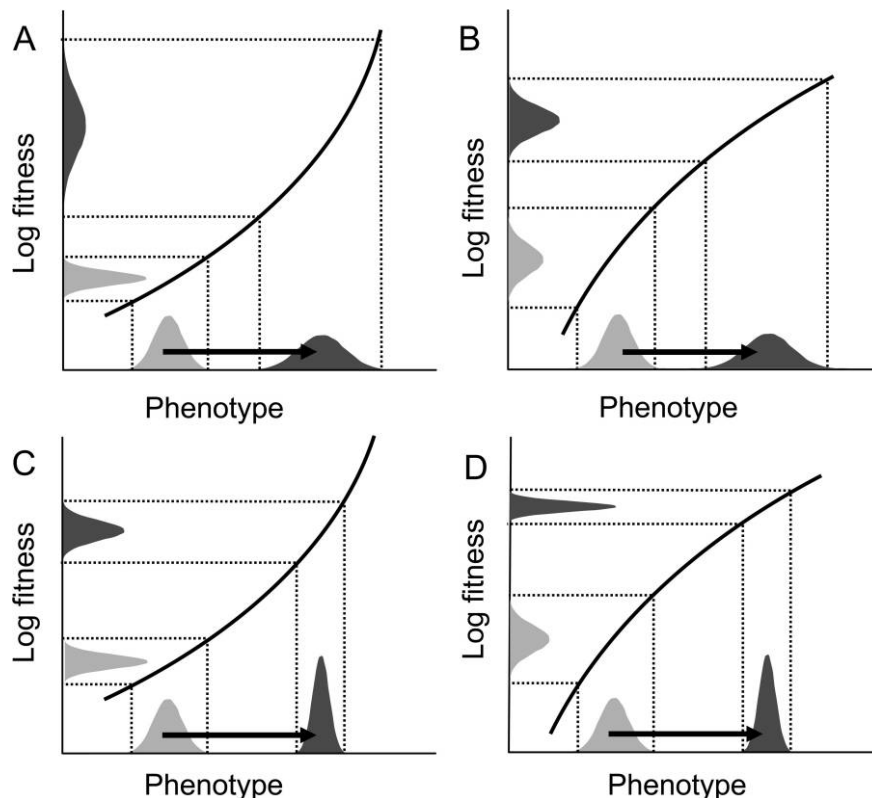


Figure 2: Schematic representation of how directional plasticity defined in equation (11) affects the way variation in the genotypic values x translates into variation in log fitness, which in turn affects the response of mean x to selection. The solid line shows the log of the fitness function $f(z)$, and the shaded areas symbolize the distributions of phenotypes and fitness without plasticity (*light shading*) and with plasticity (*dark shading*). Directional plasticity (symbolized by an *arrow*) shifts the distribution of the phenotypes but may also magnify (*A, B*) or reduce (*C, D*) the phenotypic expression of variation in x . The curvature of the log fitness function (convex in *A* and *C*; concave in *B* and *D*) determines how the shift in the mean of the distribution changes the variation in log fitness.

that is, plasticity compresses the phenotypic distribution but exposes it to a steeper part of the log fitness curve.

The above argument explains the results of Papaj (1994), who modeled the effect of learning on the rate of evolution. He assumed a negatively quadratic fitness function $f(z) = 1 - (1 - z)^2$, which is logarithmically concave ($\partial^2 \ln f(z) / \partial z^2 = -z^{-2} - (2 - z)^{-2} < 0$). Furthermore, he assumed a reaction norm of the form given in equation (11), with $y(x) = 1 - x$, and u being an increasing function of the learning rate and the number of learning opportunities. The assumption of Papaj's model thus corresponds to the situation in figure 2*D*, predicting that an increase in u due to higher learning rate or more learning opportunities should lead to a decrease in the rate of evolution, which his model indeed predicted.

