Models in Evolutionary Ecology

© Alan R. Rogers[1]

November 18, 2022

---

[1]Department of Anthropology, University of Utah, Salt Lake City, UT 84112

Contents

Chapter 1

Introduction

1.1   Overview

This text introduces some of the models and analytical methods used most frequently in evolutionary ecology. It covers:

- population genetics

- optimization (with applications to clutch size and foraging ecology)

- the covariance formulation of natural selection (with application to kin selection)

- game theory

- demography and the evolution of life histories

Extensive use is made of the Maple computer package for graphics and symbolic mathematics.

Chapter 2

Introduction to Maple

2.1 Getting into and out of Maple

The technique for invoking Maple will depend on which operating system you are using, and on whether you want the graphical user interface (GUI) or the command line interface (CLI). You can launch the GUI from the menus by clicking on the Maple icon. This works on every operating system. On linux, you can also get the GUI by entering `xmaple` at the command prompt. For the moment, this is all you need to know. I will discuss the command line interface later in the chapter.

To get out of Maple, you must either type "`quit;`" and hit return or select `exit` from the `FILE` menu.

2.2 Maple as a calculator

Maple will work just like a calculator:

```
> 1+2;
    3
```

Here, the ">" character is the prompt that Maple typed to indicate that it was ready to receive input. I typed "1+2;" and then hit the return key. The semicolon marks the end of the command, and Maple will not do *anything* until you type it. After I hit "return," Maple typed the result "3" on the line that followed.

Maple pays no attention to the line-breaks in your commands. I could have achieved the same result this way:

```
> 1

> +2;
    3
```

Remember that Maple will not do *anything* until it sees a semicolon. If you leave off the semicolon and simply hit return, Maple will just sit there waiting for you to complete the command.

Here are a few additional arithmetic commands:

```
> 3 - 1; # subtraction
   2

> 3*2; # multiplication
   6

> 3/2; # division
   3/2

> 12/8; # division again
   3/2
```

This example illustrates two new ideas. First, "#" is the comment symbol. Maple ignores everything from the "#" to the end of the line. This makes it possible to put comments into your input. Second, Maple does not automatically represent fractions in decimal notation. Instead, it represents them as ratios of integers. Note that 3/2 and 12/8 yield the same output, since they are really the same number. When it works with integers, Maple's arithmetic is exact: there is no round-off error.

For a decimal approximation to the answer, use the `evalf` function:

```
> evalf(3/2);
   1.500000000
```

You can get the same result another way:

```
> 3.0/2.0;
   1.500000000
```

Maple assumes that 3.0 and 2.0 are already decimal approximations, so there is no point in doing exact arithmetic.

Maple understands parentheses, and all of the standard mathematical functions.

```
> 2*(3+4); # using parentheses
   14

> log10(10); # logarithm to base 10
   log10(10)

> evalf(%);
   1.000000000
```

The last line introduces the "%" symbol, which stands for the result of the previous operation (`log10(10)`). I used the `evalf` function to get a decimal representation of the answer. The "%" symbol saved me some typing.

You can also type "%%" to get the result from two lines back:

```
> 1+2;
       3

> 2+3;
       5

> %%;
       3
```

Finally, you can type "%%%" to get the result from three lines back.

2.3   Defining variables

The symbol ":=" is used to assign values to variables.

```
> z := 3;
      z := 3

> z;
       3
```

This creates a new variable called z, and assigns to it the value 3.

Variables can also be given symbolic values.

```
> z := x^2;
              2
         z := x
```

Now z has the value x^2, but x has not been given a value. Therefore, z and x are both symbolic expressions.

If we give a numeric value to x, this will also give z a numeric value:

```
> x := 3;
       x := 3

> z;
       9
```

Suppose that we wanted to undo the assignment to x so that z will once again become a symbolic expression. There are two ways to do this.

Like this:

```
> x := 'x';
       x := x

> z;
        2
        x
```

or like this:

```
> unassign('x');
> z;
    2
   x
```

To tell Maple to forget everything, use the `restart` command.

The `subs` command is used to assign a value to an expression temporarily:

```
> subs(x=10,z);
   100

> z;
    2
   x
```

In the first line above, the `subs` command prints the value of z after replacing x with 10, i.e. after substituting 10 for x. Since $z = x^2$, the answer is 100. The next line shows that no permanent change has been made in either x or z.

When you are defining expressions, it is always best to define numeric values last. To see why, let us enter the formula for the area of a circle.

a is the area of a circle with radius r:

```
> a := evalf( Pi * r^2);
                      2
   a := 3.141592654 r
```

After setting $r = 3$, a gives the area of a circle with radius 3:

```
> r := 3;
   r := 3

> a;
   28.27433389
```

Here, `Pi` is Maple's builtin symbol for the mathematical constant π.

To look at a circle with a different radius, we need only give a different value to r:

```
> r := 2;
   r := 2

> a;  #area of circle with radius 2
   12.56637062
```

Now, I will illustrate a mistake.

Without unassigning any of these variables, I enter the formula for the circumference of a circle:

```
> c := evalf( 2 * Pi * r );
   c := 12.56637062
```

Next, I change the value of r in an effort to calculate the circumference of a circle with a different radius:

```
> r := 6;
   r := 6

> c;
   12.56637062
```

The value of c did not change. This is because at the time I defined c, r already had a numeric value. Therefore, c became a numeric constant rather than an algebraic expression.

Here is what I should have done:

```
> r := 'r';
   r := r

> c := evalf( 2 * Pi * r );
   c := 6.283185308 r
```

Now I can check the circumference of circles with various radii:

```
> subs(r=1, c); #radius 1
   6.283185308

> subs(r=2, c); #radius 2
   12.56637062
```

2.4   Manipulating symbolic expressions

Before proceding, enter the `restart` command to clear Maple's memory. This done, we don't need to worry about inadvertently using one of the earlier definitions.

Maple often generates complicated expressions, which become useful only after they have been simplified algebraically. A variety of commands are useful for this purpose, including `expand` and `factor`, which are illustrated below.

A product:

```
> (x-2)*(x-3);
(x - 2) (x - 3)
```

Re-expressing the product in expanded form, as a polynomial:

```
> expand(%);

 2
x  - 5 x + 6
```

Factoring recovers the original product

```
> factor(%);

(x - 2) (x - 3)
```

Other useful commands include `simplify` and `combine`.

2.5   Getting help

There is extensive on-line documentation about each command. To find out more about the `expand` command introduced in the preceding section, type `?expand` or `help(expand)` and read the resulting output. These commands are synonyms. Note that the semicolon is not needed (I have no idea why) after the `?expand` command. Use the on-line documentation to read about the following commands: `evalf`, `simplify`, `factor`, `expand`.

2.6 Making plots

Maple makes beautiful plots in two and three dimensions. The simplest facility is `plot`, which does two-dimensional plots. For example, enter the command `plot(x^2, x=0..5)`. This tells Maple to plot the function x^2 from $x = 0$ to $x = 5$. To make a 3-D plot, type `plot3d(x^0.7 * y^0.3, x=0..1, y=0..1)`. The form of these plots will depend on which version of Maple you are running.

2.7 Solving equations

Define an expression `f := (x-3)^2`, which is a function of x. To get a sense of how f behaves, plot it using `plot(f, x=0..4)`. Notice that $f = 0$ when $x = 3$. Now let us see whether Maple can discover this rather obvious fact. Type `solve(f=0, x)`.

2.8 Defining procedures

Suppose that we wanted to evaluate the function $f(x) = x^2$ at several values of x.

We might begin by defining:

```
> f := x^2
```

and then issue commands such as:

```
> subs(x=2, f);
   4

> subs(x=3, f);
   9

> subs(x=4, f);
   16
```

This works, but it takes too much typing. It is more convenient to define f as a function of x. There are two ways of doing this, which have exactly the same effect. The first method is shown below.

The first line defines f as a function of x, whose value equals x^2. Thereafter, expressions like `f(y)` are interpreted as y^2.

```
> f := (x) -> x^2;
                2
     f := x -> x

> f(2);
   4

> f(3);
   9

> f(y);
    2
   y
```

Here is another way to do the same thing:

```
> f := proc(x) x^2 end proc;
     f := proc(x) x^2 end proc
```

This has exactly the same meaning as the previous definition. The first form is often used for short function definitions, whereas the second form is preferred for long definitions.

2.9   Maple's command line interface (CLI)

Maple's GUI has a serious disadvantage. It allows you to scroll up the page and reset the value of a variable without automatically recalculating everything from there on down.

For example, suppose that you begin by typing:

```
> x := 2;
     x := 2

> x^2;
     4
```

So far, so good. But now use the arrow key to move up the page and then change the 2 to a 3. You can generate output that makes no sense:

```
> x := 3;
     x := 3

> x^2;
     4
```

You can fix things by clicking on the $\boxed{\texttt{!!!}}$ icon, which recalculates the entire worksheet, but people often forget to do this, and then get frustrated trying to figure out what is wrong.

For these reasons, I use the command-line interface (CLI) nearly all the time. I write my Maple commands into a file using a text editor. From a separate window, I then run these commands using Maple's CLI. I will explain these steps in reverse order.

You can launch Maple's CLI from within the CLI of your operating system. Type `cmaple` under Windows or `maple` under Linux or Mac OS 10. You will see something like this:

```
    |\^/|      Maple 11 (IBM INTEL LINUX)
._|\|   |/|_. Copyright (c) Maplesoft, a division of Waterloo Maple Inc. 2007
 \  MAPLE  /  All rights reserved. Maple is a trademark of
 <____ ____>  Waterloo Maple Inc.
      |        Type ? for help.
>
```

If you type a Maple command, Maple will type the answer back at you. You can edit the current line of input using the left and right arrows, the delete key, and the backspace key. You can scroll back and forth through Maple's list of previous commands using the up and down arrows.

Maple's CLI launches faster than its GUI, so I always use it for small jobs unless I need pretty graphics. For larger programs, I always put the commands into a text file, using a text editor. You can use any editor you choose. Just be sure you are producing output as plain text, including line breaks.

Use a text editor to create a file named `circle.mpl` that contains the following:

```
unassign('r');
area := proc(r)
    RETURN( Pi*r*r );
    end proc;
```
Next, read these commands into Maple, using the `read` command. The syntax of this command is very simple:

```
read "circle.mpl";
```

This command assumes that `circle.mpl` resides in the current working directory (or folder). It will fail if your file is somewhere else.

The next step is to set Maple's current working directory. Things are simplest if you launch Maple from the command line of your operating system. Before launching Maple, make sure you are in the directory that contains `circle.mpl`. Then Maple will know where to find your input file.

If you launch Maple by clicking an icon, you will have to reset Maple's current working directory from within Maple, using the `currentdir` command. The procedure is explained in Box 1.

When you want pretty graphics, you need to read your file of commands into Maple's GUI.

1. Open a windows CLI and `cd` into the directory containing your file of Maple input commands.

2. Type `cd` and hit return. Windows will tell you the full path name of this directory. Let's assume the name is `C:\full\path\name`.

3. Launch the GUI by clicking on the Maple icon in the Program menu.

4. Use Maple's "currentdir" command to reset Maple's idea of the current directory:

   ```
   currentdir("C:\\full\\path\\name");
   ```

   Note that the path name is within double quotes and that each backslash is written twice.

5. Then use the `read` command as usual.

Box 1: Importing a file of commands into the Maple GUI

2.10 Maple's Graphical User Interfaces (GUIs)

I only use Maple's GUI when I want to make pretty graphs. You may prefer to use it all the time. Either way, you need to know a little about it.

The first thing to understand is that Maple has two GUIs. The one that Maple introduced first is now called the "classic worksheet" and is much easier to use. To launch it, type `maple -cw` at the Windows command prompt. The other one is called the "Java worksheet," and it is the one you get when you click on an icon in the Windows menu system.

The problem with the new worksheet involves the way you type equations into it. In the old Maple, you typed lines like those illustrated in this chapter. In the new Maple worksheet, equations are beautifully formatted as you type them in, but you have to use the arrow keys to move from one part of the equation to another. This is appreciably slower, and it's easy to get confused.

In either GUI, you can scroll up and change the values of variables anywhere in the worksheet. Whenever you do this, click on the $\boxed{!\,!\,!}$ icon, which recalculates the entire worksheet. Otherwise, your worksheet will be in an inconsistent state.

Exercises

Use Maple for all of the calculations in the exercises below.

⋆ EXERCISE **2–1** What is 37×73?

⋆ EXERCISE **2–2** What is $37 - 73$?

⋆ EXERCISE **2–3** What is $2701/2847$ as a ratio of integers?

⋆ EXERCISE **2–4** What is the floating point approximation to $2701/2847$?

⋆ EXERCISE **2–5** What is $27 \times (19 + 88)$?

⋆ EXERCISE **2–6** A circle with radius r has area πr^2. Type this formula into Maple, and then use the `subs` command to find the areas of circle with the following radii 1, 2, 3, 4, and 5.

⋆ EXERCISE **2–7** Expand the following formula as a polynomial: $(x - 2)(x + 4)(x - 5)$.

⋆ EXERCISE **2–8** Factor this polynomial: $24 + x^4 - 10x^3 + 35x^2 - 50x$.

⋆ EXERCISE **2–9** Use Maple's help facility to look up the `solve` command, and write a paragraph explaining what it does. Feel free to quote from Maple's online documentation.

⋆ EXERCISE **2–10** Plot the function $f(x) = x^{1/3}$ from $x = 0$ to $x = 1$, and describe the output in words. What does the function look like?

⋆ EXERCISE **2–11** Plot the function $f(x) = 1 - e^{-x}$ from $x = 0$ to $x = 1$, and describe the output in words. What does the function look like?

⋆ EXERCISE **2–12** Use the `solve` command to find the value of x at which $1 - e^{-x} = 1/2$.

⋆ EXERCISE **2–13** Write a Maple procedure that calculates $f(x) = 1 - e^{-x}$. Use a text editor to write this procedure into a file. Then use Maple's `read` command to read the file. Demonstrate that the procedure gives correct answers.

Chapter 3

Calculus

Calculus is used in various ways by evolutionary ecologists. This chapter provides a (very) brief introduction. If you have never learned calculus, this chapter is unlikely to be very useful. But if you had the course several years ago and are now rusty, this chapter may refresh your memory.

3.1   Slope

Figure 3.1 shows two straight lines. When the function $f(x)$ increases with x, its slope is said to be positive. When it deceases with x, the slope is negative. If the function describes a straight line, its slope is measured by

$$\text{slope} = \frac{f(x_1) - f(x_0)}{x_1 - x_0} \tag{3.1}$$

where x_0 and x_1 are two values of x, and $f(x_0)$ and $f(x_1)$ are the corresponding values on the y-axis. For the left-hand graph, this will give a positive number since $f(x_1) - f(x_0)$ has the same sign as $x_1 - x_0$. The right-hand graph has a negative slope because the numerator and denominator have opposite signs.

