

# Hedging your evolutionary bets

Richard McElreath\*,<sup>1</sup>

<sup>1</sup>Department of Anthropology and Graduate Groups in Ecology, Population Biology and Animal Behavior, University of California, Davis CA 95616, USA

August 27, 2009

## Contents

<b>1</b>	<b>Introduction</b>	<b>2</b>
<b>2</b>	<b>Problem</b>	<b>2</b>
<b>3</b>	<b>Germination in an annual plant</b>	<b>2</b>
3.1	When the environment is static . . . . .	3
3.2	When the environment varies from year to year . . . . .	5
3.3	Geometric mean fitness . . . . .	6
3.4	Calculus bother . . . . .	9
3.5	Back to biology . . . . .	10
3.6	When dry years are very dry . . . . .	11
3.7	Predictions . . . . .	11
3.8	Extending the model: Selection on seed longevity . . . . .	11
<b>4</b>	<b>Diapause</b>	<b>13</b>
4.1	Diapause in a static but seasonal environment . . . . .	13
4.2	When Winter conditions vary . . . . .	14

---

\*mcelreath@ucdavis.edu

<b>5</b>	<b>The timing of reproduction</b>	<b>16</b>
5.1	Flowering time in an unpredictable environment . . . . .	17
5.2	Polination . . . . .	18

## 1 Introduction

## 2 Problem

Your paper will address the evolution of delayed germination in desert annual plants. These plants live in extremely challenging environments with light and unpredictable rainfall. But since these plants do not live for more than one year—that’s why they are called annuals—they cannot very well cope with this unpredictability by waiting for rainfall to make seeds. Instead, they must program their seeds themselves to cope with the unpredictable weather.

**When does selection favor always germinating, according to the model?**

**How does the rate of seed decay,  $d$ , affect the equilibrium germination rate, according to the model?**

**How does the number of seeds produced in wet years,  $y_w$ , affect the equilibrium germination rate, according to the model?**

**When is the plant population viable?**

**BONUS: If plants can invest in seed size, in order to make them decay more slowly, what is the joint equilibrium of  $d$  and  $g$ ?**

## 3 Germination in an annual plant

One of the best known models of this kind of problem is Cohen’s (1966) model of germination rate of the seeds of an annual plant. Annuals live for only one year (or less), and so continuation of their lineage depends upon maintaining seeds from year to year. Seeds can remain dormant, but dormant seeds die

or are eaten at some rate each year. Let this rate of death and predation be  $d$ , for *decay*. You might think then that the savvy plant will program all of its seeds to germinate right away, because otherwise they risk death. You'd be right, at least until we add some realistic complications to the model. But first, let's prove your intuition.

### 3.1 When the environment is static

Suppose that the environment from year to year is always the same, such that each plant produces  $y$  seeds before it withers and dies ( $y$  for *yield*). Further suppose that a parent plant can program its seeds such that each germinates with probability  $g$  (for *germinate*) at the start of each yearly cycle. This implies that the number of seeds alive from a given lineage in a given year  $t$ ,  $s_t$ , is:

$$s_t = s_{t-1} - gs_{t-1} - (1 - g)ds_{t-1} + gys_{t-1}. \quad (1)$$

Let's explain this expression. Moving from left to right, on the right-hand side, the first term is just the number of seeds that were alive in the previous year, year  $t - 1$ . From this number, we next subtract those seeds that germinate. Since there were  $s_{t-1}$  seeds from last year, and a fraction  $g$  germinate, we expect  $g \times s_{t-1}$  seeds to germinate this year. Next, of those seeds that do not germinate, a fraction  $d$  of them decay and also get removed from the seed stock, so we subtract a fraction  $(1 - g)d$  of the  $s_{t-1}$  seeds from the previous year. Finally, those seeds that do germinate produce new seeds, so the  $gs_{t-1}$  germinating seeds each produce  $y$  new seeds, added  $gys_{t-1}$  seeds to the lineage. This completes Expression (1) above.

Simplifying (1), we get:

$$s_t = s_{t-1}((1 - g)(1 - d) + gy). \quad (2)$$

This implies that the number of seeds grows (or shrinks) by a factor  $(1 - g)(1 - d) + gy$  each year. After 1 year, for example, the number of seeds is expected to be:

$$s_1 = s_0((1 - g)(1 - d) + gy),$$

where  $s_0$  is just the number of seeds in year zero, the year we started counting. Then after one more year:

$$s_2 = s_1((1 - g)(1 - d) + gy).$$

Substitute the expression for  $s_1$  into the expression above for  $s_2$ , then simplify:

$$\begin{aligned} s_2 &= s_0((1-g)(1-d) + gy)((1-g)(1-d) + gy), \\ &= s_0((1-g)(1-d) + gy)^2. \end{aligned}$$

You can probably see now, by induction, that after any number of years  $n$ , the expected number of seeds will be:

$$s_n = s_0((1-g)(1-d) + gy)^n. \quad (3)$$

We want to find the value of  $g$  that natural selection will favor in the long run. We suppose now that there are many loci that affect the value of  $g$ , so that many values between 0 and 1 are possible heritable strategies. Plants with different values of  $g$  will produce lineages of seeds with different values of  $g$ , and those lineages with the higher long-term growth rates will eventually dominate the population.