The situation depicted in figure 2*D* is likely to occur in the vicinity of an optimum. First, the fitness function is likely to be concave (and thus also logarithmically concave) around the optimum. Second, if the ability to modify

the phenotype adaptively is large relative to the distance from the optimum, most individuals will be able to show the optimal (or nearly optimal) phenotype. In such a case, individuals closer to the optimum would necessarily benefit less from learning, and evolution would be decelerated. In the extreme case, directional plasticity may allow all individuals to compensate fully for the deficiencies of their genotype and express the optimal phenotype, which will eliminate fitness differences, thus preventing evolutionary change. This is the case of "ideal deterministic learning" examined by Borenstein et al. (2006), which explains their prediction that plasticity decelerates evolution. However, Borenstein et al. (2006, p. 1,565) extrapolated from that special case to make a general statement that, "If selection takes place only within the domain of a simple fitness function that does not include multiple local optima, plasticity hinders the evolutionary process." The above analysis shows that this generalization does not hold.

Needle-in-a-Haystack Fitness Landscape

In the first model to demonstrate that plasticity may facilitate evolution, Hinton and Nowlan (1987; see also Maynard Smith 1987) considered a situation where there is a single “good” phenotype with high fitness while fitness of all other phenotypes is identical and low. Thus, using Hinton and Nowlan’s own words, the fit phenotype is like a needle in a haystack. This form of selection approximates a situation where fitness is increased by successfully performing a complex behavioral task, such as retrieving nectar from a complex flower. Performing the task halfway does not bring any reward and may even be associated with costs. Because the phenotypes closer to, but not yet at, the optimum do not have higher fitness than those farther away, natural selection is not effective in bringing the population closer to the optimum. In a multilocus individual-based simulation model, Hinton and Nowlan have shown that allowing for trial-and-error learning helps a population initially far away from the optimum to eventually reach it. Similar fitness landscape was considered in several other models studying interaction of plasticity and evolution (Fontanari and Meir 1990; French and Messinger 1994; Ance 1999, 2000).

To analyze this case within the framework of this model, we consider a version of the needle-in-a-haystack fitness landscape with a continuous phenotype, where the phenotypic fitness $f(z) = 1 + k$ for z within $\delta/2$ of the optimum and $f(z) = 1$ for z outside this interval; $k > 0$, and δ is assumed to be positive and small. Note that $f(z)$ is not differentiable, so the results of the last two sections cannot be applied directly, but the general result in equation (3) can still be used. Without loss of generality, we assume that the optimum is at $z = 0$ and that the population mean is below the optimum (i.e., $\bar{x} < 0$). We first implement plasticity as in equation (4), that is, as normally distributed developmental noise with variance V_E ; this is equivalent to Hinton and Nowlan’s random search with a single search round. The probability that genotype x produces the optimal phenotype is $\int_{x-\delta/2}^{x+\delta/2} q(z) dz$, where $q(z)$ is the probability density function of a normal distribution with zero mean and variance V_E . Because δ is small relative to V_E , this probability can be approximated as $\delta q(x)$, so the genotypic fitness can be written as

$$w(x, V_E) = 1 + \frac{k\delta}{\sqrt{2\pi V_E}} \exp\left(\frac{-x^2}{2V_E}\right). \quad (13)$$

The fitness gain gradient can then be written as

$$\frac{\partial^2}{\partial x \partial V_E} \ln w(x, V_E) = Bx \left[\sqrt{\pi}(3V_E - x^2) \exp\left(\frac{x^2}{2V_E}\right) + k\delta\sqrt{2V_E} \right], \quad (14)$$

where B stands for

$$\frac{k\delta}{\sqrt{2V_E^5} \{k\delta + \sqrt{2\pi V_E} \exp[x^2/(2V_E)]\}^2}, \quad (15)$$