In the right-hand graph, $x_0 = 0.5$, $x_1 = 0.8$, $f(x_0) = 0.5$ and $f(x_1) = 0.26$. The slope is therefore

$$\text{slope} = \frac{0.26 - 0.5}{0.8 - 0.5} = -0.8$$

● EXAMPLE **3–1**
What is the slope of the left-hand graph in Figure 3.1, whose values are $x_0 = 0.5$, $x_1 = 0.8$, $f(x_0) = 0.5$, and $f(x_1) = 0.74$.

Positive Slope   Negative Slope
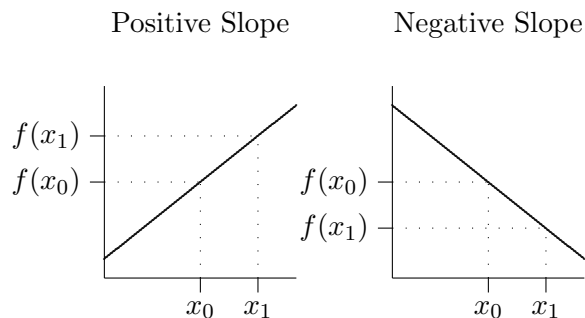
Figure 3.1: Straight lines with positive and negative slope
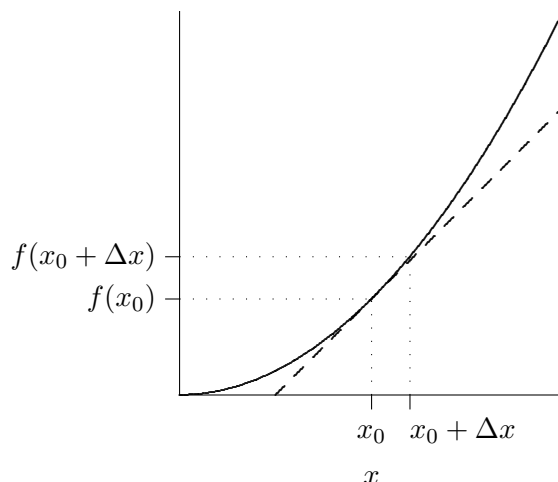
Figure 3.2: Graph of $f(x) = x^2$
The dashed line is tangent to $f(x) = x^2$ at the point $x = x_0$ and has slope $2x_0$.

∘ ANSWER
The slope is $(0.74 - 0.5)/(0.8 - 0.5) = 0.8$.

3.2 The derivative of x^2

Slope calculations run into difficulty when the line is curved, like that in figure 3.2. The slope is positive, but gets steeper as one moves from left to right across the graph. Therefore, this function has no single slope. It has a different slope for every value of x.

Suppose we wanted to calculate the slope of the curve at the point where $x = x_0$. We could approximate this slope by choosing a second value, x_1, and then applying formula 3.1. Let us choose $x_1 = x_0 + \Delta x$. Here, the symbol Δ is pronounced "delta" and represents change. Eqn. 3.1 can be re-written as

$$\frac{\Delta f(x)}{\Delta x} = \frac{f(x_0 + \Delta x) - f(x_0)}{\Delta x} \tag{3.2}$$

On both sides, the denominator is the change in x and the numerator is the corresponding change in $f(x)$. Thus, this is nothing more than the old familiar formula for the slope of a straight line. If we wished, we could substitute numbers into this formula and thus calculate the approximate slope at any point along the curve. These approximations will be most accurate if we choose points close together so that Δx and $\Delta f(x)$ are both small.

● EXAMPLE **3–2**
In figure 3.2, $x_0 = 0.5$, $x_0 + \Delta x = 0.6$, $f(x_0) = 0.25$ and $f(x_0 + \Delta x) = 0.36$. What is the approximate slope of $f(x)$ at the point where $x = x_0$?
∘ ANSWER
The approximate slope is $(0.36 - 0.25)/(0.6 - 0.5) = 1.1$.

These numerical approximations are cumbersome, so let us look for a better approach. The curve in figure 3.2 was calculated from the formula

$$f(x) = x^2$$

This means that $f(x_0) = x_0^2$ and that $f(x + \Delta x) = (x_0 + \Delta x)^2$. Plugging these into Eqn. 3.2 allows

us to re-write the approximate slope as

$$\frac{(x_0 + \Delta x)^2 - x_0^2}{\Delta x} = \frac{x_0^2 + 2x_0\Delta x + (\Delta x)^2 - x_0^2}{\Delta x}$$
$$= 2x_0 + \Delta x$$

To make our approximation accurate, we want to make Δx as small as possible. As Δx gets smaller and smaller, the approximate slope gets closer and closer to $2x_0$. This limiting value of our approximate slope is called the *derivative* and is written as $df(x)/dx$ or as dx^2/dx.

The derivative at point $x = x_0$ can be interpreted as the slope of a straight line that is tangent to the curve at point where $x = x_0$. This straight line is shown as a dashed line in figure 3.2.

Notice that dx^2/dx is not a single number, but an entire function. The function

$$\frac{dx^2}{dx} = 2x$$

tells us that at each value of x, the slope of $f(x) = x^2$ is twice the value of x. Thus, the slope is greater at large values of x than at small ones, just as figure 3.2 would suggest.

3.3 Notation

Derivatives are written in several different ways. The derivative of f with respect to x can be written either as

$$\frac{df}{dx} \qquad \text{or as} \qquad \frac{d}{dx}f \qquad \text{or as} \qquad f' \qquad \text{or as} \qquad \dot{f}$$

These all mean exactly the same thing. Sometimes they are modified to emphasize that the derivative is itself a function, becoming

$$\frac{df(x)}{dx} \qquad \text{or} \qquad \frac{d}{dx}f(x) \qquad \text{or} \qquad f'(x) \qquad \text{or} \qquad \dot{f}(x)$$

These all mean the same thing too.

Often, we wish to refer to the value of the derivative at some particular value of x. The derivative of $f(x)$ at the point where $x = 3$ is written either as

$$\left.\frac{df}{dx}\right|_{x=3} \qquad \text{or as} \qquad f'(3) \qquad \text{or as} \qquad \dot{f}(3)$$

For example, we know already that the derivative of x^2 is $2x$. When $x = 2$, this derivative is equal to 4. This fact can be expressed by writing either

$$\left.\frac{df}{dx}\right|_{x=2} = 4 \qquad \text{or as} \qquad f'(2) = 4 \qquad \text{or as} \qquad \dot{f}(2) = 4$$

● EXAMPLE **3–3**
What is the derivative of x^2 when $x = 0.5$.
○ ANSWER
$\left.\frac{dx^2}{dx}\right|_{x=0.5} = 1$.

3.4 Derivatives of other functions

To find the derivative of a function is to "differentiate" it. One can differentiate any smooth function using the method outlined above. This is the subject of the first semester of college calculus, which I will not try to replicate here. For reference, however, I will tabulate some of the main results in Box 2.

Power rule	$\frac{dkx^n}{dx}$	$=$	nkx^{n-1}
Log rule	$\frac{d\ln x}{dx}$	$=$	$1/x$
Exponential rule	$\frac{de^{kx}}{dx}$	$=$	ke^{kx}
Sum rule	$\frac{d[f(x)+g(x)]}{dx}$	$=$	$f'(x) + g'(x)$
Product rule	$\frac{df(x) \cdot g(x)}{dx}$	$=$	$f'(x)g(x) + f(x)g'(x)$
Quotient rule	$\frac{df(x)/g(x)}{dx}$	$=$	$\frac{f'(x)g(x) - f(x)g'(x)}{g^2(x)}$
Chain rule	$\frac{df(g(x))}{dx}$	$=$	$\frac{df}{dg} \cdot \frac{dg}{dx}$

Box 2: Miscellaneous derivatives. In these formulas, k is a constant; $e = 2.718$ is the base of natural logarithms; and ln is the logarithm to base e, also called the "natural log".

3.5 Optima

In evolutionary ecology we are often concerned with optimization problems. For example, in order to build a theory of reproduction, we ask (among other things) "At what age does it become advantageous for an individual to stop growing and begin reproducing?" Or in other words: What is the optimal age of reproductive maturity?

To show how such problems are solved, I will use the simplest optimization problem that I know. Suppose that we have 40 feet of fence and want to use it to enclose a rectangular area (a new garden). We wish to choose the garden's width w and length l so as to make the garden's area A as large as possible. In this effort, we are constrained by the consideration that the perimeter of our garden must equal the length of the available fencing: 40 feet. Our problem can be stated as follows:

$$\max A = lw$$
$$\text{subject to: } 2 \cdot (l + w) = 40$$

The first line above says that we wish to maximize the area A, which is equal to the product lw of length and width. The function A, which we are trying to maximize, is called the *objective function*. The second line specifies a constraint and says that the perimeter must equal 40 feet.

This is a problem of *constrained optimization*. I will illustrate here the simplest techniques for solving such problems. The first step is to use the constraint equation to eliminate one of the variables. Rearranging the constraint gives

$$l = 20 - w$$

and substituting this into the formula for A gives

$$A = 20w - w^2$$

Our problem is to choose the value of w that will make this as large as possible.

The garden's area is graphed as a function of its width in figure 3.3. Clearly, the maximum is attained by a point at or near $w = 10$. Many optimization problems can be solved simply by graphing the objective function.
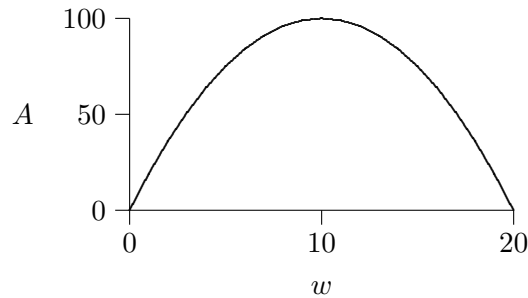
Figure 3.3: Garden's area A as a function of its width w

Another method is suggested by the observation that the derivative of the objective function is positive to the left of the maximum point, negative to the right, and zero at the optimum. Thus, we could also have found the equilibrium by solving the equation

$$\frac{dA}{dw} = 0 \tag{3.3}$$

I will take this derivative both "by hand" and also using Maple. To take it by hand, I first note that the function $A = 20w - w^2$ is a sum of $20w$ and $-w^2$. Therefore, the sum rule (Box 2) implies that

$$\frac{dA}{dw} = \frac{d}{dw}(20w) + \frac{d}{dw}(-w^2)$$

Each of the derivatives on the right-hand side can be evaluated using the power rule (Box 2). The result is

$$dA/dw = 20 - 2w$$

Setting this equal to zero gives $w = 10$, as we inferred above from the figure.

Maple gives the same answer:

```
> A := 20*w - w^2;
                 2
      A := 20 w - w

> dA := diff(A,w);

      dA := 20 - 2 w

> solve(dA = 0, w);

      10
```

The answer, which we have now obtained three ways, tells us that the garden should be 10 feet wide. To satisfy the constraint, it must also be 10 feet long. Thus, a rectangular garden must be square if it is to maximize the area enclosed by a given length of fence.

(I specified at the outset that the garden had to be rectangular. Had I relaxed this restriction, we might have found a garden that was even larger.)
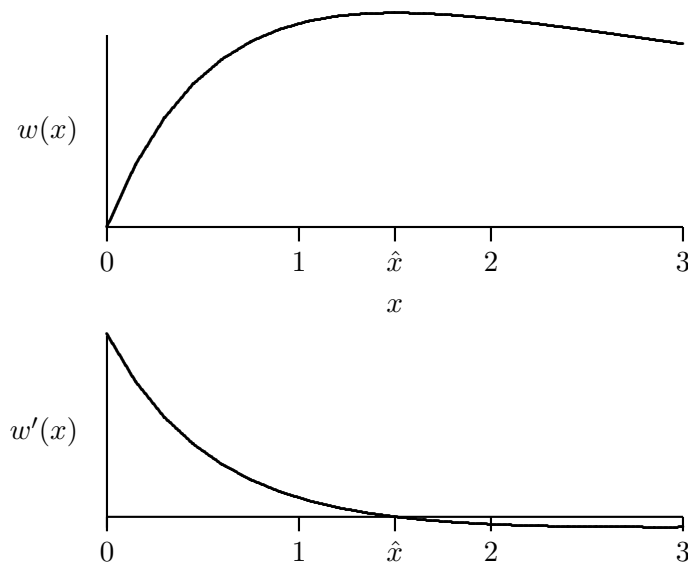
Figure 3.4: Fitness and its derivative

3.6   Second-order condition for a local maximum

Eqn. 3.3 is what is known as the *first-order condition for a local maximum*. It must be satisfied at any internal maximum, but it can also be satisfied at points that are not maxima. It is a necessary condition but not a sufficient condition. Thus far, we have relied on graphical methods to separate maxima from minima and other kinds of extreme points. This section will introduce a non-graphical method, the so-called *second-order condition for a local maximum*.

   The upper panel of figure 3.4 graphs a fitness function.[1] It shows that the extreme point at $x = 1.5$ is indeed a local maximum. But how could we have determined this without drawing a graph? The lower panel of figure 3.4 suggests an answer. Since $\hat{x}$ is a local maximum, $w'(x)$ is positive to the left of $\hat{x}$ and negative to the right of $\hat{x}$. Consequently, the slope of $w'(x)$ is negative where $x = \hat{x}$. This suggests a second means of determining whether $\hat{x}$ is a local maximum: It is a maximum only if the derivative of $w'(x)$ is negative at the point $x = \hat{x}$.

   The derivative of a derivative is called a second derivative and is written using such notation as $d^2 w(x)/dx^2$ or $w''(x)$. The second derivative at some particular value of x (say $\hat{x}$) is written either as

$$\left. \frac{d^2 w(x)}{dx^2} \right|_{x=\hat{x}} \qquad \text{or as} \qquad w''(\hat{x})$$

If we find a point $\hat{x}$ for which (1) $f'(\hat{x}) = 0$, and (2) $f''(\hat{x}) < 0$, then it is bound to be a local maximum. However, there are some maxima at which $f''(\hat{x}) = 0$. Such cases require use of additional criteria, which are beyond my scope here. You will find this issue discussed in any text on mathematical optimization [12].

3.7   A palette of functions

Functions are the building blocks of ecological models. Suppose an ecologist is interested in the probability $s(x)$, that an offspring will survive, given that it receives x units of resource. Presumably,

---

[1]You will meet this function later, in example 7–1 on p. 45.