A bit later, we'll solve this problem in a more general way, but for now you can probably see by inspection that we just need to set  $g$  to maximize  $(1-g)(1-d) + gy$ . Since the growth rate of each lineage is the same in every year, the growth rate that maximizes growth in any year will maximize it in the long run (this will not be true in the next section!). So to find the value of  $g$  favored by natural selection in the long run, we just find which value of  $g$  makes  $(1-g)(1-d) + gy$  biggest. It is conventional to denote the *evolutionarily stable* value of a continuous strategy like  $g$  by adding a star. So in this case, we are seeking the long run successful value of  $g$ ,  $g^*$ , that will out-compete all other values of  $g$ .

You could muck around with some calculus now, trying to find the maximum of this expression as a function of  $g$ . But it's easier in this case to think for a moment. The expression  $(1-g)(1-d) + gy$  is the equation for a line, if you imagine  $g$  as the variable on the horizontal axis ( $x$ -axis). Re-arranging the expression in this way, we get:

$$g(d + y - 1) + 1 - d.$$

Since  $d$  is a probability ( $0 < d < 1$ ) and  $y > 1$  (if the plant has any hope of persisting), the term  $d + y - 1$  is always positive. This means the slope of this line is increasing with  $g$  at a constant rate  $d + y - 1$ . Therefore the value of  $g$  that maximizes this expression must in fact be the maximum

value of  $g$ ,  $g^* = 1$ . Natural selection favors always programming your seeds to germinate, in this static environment, just as we intuited at the start of this section.

You might think this exercise was pretty useless, if it just confirmed your intuition. But often our intuitions are wrong, so confirming intuition is one of the important functions of mathematical biology. Also, going through this exercise has introduced the mathematics that we'll complicate in the next section. So it's really been worth the bother, believe me.

### 3.2 When the environment varies from year to year

Now suppose that not every year is the same. Some years, the weather is particularly dry, and our annual plant will have a hard time reproducing, producing fewer seeds than in wetter years. Specifically, assume that there are two types of years, wet and dry, and that each occurs with probability  $p$  and  $1 - p$ , respectively. Reproduction in each kind of year is  $y_W$  and  $y_D$ , respectively.

All of this means we have to back up and re-write the growth rate of a lineage. In any given year  $t$ , the weather is either wet or dry, but never both. For the moment, just let the number of seeds produced be  $y_t$ , which can be either  $y_W$  or  $y_D$ , depending upon the year. We'll figure out which later. For now, we just have to note that growth of a series of  $n$  years will be:

$$s_n = s_0 \prod_{t=1}^n ((1 - g)(1 - d) + gy_t). \quad (4)$$

Note that the order that the different wet and dry weather conditions occur in is irrelevant—the growth is calculated as the product of each annual growth. Population growth is a *multiplicative*, not an *additive*, process.

Now to simplify (4) for our specific model, with the dry/wet weather conditions. In any particular sequence of  $n$  years, there will be  $n_W$  wet years and  $n_D$  dry years ( $n_W + n_D = n$ ). Leave these as variables for the moment. This gives us:

$$s_n = s_0 \times ((1 - g)(1 - d) + gy_D)^{n_D} \times ((1 - g)(1 - d) + gy_W)^{n_W}. \quad (5)$$

This is just to say that the total growth over the  $n$  years is the product of  $n_D$  years of growth under dry conditions and  $n_W$  years of growth under wet conditions.

Now again, we want to find the value of  $g$  that maximizes long-term growth of the lineage. In the previous section, we could just see from inspection that the value of  $g$  that maximized growth in any particular year would also maximize growth in every other year. But that is no longer true here. Consider for example the previously optimal strategy,  $g^* = 1$ , and let  $y_D = 0$ , just to make the example stark and easy to grasp. This means we are assuming that dry years in fact are so harsh that they kill off all plants that try to grow. Only seeds remain. Now compute the growth rate of a lineage with  $g = 1$ :

$$\begin{aligned} s_n &= s_0 \times ((1 - [1])(1 - d) + [1][0])^{n_D} \times ((1 - [1])(1 - d) + [1]y_W)^{n_W}, \\ &= s_0 \times (0)^{n_D} \times (y_W)^{n_W}. \end{aligned}$$

So if there are any dry years,  $n_D > 0$ , then  $s_n = 0$ . Even one zero is enough to make the entire product zero (of course). So in the long run, a lineage that programs all of its seeds to germinate every year will necessarily go extinct. This example is pretty extreme, but even if  $y_D > 0$ , the same logic can favor smaller values of  $g$ . You'll prove that next.