which is always positive. For reasonably small values of k and δ ($k\delta \leq V_E$), the second term in the brackets of equation (14) is much smaller than the first, and the nonzero solution to $\partial^2 \ln w(x, V_E) / \partial x \partial V_E = 0$ is very close to $V_E = x^2/3$ (the solid line in fig. 3 represents the exact numerical solution). If, as assumed above, the population is below the optimum ($x < 0$), the fitness gain gradient is positive for $V_E < x^2/3$ and negative for $V_E > x^2/3$. That is, selection is predicted to be most effective in pushing the population toward the optimum if the noise variance is roughly one-third of the squared distance to the optimum. Recall that the result in equation (3) is not formulated in terms of the population mean \bar{x} but rather assumes that the sign of the fitness gain gradient is the same within the range (x_{\min}, x_{\max}) . Thus, applying this result strictly, the value of V_E that maximizes the selection differential S_x for a given distribution of x is somewhere in the interval $(x_{\min}^2/3, x_{\max}^2/3)$. In practice, it will usually be close to but not exactly at $\bar{x}^2/3$. As an example, the dotted line in figure 3A shows the values of V_E that maximize the selection coefficient S_x if the genotypic values are normally distributed with mean \bar{x} and variance $V_G = 1$. Additional numerical exploration (not shown) indicates that the fitness gain gradient approach predicts quite well the value of V_E maximizing the selection differential unless the population mean is within about $2(V_G)^{1/2}$ of the optimum. In the latter case, even a small amount of noise reduces S_x . These results offer an intuitive explanation for the finding that plasticity accelerates evolution on a needle-in-a-haystack fitness function when the population is far away from the optimum but it slows it down when the population is close to the optimum (French and Messinger 1994; Ance 2000).

Hinton and Nowlan (1987) and French and Messinger (1994) implemented trial-and-error learning with multiple random-search trials. If the probability of finding the optimal phenotype in a single trial is q , then the probability of finding it at least once in n trials is $1 - (1 - q)^n$. We now treat the number of learning trials as the focal plasticity parameter; the corresponding genotypic fitness is

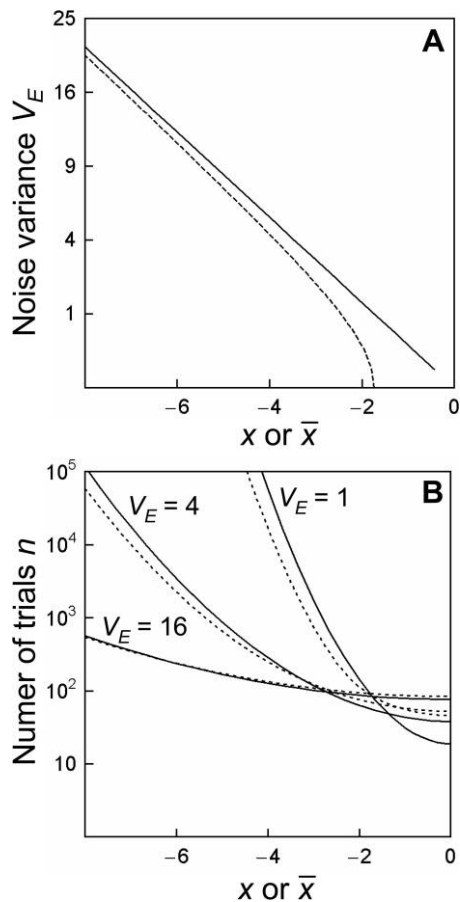


Figure 3: Numerical analysis of the fitness gain gradient approach for the needle-in-a-haystack fitness function. *A*, Random noise, corresponding to equation (13). The solid line shows a solution to $\partial^2 \ln w(x, V_E) / \partial x \partial V_E = 0$; below the line, the gain gradient is positive; above it, it is negative. The dotted line indicates the noise variance V_E that maximizes the selection differential S_x for a given genotypic population mean \bar{x} . *B*, Random search with n trials, corresponding to equation (16). The solid lines show solutions to $\partial^2 \ln w(x, n) / \partial x \partial n = 0$ for three values of V_E ; the gain gradient is positive below the lines. The dotted lines show the number of search trials n that maximizes the selection differential S_x . In both panels, $k = 1$ and $\delta = 0.1$; the variable on the horizontal axis is x for the solid lines and \bar{x} for the dotted lines. The calculations of S_x (dotted lines) assumed that the genotypic values were normally distributed with mean \bar{x} and variance $V_G = 1$.

$$w(x, n) = 1 + k \left\{ 1 - \left[1 - \frac{\delta}{\sqrt{2\pi V_E}} \exp\left(\frac{-x^2}{2V_E}\right) \right]^n \right\}. \quad (16)$$

We can now use the fitness gain gradient approach to ask how the number of learning trials affects the rate of evolution. The expression for $\partial^2 \ln w(x, n) / \partial x \partial n$ is rather unwieldy, so we do not give it here. Numerical analysis indicates that the fitness gain gradient is positive for small n but becomes negative as n increases above a threshold

value (shown as solid lines in fig. 3*B*). This threshold value of n at which the fitness gain gradient reaches its maximum for a given x predicts quite well the number of search trials that actually maximize the selection differential (dotted lines, fig. 3*B*). These results offers an intuitive explanation for the conclusion from the simulation model by French and Messinger (1994) that a small degree of learning may accelerate evolution on a needle-in-a-haystack landscape, but allowing many learning trials or high learning ability slows evolutionary change.