Increasing at decreasing rate

Sigmoid
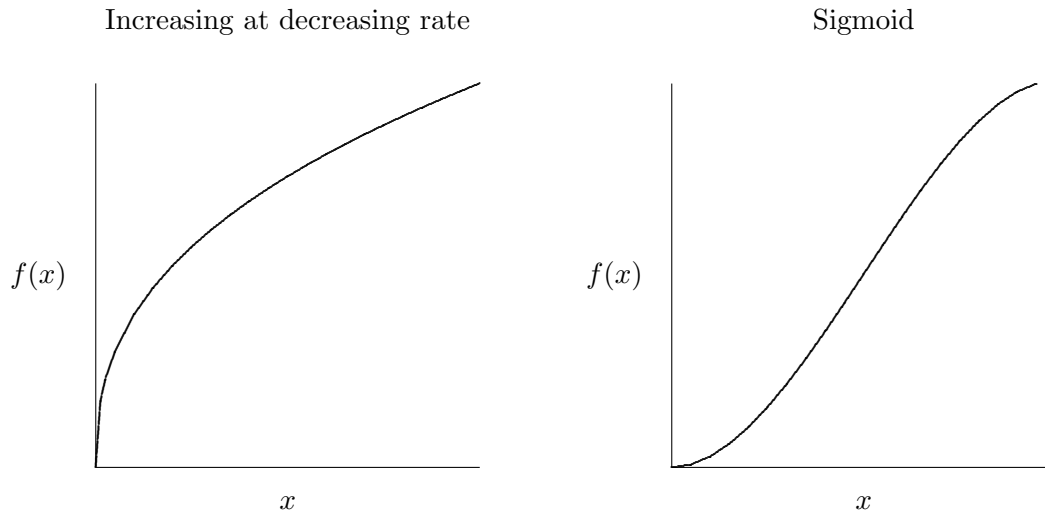
Figure 3.5: Increasing functions with different shapes

more is better, so s must increase with x. But how does it increase? Maybe each unit of resource helps more than the last. If so, then s would increase at an increasing rate, as in Figure 3.2. This, however, seems unlikely. Surely food is worth more to a hungry offspring than to one that is well nourished. Consequently, our ecologist might decide that s should increase at a decreasing rate, like the function in left panel of Figure 3.5.

On the other hand, another ecologist might argue that a starving offspring is unlikely to survive anyway, so a unit of food isn't worth much if x is very small. Consequently, this second ecologist would prefer the function in the right panel of Figure 3.5.

Once the ecologist knows what he wants to assume, his next problem is to choose a function that captures that assumption. To do so, he needs to know something about the shapes of functions. Since he thinks s increases with x, the ecologist wants a function whose first derivative is positive. This is true of all the functions in Figures 3.2 and 3.5. To distinguish between these shapes, he needs to look at the second derivative.

Consider Figure 3.2. The derivative is small when x is near 0, but gets larger as x increases. In other words, the derivative is an increasing function of x. This means that the second derivative is positive. Thus, if the ecologist wanted a function like this, he should look for one whose first and second derivatives are both positive.

Now consider the function on the left side of Figure 3.5. It's derivative is large when x is small, but decreases as x increases. In other words, the second derivative is negative. The function on the right side of Figure 3.5 combines both of the previous cases. Its second derivative is positive for small values of x, but negative for large values.

Now we know what to look for. Let us evaluate some specific functions: (a) x^a, (b) $1 - e^{-ax}$, (c) $ax^2/2 - bx^3/3$, and (d) $[x/(1+x)]^2$.

● EXAMPLE **3–4**

Calculate the first and second derivatives of each function. Based on these derivatives, describe the shape of each function in words. How do these shapes depend on the values of a and b?

Exercises

Do these exercises either by hand or using Maple's `diff` function.

⋆ EXERCISE **3–1** Use Maple's `diff` command to check formulae in Box 2.

Use the method illustrated in section 3.2 to differentiate (i.e. find the derivatives of) the following functions.

⋆ EXERCISE **3–2** $f(x) = a$, where a is a constant

⋆ EXERCISE **3–3** $f(x) = ax$

⋆ EXERCISE **3–4** $f(x) = ax^2$

⋆ EXERCISE **3–5** $f(x) = ax^3$

Your answers to the last few questions should be as predicted by the power rule (Box 2).

⋆ EXERCISE **3–6** Differentiate $f(x) = a + bx + cx^2 + dx^3$, where a, b, c, and d are constants.

⋆ EXERCISE **3–7** What is the derivative of 3?

⋆ EXERCISE **3–8** What is the derivative of $3x$?

⋆ EXERCISE **3–9** What is the derivative of $3x - 3x^2$?

⋆ EXERCISE **3–10** What is the derivative of $3x(1 - x)$?

⋆ EXERCISE **3–11** What is the derivative of $3e^{3x}$?

⋆ EXERCISE **3–12** What is the derivative of $\ln(3x)$?

⋆ EXERCISE **3–13** What is the derivative of $\frac{3 + 4x - 5x^2}{x^2 + e^x}$?

Chapter 4

Selection

Most of the theory in evolutionary ecology involves natural selection, so that is where we begin. This chapter covers the subject from the point of view of population genetics. It is about alleles and how their frequencies change. This theory provides the foundation on which other chapters will build. In addition, I use it as a vehicle for teaching a little more about Maple.

4.1   A recurrence equation

Suppose there is a locus with two alleles A_1 and A_2, whose relative frequencies in some generation are p and $q \equiv 1 - p$, respectively. What makes this locus interesting is that the three genotypes differ in their Darwinian fitness. In general, variation in fitness might be due either to differences in fertility or to differences in survival. The math is simpler in the latter case, so I will assume that our three genotypes differ only with respect to their probabilities of survival.

Let w_{ij} denote the probability that an individual of genotype A_iA_j will survive to reproduce. This probability is called the fitness of genotype A_iA_j. The genotypic fitnesses and frequencies are shown in table 4.1. The frequencies before selection are obtained from the Hardy-Weinberg formula.[1] This should give the right answer provided that the population is not too small and that mating is at random. The contribution of each genotype to the gene pool of the next generation is proportional to the product of its frequency and fitness. Dividing by $\bar{w}$, the mean fitness, ensures that the genotypic frequencies after selection sum to unity.

---

[1] Since our locus has 2 alleles, there are three genotypes A_1A_1, A_1A_2, and A_2A_2. In a large population that mates at random, the frequencies of these genotypes will be close to p^2, $2pq$, and q^2. This is called the Hardy-Weinberg principle.

Table 4.1: Genotypic fitnesses and frequencies

Genotype	A_1A_1	A_1A_2	A_2A_2
Freq. before sel'n	p^2	$2pq$	q^2
Fitness	w_{11}	w_{12}	w_{22}
Freq. after sel'n	$p^2 w_{11}/\bar{w}$	$2pq w_{12}/\bar{w}$	$q^2 w_{22}/\bar{w}$

Where $q = 1 - p$
and $\bar{w} = p^2 w_{11} + 2pq w_{12} + q^2 w_{22}$

The frequency of allele A_1 is $p = P_{11} + \frac{1}{2}P_{12}$, where P_{11} and P_{12} are the frequencies of A_1A_1 and A_1A_2.

$\star$ EXERCISE **4–1** Show that $p = P_{11} + \frac{1}{2}P_{12}$.

Applying this to the last row of table 4.1 gives the frequency of A_1 after selection.

$$p(t+1) = \frac{p^2(t)w_{11} + p(t)q(t)w_{12}}{\bar{w}(t)} \tag{4.1}$$

As usual, $p(t)$ refers to the allele frequency in generation t, while $p(t+1)$ refers to that in generation $t + 1$. Equation 4.1 tells us p's value in one generation as a function of its value in the previous generation. Such equations are called *recurrence equations*. They provide recipes for getting from one generation to the next.

4.1.1   The change per generation

Given an equation that gets us from one generation to the next, the next question to ask is: How much change is there between successive generations? To find out, we need to calculate a quantity called $\Delta p(t)$ (pronounced "delta pee of t"), which is defined to equal $p(t + 1) - p(t)$. Δp is often useful because it is positive when p is increasing, and negative when p is decreasing. When there is no tendency to change, the system is said to be at *equilibrium*, and $\Delta p = 0$. Subtracting $p(t)$ from equation 4.1 gives (after some algebra)

$$\Delta p(t) = \frac{p(t)q(t)}{\bar{w}(t)}[p(t)(w_{11} - w_{12}) + q(t)(w_{12} - w_{22})] \tag{4.2}$$

This equation provides a great deal of intuition concerning the dynamics of natural selection. First, the rate of change is proportional to pq, and will therefore be very slow when *either* allele is near zero. Response to selection should therefore be slow both when p is very rare and when it is very common. Second, if $w_{11} \approx w_{12}$, then the term in square brackets will be small when $p \approx 1$ and $q \approx 0$. This means that response will be slow when a dominant allele (one for which $w_{11} = w_{12}$) is common. Third, the same reasoning implies that response will be slow when a recessive allele is rare.

$\star$ EXERCISE **4–2** Use Maple to verify that the rate of change is fast when a dominant allele is rare, and slow when it is common.

$\star$ EXERCISE **4–3** Use Maple to verify that equation 4.2 can be derived from equation 4.1. Hint: Enter Eqns. 4.1 and 4.2. Subtract p from the first, and from this subtract the second. Then `simplify`. Don't bother with the "(t)" in $p(t)$ and $q(t)$: just write p and q.

4.2   The same result in terms of genic fitnesses

In later chapters I will often use a slightly different formulation of equation 4.2:

$$\Delta p(t) = \frac{p(t)q(t)(w_1 - w_2)}{\bar{w}(t)} \tag{4.3}$$

Here, w_1 and w_2 are the *genic fitnesses* of A_1 and A_2. The genic fitness of allele A_i is simply the average fitness of copies of A_i.

In diploid organisms, genes occur in pairs. Each copy of the A_1 allele is paired with some other gene. What is the probability that this second gene is also a copy of A_1? If mating is at random, then the second gene is (in effect) drawn at random from among all the genes in the population. Since a fraction p of these are copies of A_1, p is also the probability that the second gene is an A_1. In this case, the original copy of A_1 resides in an A_1A_1 homozygote with fitness w_{11}. On the other hand, the original A_1 might (with probability $q = 1 - p$) be paired with a copy of A_2. If so, then the initial A_1 resides in a heterozygote with fitness w_{12}. The mean fitness of allele A_1 is therefore

$$w_1 = pw_{11} + qw_{12}$$

A similar argument applies if the initial gene is a copy of A_2 and leads to

$$w_2 = qw_{22} + pw_{12}$$

To justify equation 4.3, subtract w_2 from w_1. The result is equivalent to the quantity in square brackets in equation 4.2.

⋆EXERCISE **4–4** Verify that $w_1 - w_2$ is equal to $p(w_{11} - w_{12}) + q(w_{12} - w_{22})$.

Since $p(1 - p)/\bar{w}$ is always positive, equation 4.3 implies that allele A_1 will increase whenever

$$w_1 - w_2 > 0 \tag{4.4}$$

Thus, to figure out which allele is being favored by selection, it is only necessary to calculate the genic fitnesses, w_1 and w_2. This trick is central to much of the theory in evolutionary ecology.

⋆EXERCISE **4–5** Suppose that there are three genotypes, A_1A_1, A_1A_2, and A_2A_2, with fitnesses 1, 1.5, and 0.5. Calculate the two genic fitnesses under the assumption that $p = 0.1$. Which allele is being favored by selection?

⋆EXERCISE **4–6** Repeat this calculation assuming that $p = 0.9$. Which allele is being favored by selection?

⋆EXERCISE **4–7** Equation 4.3 also tells us that equilibrium will occur when $w_1 = w_2$. Use this principle to find the equilibrium value of p from the genotypic fitnesses above.

4.3   A Worked Example

I begin by using a text editor to create a file called `selection.mpl`:

```
1    unassign('wbar,'q','p','p2','dp','w11','w12','w22');
2    p2 := (w11*p^2 + w12*p*q)/wbar;
3    dp := p*q*((w11 - w12)*p  + (w12 - w22)*q)/wbar;
4    wbar := w11*p^2 + w12*2*p*q + w22*q^2;
5    q := 1-p;
6    w11 := 1.0;
7    w12 := 1.4;
8    w22 := 0.7;
```

The first few lines (1) initialize parameters, just in case they have been given values before the file is read into Maple, (2) define `p2` using Eqn. 4.1, (3) define `dp` using Eqn. 4.2, and (4) define $\bar{w}$ as shown in Table 4.1. The remaining lines define `q` and the three genotypic fitnesses.

Having created this disk file, I invoked Maple and typed `read 'selection.mpl'`:

```
> read 'selection.mpl';

                2
        w11 p  + w12 p q
    p2 := ----------------
                wbar


          p q ((w11 - w12) p + (w12 - w22) q)
    dp := -----------------------------------
                        wbar

                2                      2
    wbar := w11 p  + 2 w12 p q + w22 q
    q := 1 - p
    w11 := 1.0
    w12 := 1.4
    w22 := .7
```

Maple did not echo my commands, but did print its response to each command. At this point `dp` looks like this:
```
> dp;
         p (1 - p) ( - 1.1 p + .7)
    ----------------------------------
        2                        2
    1.0 p  + 2.8 p (1 - p) + .7 (1 - p)
```
Note that, `dp` is a function of a single variable, p. Thus, if we give p a numeric value (say $p = 0.5$), then `dp` will also have a numeric value:

```
> subs(p=.5,dp);
    .03333333333
```

This is a useful check to make sure that you have made no mistake. When you have assigned numeric values to all the quantities in the model, your results should not contain non-numeric symbols.

The easiest way to find equilibria (i.e. points at which $\Delta p = 0$) is to plot Δp against p. This can be done in Maple by typing `plot(dp,p=0..1)`, and should give you a graph that looks like Fig. 4.1. The figure shows that $\Delta p = 0$ when $p = 0$, when $p = 1$, and also when p is at an intermediate value of approximately 0.6. These are the three equilibria of this equation. The equilibria at $p = 0$ and $p = 1$ are easy to understand. A population can respond to selection only if it contains variation. This, after all, was one of Darwin's original premises. But variation is lacking when $p = 0$ or $p = 1$, for then there is only one allele in the population. The equilibria at 0 and 1 can also be understood algebraically. As I noted above, equation 4.2 tells us that Δp is proportional to $p(t)q(t)$, and must
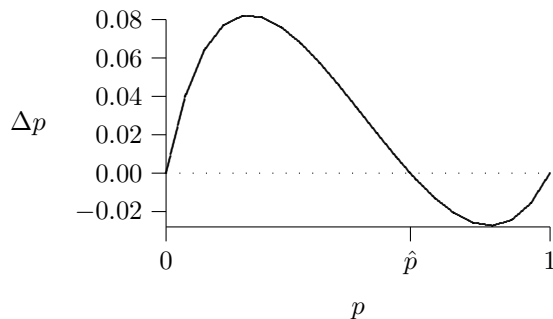
Figure 4.1: Change per generation under natural selection. The curve is drawn from Eqn. 4.2. The equilibrium value, $\hat{p}$, is the point at which $\Delta p = 0$.

therefore equal zero if p is equal either to 0 or to 1. This accounts for the fact that the curve in Fig. 4.1 equals zero at the left and at the right. But what about the equilibrium in the middle?

We can find this equilibrium more accurately by solving the equation $\Delta p = 0$:

```
> solve(dp=0, p);
     0, 1., .6363636364
```

Maple gives three solutions to the equation, which correspond to the three points at which Δp equals zero in Fig. 4.1. Thus, we see that the internal equilibrium occurs at the value $\hat{p} = 0.6363636364$.