### 3.3 Geometric mean fitness

What we've stumbled upon here is a basic fact about population growth and natural selection: average fitness through time is not what natural selection maximizes. The reason is that population growth is multiplicative, such that if the population ever experiences a crash in numbers, it will take it some time to recover. Thus natural selection is influenced by both mean fitness and the variance in fitness, through time.

Let's take the growth rate over  $n$  years, Expression (5), and extrapolate it out to the truly long-run growth rate. What we actually want to calculate is the expected annual growth rate. The strategy  $g^*$  that maximizes this value will be the one natural selection favors. But now that growth varies from year to year, we'll need a slightly messier expression. Here are two ways to arrive at the result: first, a somewhat dodgy intuitive approach, and second, a more rigorous approach that gets us to the conventional approach in theoretical evolutionary ecology.

First, suppose  $n$  is a very large number indeed, like 10-thousand years. Over such a long period of time, the realized numbers of wet and dry years will come very close to the probabilities of each kind of year,  $p$  and  $1 - p$ .

This implies that the total growth is:

$$s_n = s_0 \times ((1 - g)(1 - d) + gy_D)^{n(1-p)} \times ((1 - g)(1 - d) + gy_W)^{np}.$$

For a particular lineage (creatively called “1”) with  $g = g_1$ :

$$s_{1,n} = s_0 \times ((1 - g_1)(1 - d) + g_1y_D)^{n(1-p)} \times ((1 - g_1)(1 - d) + g_1y_W)^{np}.$$

Comparing this to another lineage, “2,” with  $g = g_2$ , the lineage 1 out-competes lineage 2 if  $s_{1,n} > s_{2,n}$ , which implies:

$$\begin{aligned} & ((1 - g_1)(1 - d) + g_1y_D)^{n(1-p)} \times ((1 - g_1)(1 - d) + g_1y_W)^{np} > \\ & ((1 - g_2)(1 - d) + g_2y_D)^{n(1-p)} \times ((1 - g_2)(1 - d) + g_2y_W)^{np}. \end{aligned}$$

Taking the  $n$ -th root of both sides, to get rid of the number of years  $n$ , we have:

$$\begin{aligned} & ((1 - g_1)(1 - d) + g_1y_D)^{(1-p)} \times ((1 - g_1)(1 - d) + g_1y_W)^p > \\ & ((1 - g_2)(1 - d) + g_2y_D)^{(1-p)} \times ((1 - g_2)(1 - d) + g_2y_W)^p. \end{aligned}$$

So now you can see that the lineage with a value of  $g = g^*$  that maximizes the expression:

$$((1 - g^*)(1 - d) + g^*y_D)^{(1-p)} \times ((1 - g^*)(1 - d) + g^*y_W)^p \quad (6)$$

will out-compete all other lineages.

So now you could do some calculus to find the value  $g^*$ , but we’ll do that in a bit. For now, notice that this expression, the long-term average growth rate over all the years  $n$ , is not an *arithmetic* mean, but instead a *geometric* mean. The geometric mean of a series of  $n$  numbers  $x_1, x_2, \dots, x_n$  is defined as  $(\prod_i x_i)^{1/n}$ . That is, it is the  $n$ -th root of the product of the  $n$  numbers. So what we have shown above is that the expected annual growth of the lineage is the geometric mean of growth rates in the different environments.

The second, more rigorous, way to motivate the same conclusion is to take the limit of the expected annual growth rate as  $n$  approaches infinity. Here’s how this is traditionally done. First, take the logarithm of  $s_n$ :

$$\log[s_n] = \log[s_0 \cdot ((1 - g^*)(1 - d) + g^*y_D)^{n_D} \cdot ((1 - g^*)(1 - d) + g^*y_W)^{n_W}].$$

Using the logarithm rules you learned in high-school:

$$\begin{aligned}\log[s_n] &= \log[s_0] + n_D \log [(1 - g^*)(1 - d) + g^*y_D] \\ &\quad + n_W \log [(1 - g^*)(1 - d) + g^*y_W].\end{aligned}$$

