# The evolution of age-dependent plasticity

- Online Supplementary Material -

Barbara Fischer, G. Sander van Doorn, Ulf Dieckmann and Barbara Taborsky

# Contents

Appendix A	Details of mathematical analysis Notational conventions				
	Derivation of the Bayesian update rule	1			
	From individual reproductive success to long-term average fitness	3			
	Weak selection approximation	4			
	Calculation of the width of the reaction norm's plateau	8			
	Nonlinear fitness functions	9			
Figure A1	Dependence of optimal plasticity patterns on nonlinearities in the fitness function	10			
Figure A2	A2 Optimal reaction norms for non-linear fitness functions				
Table A1	ble A1 Life-history parameters for the iteroparous life history in figure 6				

## Appendix A – Details of mathematical analysis

#### Notational conventions

Two important notational conventions are used consistently throughout this paper. First, we carefully distinguish between a random variable itself and its realization by the use of upper- and lowercase letters. For example, the state of the environment at age *t* is a random variable, denoted by  $E_t$ , which can take two values (A or B); the value that is realized for a particular organism under consideration will be denoted  $e_t$ . Accordingly,  $E_t = e_t$  should be read as 'the event that the environment is in state  $e_t$  at time *t*'.

Second, boldface will be used to represent a sequence of environmental states or observations. For example,  $E = (E_1, E_2, ..., E_T)$  represents the sequence of environmental states over the entire lifetime of the organism. The use of a subscript *t* in combination with a boldface symbol indicates that we only consider the sequence of environmental states or observations up to age *t*. Upper- and lowercase boldface letters distinguish between sequences of random variables and their realizations. Finally, the shorthand notation  $O_t = o_t$  stands for a sequence of events  $O_1 = o_1 \cap O_2 = o_2 \cap ... \cap O_t = o_t$ .

#### Derivation of the Bayesian update rule

Although the organism cannot observe the state of the environment directly, it can determine the likelihood that the environment is in a particular state based on the sequence of observations it has made so far. The distribution of environmental states given an individual's history of observations  $o_t = (o_1, ..., o_t)$  up to the present point in time is characterized by the conditional probability  $p_t = P[E_t = A | O_t = o_t]$ , which we will refer to as the individual's personal 'estimate' of the state of the environment at age *t*. We note that the estimate  $p_t$  integrates the information acquired so far by the organism into a single state variable, which we will later assume to be one of the inputs of the reaction norm governing phenotypic adjustment.

Equation (2) in the main text specifies how the estimate  $p_t$  depends on current and prior information contained in, respectively, the most recent observation  $o_t$  and the estimate  $p_{t-1}$  at the previous time step. To derive this result, we start from the joint probability of the events  $E_t = A$  and  $O_t = o_t$ , which we rewrite in two different ways, using the standard rules for probabilities of statistically dependent events:

$$P[E_{t}=A \cap \boldsymbol{O}_{t}=\boldsymbol{o}_{t}] = P[E_{t}=A \cap (O_{t}=\boldsymbol{o}_{t} \cap \boldsymbol{O}_{t-1}=\boldsymbol{o}_{t-1})] =$$

$$P[\boldsymbol{O}_{t-1}=\boldsymbol{o}_{t-1}] P[E_{t}=A \cap O_{t}=\boldsymbol{o}_{t} | \boldsymbol{O}_{t-1}=\boldsymbol{o}_{t-1}] =$$

$$\begin{cases} P[\boldsymbol{O}_{t-1}=\boldsymbol{o}_{t-1}] P[O_{t}=\boldsymbol{o}_{t} | \boldsymbol{O}_{t-1}=\boldsymbol{o}_{t-1}] P[E_{t}=A | \boldsymbol{O}_{t}=\boldsymbol{o}_{t}] \end{cases}$$
(A1)
$$P[\boldsymbol{O}_{t-1}=\boldsymbol{o}_{t-1}] P[E_{t}=A | \boldsymbol{O}_{t-1}=\boldsymbol{o}_{t-1}] P[O_{t}=\boldsymbol{o}_{t} | E_{t}=A \cap \boldsymbol{O}_{t-1}=\boldsymbol{o}_{t-1}]$$
(b)

Here, we have repeatedly used the fact that the composite event  $O_t = o_t$  can be written as the intersection of two events,  $O_t = o_t \cap O_{t-1} = o_{t-1}$ .