Discussion

The modeling approach developed in this article offers a general theoretical framework to study the effect of plasticity and related processes, such as developmental noise and learning, on the rate of evolution under directional selection. This framework is formulated in terms of the effect of plasticity on the shape of the relationship between genotype and fitness, as defined by the fitness gain gradient (eq. [3]). If the proportional gain of fitness due to plasticity is greater for genotypes that are already fitter, plasticity will magnify the selection differential and thus should accelerate evolution. Conversely, if genetically less fit individuals gain proportionally more fitness from plasticity, the relative differences in fitness between genotypes become reduced and the selection becomes effectively weaker. These conditions assume that the sign of the fitness gain gradient is constant throughout the range of variation in genotypic values in the population. If the fitness gain gradient changes the sign (e.g., has a maximum) well within the range of the genotypic variability of the population, our qualitative approach does not predict the overall effect on the rate of evolution. Thus, the maximum of the fitness gain gradient with respect to a plasticity parameter does not predict the exact value of the parameter that would maximize the rate of evolution. Our numerical examples demonstrate, however, that this value will usually be close to the value for which the corresponding fitness gain gradient evaluated at the population mean changes its sign from positive to negative.

Despite this limitation, our approach provides a general heuristic explanation for the qualitative effects of plasticity on evolution under directional selection, irrespective of mechanisms generating plasticity and the underlying genetic details. We show that it explains a number of disparate theoretical and simulation results published previously. It can also potentially be used to interpret experimental results. Finally, it can be applied to artificial intelligence and robotics, where combinations of simulated evolution (genetic algorithms) and plasticity (random noise and learning) are applied to solve complex problems (e.g., Nolfi 1999; Harvey et al. 2005).

The fitness gain gradient reflects the combination of two elements: the effect of plasticity on the phenotype (described by the reaction norm) and the relationship between phenotype and fitness (the fitness function). This allowed us to express the conditions for an accelerating versus decelerating effect of plasticity on evolution for some simple forms of the reaction norm. These conditions are expressed in a rather straightforward way in terms of mathematical properties of the fitness landscape. In particular, if the change of the phenotype due to plasticity is adaptive (i.e., toward higher fitness) and its magnitude is similar for all genotypes, evolution is predicted to be accelerated if the logarithm of the fitness function is convex and decelerated when it is concave. More generally, a fitness landscape with a convex logarithm is more likely than one with a concave logarithm to result in evolution being accelerated due to directional plasticity, even if the effect of learning on the phenotype differs among genotypes. We could also formulate the conditions for evolution to be accelerated versus slowed by adding a nongenetic variance component (developmental noise) to the phenotype without changing the mean. In particular, in contrast to results of Anderson (1995) and Ancel (2000), we could find conditions under which symmetrically distributed noise accelerated evolution. This is possible under some highly nonlinear fitness functions, where noise makes genotypes closer to the optimum much less likely than genotypes farther away from the optimum to produce particularly unfit phenotypes or much more likely to produce exceptionally fit phenotypes. An extreme case of the latter situation is the needle-in-a-haystack fitness landscape, assumed in simulation models that predicted an accelerating effect of both noise and learning on the rate of evolution (Hinton and Nowlan 1987; Fontanari and Meir 1990; French and Messinger 1994). Our results also show that the influence of plasticity on the rate of evolution is sensitive to the choice of specific model of plasticity. In par-

ticular, directional effects will generally have a different influence than developmental noise.