Now divide both sides by  $n$ , to get the average (this is the same as taking the  $n$ -th root before):

$$\begin{aligned}\frac{\log[s_n]}{n} &= \frac{\log[s_0]}{n} + \frac{n_D}{n} \log [(1 - g^*)(1 - d) + g^*y_D] \\ &\quad + \frac{n_W}{n} \log [(1 - g^*)(1 - d) + g^*y_W].\end{aligned}$$

Now we take the limit as  $n \rightarrow \infty$ , because we want to exact long-term average growth rate, not just one that is approximate for a large  $n$ . The first term on the right goes to zero:

$$\lim_{n \rightarrow \infty} \frac{\log[s_0]}{n} = 0.$$

The factors  $n_W/n$  and  $n_D/n$  converge to  $p$  and  $1 - p$ , respectively, because as  $n$  gets very large, these proportions become identical to the probabilities of each environment. All together, we end with:

$$\begin{aligned}\lim_{n \rightarrow \infty} \frac{\log[s_n]}{n} &= (1 - p) \log [(1 - g^*)(1 - d) + g^*y_D] \\ &\quad + p \log [(1 - g^*)(1 - d) + g^*y_W].\end{aligned}$$

And this is just the logarithm of Expression (6). So finding the value  $g^*$  that maximizes the above is the same as finding the value of  $g^*$  that maximizes (6)—it is just usually easier to work with logarithms, so this is how the theory is usually presented.

In general, for any number of environmental states, each of which is indexed by  $i$ , the average fitness is given by:

$$\lim_{n \rightarrow \infty} \frac{\log[s_n]}{n} = \sum_i p_i \log [(1 - g^*)(1 - d) + g^*y_i],$$

where  $y_i$  is the number of seeds produced by a plant in a year that the environment is in state  $i$ ,  $p_i$  is the probability the environment takes state  $i$  in any given year, and  $\sum_i p_i = 1$ . As you might expect, the above is exactly the logarithm of the geometric mean fitness. The geometric mean fitness itself is simply  $\prod_i ((1 - g^*)(1 - d) + g^*y_i)^{p_i}$ .



### 3.4 Calculus bother

Okay, so up to this point we have the mean fitness:

$$\lim_{n \rightarrow \infty} \frac{\log[s_n]}{n} = (1-p) \log [(1-g^*)(1-d) + g^*y_D] \\ + p \log [(1-g^*)(1-d) + g^*y_W],$$

and we need to find the value  $g^*$  that maximizes this expression. We proceed by taking the derivative of the above, with respect to  $g^*$ . If you haven't done calculus in some time, this might look a bit daunting. But the only trick is to use the chain rule and to know that  $d \log[x]/dx = 1/x$ . I'll do the second term on the right-hand side, and you can repeat the steps for the other half and add the two derivatives together (because the derivative of a sum is equal to the sum of the derivatives).

We want to compute:

$$\frac{d}{dg^*} p \log [(1-g^*)(1-d) + g^*y_W] = p \frac{d}{dg^*} \log [(1-g^*)(1-d) + g^*y_W].$$

Now, according to the chain rule:

$$\frac{d}{dg^*} \log [(1-g^*)(1-d) + g^*y_W] = \frac{d}{dg^*} \log [X] = \frac{d \log[X]}{dX} \frac{dX}{dg^*}.$$

The  $X = (1-g^*)(1-d) + g^*y_W$  is just to shorten the notation. This gives us:

$$\frac{d}{dg^*} \log [X] = \frac{1}{X} \frac{dX}{dg^*} = \frac{1}{X} (d-1+y_W) = \frac{d-1+y_W}{(1-g^*)(1-d) + g^*y_W}.$$

So the answer is:

$$\frac{d}{dg^*} p \log [(1-g^*)(1-d) + g^*y_W] = p \frac{d-1+y_W}{(1-g^*)(1-d) + g^*y_W}.$$

Using the same procedure on the other half of the expression, the total derivative is:

$$\frac{d}{dg^*} \lim_{n \rightarrow \infty} \frac{\log[s_n]}{n} = \frac{p(d-1+y_W)}{(1-g^*)(1-d) + g^*y_W} + \frac{(1-p)(d-1+y_D)}{(1-g^*)(1-d) + g^*y_D}.$$

### 3.5 Back to biology

Now that the calculus is out of the way, we can solve for the value  $g^*$  that maximizes fitness.