The equality between (a) and (b) in equation (A1) implies that

$$P[E_{t}=A | \boldsymbol{O}_{t}=\boldsymbol{o}_{t}] = P[E_{t}=A | \boldsymbol{O}_{t-1}=\boldsymbol{o}_{t-1}] \frac{P[O_{t}=o_{t} | E_{t}=A \cap \boldsymbol{O}_{t-1}=\boldsymbol{o}_{t-1}]}{P[O_{t}=o_{t} | \boldsymbol{O}_{t-1}=\boldsymbol{o}_{t-1}]},$$
(A2)

which is a form of Bayes' theorem with probabilities conditioned on  $O_{t-1} = o_{t-1}$ .

The term on the left-hand side of equality (A2) has previously been defined as the estimate  $p_t$ . On the right-hand side are three conditional probabilities that can be expressed in terms of the sampling accuracy, the environmental switching rates and the previous estimate  $p_{t-1}$ . First, the term in the numerator of the last term can be simplified by realizing that the sequence of previous observations provides no information about the current observation  $O_t$  if the state of the environment is known. This means that the events  $O_t = o_t$  and  $O_{t-1} = o_{t-1}$  are conditionally independent given  $E_t = A$ . In other words,

$$P[O_t = o_t | E_t = A \cap O_{t-1} = o_{t-1}] = P[O_t = o_t | E_t = A],$$
(A3)

which is equal to either *a* (the sampling accuracy) or 1 - a, depending on whether the current state of the environment is perceived correctly ( $o_t = A$ ) or not ( $o_t = B$ ).

Second, we rewrite the conditional probability in the denominator of the last term of equation (A2), in order to make explicit how the probability of making observation  $o_t$  depends on the state of the environment. This can be done as follows

$$P[O_{t}=o_{t} | \boldsymbol{O}_{t-1}=\boldsymbol{o}_{t-1}] = P[O_{t}=o_{t} \cap (E_{t}=A \cup E_{t}=B) | \boldsymbol{O}_{t-1}=\boldsymbol{o}_{t-1}]$$

$$= P[O_{t}=o_{t} \cap E_{t}=A | \boldsymbol{O}_{t-1}=\boldsymbol{o}_{t-1}] + P[O_{t}=o_{t} \cap E_{t}=B | \boldsymbol{O}_{t-1}=\boldsymbol{o}_{t-1}]$$

$$= P[E_{t}=A | \boldsymbol{O}_{t-1}=\boldsymbol{o}_{t-1}] P[O_{t}=o_{t} | E_{t}=A \cap \boldsymbol{O}_{t-1}=\boldsymbol{o}_{t-1}]$$

$$+ P[E_{t}=B | \boldsymbol{O}_{t-1}=\boldsymbol{o}_{t-1}] P[O_{t}=o_{t} | E_{t}=B \cap \boldsymbol{O}_{t-1}=\boldsymbol{o}_{t-1}]$$

$$= P[E_{t}=A | \boldsymbol{O}_{t-1}=\boldsymbol{o}_{t-1}] P[O_{t}=o_{t} | E_{t}=A]$$

$$+ (1 - P[E_{t}=A | \boldsymbol{O}_{t-1}=\boldsymbol{o}_{t-1}]) P[O_{t}=o_{t} | E_{t}=B],$$
(A4)

where, in the final step, we have again used the fact that  $O_t = o_t$  and  $O_{t-1} = o_{t-1}$  are conditionally independent given that the state of the environment is known.

One final consideration needed for deriving the right-hand side of the Bayesian update rule (2) is that  $P[E_t = A | O_{t-1} = o_{t-1}]$ , which appears as the first factor on the right-hand side of equation (A2) and again on the final lines of result (A4), is related to the previous estimate  $p_{t-1}$ . The exact relationship can be derived as follows:

$$P[E_{t}=A | \boldsymbol{O}_{t-1}=\boldsymbol{o}_{t-1}] = P[E_{t}=A \cap (E_{t-1}=A \cup E_{t-1}=B) | \boldsymbol{O}_{t-1}=\boldsymbol{o}_{t-1}] \\ = P[E_{t}=A \cap E_{t-1}=A | \boldsymbol{O}_{t-1}=\boldsymbol{o}_{t-1}] + P[E_{t}=A \cap E_{t-1}=B | \boldsymbol{O}_{t-1}=\boldsymbol{o}_{t-1}] \\ = P[E_{t-1}=A | \boldsymbol{O}_{t-1}=\boldsymbol{o}_{t-1}] P[E_{t}=A | E_{t-1}=A \cap \boldsymbol{O}_{t-1}=\boldsymbol{o}_{t-1}] \\ + P[E_{t-1}=B | \boldsymbol{O}_{t-1}=\boldsymbol{o}_{t-1}] P[E_{t}=A | E_{t-1}=B \cap \boldsymbol{O}_{t-1}=\boldsymbol{o}_{t-1}] \\ = P[E_{t-1}=A | \boldsymbol{O}_{t-1}=\boldsymbol{o}_{t-1}] P[E_{t}=A | E_{t-1}=B] \\ + P[E_{t-1}=B | \boldsymbol{O}_{t-1}=\boldsymbol{o}_{t-1}] P[E_{t}=A | E_{t-1}=B] \\ + P[E_{t-1}=B | \boldsymbol{O}_{t-1}=\boldsymbol{o}_{t-1}] P[E_{t}=A | E_{t-1}=B] \\ = p_{t-1}(1-\beta) + (1-p_{t-1})\alpha.$$

The crucial step in this derivation (marked by <sup>(\*)</sup>) is to recognize that the events  $E_t = A$  and  $O_{t-1} = o_{t-1}$  are conditionally independent given that the state of the environment at age t-1 is known. Substituting the relationships (A3)-(A5) into equation (A2) and evaluating the resulting expression for either  $o_t = A$  or  $o_t = B$ , yields

$$p_{t} = \begin{cases} \frac{a((1-\beta)p_{t-1} + \alpha(1-p_{t-1}))}{a+(1-2a)(\beta p_{t-1} + (1-\alpha)(1-p_{t-1}))} & \text{if } o_{t} = A, \\ \frac{a(\beta p_{t-1} + (1-\alpha)(1-p_{t-1}))}{1-\frac{a(\beta p_{t-1} + (1-\alpha)(1-p_{t-1}))}{a+(1-2a)((1-\beta)p_{t-1} + \alpha(1-p_{t-1}))} & \text{if } o_{t} = B, \end{cases}$$
(A6)

corresponding to the definition of the Bayesian update rule  $U(p_{t-1}, o_t)$  provided in equation (2) in the main text.

#### From individual reproductive success to long-term average fitness

The phenotype sequence of an individual is fully determined by its reaction norm and the sequence of observations  $o = (o_1, o_2, ..., o_T)$  it has made about the state of the environment. This follows from equations (2) and (3) in the main text, which, for t > 0, define  $x_t$  by the recursions

$$p_t = U(p_{t-1}, o_t)$$
 and  $x_t = x_{t-1} + h_t(x_{t-1}, p_t)$ . (A7)

Unless the sampling accuracy is equal to one, every possible observation sequence will occur in every environment, although generally not with equal frequency.

Let  $R_t(o, e)$  denote the lifetime reproductive success from age *t* onwards of an individual with observation sequence *o* in environment  $e = (e_1, e_2, ..., e_T)$ . This measure of future reproductive success satisfies the recursion

$$R_{t}(\boldsymbol{o},\boldsymbol{e}) = S_{t}^{e_{t}}(x_{t},\Delta x_{t}) \Big[ F_{t}^{e_{t}}(x_{t},\Delta x_{t}) + R_{t+1}(\boldsymbol{o},\boldsymbol{e}) \Big].$$
(A8)

Here,  $F_t^{e_t}$  and  $S_t^{e_t}$ , respectively, denote the age-specific fecundity and survival probability. The terminal reward  $R_{T+1}(o, e)$  is defined to be zero for all o and e. We assume that viability selection occurs after phenotypic adjustment but before offspring production. As reflected by our notation, the age-specific fecundities and survival probabilities depend on the current state of the environment  $e_t$ , the current phenotype and the magnitude of phenotypic adjustment at age t,  $\Delta x_t = |x_t - x_{t-1}|$  (this is to incorporate costs of plasticity).