In this article, we concentrated on the qualitative effect (acceleration vs. deceleration) and did not systematically investigate its magnitude. The accelerating effect of plasticity should tend to be greatest under strongly nonlinear fitness landscapes, like the needle-in-a-haystack landscape considered in the previous section. However, the accelerating effect of plasticity on evolution may also be large if the effect of plasticity on the phenotype is much larger in individuals that already show some genetic predisposition toward that phenotype. Conversely, if the fitness gain gradient is strongly negative, that is, when genetically poorly fit individuals show a particularly strong adaptive plasticity, plasticity might even reverse the ranking of genotypes with respect to fitness. This would result in selection pushing the population away from the genetic optimum, which could be one way for a population to cross an “adaptive valley” to reach a higher “adaptive peak.” It is unclear how biologically likely such a scenario is. More work is needed to evaluate systematically the magnitude of the effect of plasticity and learning on the rate of evolution. Nonetheless, the experimental results of Mery and Kawecki (2004) indicate that the effect may be large enough to be of biological relevance. We hope to stimulate further theoretical and experimental studies on the importance of plasticity in driving or inhibiting evolutionary change.

Acknowledgments

We are grateful to V. Barger, T. Clerc, T. Han, V. Khare, F. Mery, R. J. N. Phillips, M. Stein, M. Vaz, and two reviewers for comments. I.P. and B.S. want to thank E. Koerner for his support. I.P. wants to thank H. Schmeck for his support. This work was supported in part by the Swiss National Science Foundation (grant to T.J.K.).

APPENDIX

Proof of Result (3)

Expression (3) of the main text can be written more formally as

$$\text{if, for all } x \in (x_{\min}, x_{\max}), \begin{cases} \frac{\partial^2}{\partial x \partial a} \ln w(x, a) > 0, & \text{then } \partial S_x / \partial a > 0 \text{ (A1a)} \\ \frac{\partial^2}{\partial x \partial a} \ln w(x, a) < 0, & \text{then } \partial S_x / \partial a < 0 \text{ (A1b)} \\ \frac{\partial^2}{\partial x \partial a} \ln w(x, a) = 0, & \text{then } \partial S_x / \partial a = 0 \text{ (A1c)} \end{cases}, \tag{A1}$$

where S_x is defined in equation (2). Here we show that this proposition holds true under the assumptions of the model. We define

$$Q(x_0, x_1) = \left(\int_{x_0}^{x_1} p(x)w(x, a)dx \right) \left(\int_{x_0}^{x_1} xp(x) \frac{\partial w(x, a)}{\partial a} dx \right) - \left(\int_{x_0}^{x_1} xp(x)w(x, a)dx \right) \left(\int_{x_0}^{x_1} p(x) \frac{\partial w(x, a)}{\partial a} dx \right) \tag{A2}$$

Note that

$$\frac{\partial S_x}{\partial a} = \frac{Q(x_{\min}, x_{\max})}{\bar{w}^2}. \tag{A3}$$

The sign of $\partial S_x / \partial a$ is thus determined by the sign of $Q(x_{\min}, x_{\max})$, and proving proposition (A1) reduces to showing that $Q(x_{\min}, x_{\max})$ has the same sign as $\partial^2 w(x, a) / \partial x \partial a$. In the following, we first show that the sign of $\partial^2 \ln w(x, a) / \partial x \partial a$ determines the sign of the corresponding expression defined for a narrow interval within the distribution of x , $Q(x_0, x_0 + \delta)$, where $x_{\min} < x_0 < x_0 + \delta < x_{\max}$ and δ is small enough for the functions to be treated as linear. Then we show that, for any $x_{\min} \leq x_0, x_1 \leq x_{\max}$, widening the interval (i.e., increasing x_1 or decreasing x_0) does not change the sign of $Q(x_0, x_1)$, and so $Q(x_{\min}, x_{\max})$ has the same sign as $Q(x_0, x_0 + \delta)$.

Within a narrow interval $(x_0, x_0 + \delta)$, we can make the following linear approximations:

$$w(x, a) = w(x_0, a) + (x - x_0) \frac{\partial w(x_0, a)}{\partial x}, \tag{A4a}$$

$$\frac{\partial w(x, a)}{\partial a} = \frac{\partial w(x_0, a)}{\partial a} + (x - x_0) \frac{\partial^2 w(x_0, a)}{\partial x \partial a}, \tag{A4b}$$

where $\partial w(x_0, a) / \partial x$ denotes $\partial w(x, a) / \partial x$ evaluated at $x = x_0$. The function $p(x)$ could be linearized in an analogous way, but it is simpler to express it as

$$p(x) = p(x_0) + \frac{x - x_0}{\delta} [p(x_0 + \delta) - p(x_0)]. \tag{A4c}$$

With these substitutions, after carrying out the integration and rearranging of terms,

$$Q(x_0, x_0 + \delta) = \frac{\delta^2}{72} [p^2(x_0) + 4p(x_0)p(x_0 + \delta) + p^2(x_0 + \delta)] \times \left(w(x_0, a) \frac{\partial^2 w(x_0, a)}{\partial x \partial a} - \frac{\partial w(x_0, a)}{\partial x} \frac{\partial w(x_0, a)}{\partial a} \right). \tag{A5}$$