There are two interesting biological situations. First, are there any conditions under which  $g^* = 1$ , as it did in the static environment? To answer this question, ask when fitness is increasing at  $g = 1$ :

$$\left. \frac{d}{dg^*} \lim_{n \rightarrow \infty} \frac{\log[s_n]}{n} \right|_{g^*=1} > 0 \implies \frac{(1-p)(d-1+y_D)}{y_D} + \frac{p(d-1+y_W)}{y_W} > 0.$$

This requires that:

$$p > \frac{y_W(1-d-y_D)}{(1-d)(y_W-y_D)}. \quad (7)$$

So the more likely the wet weather, the more likely selection will favor always germinating. We can further ask when the above will always be satisfied, by asking when the right-hand side will be less than zero. Since  $p$  must be between zero and one, if the right-hand side above is negative, then selection always favors 100% germination, regardless of the value  $p$ . By inspecting the top, this requires:

$$y_D > 1 - d.$$

When reproduction under the bad weather is sufficiently high, selection again always favors germination.

The second case is that  $g^* < 1$ . We know that  $g^* > 0$ , always, because if the lineage never germinates any of its seeds, eventually all the seeds will die in the ground and the lineage will go extinct. So if Condition (7) for  $g^* = 1$  does not hold, the optimal germination rate must be greater than zero but less than one. This is found where:

$$\frac{d}{dg^*} \lim_{n \rightarrow \infty} \frac{\log[s_n]}{n} = 0.$$

After some high-school algebra:

$$g^* = \frac{(1-d)(py_W + (1-p)y_D + d - 1)}{(1-d-y_D)(y_W + d - 1)}. \quad (8)$$

Because  $y_D < 1 - d$  (we showed above that this is necessary for  $g^* < 1$ ), the above is positive.

### 3.6 When dry years are very dry

To see the impact of all this math, consider the extreme case in which  $y_D = 0$ . This gives us:

$$g^* = \frac{py_W + d - 1}{y_W + d - 1} = 1 - \frac{(1 - p)y_W}{y_W + d - 1}.$$

Consider the effect of each parameter.

As  $p$  increases,  $g^*$  approaches 1. This is because as the wet condition becomes increasingly common, the risk of dry-time disaster is reduced. Postponing germination is a strategy for weathering (pun intended) bad weather, so as bad weather becomes rarer, germination becomes more frequent. In the limit of  $p = 1$ , it is again best to always germinate.

As  $d$  increases,  $g^*$  also increases. This is because as the death rate of waiting seeds goes up, selection will favor less waiting. In the theoretical limit of  $d = 1$  ( $d$  cannot actually be one, because then of course selection would never favor postponing germination),  $g^* = p$ . As  $y_W \rightarrow \infty$ ,  $g^* \rightarrow p$ , as well.

### 3.7 Predictions

So if this model tells us much about reality, what should we expect?

First, the model predicts that some seeds will not germinate in their first year, even under good conditions. However, these same seeds may germinate in later years, under the same weather conditions. This is because germination is randomized. Second, if sites vary in the frequency of wet/dry conditions, then germination rates should correlate with  $p$ . Third, each parent plant should produce seeds that germinate in different years, owing to the randomization of germination.

Using data from a desert annual plant, like many of those found in the American southwest, should shed some light on these predictions.

### 3.8 Extending the model: Selection on seed longevity

The only parameter we have considered to be under the influence of natural selection so far is  $g$ , the rate of germination. Now suppose that the longevity of seeds  $d$  is also under partial genetic control. Plants can sacrifice some fertility to make larger and more durable seeds that can survive longer before

germinating. Specifically, assume that seed longevity is given by  $d$  still, but that fertility is adjusted to  $y_t(1 - c(1 - d))$ . That is, yield in any year  $t$  is reduced by a fraction  $c(1 - d)$ , where  $c$  is a parameter that determines the rate at which investment in seed longevity reduces seed number. For example, if  $c = 1/2$ , then if seeds are made immortal ( $d = 0$ ), yield will be halved. If  $c = 0$ , then there is no tradeoff and this model reduces to the previous model.

We wish to find the joint evolutionary dynamics of  $g$  and  $d$ . Again, we write a geometric mean fitness:

$$\lim_{n \rightarrow \infty} \frac{\log[s_n]}{n} = p \log[(1 - g)(1 - d) + gy_W(1 - c(1 - d))] + (1 - p) \log[(1 - g)(1 - d) + gy_D(1 - c(1 - d))].$$

For simplicity, we will assume  $y_D = 0$  as in the extreme case before. So we will work with:

$$\lim_{n \rightarrow \infty} \frac{\log[s_n]}{n} = p \log[(1 - g)(1 - d) + gy_W(1 - c(1 - d))] + (1 - p) \log[(1 - g)(1 - d)].$$

Now we want two derivatives. First, with respect to  $g$ , as before:

$$\frac{\partial}{\partial g} \lim_{n \rightarrow \infty} \frac{\log[s_n]}{n} = \frac{p(d - 1 + (1 - c(1 - d))y_W)}{(1 - d)(1 - g) + (1 - c(1 - d))gy_W} - \frac{(1 - p)(1 - d)}{(1 - d)(1 - g)}.$$