Equilibria are interesting only if the population evolves toward them reasonably rapidly, and this is not always the case. For example, mutation is such a weak force that allele frequencies may take many millions of years to reach the equilibria implied by their mutation rates. When progress is that slow, one must think seriously about other forces such as genetic drift. The present model is about selection, which is often much faster than mutation. But not always, so we should check. The Maple output below shows a numerical experiment that begins with $p(0) = 0.1$ and then uses the recurrence equation above to get from each generation to the next. I will supress blank lines to save space:

```
> subs(p=0.1, p2);   # Find p(1)
     .1640530760
> subs(p=%, p2);      # Now p(2)
     .2432133616
> subs(p=%, p2);      # Now p(3)
     .3248181678
> subs(p=%, p2);      # Now p(4)
     .3971767598
> subs(p=%, p2);      # Now p(5)
     .4553692119
> subs(p=%, p2);      # Now p(6)
     .4998761659
> subs(p=%, p2);      # Now p(7)
     .5332408700
> subs(p=%, p2);      # p(8)
     .5581434063
> subs(p=%, p2);      # p(9)
     .5767780168
> subs(p=%, p2);      # p(10)
     .5907937712
```

Figure 4.2: Time path of allele frequency. The curve is drawn using the results of the numerical example in section 4.3. The dotted line indicates the equilibrium value, $\hat{p} \approx 0.636$.
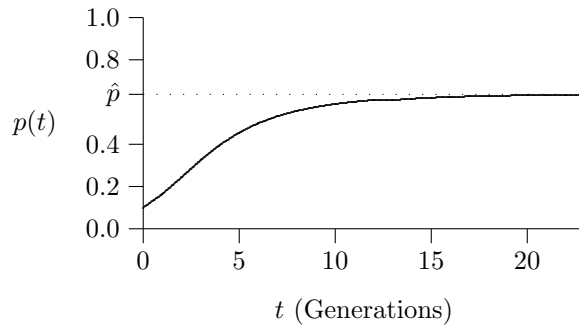
```
> subs(p=%, p2);      # p(11)
     .6013945473
> subs(p=%, p2);      # p(12)
     .6094541224
```

Another 10 generations the quick way. (If you're curious, look up "for" in the Maple documentation.)

```
> pp := %;  # For intermediate results
     pp := .6094541224
> for i from 1 by 1 to 10
  do pp := subs(p=pp,p2) od;
     pp := .6156092326  #p(13)
     pp := .6203275030  #p(14)
     pp := .6239553742  #p(15)
     pp := .6267516688  #p(16)
     pp := .6289111975  #p(17)
     pp := .6305815324  #p(18)
     pp := .6318750562  #p(19)
     pp := .6328777273  #p(20)
     pp := .6336555245  #p(21)
     pp := .6342592332  #p(22)
```

Note that the allele frequency is converging toward the equilibrium at $\hat{p} = 0.6363636364$. A better view of this process is obtained by graphing the results, as shown in Fig. 4.2. There, it is clear that p approaches its equilibrium value within a small number of generations. Thus, the equilibrium in this model is one worth considering.

4.4   Wright's equation

Sewall Wright [79] often expressed equation 4.2 as

$$\Delta p(t) = \frac{p(t)q(t)}{2\bar{w}(t)} \cdot \frac{d\bar{w}(t)}{dp(t)} \tag{4.5}$$

This equation is equivalent to 4.2, but provides a different kind of insight. Notice that $pq/\bar{w}$ is always positive when p is neither zero nor unity. This means that p will increase when $d\bar{w}/dp > 0$ and will decrease when this derivative is negative. In other words, the allele frequency always moves in a direction that increases $\bar{w}$, the mean fitness. The allele frequency is always moving "uphill" on the function $\bar{w}$. This suggests that selection will tend to maximize mean fitness. Fig. 4.3 shows that this is indeed the case.
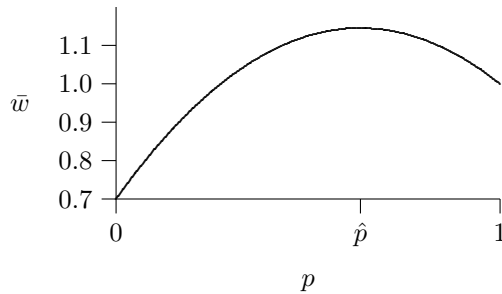
Figure 4.3: $\bar{w}$ as a function of p. The curve is drawn from equation 4.2. The equilibrium value, $\hat{p}$, is the point at which $\bar{w}$ is maximized, which is also the point at which $\Delta p = 0$.

⋆ EXERCISE **4–8** Repeat all steps of the worked example, assuming that genotypes A_1A_1, A_1A_2, and A_2A_2 have fitnesses 0.1, 1, and 1, and that $p(0) = 0.1$.

⋆ EXERCISE **4–9** Verify (using Maple if you wish) that equations 4.5 and 4.2 are equivalent.

4.5 Monomorphic and polymorphic equilibria

Equilibria are points at which there is no tendency for allele frequencies to change, i.e. values of p for which $\Delta p = 0$. Equation 4.5 shows that there are two kinds of equilibrium:[2]

Monomorphic equilibria These occur when $p = 0$ or $p = 1$, i.e. when there is only one allele in the population.

Polymorphic equilibria According to Wright's formula, these occur when $d\bar{w}/dp = 0$, or in other words, when p is at a maximum point on the function relating mean fitness $\bar{w}$ to allele frequency[3].

Wright's formula provides a new method for finding evolutionary equilibria: They are points at which mean fitness is maximal.

⋆ EXERCISE **4–10** Define $w_{11} = 1$, $w_{12} = 1.1$, and $w_{22} = 0.8$. Find the equilibria in this system by graphing Δp as a function of p.

⋆ EXERCISE **4–11** Find the equilibria by using Maple to solve the equation $\Delta p = 0$.

⋆ EXERCISE **4–12** Find the equilibria by graphing $\bar{w}$ against p and searching for the maximum by eye.

⋆ EXERCISE **4–13** Find the equilibria by solving the equation $d\bar{w}/dp = 0$.

In these exercises, you will find that the methods using Δp give more equilibria than those using $\bar{w}$, since they tell you about the "trivial" equilibria at which $p = 0$ and $p = 1$.

---

[2]We will encounter additional kinds of equilibria when we allow fitnesses to change in response to allele frequencies.
[3]For an explanation of this point, see section 3.5 of Ch. 3.

4.6   Numerical Example Revisited

Wright's equation (equation 4.5) suggests that internal equilibria are the point(s) at which $\bar{w}$ reaches its maximal value(s).

The graph of $\bar{w}$ has a peak at about 0.6—very close to the equilibrium value that we found before. To see whether it is indeed the same equilibrium, we solve the equation $d\bar{w}/dp = 0$:

```
> solve(diff(wbar, p) = 0, p);
    .6363636364
```

4.7   Summary

This chapter has introduced two ways to describe evolutionary change: recurrence equations, and difference equations. The variant formulation involving genic fitnesses will turn up in later chapters. Equilibria occur where $\Delta p = 0$. We are particularly interested in stable equilibria, because these are the ones we are likely to see in nature. Wright's formulation justifies the idea of an adaptive landscape. In that conception, populations move uphill, and equilibria are the tops of peaks in the adaptive landscape.

Chapter 5

More Maple

This chapter begins with a digression on using Maple to calculate derivatives and integrals. The rest of the chapter shows how to use Maple to study the equations introduced in Ch. 4.

5.1 Calculus

Maple can take derivatives:

```
> diff(x^2, x);
     2 x
```

Here, `diff(f,x)` is Maple's way of representing the derivative of f with respect to x. The preceding dialog is Maple's way of saying that

$$\frac{dx^2}{dx} = 2x$$

Maple can also do integrals:

```
> int(x, x);
         2
    1/2 x
```

This says that the indefinite integral of x is equal to $x^2/2$. This answer is not quite complete. In calculus, we learn that

$$\int x\,dx = x^2/2 + C$$

where C is an arbitrary constant of integration. Maple omits the constant of integration.

⋆ EXERCISE **5–1** Use Maple's `diff` and `solve` commands to find the value of x that maximizes $f(x) = x - 2x^2$. Don't forget to check the 2nd-order condition.

⋆ EXERCISE **5–2** Repeat the preceding exercise for the function $f(x) = xe^{-x}$.

5.2 Writing a Maple procedure

Chapter 4 introduced the idea of a recurrence equation. It showed how one can calculate $p(1)$ from $p(0)$, then $p(2)$ from $p(1)$, and so on. In principle, one could use this method to find $p(1000)$, but it would be tedious.

To automate this procedure, we can write a Maple procedure. (These were introduced in Ch. 2, p. 7.) Here is a Maple procedure that implements Eqn. 4.1.

```
unassign('w11', 'w12', 'w22');
p := proc(t)
     if t <= 0 then RETURN(p0) end if;
     pp := p(t-1);
     wbar := pp^2*w11+2*pp*(1-pp)*w12+(1-pp)^2*w22;
     (pp^2*w11+pp*(1-pp)*w12) / wbar;
     end proc;
p0 := 0.1;  # initial value of p
```

Compare this code to Eqn. 4.1 on page 20. You should be able to recognize the code for mean fitness and the code for the recurrence equation itself. You may however be puzzled by the line that reads `pp := p(t-1)`. In this line, the procedure calls itself.

The reason for this is straightforward. When I invoke the procedure by typing $p(2)$, I am asking for the allele frequency in generation 2. However, the code only knows how to find $p(2)$ from $p(1)$, which it has to obtain somehow. It finds this value via the call to `p(t-1)`. Thus, a call to $p(2)$ generates a call to $p(1)$, which in turn generates a call to $p(0)$. At this point, the cascade of procedure calls stops. (Can you see why?)

This procedure is long enough that I didn't want to enter it directly into Maple, so I used a text editor to create a text file named seln1.mpl. Then I launched Maple and entered the command `read 'seln1.mpl'`. In response, Maple printed out just what it would have printed had I entered the commands by hand. Next, I entered the genotypic fitnesses interactively:

```
> w11:=1; w12:=1.2; w22:=0.1;
    w11 := 1

    w12 := 1.2

    w22 := 0.1
```

Finally, I used the procedure to calculate $p(t)$ at several value of t.

```
> p(1);
    0.3843648208

> p(10);
    0.8176882252

> p(100);
    0.8461538431

> p(1000);
bytes used=4001180, alloc=1244956, time=0.06
    0.8461538448
```

This output tells me the value of p in generations 1, 10, 100, and 1000.

There is a pitfall in this sort of code. Consider the following procedure, which implements a different recurrence equation.

```
p := proc(t) (1.0-u)*p(t-1) + v*(1 - p(t-1)) end proc;
```

```
p(0) := 0.1; u := 0.00001; v := 0.000002;
```

If you enter this into Maple, you will discover that `p(1)` and `p(10)` work fine, but `p(100)` fails miserably. Instead of an answer, you get a succession of messages indicating that Maple is gobbling up an ever larger fraction of your computer's memory.

The problem is that the procedure calls $p(t-1)$ *twice*. When I type $p(2)$, each of these references generates a separate call for $p(1)$. Then, each call to $p(1)$, generates two calls to $p(0)$. We end up with one call to $p(2)$, two to $p(1)$, and four to $p(0)$, a total of seven. When I call $p(100)$, however, I generate about 10^{31} procedure calls. No wonder the procedure never finishes!

This problem is even more acute in the procedure above for selection, where `pp` appears seven times. If each of these had been a call to `p(t-1)`, there would have been hell to pay. Instead, the code calls `p(t-1)` just once and stores the result in a variable (`pp`), which it can use over and over.

However, there is still room for improvement. Suppose I call `p(1000)`. This invokes the procedure 1000 times. (`p(1000)` calls `p(999)`, which calls `p(998)` and so on.) If I then call `p(1001)`, Maple will have to do this work all over again. It would be better, wouldn't it, if the Maple could simply remember these values, and re-use them, rather than calculating them over and over again? Here is a re-write that shows how this can be done.

```
unassign('w11', 'w12', 'w22');
p := proc(t)
    option remember;
    if t <= 0 then RETURN(p0) end if;
    pp := p(t-1);
    wbar := pp^2*w11+2*pp*(1-pp)*w12+(1-pp)^2*w22;
    (pp^2*w11+pp*(1-pp)*w12) / wbar;
    end proc;
p0 := 0.1;  # initial value of p
```

The command `option remember` tells Maple to remember the results of all calls to procedure p, and to re-use them when needed. You will discover that it takes awhile to calculate $p(1000)$. But having done so, the calculation of $p(1001)$ is almost instantaneous.

The `remember` option has one drawback: once Maple has remembered the value of $p(10)$, say, it doesn't forget that value when you change the parameters such as w_{11}, w_{12}, and w_{22}. The simplest solution is to re-define your function every time you want to change parameter values. This is easy if you keep the function definition in an external file, as described in Section 2.9 of Chapter 2.

⋆ EXERCISE **5–3** Create a file that contains the Maple procedure shown above, and use it in the following exercises. (Avoid the "remember" option so that you won't need to worry about restarting Maple after each change in parameter values.) Set $p(0) = 0.1$, $w_{11} = 1$, $w_{12} = 0.8$, and $w_{22} = 0.6$. What is $p(t)$ for $t = 1, 2, 3, 10$, and 100.

⋆ EXERCISE **5–4** Repeat the preceding exercise, but this time with $w_{12} = 1.2$.

5.3 The "for loop"

Suppose you wanted to calculate the sum of the even integers from 2 through 10. You might write

```
1    s := 0;
2    s := s+2;
3    s := s+4;
4    s := s+6;
5    s := s+8;
6    s := s+10;
```

Here is a program that does the same calculation using Maple's `for` statement:

```
1    s := 0;
2    for i from 2 by 2 to 10 do
3       s := s + i;
4       end do;
```

Line 1 initializes `s` (gives it an initial value). Lines 2–4 constitute the "for loop." When this runs, line 3 will execute five times. The variable `i` is set equal to 2 the first time, to 4 the second time, and so on. At the end, `s` equals 30 and `i` equals 12.

The loop above contains only one statement (line 3), but that was just to keep things simple. Maple can accomodate many lines of code between `for` and `end do`.

The `for` statement has many variants, which you can read about in Maple's online documentation. An online tutorial can be found at `http://www.mth.uea.ac.uk/~h007/MTH-2G22/` `mapleprog.pdf`. The easiest way to learn is to study examples, so here are a few that I copied from Maple's online documentation.

```
1    # Find the sum of all two-digit odd numbers from 11 to 99.
2
3    > tot := 0:
4    > for i from 11 by 2 while i < 100 do
5    >    tot := tot + i
6    > end do:
7    > tot;
8
9                                        2475
```

```
1    # Add together the contents of a list.
2
3    > tot := 0:
4    > for z in [1, x, y, q^2, 3] do
5    >   tot := tot + z
6    > end do:
7    > tot;
8
9                                              2
10                                    4 + x + y + q
```

```
1    # Multiply the entries of an expression sequence.
2
3    > tot := 1:
4    > for z in 1, x, y, q^2, 3 do
5    >    tot := tot * z
```

```
6   > end do:
7   > tot;
8
9                                                          2
10                                             3 x y q
```

5.4 Sums and products

For sums and products, Maple makes things even easier. You can calculate the sum

$$1/1 + 1/2 + 1/3 + \cdots + 1/100$$

like this: `sum(1.0/i, i=1..100)`. The product of the integers from 1 to 10 is

```
1   > product(i, i=1..10);
2                                        3628800
```

These expressions work even if the limits are variables. For example:

```
1   > sum(k,k=0..n-1);
2
3                                         2
4                                  1/2 n   - 1/2 n
```

⋆EXERCISE **5–5** Calculate the product of $1/i^2$ from $i = 1$ to $i = 10$ using (a) a `for` loop, and (b) the `product` statement.