The term in the first set of brackets is positive, so the sign of the expression on the right-hand side of equation (A5) depends on the term after the times sign. Note, however, that

$$\frac{\partial^2}{\partial x \partial a} \ln(w(x, a)) = \frac{1}{(w(x, a))^2} \left(w(x, a) \frac{\partial^2 w(x, a)}{\partial x \partial a} - \frac{\partial w(x, a)}{\partial x} \frac{\partial w(x, a)}{\partial a} \right). \quad (\text{A6})$$

Thus, the sign of $Q(x_0, x_0 + \delta)$ is the same as the sign of $\partial^2 \ln(w(x, a))/\partial x \partial a$ evaluated at x_0 .

Proposition (A1) requires that the same holds for $Q(x_{\min}, x_{\max})$, assuming that the sign of $\partial^2 \ln(w(x, a))/\partial x \partial a$ is constant throughout the interval. In other words, we need to show that as the limits of the integrals in $Q(x_0, x_1)$ are extended from $(x_0, x_0 + \delta)$ to (x_{\min}, x_{\max}) , the sign of $Q(x_0, x_1)$ does not change. Consider first extending the upper limit x_1 :

$$\begin{aligned} \frac{\partial Q(x_0, x_1)}{\partial x_1} &= x_1 p(x_1) \frac{\partial w(x_1, a)}{\partial a} \int_{x_0}^{x_1} p(x) w(x, a) dx + p(x_1) w(x_1, a) \int_{x_0}^{x_1} x p(x) \frac{\partial w(x, a)}{\partial a} dx \\ &\quad - p(x_1) \frac{\partial w(x_1, a)}{\partial a} \int_{x_0}^{x_1} x p(x) w(x, a) dx - x_1 p(x_1) w(x_1, a) \int_{x_0}^{x_1} p(x) \frac{\partial w(x, a)}{\partial a} dx. \end{aligned} \quad (\text{A7})$$

This can be simplified by extracting $p(x_1)$, placing all other terms under a single integral and rearranging:

$$\begin{aligned} \frac{\partial Q(x_0, x_1)}{\partial x_1} &= p(x_1) \left[\frac{\partial w(x_1, a)}{\partial a} \int_{x_0}^{x_1} (x_1 - x) p(x) w(x, a) dx - w(x_1, a) \int_{x_0}^{x_1} (x_1 - x) p(x) \frac{\partial w(x, a)}{\partial a} dx \right] \\ &= p(x_1) \int_{x_0}^{x_1} (x_1 - x) p(x) \left(w(x, a) \frac{\partial w(x_1, a)}{\partial a} - w(x_1, a) \frac{\partial w(x, a)}{\partial a} \right) dx \\ &= p(x_1) w(x_1, a) \int_{x_0}^{x_1} (x_1 - x) p(x) w(x, a) \left(\frac{1}{w(x_1, a)} \frac{\partial w(x_1, a)}{\partial a} - \frac{1}{w(x, a)} \frac{\partial w(x, a)}{\partial a} \right) dx \\ &= p(x_1) w(x_1, a) \int_{x_0}^{x_1} (x_1 - x) p(x) w(x, a) \left(\frac{\partial \ln w(x_1, a)}{\partial a} - \frac{\partial \ln w(x, a)}{\partial a} \right) dx. \end{aligned} \quad (\text{A8})$$

For $x = x_1$, the function under the last integral equals zero; for $x < x_1$, its sign is determined by the term in the last parentheses, which has the same sign as $\partial^2 \ln w(x, a)/\partial x \partial a$. (Note that $\partial^2 \ln w(x, a)/\partial x \partial a > 0$ implies $\partial \ln w(x_1, a)/\partial a > \partial \ln w(x, a)/\partial a$ for $x < x_1$, and vice versa.) Hence, the sign of $\partial Q(x_0, x_1)/\partial x_1$ is the same as the sign of $\partial^2 \ln w(x, a)/\partial x \partial a$, assuming that the sign of the latter is constant within interval (x_0, x_1) and that $p(x_1)w(x_1, a) > 0$.