The second with respect to  $d$ :

$$\frac{\partial}{\partial d} \lim_{n \rightarrow \infty} \frac{\log[s_n]}{n} = \frac{p(g - 1 + cgy_W)}{(1 - d)(1 - g) + (1 - c(1 - d))gy_W} - \frac{(1 - p)(1 - g)}{(1 - d)(1 - g)}.$$

Solving for when both are equal to zero results in messy roots that are hard to interpret. But we can recover all the insight in this case by taking the limit of the above derivatives, as  $y_W \rightarrow \infty$ :

$$\begin{aligned} \lim_{y_W \rightarrow \infty} \frac{\partial}{\partial g} \lim_{n \rightarrow \infty} \frac{\log[s_n]}{n} &= \frac{p - g}{(1 - g)g}, \\ \lim_{y_W \rightarrow \infty} \frac{\partial}{\partial d} \lim_{n \rightarrow \infty} \frac{\log[s_n]}{n} &= \frac{p - 1 + c(1 - d)}{(1 - c(1 - d))(1 - d)}. \end{aligned}$$

Solving for when these limits are both equal to zero tells us where natural selection comes to rest:

$$g^* = p,$$

$$d^* = \frac{c + p - 1}{c} = 1 - \frac{1 - p}{c}.$$

When favorable (wet) weather is common, germination is common, as before. However, wet weather also reduces investment in durable seeds, because there is little point in making tough long-lived seeds if they will germinate right away. Obviously, the more costly tough seeds are, the less selection favors investing in tough seeds. So as  $c$  increases,  $d^*$  decreases. When  $c = 1$ , for example,  $d^* = g^* = p$ , and both evolutionary parameters converge to the long-run frequency of wet weather. If  $c = 0$ , then investment in seed toughness is essentially free, and selection will favor the highest value of  $d$  possible.

So now we should predict that the rate at which seeds die before germination will be correlated with their rates of germination. The real value of  $c$  will affect the strength of this correlation, but there should be a detectable correlation, unless  $c$  is very small.

## 4 Diapause

Now consider an insect, instead of a plant, that can reproduce twice in each year. In the Spring, the insect mates and lays eggs. These eggs hatch into larvae during the Summer months. In the Fall, larvae can either (1) develop into adults and mate or (2) enter a state of hibernation known as *diapause* and wait until the next Spring to emerge as adults. Eggs laid in the Fall develop into larvae that grow during the Winter, emerging as adults in the Spring.

### 4.1 Diapause in a static but seasonal environment

Let's define the rates of transition among these states. Suppose that each adult in the Spring produces  $f_1$  eggs. (The subscript 1 indicates a parameter for the first part of the year.) Each of these eggs has a probability  $s_1$  of surviving through the Summer. Each adult who emerges in the Fall produces  $f_2 < f_1$  eggs, and each of these eggs has a chance  $s_2 < s_1$  of surviving through

the Winter. Larvae who instead enter diapause in the Fall have a chance  $s_d > s_2$  of surviving through diapause and emerging as adults in the next Spring.

With these assumptions, let's model the biology. We'll consider whether natural selection favors diapause or not. So let  $D$  be the probability a Fall larva enters diapause. This let's us write a recurrence equation for the number of Spring adults in year  $t + 1$ :

$$\begin{aligned} a_{t+1} &= (1 - D)f_1s_1f_2s_2a_t + Ds_1f_1s_da_t, \\ &= a_tf_1s_1((1 - D)f_2s_2 + Ds_d). \end{aligned} \tag{9}$$

So again, we can write the expression for the number of adults after  $n$  years:

$$a_n = a_0[f_1s_1((1 - D)f_2s_2 + Ds_d)]^n.$$

To find out whether natural selection favors diapause in the Fall, we ask when increasing  $D$  increases  $(1 - D)f_2s_2 + Ds_d$ . Again factoring into the form of a line, we want to know when the slope of  $D(s_d - f_2s_2) + f_2s_2$  is positive. That will be whenever  $s_d > f_2s_2$ . So if the rate at which diapaused larvae survive the Winter is greater than the expected number of offspring these larvae could instead produce during the winter, selection favors diapause. Seems obvious in hind-sight, perhaps.