⋆EXERCISE **5–6** To calculate variances and covariances, the first step is to calculate means, mean of squares, and the mean of cross products. In other words, given data like this:

X	Y
3	6
2	8
12	4
12	8

We need to calculate $\sum X/N$, $\sum Y/N$, $\sum X^2/N$, $\sum Y^2/N$, and $\sum XY/N$, where N is the number of observations. Do this in Maple using a single for loop.

Chapter 6

Clutch Size

6.1   Natural selection as a "strategist"

Plants and animals are marvels of engineering. Many have justified religious beliefs on the argument that where there is design, there must be a designer. However, evolutionists draw a different conclusion. The appearance of design could have been produced by natural selection. When the fitnesses of genotypes do not depend on their frequency, selection will cause a population to crawl uphill on a surface that relates mean fitness to allele frequencies. Over time, this process leads to adaptation—the appearance of design. Selection favors those designs that are best able to get their genes into subsequent generations. It is often illuminating to ask whether the characteristics of plants, animals, or people make sense as part of a strategy designed for survival and reproduction. This usage is strictly metaphorical—it does not imply that plants, animals, or people are conscious of their strategies.

6.2   The phenotypic gambit

In studying a particular character such as parental care, we confront a dilemma. We might begin with a model from population genetics, such as that outlined in Ch. 4. This would involve expressing the fitness of each genotype as a function of the sort of care that a parent provides. The various genotypes might adopt differing strategies and thus earn differing fitnesses. But in the end, this approach would seem unsatisfactory because our results would depend so heavily on something that we do not know: the genetic system that underlies parental care.

Alternatively, we could abandon the effort to work out the entire time path of allele frequency change and concentrate instead on a simpler questions: Under what circumstances can a gene encouraging one strategy invade a population dominated by another? When an allele is rare, nearly all carriers are heterozygotes, and nearly all carriers of the other allele are homozygotes. If the rare heterozygote outcompetes the common homozygote, then the rare allele can invade. Thus, we only need to compare two genotypes.

This makes the problem much simpler, but we can make it simpler still. Sewall Wright's theory of natural selection (Eqn. 4.5) shows that, so long as genotypic fitnesses are constant, selection will tend to increase the mean fitness of the population. If there is no shortage of genetic variation in, say, clutch size, then selection will adjust clutch size toward the value that maximizes mean fitness. If we are prepared to *assume* an ample supply of genetic variance, then we can simply ask which clutch size maximizes fitness. We don't need to worry about genes at all.

Alan Grafen [26] calls this the "phenotypic gambit," and it is somewhat controversial. There

is, after all, no guarantee that a population can reach the optimal character value. The genes that affect clutch size may also be intimately connected with something else, so that selection cannot improve clutch size without doing damage elsewhere. There is also the problem of maladaptive intermediates. The population's current physiology may be far from the global optimum, but it will stay where it is if it is on a local optimum. For example, squid do not swim as efficiently as fish. At any given speed, they require more energy than a fish of comparable size. Presumably, they would be better off with tails like fish yet do not evolve them because the intermediate forms would be maladaptive [1, p. 149].

Nonetheless, the phenotypic gambit is a useful way to generate hypotheses. If often yields predictions that are in good agreement with Nature. When it fails, we must turn to the detailed models of population genetics.

6.3 Components of an optimization model

6.3.1 A currency

To carry out this program, one must first decide what variable to maximize. Fitness seems an obvious choice, but there are others. For example, perhaps animals forage in a way that maximizes their rate of energy capture. This assumption is favored by foraging theorists because it avoids the issue of how nutrition affects fitness. Foraging theory simply assumes that fitness increases with energy acquired, so that maximizing the one amounts to maximizing the other. The variable that is optimized is the "currency" of the optimization problem.

6.3.2 Constraints

Not all strategies are feasible. For example, time spent searching for mates cannot also be used to care for children. Consequently, the better one cares for one's children, the lower one's mating success. Life is full of such constraints, and they play a major role in evolutionary ecology. The constraints introduced into a model determine which strategies are feasible. The problem is then to search among these feasible strategies for the one(s) that maximize the currency.

6.4 The tradeoff between quantity and quality of offspring

To illustrate these ideas, I begin with the model of Smith and Fretwell [66], which deals with the tradeoff between the quantity of one's children and the probability that each will survive. Smith and Fretwell chose to optimize fitness itself, which they expressed as

$$w = s(x)b$$

where w is fitness, b is the number of offspring born, x is the amount of resource allocated by the parent to each offspring, and $s(x)$ is the probability than an offspring who receives x units of resource will survive. Fitness is thus the product of quantity and quality, where quantity is the number of offspring and quality is is the probability that each one survives.

What can we say about the function $s(x)$? It seems safe to assume that offspring cannot survive at all unless they get at least a critical minimum (x_0) of nutrition. Beyond this point, more food is better than less, so s should increase with x, but not at a constant rate. A unit of food is worth more to a hungry offspring than to one that is full, so s should increase with x at a decreasing rate. Thus, we assume that $s'(x)$ is 0 when $x \leq x_0$ but positive when $x > x_0$. At these larger values of x, we also assume that $s''(x) < 0$.

To make this problem interesting, we must introduce some constraint. If each of the b offspring gets the same quantity x of resources, then the total expended is xb. Smith and Fretwell assumed that this product equals a constant T, the total amount of the parent's resources that is available for reproduction. Our constrained optimization problem is thus

$$\max w = s(x)b$$
$$\text{subject to: } xb = T$$

This problem can be solved using the method introduced in section 3.5. The first step is to rearrange the constraint equation as $b = T/x$. Next, substitute b into w, producing

$$w = Ts(x)/x \tag{6.1}$$

This makes fitness a function of only one variable (x), so the optimization problem is easy. A maximum (if one exists) will require that $dw/dx = 0$. But rather than solving this problem directly, it is slightly easier to solve $d \log w/dx = 0$.[1]

Using Maple, I solved the problem as follows:

```
> w := T * s(x) / x;
          T s(x)
    w := ------
            x
> expand(diff(log(w),x));
    d
    -- s(x)
    dx
    ------- - 1/x
    s(x)
```

(Here, **expand** makes Maple's answer more readable.) This derivative is zero when

$$\frac{ds(x)}{dx} = \frac{s(x)}{x} \tag{6.2}$$

This condition must hold at a maximum point, but it would also hold at a minimum point. The final step is to determine which type of point we have found.

There are several ways to do this. In Ch. 3, I discussed the "second-order" condition for a local maximum, but I will not use that here. Instead, I use a graphical method. Consider the two lines drawn in figure 6.1. Point A, at which the straight and curved lines touch, is $\hat{x}$ units to the right of the vertical axis, and $s(\hat{x})$ units above the horizontal axis. Thus, the straight line has slope $s(\hat{x})/\hat{x}$. Furthermore, at this point the two lines have equal slope, so Eqn. 6.2 holds when $x = \hat{x}$. This shows that the figure corresponds to the equation. But is it really a local maximum?

To see that it is, note that the slope of the straight line is steeper than that of any other straight line from the origin to a point on the curve. Thus, point A identifies the largest possible slope, $s(x)/x$. Now according to Eqn. 6.1, this slope is proportional to fitness. Consequently, $\hat{x}$ maximizes not only the straight line's slope, but also the parent's fitness.

It is surprising that the optimal investment, $\hat{x}$, does not depend on the wealth of the parent. Poor parents (who have low values of T) should allocate the same to each offspring as do rich parents (with high values of T). This seems counterintuitive, but it may account for the remarkable fact

---

[1]This gives the same answer because, since $\log x$ increases monotonically with x, the value of x that maximizes $\log w$ will also maximize w.
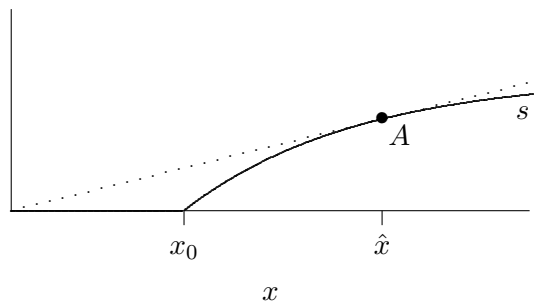
Figure 6.1: Smith & Fretwell model. Offspring cannot survive if $x < x_0$. Beyond x_0, s increases with x at a decreasing rate. The slope of a straight line from the origin through $s(x)$ is maximized when $x = \hat{x}$. In maximizing this slope, $\hat{x}$ also maximizes the parent's fitness. Since the two lines have equal slope at point A, Eqn. 6.2 is satisfied.

that many plant species produce seeds of nearly uniform size regardless of the size of the parent. Some species have seeds of such uniform size that in ancient times they were used as units of measure.

On the other hand, other species produce young of varying size. The model of Smith and Fretwell does not tell the whole story, as we will see in more detail below.

⋆ EXERCISE **6–1** In the standard model of Smith and Fretwell, suppose that

$$s(x) = x^3/(1+x)^3$$

Find the stable equilibrium graphically and then using calculus. Demonstrate that it is stable.

⋆ EXERCISE **6–2** In the standard model of Smith and Fretwell, suppose that

$$s(x) = 1 - e^{-(x-1)}$$

Find the equilibria graphically and then using calculus. Which equilibria are stable?

6.5   Evaluating the Smith-Fretwell model against data

The Smith-Fretwell model makes a very detailed prediction about the amount of resource devoted to each offspring and about the number of offspring produced. Ideally, one would evaluate the model by comparing this detailed prediction to data from particular species. Unfortunately, this effort is unlikely to succeed. It would need detailed knowledge about $s(x)$, which is seldom available.

Fortunately, the Smith-Fretwell model also makes another prediction: the size of an offspring should not depend on the resources available to the mother. Ordinarily, large mothers will be able to mobilize more resources than small ones, yet the result predicts that large and small mothers should produce offspring of equal size.

Figure 6.2 compares the volumes of eggs against the sizes of mothers in a series of species from two different taxonomic groups—crustaceans and fish. In both groups, large species tend to produce large eggs. This appears to contradict the prediction of the Smith-Fretwell model.

There is good reason, however, to be skeptical of this test. Each point in figure 6.2 refers to a different species, and there is no reason to think the these different species share a common function, $s(x)$. The pattern seen in the figure could reflect some variation across species in the shape of this function. To test the model, we really need to look at variation across individuals within a single species.

This is done in figure 6.3, which plots the mean birth weight of Red Deer offspring against the mean weight of their mother's cohort. Some cohorts are larger than others, presumably because
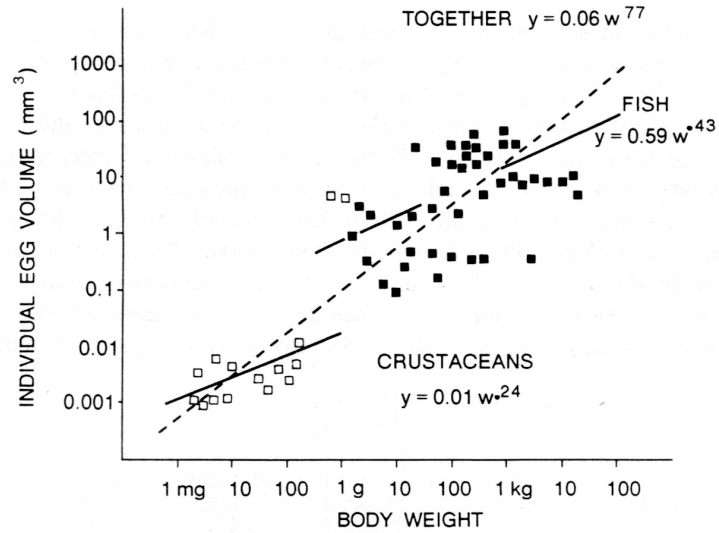
Figure 6.2: Individual egg volume versus maternal weight for fish, and for crustaceans. Straight lines show regression of log egg volume on log maternal weight for each group separately and for both groups combined. [3], cited in [14, p. 60]
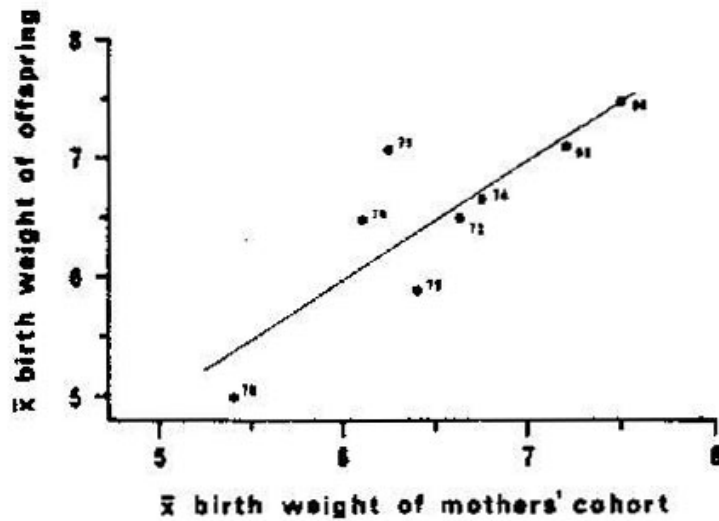
Figure 6.3: Mean birth weight of Red Deer hinds versus mean weight of mother's cohort. [13, Fig. 21.7]

they had better nutrition as they were growing. The hinds of large cohorts presumably have more resources to allocate toward reproduction. The Smith-Fretwell model predicts that this should not affect the size of the offspring. Clearly, the prediction fails for Red Deer.

6.6 Modifying the model

How should the model be modified? Perhaps offspring from large clutches waste resources in conflict with other siblings [53]. Sibling rivalry might affect the fitness of offspring in several ways. Most obviously, the offspring may kill one another. Yet rivalry can be costly even if they do not, for the more energy one spends on conflict, the less one can spend on growth, defense against pathogens and so on. Here, I focus on this second cost.

One approach is to leave $s(x)$ alone and modify the constraint function. Suppose that each offspring wastes c units of resource in conflict with each of its $b - 1$ siblings. Then the constraint function becomes

$$xb + cb(b - 1) = T$$

What is the optimal value of c? In other words, how much energy should an offspring allocate to conflict? The answer depends on what other offspring are doing. The more one allocates to conflict, the more the others should allocate. This kind of problem requires evolutionary game theory, which I introduce later in this book (Ch. 8). Here, I cheat by treating c as an *exogeneous constant*. I assume, in other words, that the value of c is determined in advance. This approach may give the wrong answer if (in a more realistic model) c varies in response to T.