Individuals subject to the same sequence of environmental states may develop differently as a result of errors in the assessment of environmental cues. In order to calculate the long-term fitness of the reaction norm h, it is therefore necessary to first average the lifetime reproductive success over the possible realizations of the sequence of observations in a given environment. Next, we need to average over all realizations of the sequence of environmental states that populations may be exposed to. As explained in the main text, the fitness function W to be optimized is therefore given by

$$W = \prod_{e} \left( \sum_{o} P \left[ \mathbf{O} = \mathbf{o} \middle| \mathbf{E} = \mathbf{e} \right] R_{I}(\mathbf{o}, \mathbf{e}) \right)^{P[\mathbf{E} = \mathbf{e}]},$$
(A9)

where P[E = e] is the probability that an individual will experience the sequence of environmental states *e*, and P[O = o | E = e] is the probability that such an individual will adjust its phenotype based on the observation sequence *o* in that environment.

#### Weak selection approximation

In order to facilitate the optimization of fitness function (A9), we will assume that selection is weak. This implies that the fitness variability among individuals in different environments or expressing different phenotypes is modest, such that the fecundities and survival probabilities can be expressed in terms of small deviations from a suitably chosen reference life history. In other words, we write:

$$F_{t}^{e_{t}}(x_{t},\Delta x_{t}) = \overline{F}_{t}\left(1 + \partial F_{t}^{e_{t}}(x_{t},\Delta x_{t}) + \partial F_{t}''(i(e_{t}) - \frac{\alpha}{\alpha+\beta})\right),$$

$$S_{t}^{e_{t}}(x_{t},\Delta x_{t}) = \overline{S}_{t}\left(1 + \partial S_{t}^{e_{t}}(x_{t},\Delta x_{t}) + \partial S_{t}''(i(e_{t}) - \frac{\alpha}{\alpha+\beta})\right),$$
(A10)

where  $\overline{F}_t$  and  $\overline{S}_t$  represent, respectively, the average fecundity and survival probability across environments of a hypothetical reference individual with fixed phenotype  $x_t = z$  throughout its entire life. The functions  $\partial F_t^{e_t}$  and  $\partial S_t^{e_t}$  represent relative deviations from these benchmark values for an individual with a plastic phenotype. The first one of these functions is defined as  $\partial F_t^{e_t}(x_t, \Delta x_t) = \left(F_t^{e_t}(x_t, \Delta x_t) - F_t^{e_t}(z, 0)\right)/\overline{F}_t$ , which quantifies the relative fecundity effect of expressing the adjusted phenotype  $x_t$  instead of the fixed phenotype z, as well as the fecundity cost of adjusting the phenotype from  $x_{t-1}$  to  $x_t$ . The definition for  $\partial S_t^{e_t}$  and its interpretation are analogous.

The deviations  $\partial F_t^{e_t}$  and  $\partial S_t^{e_t}$  measure fitness differences within environments. Equations (A10) are therefore made consistent by including additional deviation terms with coefficients  $\partial F_t'' = (F_t^A(z,0) - F_t^B(z,0))/\overline{F_t}$  and  $\partial S_t'' = (S_t^A(z,0) - S_t^B(z,0))/\overline{S_t}$  that quantify the relative fecundity and survival differences between environment A and B of individuals with phenotype z. These coefficients are multiplied by a weighting term in which the indicator function  $i(e_t)$  takes the value 1 if  $e_t = A$  and the value 0 if  $e_t = B$ . Note that the term  $\alpha / (\alpha + \beta)$  corresponds to the probability that the environment is in state A at any point in time.

For the purpose of linearization, we now introduce a positive parameter  $\varepsilon$ , which we choose as small as possible, but such that it bounds the absolute values of all fitness deviations between the plastic and reference life histories. If selection is weak,  $\varepsilon$  is much smaller than 1, allowing us to ignore terms of  $\mathcal{O}[\varepsilon^2]$ . The resulting approximation errors are minimized by choosing the reference phenotype *z* equal to the value that maximizes lifetime reproductive success for an individual with a fixed phenotype.