Similarly, the effect of changing the lower limit x_0 is described by

$$\frac{\partial Q(x_0, x_1)}{\partial x_0} = p(x_0) w(x_0, a) \int_{x_0}^{x_1} (x_0 - x) p(x) w(x, a) \left(\frac{\partial \ln w(x, a)}{\partial a} - \frac{\partial \ln w(x_0, a)}{\partial a} \right) dx. \quad (\text{A9})$$

For $x > x_0$, the term in the last parentheses has the same sign as $\partial^2 \ln w(x, a)/\partial x \partial a$; however, the term $x_0 - x$ is negative, so the function under the integral has the opposite sign from $\partial^2 \ln w(x, a)/\partial x \partial a$ for $x_0 < x < x_1$.

The above argument proves proposition (A1) as follows. Consider first the case (A1a). If $\partial^2 \ln w(x, a)/\partial x \partial a > 0$ for all $x \in (x_{\min}, x_{\max})$, then $Q(x_0, x_1) > 0$ for any interval (x_0, x_1) of width δ within (x_{\min}, x_{\max}) ; eq. [A5]. Furthermore, $\partial Q(x_0, x_1)/\partial x_0 \leq 0$ and $\partial Q(x_0, x_1)/\partial x_1 \geq 0$, so as the interval is extended in either direction (increasing x_1 toward x_{\max} or decreasing x_0 toward x_{\min}), $Q(x_0, x_1)$ remains positive (eqq. [A8], [A9]). Hence, $Q(x_{\min}, x_{\max}) > 0$ and $\partial S_x/\partial a > 0$ (eq. [A3]), which proves proposition (A1a). Proof of case (A1b) is analogous: if $\partial^2 \ln w(x, a)/\partial x \partial a > 0$ for all $x \in (x_{\min}, x_{\max})$, then $Q(x_0, x_1) < 0$ for a narrow interval of width δ ; furthermore, $\partial Q(x_0, x_1)/\partial x_0 \geq 0$ and $\partial Q(x_0, x_1)/\partial x_1 \leq 0$; hence, $Q(x_{\min}, x_{\max}) < 0$ and $\partial S_x/\partial a < 0$. Finally, for the case (A1c), if $\partial^2 \ln w(x, a)/\partial x \partial a = 0$ for all $x \in (x_{\min}, x_{\max})$, then $Q(x_0, x_1) = 0$ for a narrow interval of width δ , and it remains zero as the interval is broadened because $\partial Q(x_0, x_1)/\partial x_0 = 0$ and $\partial Q(x_0, x_1)/\partial x_1 = 0$; hence, $Q(x_{\min}, x_{\max}) = \partial S_x/\partial a = 0$.