The new model is studied in Box 3, leading to

$$\frac{ds(x)}{dx} = \frac{s(x)}{[(x - c)^2 + 4cT]^{1/2}} \tag{6.3}$$

When there is no sibling rivalry, $c = 0$, and Eqn. 6.3 reduces to Eqn. 6.2. This is reassuring, for when $c = 0$ there is no difference between the two models.

The constraint can be written in the form of a quadratic equation: $cb^2 + (x - c)b - T = 0$, and solved using the quadratic formula. The equation has two roots, one of which is negative and thus irrelevant. The positive root is $b = [-(x - c) + ((x - c)^2 + 4cT)^{1/2}]/2c$. The fitness function is $w = bs(x)$, so

$$\begin{aligned}
\frac{d \log w}{dx} &= \frac{1}{b}\frac{db}{dx} + \frac{1}{s(x)}\frac{ds(x)}{dx} \\
&= -\frac{1}{[(x - c)^2 + 4cT]^{1/2}} + \frac{s'(x)}{s(x)}
\end{aligned}$$

At the optimal value of x, this derivative will equal 0. Setting it to 0 and rearranging gives Eqn. 6.3.

Box 3: Derivation of Eqn. 6.3

It is also reassuring that T appears in the answer, for we are trying to understand how T affects $\hat{x}$. But it is hard to tell, by staring at Eqn. 6.3, whether $\hat{x}$ increases or decreases with T. In Box 4, I prove that $d\hat{x}/dT > 0$. In other words, females with ample resources (large T) should produce larger offspring (large $\hat{x}$). However, don't bother with Box 4 unless you are comfortable with calculus. Look instead at Box 5, which makes the same point using a numerical example. Both

If T changes, what happens to the optimal value of x? This box will get the answer with calculus. *Skip to the next box unless you are comfortable with caculus.* We begin with the equation $\partial \log w(\hat{x}, T)/\partial \hat{x} = 0$, which holds because $\hat{x}$ maximizes w. This equation defines $\hat{x}$ as an implicit function of T. Consequently, we find $d\hat{x}/dT$ using the rule for implicit functions (see Appendix E). This rule allows us to write

$$\frac{d\hat{x}}{dT} = -\left.\frac{\partial^2 \log w/\partial x \partial T}{\partial^2 \log w/\partial x^2}\right|_{x=\hat{x}}$$

In section 3.6 (p. 16), I discussed the second-order condition for x to be a local maximum. As that section explains, the denominator of the expression above is negative provided that $\hat{x}$ is a local maximum. Thus, the sign of $d\hat{x}/dT$ is the same as the sign of the numerator,

$$\partial^2 \log w/\partial x \partial T = 2c((x-c)^2 + 4cT)^{-3/2}$$

Since everything within this expression is positive, the expression must be positive too. This proves that $d\hat{x}/dT > 0$. In other words, females with ample resources should produce larger offspring.

Box 4: Proof that $\hat{x}$ increases with T in the model of sibling rivalry.

Suppose that $s(x) = 0$ if $x \leq x_0$ and equals $(x - x_0)^{1/2}$ if $x > x_0$. This function equals 0 when $x < x_0$ and increases at a decreasing rate when $x > x_0$. In addition, I assume arbitrarily that $x_0 = 1$ and $c = 1$. Here is a Maple program that evaluates this model:

```
interface(echo=4);

# Definitions
w  := b*s(x);
constr := c*b^2 + (x-c)*b - T = 0;
s  := proc(x) (x-1)^(1/2) end proc;
c  := 1;

# Solve constraint for b using positive root of quadratic formula
b := (-(x-c) + sqrt((x-c)^2 + 4*c*T))/(2*c);

# Derivative of log of fitness
dlw := diff(log(w), x):

# Find value of x that makes dlw equal zero.
solve(dlw=0, x);
```

When you run this program, the last line of output should contain two formulas, which represent the two values of x for which $d \log w/dx = 0$. One of these is negative unless T is very small and is thus irrelevant to our problem. The other is equivalent to $\hat{x} = 1 + \sqrt{4T/3}$ and implies the $\hat{x}$ increases with T. If you have trouble seeing this, then calculate a few numerical values using different values of T. The result predicts that well-fed mothers (large T) should produce larger offspring (large $\hat{x}$).

Box 5: Numerical example of model of clutch size with sibling rivalry

boxes show that the model predicts what we observe in red deer: females with ample resources produce large offspring.

Although the model predicts what we observe, this does not prove that it is correct. It could be giving the right answer for the wrong reason. To test it further, it would be interesting to compare species with and without sibling rivalry. We should find a positive relationship between $\hat{x}$ and T where sibling rivalry is important, but not where it is absent. This is a new prediction that could be tested through empirical research.

6.7   On abstraction

In this chapter, the focus has been on abstract models rather than living, breathing organisms. For many people, abstract models are hard. I might have taken a different approach, discussing a particular organism, the surival chances of its offspring, and how these are affected by provisioning and sibling rivalry. This would have made the argument less abstract and perhaps easier to grasp. But this advantage would have come at a cost.

First, it is never possible to estimate anything without error, and we would have had to consider how these errors affected our answers. The model would have needed a statistical component, a layer of complexity that the abstract models above did not need. Furthermore, the model would have been less general. If it were tuned to estimates from a single population, then it would tell us only about that population. No general insights would have emerged. Thus, abstraction has benefits, and theoreticians often make their models as abstract as possible.

6.8   Conclusions

This chapter has illustrated the interaction between modeling and data analysis. We started with a simple model that seemed to explain some things but failed to explain others. This led to a second model, which explained what was not explained before and which suggested other kinds of empirical research.

$\star$ EXERCISE **6–3** Box 5 assumes that $s(x) = (x - 1)^{1/2}$. Use this model to determine $\hat{x}$ in (a) the Smith-Fretwell model and (b) the sibling rivalry model, with $c = 1$. In both cases, assume that $T = 20$. In which model is $\hat{x}$ larger?

Chapter 7

Foraging

7.1 Diet choice

The next problem on our agenda is that of diet choice, and illustrates an optimization problem that does not require calculus. Suppose that a foraging animal encounters potential prey of several different types. At each encounter, it must decide whether to attack the prey or pass it up. How should such decisions be made? Which species of prey should be attacked, and which ignored?

I shall simplify this problem by assuming that there are only two species of prey, and that each of these is attacked either never or else at each opportunity. I am ruling out, in other words, the possibility that the forager might attack some species on 40% of encounters, but ignore it the rest of the time. This is true of any optimal strategy, but I will not prove that here. For details, see [69, p. 18].

In this problem, fitness would be a poor currency. As explained in section 6.3.1, foraging theorists usually assume that selection favors the strategy that maximizes the animal's supply of some nutrient such as energy. This supply is often measured by the rate at which energy is acquired,

$$R = A/T$$

where A is the amount of energy acquired and T the total time spent foraging.

It is easy to misunderstand the meaning of this assumption. Suppose that I live in a habitat in which I can expect to acquire, on average, 10 calories per minute while foraging. But one morning, 1/2 minute after I leave home, some kind person gives me a piece of chocolate cake containing 400 calories. Then my rate of energy capture is already 800 calories per minute. If I continue foraging at 10 calories per minute, my rate of energy capture will fall. Therefore, if I were really trying to maximize R, I should stop foraging. Furthermore, I should never forage again. This program can hardly be optimal, for if I followed it I would eventually starve to death.

But this argument misunderstands what is meant by rate maximization. By maximizing R, we usually mean either: (a) maximizing the amount of energy acquired over some fixed period of time, (b) minimizing the time required to acquire some fixed amount of energy, or (c) maximizing the rate over a long but unspecified period of time. In what follows, I use the first of these interpretations.

For simplicity, suppose that prey species 1 is better than species 2, in that a forager who took only species 1 would acquire more energy that a forager who took only species 2. Our only question, then, is: Should the forager take only species 1 or should it take both species? The answer depends on two quantities: A_1 (the energy acquired by taking species 1 only) and A_2 (that acquired by taking both species). If $A_2 > A_1$ then the forager should take both species. Otherwise it should take only species 1.

These amounts will depend on the following variables.

$$T_s \quad = \quad \text{"search time," the time spent searching for prey}$$

$$T_h \quad = \quad \text{"handling time," the time spent pursuing, attacking, and eating prey;}$$

$$\lambda_i \quad = \quad \text{the number of prey species } i \text{ encountered per minute;}$$

$$e_i \quad = \quad \text{net energy acquired from an individual of prey species } i;$$

$$h_i \quad = \quad \text{handling time per individual of prey species } i.$$

Suppose the forager takes only species 1. In T_s minutes of searching, it will encounter $T_s \lambda_1$ prey of species 1, which will provide a total of

$$A_1 = T_s \lambda_1 e_1$$

calories. To determine the magnitude of A_1, we must first find T_s. In other words, we must find out how much of the total T minutes is spent searching. T_s can be calculated from the following constraint:

$$T_s + T_h = T$$

which says that search time and handling time must together equal the total foraging time T. Handling time is $T_h = T_s \lambda_1 h_1$, since the $T_s \lambda_1$ prey items each require h_1 minutes. Substituting this back into the constraint equation yields $T_s = T/(1 + \lambda_1 h_1)$ and

$$A_1 = T \frac{\lambda_1 e_1}{1 + \lambda_1 h_1} \tag{7.1}$$

A similar argument shows that a forager taking both species will acquire

$$A_2 = T \frac{\lambda_1 e_1 + \lambda_2 e_2}{1 + \lambda_1 h_1 + \lambda_2 h_2} \tag{7.2}$$

calories. Note that A_1 and A_2 are both proportional to T, the total foraging time. If we divided by T on both sides of each equation, we would obtain the expressions for rates of return that are usually studied in models of foraging ecology.

The forager should take both prey species if $A_2 > A_1$, i.e. if

$$\frac{\lambda_1 e_1 + \lambda_2 e_2}{1 + \lambda_1 h_1 + \lambda_2 h_2} > \frac{\lambda_1 e_1}{1 + \lambda_1 h_1}$$

which is true provided that[1]

$$\frac{e_2}{h_2} > \frac{\lambda_1 e_1}{1 + \lambda_1 h_1} = R_1$$

---

[1]Here is the algebra behind this claim:

$$\frac{\lambda_1 e_1 + \lambda_2 e_2}{1 + \lambda_1 h_1 + \lambda_2 h_2} \quad > \quad \frac{\lambda_1 e_1}{1 + \lambda_1 h_1}$$

$$\frac{\lambda_2 e_2}{1 + \lambda_1 h_1 + \lambda_2 h_2} \quad > \quad \lambda_1 e_1 \left(\frac{1}{1 + \lambda_1 h_1} - \frac{1}{1 + \lambda_1 h_1 + \lambda_2 h_2} \right)$$

$$\frac{\lambda_2 e_2}{1 + \lambda_1 h_1 + \lambda_2 h_2} \quad > \quad \lambda_1 e_1 \left(\frac{(1 + \lambda_1 h_1 + \lambda_2 h_2) - (1 + \lambda_1 h_1)}{(1 + \lambda_1 h_1)(1 + \lambda_1 h_1 + \lambda_2 h_2)} \right)$$

$$\lambda_2 e_2 \quad > \quad \frac{\lambda_1 e_1 \cdot \lambda_2 h_2}{1 + \lambda_1 h_1}$$

$$\frac{e_2}{h_2} \quad > \quad \frac{\lambda_1 e_1}{1 + \lambda_1 h_1} = R_1$$

Figure 7.1: Choice of an optimal prey set. Open circles (○) show e_i/h_i for prey type i, with the prey types ranked in order of descending e_i/h_i. Bullets (●) show R_i for a diet consisting of prey types $1, 2, \ldots, i$. The optimal diet consists of all prey types for which $e_i/h_i > R_{i-1}$, i.e. all prey types to the left of the vertical line.

Prey types ranked by decreasing e_i/h_i
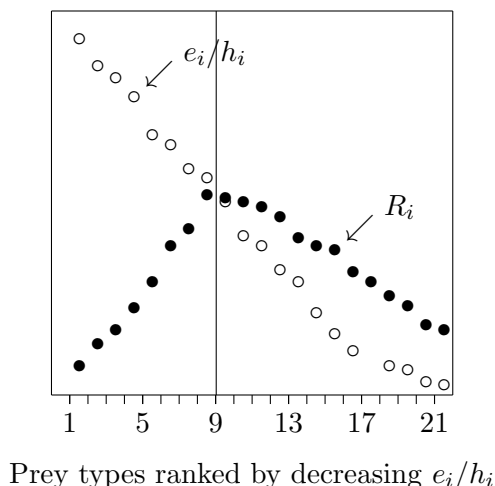
Table 7.1: Data on three hypothetical prey species

Prey	λ_i	e_i	h_i	e_i/h_i
1	0.1	20	10	2.00
2	1.0	3	2	1.50
3	2.0	10	15	0.67

This tells us that species 2 should be included in the diet only if e_2/h_2 exceeds the overall rate of return for a forager specializing in species 1.

This model can be generalized to deal with more than two species of prey. In this more general context, prey species i should be added to the diet if e_i/h_i exceeds the average rate of return, R_{i-1}, for the items already in the diet [69, p. 22]. This makes intuitive sense: When a prey of type i is attacked, the rate of return during the attack is e_i/h_i. If the prey is not attacked, the predator can expect to gain calories at the usual rate, R_{i-1}. Thus, it makes sense to attack if $e_i/h_i > R_{i-1}$.

This decision rule is illustrated graphically in Fig. 7.1. The optimal diet there consists of all prey types to the left of the vertical line.

Numerical example Suppose that a forager is confronted with the three types of prey described in table 7.1. If it eats prey type 1 alone, its rate of return will be

$$R_1 = \frac{0.1 \times 20}{1 + 0.1 \times 10} = 1$$

Should it add prey type 2 to its diet? The answer is yes, since $e_2/h_2 = 1.5$, and this exceeds $R_1 = 1$. After adding prey species 2, the overall rate becomes

$$R_2 = \frac{0.1 \times 20 + 1 \times 3}{1 + 0.1 \times 10 + 1 \times 2} = 1.25$$

This exceeds R_1, as expected. Having added prey species 2, should the forager now add species 3? The answer is no, since $e_3/h_3 = 0.67$, which is less than the new overall rate, $R_2 = 1.25$. To check this conclusion, let us calculate the rate that all three species would provide.

$$R_3 = \frac{0.1 \times 20 + 1 \times 3 + 2 \times 10}{1 + 0.1 \times 10 + 1 \times 2 + 2 \times 15} = 0.74$$

This rate is less that $R_2 = 1.25$, as expected.