We now seek to express the reproductive success of a plastic individual in terms of a small marginal deviation  $\partial R_t$  of  $\mathcal{O}[\varepsilon]$  from the reproductive success of an individual following the reference life history. In other words,  $R_t(o, e)$  is approximated by

$$R_t(\boldsymbol{o},\boldsymbol{e}) \approx \overline{R}_t \left( 1 + \partial R_t(\boldsymbol{o},\boldsymbol{e}) \right), \tag{A11}$$

where  $\overline{R}_{t}$ , the average reproductive success associated with the reference life history, is defined by the recursion

$$\overline{R}_{t} = \overline{S}_{t} \left( \overline{F}_{t} + \overline{R}_{t+1} \right) \tag{A12}$$

and  $\overline{R}_{T+1} = 0$ .

An expression for  $\partial R_t$  can be found by substituting the expressions (A10) and (A11) into equation (A8), writing out the products, and retaining only the terms of  $\mathcal{O}[\mathcal{E}]$  (this means ignoring all terms that depend only on the characteristics of the reference life history, and dropping all interaction terms involving products of the deviations  $\partial F_t^{e_t}$ ,  $\partial S_t^{e_t}$ ,  $\partial F_t''$ ,  $\partial S_t''$ ,  $\partial R_t$  and  $\partial R_{t+1}$ ). After some rearrangement, we obtain

$$\partial R_t(\boldsymbol{o},\boldsymbol{e}) = \frac{\overline{R}_{t+1}\overline{S}_t}{\overline{R}_t} \partial R_{t+1}(\boldsymbol{o},\boldsymbol{e}) + \partial S_t^{e_t}(x_t,\Delta x_t) + \frac{\overline{F}_t\overline{S}_t}{\overline{R}_t} \partial F_t^{e_t}(x_t,\Delta x_t) + \left(i(e_t) - \frac{\alpha}{\alpha+\beta}\right) \left[\partial S_t'' + \frac{\overline{F}_t\overline{S}_t}{\overline{R}_t} \partial F_t''\right].$$
(A13)

In this expression, effects on fecundity are weighted with respect to the relative contribution of current reproductive success ( $\overline{F}_t \, \overline{S}_t$ ) to the remaining reproductive success from age *t* onwards ( $\overline{R}_t$ ). Similarly, the future fitness effect is weighted by the relative contribution of future fitness ( $\overline{R}_{t+1} \, \overline{S}_t$ ) to the remaining reproductive success.

The next step is to linearize the fitness function (A9) by substituting an approximation for  $R_1(o, e)$  as in equation (A11). This yields,

$$W \approx \prod_{e} \left( \sum_{o} P[O=o|E=e] \overline{R}_{1} (1+\partial R_{1}(o,e)) \right)^{P[E=e]}$$

$$= \overline{R}_{1} \prod_{e} \left( 1+\sum_{o} P[O=o|E=e] \partial R_{1}(o,e) \right)^{P[E=e]}$$

$$\approx \overline{R}_{1} \left( 1+\sum_{e} P[E=e] \sum_{o} P[O=o|E=e] \partial R_{1}(o,e) \right)$$

$$= \overline{R}_{1} \left( 1+\sum_{e,o} P[E=e \cap O=o] \partial R_{1}(o,e) \right).$$
(A14)

The final line of this result indicates that the fitness of the reaction norm can be approximated by averaging the fitness effects on lifetime reproductive success over the joint distribution of environmental states and observation sequences. Accordingly, the environmental effect term in the expression for the future reproductive success (the final term on the second line of equation (A13)) averages out in the fitness function, because  $i(e_i)$  takes the value 1 in a fraction  $\alpha / (\alpha + \beta)$  of the cases, and the value 0 in the remaining ones. From here on, we will therefore omit this contribution to the reproductive success.

To accomplish the averaging over the joint distribution of E and O for the remaining contributions to  $\partial R_1$ , we use a stepwise procedure that is based on the recursive definition of fitness in equation (A13). First, we focus on an individual at age *t*, after it has sampled the state of the environment but before has adjusted its phenotype. The state of such an individual is summarized by two variables:  $x_{t-1}$  and  $p_t$ , which defines the sequence of observations made up to age *t* as well as the distribution of environmental states at time *t*. This follows because  $P[E_t = A | O_t = o_t] = p_t$  by definition.