Literature Cited

- Ancel, L. W. 1999. A quantitative model of the Simpson-Baldwin effect. *Journal of Theoretical Biology* 196:197–209.
- . 2000. Undermining the Baldwin expediting effect: does phenotypic plasticity accelerate evolution? *Theoretical Population Biology* 58:307–319.
- Anderson, R. W. 1995. Learning and evolution: a quantitative genetics approach. *Journal of Theoretical Biology* 175:89–101.
- Baldwin, J. M. 1896. A new factor in evolution. *American Naturalist* 30:441–451, 536–553.
- Borenstein, E., I. Meilijson, and E. Ruppin. 2006. The effect of phenotypic plasticity on evolution in multi-peaked fitness landscapes. *Journal of Evolutionary Biology* 19:1555–1570.
- Cavalli-Sforza, L. L., and M. W. Feldman. 1976. Evolution of continuous variation: direct approach through joint distribution of genotypes and phenotypes. *Proceedings of the National Academy of Sciences of the USA* 73:1689–1692.
- Fontanari, J. F., and R. Meir. 1990. The effect of learning on the evolution of asexual populations. *Complex Systems* 4:401–414.
- French, R. M., and A. Messinger. 1994. Genes, phenes and the Baldwin effect. Pages 277–282 in R. Brooks and P. Maes, eds. *Artificial life IV*. MIT Press, Cambridge, MA.
- Gordon, D. M. 1992. Phenotypic plasticity. Pages 255–262 in E. A. Lloyd and E. F. Kell, eds. *Keywords in evolutionary biology*. Harvard University Press, Cambridge, MA.
- Hall, B. K. 2001. Organic selection: proximate environmental effects on the evolution of morphology and behaviour. *Biology and Philosophy* 16:215–237.
- Harvey, I., E. Di Paolo, R. Wood, M. Quinn, and E. Tuci. 2005. Evolutionary robotics: a new scientific tool for studying cognition. *Artificial Life* 11:79–98.
- Hinton, G. E., and S. J. Nowlan. 1987. How learning can guide evolution. *Complex Systems* 1:495–502.
- Huey, R. B., P. E. Hertz, and B. Sinervo. 2003. Behavioral drive versus behavioral inertia in evolution: a null model approach. *American Naturalist* 161:357–366.
- Johnston, T. D. 1982. Selective costs and benefits in the evolution of learning. *Advances in the Study of Behavior* 12:65–106.
- Keesing, R., and D. G. Stork. 1991. Evolution and learning in neural networks: the number and distribution of learning trials affect the rate of evolution. Pages 804–810 in R. P. Lippmann, J. E. Moody, and D. S. Touretzky, eds. *Proceedings of the 1990 Neural Information Processing Systems Conference (NIPS-3; Denver)*. Morgan Kaufman, San Mateo, CA.
- Kirkpatrick, M. 1982. Quantum evolution and punctuated equilibria in continuous genetic characters. *American Naturalist* 119:833–848.
- Lande, R. 1979. Quantitative genetic analysis of multivariate evolution, applied to brain-body size allometry. *Evolution* 33:402–416.
- Mayley, G. 1997. Guiding or hiding: explorations into the effects of learning on the rate of evolution. Pages 135–144 in P. Husbands and I. Harvey, eds. *Proceedings of the Fourth European Conference on Artificial Life*. Bradford/MIT Press, Cambridge, MA.
- Maynard Smith, J. 1987. When learning guides evolution. *Nature* 329:761–762.
- Mery, F., and T. J. Kawecki. 2004. The effect of learning on experimental evolution of resource preference in *Drosophila melanogaster*. *Evolution* 58:757–767.
- Nolfi, S. 1999. How learning and evolution interact: the case of a learning task which differs from the evolutionary task. *Adaptive Behavior* 7:231–236.
- Osborn, H. F. 1896. Ontogenetic and phylogenetic variation. *Science* 4:786–789.
- Papaj, D. R. 1994. Optimizing learning and its effect on evolutionary change in behavior. Pages 133–154 in L. A. Real, ed. *Behavioral mechanisms in evolutionary biology*. University of Chicago Press, Chicago.
- Pigliucci, M., and C. J. Murren. 2003. Genetic assimilation and a possible evolutionary paradox: can macroevolution sometimes be so fast as to pass us by? *Evolution* 57:1455–1464.
- Price, T. D., A. Qvarnstrom, and D. E. Irwin. 2003. The role of phenotypic plasticity in driving genetic evolution. *Proceedings of the Royal Society B: Biological Sciences* 270:1433–1440.
- Rice, S. H. 1998. The evolution of canalization and the breaking of von Baer's laws: modeling the evolution of development with epistasis. *Evolution* 52:647–656.
- Robinson, B. W., and R. Dukas. 1999. The influence of phenotypic modifications on evolution: the Baldwin effect and modern perspectives. *Oikos* 85:582–589.
- Simpson, G. G. 1953. The Baldwin effect. *Evolution* 7:110–117.
- Waddington, C. H. 1952. Selection for a genetic basis of an acquired character. *Nature* 169:278–278.
- . 1953. The "Baldwin effect," "genetic assimilation" and "homeostasis." *Evolution* 7:386–387.
- Whitlock, M. C. 1997. Founder effects and peak shifts without genetic drift: adaptive peak shifts occur easily when environments fluctuate slightly. *Evolution* 51:1044–1048.