Another way of arranging the condition is  $f_2 < s_d/s_2$ . So if the fecundity of Fall adults is less than the ratio of Winter diapause survival and Winter active larvae survival, selection favors diapause. Consider the limiting case, for example, when  $s_d = s_2$ . Then Fall fecundity must be less than 1, for selection to favor diapause. This means that if Fall reproduction doesn't even meet replacement levels, it is better to diapause. Now as  $s_d$  increases relative to  $s_2$ , this condition is easier to satisfy. Suppose larvae in diapause are twice as likely to survive the Winter as active larvae. Then the condition is  $f_2 < 2$ . If active larvae are very unlikely to survive the Winter, because of freezing conditions for example, then as  $s_2 \rightarrow 0$ , the condition approaches  $f_2 < \infty$ , which is pretty easy to satisfy, I think you'll agree.

## 4.2 When Winter conditions vary

But there is no bet hedging above. To re-introduce obvious bet hedging, we need to assume that some Winters are worse than others, but that this is largely unpredictable to the insects. Specifically assume that harsh blizzards

occur with probability  $p$  each Winter. During a blizzard, all active larvae die. Larvae in diapause survive a blizzard with probability  $b < s_d$ . In non-blizzard years, survival is still  $s_2$  for active larvae and  $s_d > s_2$  for diapause larvae.

Now we have:

$$a_{t+1} = a_t f_1 s_1 [(1-p)((1-D)f_2 s_2 + D s_d) + p D b]. \quad (10)$$

Over  $n$  years with  $n_b$  blizzard and  $n_0$  non-blizzard years:

$$a_n = a_0 [f_1 s_1 ((1-D)f_2 s_2 + D s_d)]^{n_0} [f_1 s_1 D b]^{n_b}. \quad (11)$$

So using the same logic as before, average fitness is given by:

$$\begin{aligned} \lim_{n \rightarrow \infty} \frac{\log[a_n]}{n} &= (1-p) \log[f_1 s_1 ((1-D)f_2 s_2 + D s_d)] + p \log[f_1 s_1 D b], \\ &= \log[f_1 s_1] + (1-p) \log[(1-D)f_2 s_2 + D s_d] + p \log[D b]. \end{aligned} \quad (12)$$

We maximize the above as a function of  $D$ , to find the value  $D^*$  that natural selection will favor in the long run. Consider when selection favors our previous result,  $D^* = 1$ . This requires:

$$\begin{aligned} \frac{d}{dD} \lim_{n \rightarrow \infty} \frac{\log[a_n]}{n} \Big|_{D=1} &> 0, \\ p + (1-p) \frac{s_d - f_2 s_2}{s_d} &> 0, \\ s_d &> (1-p) f_2 s_2. \end{aligned}$$

When this condition is met, selection always favors diapause. Notice that as  $p$  increases and blizzards become more common, this is increasingly easy to satisfy. When  $p = 1$ , it is in fact always satisfied, for any value  $s_d > 0$ . In more moderate cases, the fact of blizzards increases the range of conditions that favor always entering diapause—this is a bet hedging effect already. Suppose for example that  $s_d = s_2$ . Before we found that selection will favor diapause (in non-blizzard years) when  $s_d/s_2 > f_2$ . For our example, this requires  $1 > f_2$ . Even when this is not satisfied, it might be easy to satisfy  $s_d/s_2 > (1-p)f_2 \implies 1 > (1-p)f_2$ , the condition we just derived, when  $p$  is large and blizzards are common. Thus unpredictable blizzards can lead selection to favor always playing it safe and entering diapause, even though

if insects could tell when the blizzards were coming, they would be better off staying active in non-blizzard years.

But suppose the above is not satisfied, so that selection favors sometimes entering diapause, but not always ( $D^* < 1$ ). This implies selection will have to hedge its bets across offspring, because insects can't tell when a blizzard year will hit. So we assume  $s_d < (1 - p)f_2s_2$  and solve:

$$\frac{d}{dD} \lim_{n \rightarrow \infty} \frac{\log[a_n]}{n} = \frac{(1 - p)(s_d - f_2s_2)}{(1 - D)f_2s_2 + Ds_d} + \frac{p}{D} = 0.$$

for  $D$ . This yields:

$$D = D^* = \frac{pf_2s_2}{f_2s_2 - s_d}.$$

Notice that the solution does not depend upon the value of  $b$ ! The mathematical reason is that  $b$  vanishes from  $d \log(Db)/dD = d \log(D)/dD + d \log(b)/dD = 1/D$ . What is the biological reason that the rate of change in fitness as  $D$  changes is independent of  $b$ ?

Most of the action here lies in  $p$ . As  $p$  increases, the chance of entering diapause increases, obviously. But consider the limiting case where  $s_d = s_2$  and  $f_2$  is very large. Then the above simplifies to approximately  $p$ . Selection will favor a chance of entering diapause that is proportional to the chance of a blizzard. This is directly analogous to the result we found for delayed germination.