The *expected* reproductive success of the focal individual from age *t* onwards (again, relative to the reference life history) is denoted by  $\partial \overline{R}_t(x_{t-1}, p_t)$ . Following equations (A13) and (A14), we define this function as

$$\partial \overline{R}_{t}(x_{t-1}, p_{t}) = \sum_{\substack{e \in \{A, B\}\\ o \in \{A, B\}}} P[E_{t} = e \cap O_{t+1} = o \mid \boldsymbol{O}_{t} = \boldsymbol{o}_{t}] \left( \frac{\overline{R}_{t+1} \overline{S}_{t}}{\overline{R}_{t}} \partial \overline{R}_{t+1}(x_{t}, p_{t+1}) + \partial S_{t}^{e}(x_{t}, \Delta x_{t}) + \frac{\overline{F}_{t} \overline{S}_{t}}{\overline{R}_{t}} \partial F_{t}^{e}(x_{t}, \Delta x_{t}) \right)$$
(A15)

such that the relative fitness difference between an individual with reaction norm h and the reference life history can be written as

$$\partial W = \frac{W - R_1}{\overline{R}_1} = \sum_{o \in \{A,B\}} P[O_1 = o] \partial \overline{R}_1(x_0, U(p_0, o)), \qquad (A16)$$

with  $p_0 = \alpha / (\alpha + \beta)$ , as explained in the main text.

As indicated by equation (A15), the expected future reproductive success in a given state  $(t, x_{t-1}, p_t)$  is a weighted average of the reproductive success along four possible life-history paths that emerge from the two possible states of the environment at time *t* and the two possible observations in the next time step (note that the realization of  $O_{t+1}$  determines the next estimate  $p_{t+1}$ ). Writing out  $x_t$  and  $p_{t+1}$  as specified by equation (A7), equation (A15) is expanded as

$$\partial \overline{R}_{t}(x_{t-1}, p_{t}) = \frac{\overline{R}_{t+1}\overline{S}_{t}}{\overline{R}_{t}} \sum_{o \in \{A,B\}} P[O_{t+1}=o \mid \boldsymbol{O}_{t}=\boldsymbol{o}_{t}] \partial \overline{R}_{t+1}(x_{t-1}+h_{t}(x_{t-1}, p_{t}), U(p_{t}, o)) + \sum_{e \in \{A,B\}} P[E_{t}=e \mid \boldsymbol{O}_{t}=\boldsymbol{o}_{t}] \quad \partial S_{t}^{e}(x_{t-1}+h_{t}(x_{t-1}, p_{t}), |h_{t}(x_{t-1}, p_{t})|) + \frac{\overline{R}_{t}\overline{S}_{t}}{\overline{R}_{t}} \sum_{e \in \{A,B\}} P[E_{t}=e \mid \boldsymbol{O}_{t}=\boldsymbol{o}_{t}] \quad \partial F_{t}^{e}(x_{t-1}+h_{t}(x_{t-1}, p_{t}), |h_{t}(x_{t-1}, p_{t})|).$$
(A17)

Expressions for the probabilities  $P[E_t = e | O_t = o_t]$  and  $P[O_{t+1} = o | O_t = o_t]$  that appear in this result, follow directly from the definition of  $p_t$  or from equations (A4) and (A5):

$$P[E_{t}=A | \boldsymbol{O}_{t}=\boldsymbol{o}_{t}] = p_{t}, \quad P[E_{t}=B | \boldsymbol{O}_{t}=\boldsymbol{o}_{t}] = 1 - p_{t},$$

$$P[O_{t+1}=A | \boldsymbol{O}_{t}=\boldsymbol{o}_{t}] = a P[E_{t+1}=A | \boldsymbol{O}_{t}=\boldsymbol{o}_{t}] + (1-a)(1-P[E_{t+1}=A | \boldsymbol{O}_{t}=\boldsymbol{o}_{t}])$$

$$= a((1-\beta) p_{t} + \alpha(1-p_{t})) + (1-a)(\beta p_{t} + (1-\alpha)(1-p_{t}))$$

$$= a + (1-2a)(\beta p_{t} + (1-\alpha)(1-p_{t})), \quad (A18)$$

$$P[O_{t+1}=B | \boldsymbol{O}_{t}=\boldsymbol{o}_{t}] = (1-a) P[E_{t+1}=A | \boldsymbol{O}_{t}=\boldsymbol{o}_{t}] + a(1-P[E_{t+1}=A | \boldsymbol{O}_{t}=\boldsymbol{o}_{t}])$$