Table 7.2: Parameters of the bluegill experiments of Werner and Hall [74]

| | | | | Prey Density | | | | | |
| | | | | Low | | Medium | | High | |
Prey	e_i	h_i	e_i/h_i	λ_i	R_i	λ_i	R_i	λ_i	R_i
1	371	1.2	309.167	0.035	12.496	0.207	61.564	1.057	172.881
2	108	1.2	90.000	0.025	14.658	0.147	65.086	0.751	149.327
3	18	1.2	15.000	0.009	14.661	0.056	62.833	0.285	136.226

Note: The λ_i are based on Fig. 4 of [74].
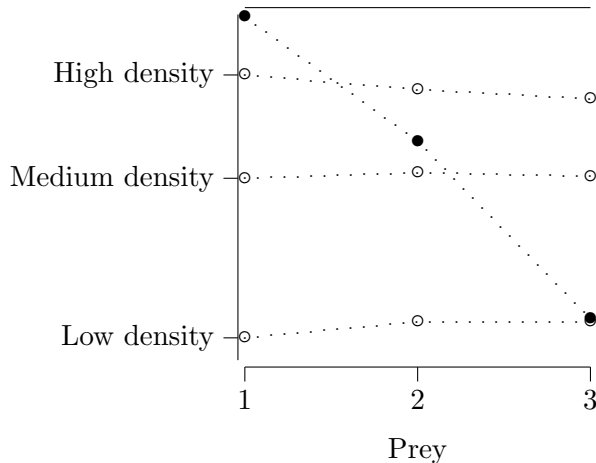
Figure 7.2: Choice of an optimal prey set in the experiment of Werner and Hall [74]. The vertical axis is on a logarithmic scale. Filled circles ($\bullet$) show e_i/h_i for prey type i, with the prey types ranked in order of descending e_i/h_i. Open circles ($\circ$) show $R_i = A_i/T$ for a diet consisting of prey types $1, 2, \ldots, i$. For each density, the optimal diet consists of all prey types for which $e_i/h_i \geq R_{i-1}$. Figure is based on data in Table 7.2

7.2 Diet choice among bluegill sunfish

The diet breadth model has been tested many times, often with great success. One such test was published by Werner and Hall [74], using data from the bluegill sunfish (*Lepomis macrochirus*) feeding on the zooplankton (*Daphnia magna*). In their experiments, a mixture of *Daphnia* of varying size was stirred into the pool that contained the sunfish. The fish were later sacrificed to find out how many *Daphnia* in each size class were taken. Handling time was the same for all of the size categories of prey, but e_i was largest for large prey. Thus, the theory predicts that the sunfish should take all prey sizes if prey are rare but should take only large prey if prey are common. To test this prediction, Werner and Hall repeated their experiment several times with different concentrations of prey. The parameters of their experiments are shown in table 7.2.

In the table, R_1 is the rate of return from a diet that includes only prey type 1, R_2 is the return from a diet including types 1 and 2, and R_3 is the return from a diet including all three types. The table shows that when prey density is low, R_3 greater than R_1 or R_2. Consequently, the optimal diet includes all three types of prey. We should expect the proportion of each prey type in the bluegill's diet to mirror the frequencies (the λ_i) with which they are encountered. At medium density, R_2 is larger than R_1 or R_3, so the optimal diet includes prey types 1 and 2 but not type 3. At high density, the optimal diet includes only prey type 1.

These predictions may be easier to grasp when the data in table 7.2 are presented in graphical form, as is done in Fig. 7.2. There, the e_i/h_i value for each prey species is shown as a filled circle and the R_i values are shown as open circles. For each level of prey density, the optimal diet consists of those prey items for which $e_i/h_i > R_{i-1}$. This is true of all three prey items at low density, of items 1 and 2 at medium density, but only of item 1 at high density. Thus, we end up with the same predictions from the graph as from the table.
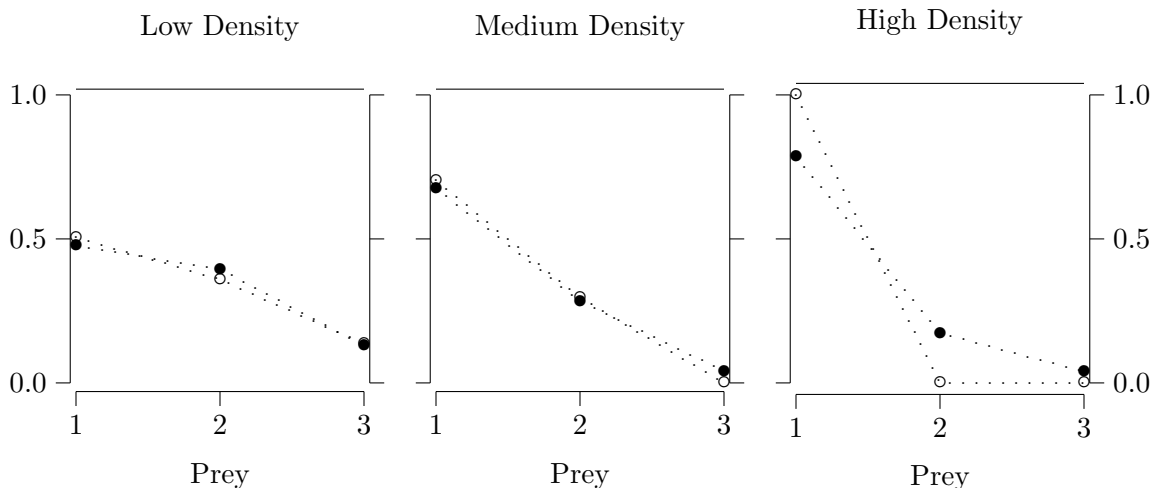
Figure 7.3: Results from bluegill experiment of Werner and Hall [74]. Filled circles show the frequencies of prey types in the stomach contents of the fish. Open circles show the frequencies predicted by the model.

So what did the bluegills actually do? The results are shown in Fig. 7.3. The open circles there show the prey-type frequencies that were predicted by the model and the filled circles show the frequencies among the prey that were recovered from the stomachs of the fish. For the low- and medium-density experiments, the agreement between prediction and observation is remarkable. In the high-density experiment, the agreement is also good, but it is not perfect. The fish took more prey in size-class 2 than the model says they should have. Similar discrepancies have appeared in experiments published by other authors, and these authors often make the same suggestion: a forager cannot make optimal decisions without sampling its environment. Perhaps these bluegills were taking the occasional prey from size-category 2 in order to make sure that they were not missing out on something good.

7.3   The patch model

We turn now to a different problem. Many resources occur in patches. Patches exist on many different scales. For example, a pine cone is a patch to a chickadee, but not to a human forager. For the human, an entire grove of pinyon pines might represent a patch. But regardless of these differences in scale, these foragers all face the same decision: How long should one forage in a patch before moving to another?

As before, I assume that selection favors the strategy that maximizes the energy obtained in some fixed time T. Let $g(x)$ denote the energy acquired from a patch if one forages there for x minutes. Then the total energy gained is $Ng(x)$, where N is the number of patches visited. If it takes on average a minutes to travel from patch to patch, then the number of patches visited must satisfy $N \cdot (x + a) = T$. The optimization problem is

$$\max w = Ng(x)$$
$$\text{subject to: } N \cdot (x + a) = T$$

In the literature, this problem is referred to as the *Marginal Value Theorem* after a paper by Eric Charnov [10]. It is similar to the Smith-Fretwell model and can be solved by repeating the steps
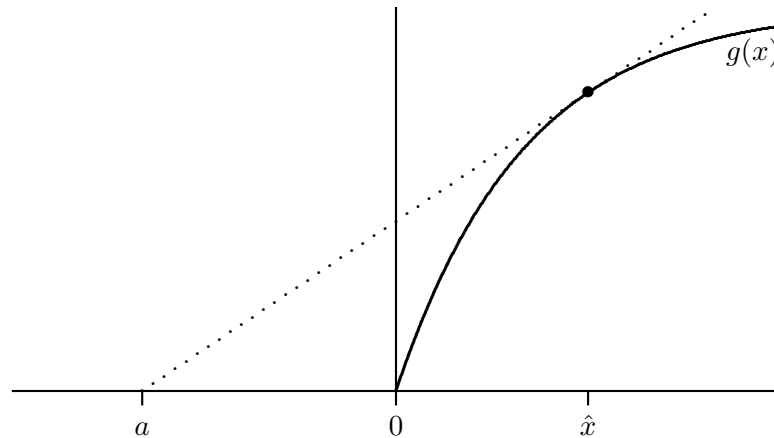
Figure 7.4: The patch model

Between-patch travel time, a, is measured from the vertical line towards the left; in-patch foraging time, x, is measured from the vertical line towards the right. Curve $g(x)$ measures the energetic gain from in-patch foraging. The dotted line shows that $\hat{x}$ satisfies the first-order condition for a maximum, Eqn. 7.4.

used there (see Ch. 6). This gives

$$\ln w = \ln T + \ln g(x) - \ln(x + a)$$
$$\frac{d \ln w}{dx} = \frac{g'(x)}{g(x)} - \frac{1}{x + a} \tag{7.3}$$

Setting the derivative equal to zero gives the first-order condition for a local maximum:

$$g'(x) = \frac{g(x)}{x + a} \tag{7.4}$$

The graph illustrating this solution is shown in figure 7.4. Note that the straight dotted line with slope $g(\hat{x})/(\hat{x} + a)$ is tangent to $g(x)$ at $\hat{x}$.

● EXAMPLE **7–1**

If $g(x) = 1 - e^{-x}$ and $\hat{x} = 1.5$, then what is the travel time a?

○ ANSWER

First define the gain function:

```
> g := proc(x)  1.0 - exp(-x) end;

   g := proc(x) 1.0 -exp(-x) end
```

Then solve for the equilibrium:

```
> solve(subs(x=1.5,
>  diff(g(x), x)) = g(1.5)/(1.5+a), a);

   1.981689071
```

Eqn. 7.4 would be satisfied not only by a local maximum, but also by a local minimum. To separate the two possibilities, we need the second-order condition discussed in section 3.6. The

Table 7.3: Results from Cowie's [16] experiment on foraging by Great Tits

Variable	Hard environment	Easy environment
Time in patch	73.67	45.80
Travel time	21.03	4.76

answer is a local maximum only if the second derivative of $\ln w$ is negative. Differentiating Eqn. 7.3 gives

$$\frac{d^2 \ln w}{dx^2} = \frac{g''(x)}{g(x)} - \left(\frac{g'(x)}{g(x)}\right)^2 + \frac{1}{(x+a)^2}$$

Now substitute the first-order condition (Eqn. 7.4) to obtain

$$\left.\frac{d^2 \ln w}{dx^2}\right|_{x=\hat{x}} = \frac{g''(\hat{x})}{g(\hat{x})}$$

Since $g(x)$ is always positive, the sign of this result is the same as the sign of $g''(\hat{x})$. Thus, $\hat{x}$ is a local maximum if g increases at a decreasing rate because in that case it will have a negative second derivative.

$\star$ EXERCISE **7–1** In the patch model, suppose that $a = 1$, and that the gain function is $g(x) = 1 + x^2$. (1) Plot this function. (Hint: In your plot command, specify the vertical range using a command such as `plot(g(x), x=0..2, g=0..5)`.) (2) At what value of x is the first-order condition (Eqn. 7.4) satisfied? (3) Use the second-order condition to determine whether this value is a local maximum. (4) Confirm your answer by plotting the fitness function.

7.4   How long do great tits stay in patches?

The patch model has been used many times in field studies and experiments. In one series of experiments, Richard Cowie [16] studied great tits (*Parus major*) foraging in an artificial environment. Cowie had constructed a series of artificial trees (really just arrangements of poles), each containing several patches consisting of holes filled with sawdust and meal worms.

The patch model predicts that the tits will forage longer in each patch if the travel time between patches is large. Cowie wanted to test this prediction, but had a problem. The obvious way to make travel time large is to move the patches far apart, but Cowie's aviary was too small. He tried another approach. Each patch was fitted with a lid, which the birds had to pry out before they could begin foraging. In some experiments, the lids were loose and easy to remove. In others, the lids were tight and had to be pried out. Cowie called these the "easy environment" and the "hard environment," respectively. Cowie's results, shown in table 7.3, are just as the patch model predicts. Tits spend more time foraging within patches when travel time is large.

These results are in qualitative agreement with the patch model, but are they also in quantitative agreement? To find out, Cowie considered a gain curve of form

$$g(x) = Q[1 - e^{-mx}]$$

where x is the time in patch, and Q and m are constants that determine the shape of the curve. Cowie estimated these constants from his data, using a method that he did not describe in detail.

He then used Eqn. 7.4 to predict the optimal time in patch in each experiment. It turned out that the birds usually stayed longer in patches than the model predicted.

What might account for this discrepancy? Cowie proposed adding the energetic costs of travel and search into the model. Can you think of other alternatives?

Chapter 8

A Graphical Approach to Evolutionary Game Theory

The optimization models discussed in in previous chapters all assume that the value of a strategy to one individual does not depend on the strategies of others. This is often a poor assumption. For example, a female should produce male offspring if other females are producing females. If you are cooperative, then perhaps I should cooperate too. If one tennis player serves consistently to the left side of the court, then the other should position himself to defend against such serves. If everyone else drives on the left side of the road, then that is where I should drive too. In all these cases, payoffs are said to be *frequency dependent* because the optimal strategy depends on the frequencies of strategies within the population. *Evolutionary game theory* was developed to find evolutionary equilibria in such cases [49].

To keep the exposition simple, I will stick to the simplest possible genetic system: that of asexual, haploid inheritance. In this system, each individual has just one parent, and each daughter inherits her mother's strategy. The results, however, will also apply (with restrictions) to organisms with diploid inheritance. I will also consider only two alleles, A and B. These alleles encourage their bearers to adopt behaviors called strategy A and strategy B. As usual, the frequencies of A and B are written as p_A and $p_B = 1 - p_A$. Under haploid inheritance these are not only allele frequencies, they are also the frequencies with which strategies A and B are exhibited in the population.

In evolutionary game theory, an equilibrium is called an *evolutionarily stable strategy*, or ESS. An ESS is a strategy that is unbeatable: Once fixed in the population, an ESS can resist invasion by all alternative strategies.

How can such equilibria be found? It will not help to search for the strategy that maximizes mean fitness, for with frequency-dependent fitnesses there is no longer any guarantee that selection will maximize mean fitness. Indeed, the ESS may even minimize mean fitness. Nonetheless, inequality 4.4 still holds (see p. 21). In the present context, it says that p_A will increase at the expense of p_B whenever

$$w_A > w_B \tag{8.1}$$

where w_A and w_B are the genic fitnesses of the two alleles. This principle underlies evolutionary game theory.