## 5 The timing of reproduction

For a final example of the concept, consider another annual plant that lives in a highly seasonal environment. The problem we will examine now is how long the plant should spend growing, before flowering and reproducing. Plants which spend more time growing can potentially produce more seeds and have higher reproductive success. But if the timing of the onset of bad weather is early, then such plants might have zero reproductive success. Conversely, plants which spend a short time growing are sure to reproduce some, but they will not produce very many or very large flowers, and therefore will have lower reproductive success than a larger plant would.



## 5.1 Flowering time in an unpredictable environment

Specifically, assume that an annual plant must reproduce before first frost, or it will die and lose all of its seeds. Each year, first frost comes on an expected (late) date with probability  $p$  but occurs earlier sometimes with probability  $1 - p$ . Call these dates late and early, respectively. A plant that reproduces just before the late date can grow larger and produces  $y_L$  seeds on average. A plant that instead reproduces just before the early date produces  $y_E < y_L$  seeds. If frost comes early, however, then a plant that reproduces late gets zero seeds, while a plant that reproduces early still gets  $y_E$ . Unlike the delayed germination example, we will assume here that seeds always germinate each year.

Now let  $\ell$  be the probability that a seed grows to reproduce late, rather than early. The number of seeds after one year is:

$$s_{t+1} = s_t(p(\ell y_L + (1 - \ell)y_E) + (1 - p)(\ell(0) + (1 - \ell)y_E)).$$

Using the same logic we've used on both previous examples, the long run fitness is:

$$\lim_{n \rightarrow \infty} \frac{\log[s_n]}{n} = p \log[\ell y_L + (1 - \ell)y_E] + (1 - p) \log[(1 - \ell)y_E].$$

We find the value of  $\ell$  that natural selection favors in the long run by the usual derivative approach:

$$\frac{d}{d\ell} \lim_{n \rightarrow \infty} \frac{\log[s_n]}{n} = \frac{p(y_L - y_E)}{(1 - \ell)y_E + \ell y_L} - \frac{1 - p}{1 - \ell} = 0.$$

Solving the above for  $\ell$  yields:

$$\ell = \ell^* = \frac{p y_L - y_E}{y_L - y_E}.$$

As usual, it's instructive to consider limiting cases. When  $p = 1$ , such that the environment is completely predictable, then  $\ell^* = 1$ . Selection favors always growing longer. On the other extreme,  $\ell^* = 0$ , and selection favors never reproducing late, when  $p \leq y_E/y_L$ . If the late frost is rare enough, as defined by the relative seed production rates of early and late growing plants, then selection will not favor risking late reproduction.

## 5.2 Polination

Of course for plants—and other organisms—reproduction is a kind of coordination game, because you need to mature when other individuals are also mature and ready to mate. This is relevant to our flowering problem, because if you are the first plant to reproduce late, you won't be able to find anyone to pollinate and be pollinated by. The same thing goes for the first early reproducer.

Here's one way to tackle this problem. Let  $\ell^*$  represent now the common value of  $\ell$  in the population. We want to know if an invader with value  $\ell = \ell^* + \delta$  can invade, where  $\delta$  is a small positive or negative number. Instead of using a detailed pollination model, we'll just assume for the sake of qualitative understanding that seed production is reduced for late flowering plants by a factor  $\ell^*$ , because more plants flower early, it is harder to get pollinated completely. At the extreme  $\ell^* = 0$ , a late flowering plant doesn't produce any seeds, because  $\ell^* = 0$ , and  $y_L \times 0 = 0$ . Visa versa for the early flowering plants.

The fitness of an invader is:

$$w(\ell, \ell^*) = p \log[\ell \ell^* y_L + (1 - \ell)(1 - \ell^*) y_E] + (1 - p) \log[(1 - \ell)(1 - \ell^*) y_E].$$

Now things get fun, because it turns out that this system has two equilibria, one with more early flowering and one with more late flowering. What we do is take the derivative with respect to  $\ell$  and then set  $\ell = \ell^*$ . Then solve the resulting expression for  $\ell^*$ . You get a quadratic, so it has two roots:

$$\ell^* = \frac{2y_E + py_L \pm \sqrt{y_L} \sqrt{p^2 y_L - 4(1 - p)y_E}}{2(y_E + y_L)}.$$

This looks pretty messy. You can get a clearer grasp on what is going on, though, if you take the limit of each root as  $y_L \rightarrow \infty$ . What this means is that we are examining the tendency of each equilibrium as the difference between early and late reproduction grows increasingly large. Taking these limits, we get  $\ell^* = \{0, p\}$ . One equilibrium is the early-flowering safety equilibrium. The other is the late-flowering risky equilibrium.