$$= a(\beta p_{t} + (1-\alpha)(1-p_{t})) + (1-a)((1-\beta) p_{t} + \alpha(1-p_{t}))$$

$$= a + (1-2a)((1-\beta) p_{t} + \alpha(1-p_{t})).$$

Considering that  $\overline{R}_{t+1} \overline{S}_t / \overline{R}_t = 1 - \overline{F}_t \overline{S}_t / \overline{R}_t$ , equation (5) in the main text is a special case of equation (A17) for linear fecundity and survival functions,

$$F_t^{e_t}(x_t, \Delta x_t) = \overline{F}_t \left( 1 + f_t^{e_t}(x_t - z) - f_t' \Delta x_t + f_t'' (i(e_t) - \frac{\alpha}{\alpha + \beta}) \right) \quad \text{and}$$

$$S_t^{e_t}(x_t, \Delta x_t) = \overline{S}_t \left( 1 + s_t^{e_t}(x_t - z) - s_t' \Delta x_t + s_t'' (i(e_t) - \frac{\alpha}{\alpha + \beta}) \right), \quad (A19)$$

where the marginal costs of plasticity do not depend on the state of the environment.

## Calculation of the width of the reaction norm's plateau

The optimal reaction norm exhibits a plateau at intermediate values of  $p_t$  if the survival and fecundity functions are linear in  $x_t$  and  $\Delta x_t$ . The width of this plateau is easily calculated for t = T. Consider, for example, the semelparous life history introduced in the main text. At age *T*, the net benefit of plasticity is given by (equation (8) in the main text)

$$\partial \overline{R}_{T}(x_{T-1}, p_{T}) = -c \left| h_{T}(x_{T-1}, p_{T}) \right| + s \left( 2p_{T} - 1 \right) \left( x_{T-1} + h_{T}(x_{T-1}, p_{T}) - z \right).$$
(A20)

This function is piecewise linear and continuous, because of the absolute value in the cost part. For  $h_T < 0$ , the fitness function can thus be written as  $\partial \overline{R}_T(x_{T-1}, p_T) = a + b_- h_T(x_{T-1}, p_T)$  with  $b_- = c + s (2 p_T - 1)$ . Similarly, for  $h_T > 0$ , the slope of the fitness function is  $b_+ = -c + s (2 p_T - 1)$ .

The maximum of the piecewise linear fitness function can be located either at  $x_T = 1$ , at  $x_T = 0$  or at an intermediate phenotype value, where  $x_T = x_{T-1}$ . The latter option, which is realized on the plateau of the reaction norm, requires that  $b_-$  is positive and that  $b_+$  is negative. Hence, it follows that the optimal phenotypic adjustment is zero whenever the estimate  $p_t$  satisfies the conditions

$$c + s(2p_T - 1) > 0 > -c + s(2p_T - 1) \iff \frac{1}{2} - \frac{1}{2}c/s < p_T < \frac{1}{2} - \frac{1}{2}c/s$$
(A21)

Note that the condition for the existence of a plateau depends only on the forecasting probability, not on the current phenotype. Subtracting the lower from the upper threshold in the above condition gives the result that the width of the plateau is c / s.

#### Nonlinear fitness functions

In this final section of the supplementary material, we concentrate once more on semelparous life histories, and briefly examine the robustness of the results with respect to nonlinearities in the fitness function. Such nonlinearities may arise, for example, if the combined cost of two (or more) smaller phenotypic adjustments  $x \rightarrow y \rightarrow z$  (with y intermediate between x and z) is different from the cost of an equivalent adjustment  $x \rightarrow z$  in a single step. Alternatively, selection on the phenotype in either one or both of the environments may be stabilizing rather than directional, generating nonlinear benefits of expressing an adaptively adjusted phenotype.

Figure A1 presents two cases where nonlinearities have a pronounced effect on the realized phenotype sequences. The first occurs when the cost of plasticity increases disproportionally with the amount of phenotypic adjustment. Even if the associated nonlinearity is weak, such supermultiplicative costs have several clear effects on the optimal reaction norm (figure A1*a*, grey phenotype tree in the middle panel; figure A2*a*): the optimal initial phenotype is at an intermediate value, and larger phenotypic changes are broken down into a series of smaller adjustments (individuals change their phenotype each timestep, sometimes in a direction that is inconsistent with their last observation). As a result, peaks in the life history schedule of plasticity (figure A1*a*; right) are smoothed out over several age classes. As an aside, we note that the optimal reaction norm for a multiplicative fitness function (figure A1*a*, left) is indistinguishable from the result obtained for linear fitness functions as in equation (7). This is to be expected under weak selection, as the two types of fitness functions are identical up to first order in *s* and *c*.

Figure A1*b* compares the phenotype trees calculated for two fitness functions with diminishing benefits of expressing an adjusted phenotype, one with multiplicative (black) and the other with super-multiplicative (grey) costs of plasticity. In this case, the realized phenotype distribution no longer peaks at the extreme values, because the cost of phenotypic adjustments is at some point no longer compensated by the benefit of expressing a more perfectly adapted phenotype. When the costs of plasticity are multiplicative (figure A1*b*; left), individuals sometimes do not change their phenotype, indicating that the reaction norm has a plateau at intermediate values of  $p_t$ , as for linear fitness functions (figure A2*b*).



Figure A1. Dependence of optimal plasticity patterns on nonlinearities in the fitness function.

Trees of phenotypes resulting from the optimal reaction norm for a semelparous life history with nonlinear survival functions  $S_t^{B}(x_t, \Delta x_t) = S_t^{A}(1-x_t, \Delta x_t) = \exp(-2^{\eta-1}s x_t^{\eta} - 2^{\theta-1}c \Delta x_t^{\theta})$ . (a) Comparison between a multiplicative fitness function ( $\eta = \theta = 1$ ; left column) and one with super-multiplicative costs of plasticity ( $\theta = 1.2$ ; right column); in the latter case, large phenotypic adjustments are disproportionally costly. (b) Comparison between fitness functions with diminishing benefits of expressing an adapted phenotype ( $\eta = 1.2$ ) with either multiplicative ( $\theta = 1$ ; left column) or super-multiplicative ( $\theta = 1.2$ ; left column) costs of plasticity. The right column shows the average absolute phenotypic adjustment, with black bars corresponding to the left column and grey bars to the middle column. Parameters: T = 8,  $\alpha = 0.16$ ,  $\beta = 0.14$ , a = 0.8, s = 0.03, and c = 0.05.



Figure A2. Optimal reaction norms for nonlinear fitness functions.

Nonlinearities in the fitness function introduce a dependence of the phenotype  $x_t$  after adjustment on the original phenotype  $x_{t-1}$ . The colored nearly opaque surfaces show optimal reaction norms at age 1 for the nonlinear survival functions used in figure A1. In each panel, the optimal reaction norm for a multiplicative fitness function with  $\eta = \theta = 1$  is shown as a colored nearly transparent surface. Contours are drawn at the levels  $\pm 0.8$ ,  $\pm 0.6$ ,  $\pm 0.4$ ,  $\pm 0.2$ , and  $\pm 0.001$ . Parameters: (a)  $\eta = 1$ ,  $\theta = 1.2$ ; (b)  $\eta = 1.2$ ,  $\theta = 1$ ; (c)  $\eta = \theta = 1.2$ .

t	${\pmb{\varphi}_t}^{\mathrm{b}}$	$\sigma_t^{c}$	$R_t^{\rm d}$	t	$\varphi_t$	$\sigma_t$	$R_t$
1	0.000	1.000	1.000	7	0.153	0.919	0.330
2	0.057	0.977	1.000	8	0.103	1.000	0.205
3	0.241	1.000	0.966	9	0.047	0.970	0.102
4	0.208	0.976	0.725	10	0.046	0.849	0.058
5	0.117	1.000	0.535	11	0.030	0.750	0.023
6	0.134	0.903	0.418	12	0.000	0.619	0.000

# Table A1 – Life-history parameters for the iteroparous life history in figure 6<sup>a</sup>

<sup>a</sup> Parameters are based on published estimates for the estuarine polychaete *Streblospio benedicti* (Levin et al. 1996; figure 5).

<sup>b</sup> Fecundity, normalized such that the expected lifetime reproductive success equals 1.

<sup>c</sup> Survival probability from age t - 1 to age t.

<sup>d</sup> Residual reproductive success at age *t*, calculated from the other two columns by means of the recursion  $R_t = \sigma_t (\varphi_t + R_{t+1})$ .