There are two types of ESS to consider: those at which the population is dominated by one of the two strategies, and those at which both strategies are maintained at some intermediate frequency. The former are called *pure ESSes*, and the latter *mixed ESSes*.
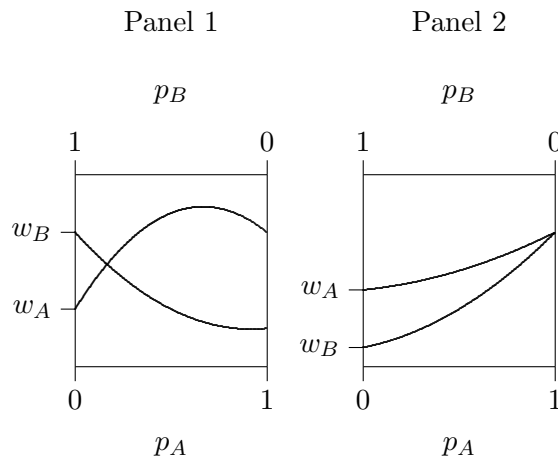
Figure 8.1: Two Pure ESSes at $p_A = 1$

8.1 Pure ESSes

It is often helpful to plot the fitnesses of both strategies against the frequency of one strategy. This is done for two hypothetical cases in Figure 8.1. In each panel, fitnesses are plotted on the vertical axis. As one moves from left to right, p_A increases from 0 to 1, and p_B decreases from 1 to 0. Let us use these graphs to figure out whether each of two strategies (A and B) can resist invasion by the other.

A resists invasion by B if B is selected against when A is common. To figure out whether this is so, we need to look at the right edge of each graph, where $p_A = 1$ and $p_B = 0$. In panel 1, $w_A > w_B$ when $p_A = 1$. Thus, in a population dominated by A, a rare B would have lower fitness. In other words, A resists invasion by B. Now consider strategy B. It is common at the left edge of the graph, where $p_A = 0$ and $p_B = 1$. At that edge, $w_B > w_A$. Consequently, a population dominated by B would resist invasion by A. Either of these strategies would resist invasion by the other.

These results illustrate the general principal that:

Result 8.1 *Strategy X resists invasion by Y if $w_X > w_Y$ when $p_X = 1$.*

Panel 2 of Fig. 8.1 illustrates a case in which this result is not useful. At the right edge of that graph, both strategies have the same fitness. Yet it is clear that A does resist invasion by B: if $p_A < 1$ then $w_A > w_B$, and p_A must increase toward $p_A = 1$. This is always the case when the derivative of w_B exceeds that of w_A where $p_A = 1$. Thus, we have another way to recognize an ESS:

Result 8.2 *A strategy X resists invasion by another strategy Y if $w_X = w_Y$ when $p_X = 1$, but $w_X > w_Y$ when p_X is slightly less than 1.*

Results 8.1 and 8.2 state conditions under which one strategy resists invasion by another. A strategy that resists invasion by *all* alternate strategies is called an *evolutionarily stable strategy*, or ESS. Results 8.1 and 8.2 make is possible to find a ESS algebraically, even when it isn't possible to draw graphs such as those the figure.

⋆EXERCISE **8–1** Use results 8.1 and 8.2 to show that $p_A = 0$ is an ESS in both panels of figure 8.2, assuming that A and B are the only feasible strategies.
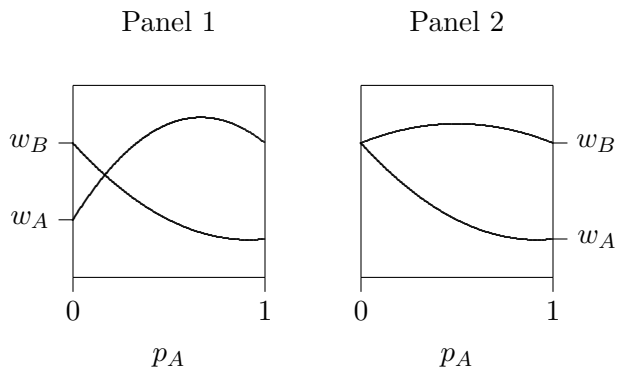
Figure 8.2: Two Pure ESSes at $p_A = 0$
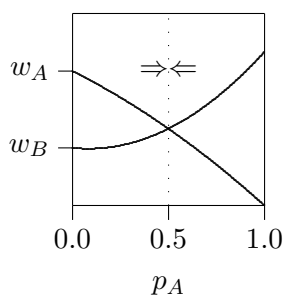
Figure 8.3: A Mixed ESS at $p_A = 0.5$

8.2   Mixed ESSes

A mixed ESS is an equilibrium at which both strategies are maintained at nonzero frequencies. Figure 8.3 illustrates such a case. In this figure, the two fitness curves cross. At the point of intersection, the two strategies have equal fitness and there is no tendency to change. Thus, the intersection is an equilibrium. To figure out whether this equilibrium is stable, we need to ask what would happen to a population whose value of p_A differed slightly from the equilibrium value. To the left of the equilibrium, $w_A > w_B$, so p_A tends to increase. To the right, the reverse is true, so p_A decreases. Either way, the the movement is toward the equilibrium. Thus, if p_A were perturbed away from this equilibrium, it would tend to return. This is a stable equilibrium.

The conditions for a mixed ESS may be summarized as follows:

Result 8.3 *The point $p_A = \hat{p}_A$ is a mixed ESS if the $w_A - w_B$ is positive to the left of $\hat{p}$, zero at $\hat{p}$, and negative to the right.*

8.3   Pairwise Interactions

Consider now the special case of interactions between pairs of individuals. Frequency dependence arises because the fitness of each individual depends on the opponents he encounters, and the likelihood that he will encounter one type rather than another depends on the frequencies of those types.

As before, I consider two strategies only, and label them A and B. To distinguish between the two individuals in an interaction, I will label them Ego and Opponent. Ego's fitness depends both on his own strategy and on that of his opponent. Let $\Pi(i, j)$ denote the payoff to Ego when Ego

Table 8.1: Payoffs from pairwise interactions: general case

Ego		Opponent	
		A	B
	A	a	b
	B	c	d
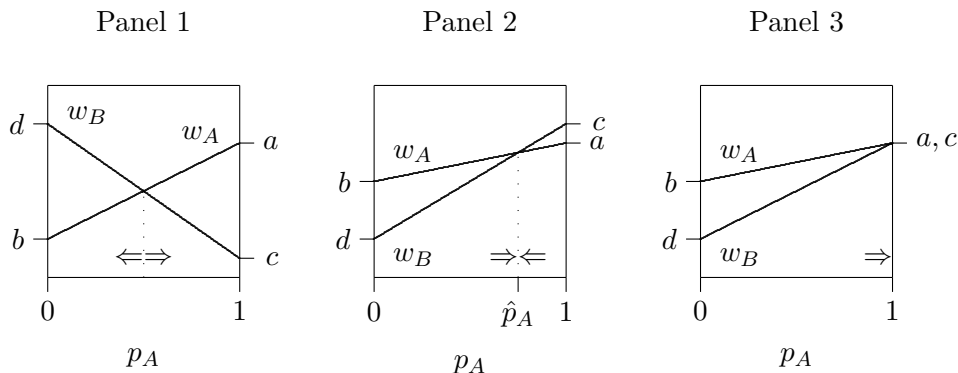
Figure 8.4: Pairwise interactions. Panel 1: Pure ESSes at $p_A = 0$ and $p_A = 1$. Panel 2: Mixed ESS at $\hat{p}_A$.

plays strategy i against an opponent who plays j. The various payoffs can be arranged into a matrix whose ijth entry is $\Pi(i,j)$. One such matrix is shown in Table 8.1, where $\Pi(A, A) = a$, $\Pi(A, B) = b$, and so on. Ego encounters opponents of types A and B with probability p_A and $1 - p_A$, respectively. Thus, if Ego adopts strategy A his fitness will be

$$
\begin{aligned}
w_A &= ap_A + b(1 - p_A) \\
&= b + p_A(a - b)
\end{aligned}
\tag{8.2}
$$

Similarly, if Ego adopts strategy B his fitness will be

$$
\begin{aligned}
w_B &= cp_A + d(1 - p_A) \\
&= d + p_A(c - d)
\end{aligned}
\tag{8.3}
$$

This gives us simple expressions for the fitnesses of the two strategies as a function of p_A. What makes these expressions interesting is that they are linear—straight lines—as shown in figure 8.4. These graphs are read the same way as the others. In panel 1, there are ESSes at $p_A = 0$ and at $p_A = 1$. There is an unstable equilibrium where the lines cross. In panel 2, the two pure strategies are unstable, but there is a mixed ESS where the lines cross. In panel 3, A is stable and B unstable.

8.4 The Hawk-Dove Game

The Hawk-Dove game models conflicts between individuals over resources. Animals often fight, yet injuries are uncommon, even in carnivore species such as the wolves in figure 8.5. It used to be argued, by way of explanation, that if animals fought more aggressively the resulting injuries would threaten the survival of the species. Evolutionary game theory was introduced in an effort to provide an alternative explanation, which relies on selection among individuals rather than among groups [50, 48]. The version of the model discussed here is based on [49].

Imagine that two individuals both want the same resource. Each must choose between two strategies: The "hawk" strategy fights for the resource; the "dove" strategy gives up without a

Figure 8.5: Wolves *(Canis lupus)* fighting

Table 8.2: Payoffs to Ego in the Hawk-Dove Game

		Opponent	
		Hawk	Dove
Ego	Hawk	$(V - C)/2$	V
	Dove	0	$V/2$

fight. The resource has value: it adds V to the fitness of the winner. Furthermore, the loser of the fight may be injured, and this possibility subtracts C from the loser's fitness. The fitness effects of this interaction are shown in table 8.2.

The entries in this table reflect the following reasoning. (1) When hawk meets hawk, they fight and each has an equal chance of winning (thus gaining V) or losing (thus gaining $-C$). The expected fitness benefit is $(V - C)/2$. (2) When hawk meets dove, the dove withdraws and the hawk wins V without a fight. (3) When dove meets hawk, the dove withdraws without a fight, and there is no effect on its fitness. (4) When dove meets dove, there is no fight, and each individual has an equal chance of receiving the benefit V. The expected effect on each dove's fitness is thus $V/2$. If fitnesses are graphed against the frequency p_H of hawks, the diagram describing this interaction is as shown in figure 8.6.

The figure's left panel illustrates the case in which $V > C$, so that the hawk strategy has a positive fitness effect even when hawks are common. In this case, hawk is the only ESS. The right panel illustrates the case in which $V < C$, so that hawk versus hawk yields a negative contribution to fitness. In this case, there is a mixed ESS with an intermediate frequency of hawks and doves.

⋆ EXERCISE **8–2** Draw a graph of the Hawk-Dove game for the case in which $V = 1$ and $C = 2$.

⋆ EXERCISE **8–3** Use Result 8.3 to prove that the mixed equilibrium in the Hawk-Dove game is stable.

The lesson taught by this model is that selection between individuals can favor a strategy other than "Hawk." It leads to an equilibrium at which individuals fight over resources only occasionally.
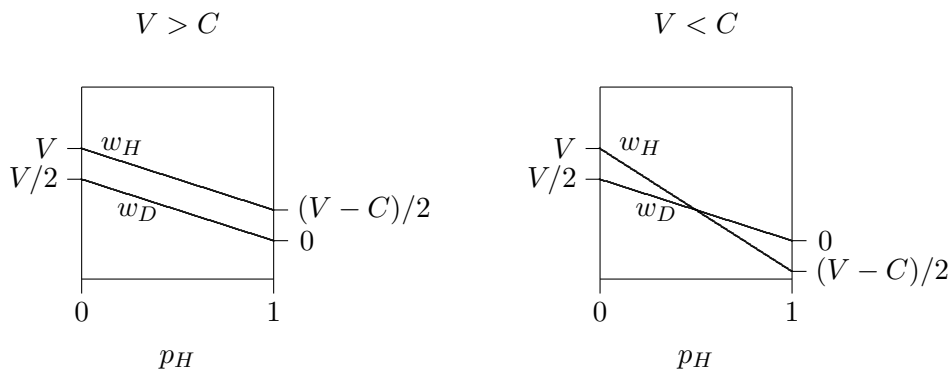
Figure 8.6: Hawk-Dove Model
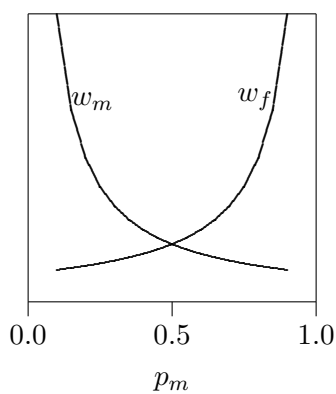
Figure 8.7: Sex Ratio. p_m is the fraction of the population that is male, w_m is the fitness of a male and w_f is the fitness of a female.

At other times, one individual gives way to the other without a fight. The frequency of serious fights diminishes still further when there is some role-asymmetry between the contestants. This may explain why, in disputes over territories, the owner of the territory usually defeats the intruder.

8.5   The Sex Ratio

When a mother produces offspring, what fraction should be boys and what fraction girls? This question puzzled Darwin because the ratio of boys to girls need have no affect on the number of the mother's offspring. Ronald Fisher solved the puzzle by counting grandchildren rather than children [23]. The idea is not hard to grasp. Suppose that a fraction p_m of the offspring in some generation are sons, the remaining $p_f = 1 - p_m$ being daughters. If N offspring are produced in all, then there are Np_m sons and Np_f daughters. If N grandchildren are produced, then mothers average $N/(Np_m) = 1/p_m$ grandchildren per son and $1/p_f$ grandchildren per daughter. If $p_m > p_f$, then daughters are more productive than sons and mothers would gain by biasing their progeny in favor of daughters. On the other hand, if $p_m < p_f$, mothers should bias their progeny in favor of sons. This moves the population toward a balanced sex ratio.[1] The graph of this process is shown in Figure 8.7.

$\star$ EXERCISE **8–4** Suppose that sons were twice as expensive to produce as daughters, so that mothers were forced to choose between producing one son or two daughters. Then the fitness of a mother who

---

[1]There is a subtle assumption here that mating is at random. See Charnov [11].

produces sons is $w_m = 1/p_m$, whereas that of a mother who produces daughters is $w_f = 2/(1-p_m)$. Draw a graph using these fitnesses, and use it find the ESS.

Chapter 9

The Algebra of Evolutionary Game Theory

A strategy that resists invasion by all rare mutants is called an *evolutionarily stable strategy* or ESS. This chapter, like the last one, is about methods for finding such strategies. The previous chapter focused on graphical methods, which are excellent for building intuition. They fail however when there are more than two strategies. Furthermore, one cannot make graphs without assigning numerical values to all parameters. Consequently, the graphical approach makes it hard to find general answers. To relax these restrictions, we turn in this chapter to algebraic methods.

These algebraic methods deal exclusively with rare invaders, and in doing so they lose something. The graphical methods of the previous chapter could deal with common alleles and rare alleles with equal ease. The methods developed below work only when one allele is very rare. Consequently, they tell only part of the story.

The sections that follow will introduce methods used for several types of game. The first section deals with pairwise games with discrete strategies, the second with N-person games with discrete strategies, and the last deals with games with continuous strategies.

9.1 Pairwise games with discrete strategies

A pairwise game is one in which there are two opponents. A game with discrete strategies is one in which, on each turn, each player must choose between a limited set of alternative strategies. For example, in the Hawk-Dove game discussed below there are only two strategies: Hawk and Dove.

9.1.1 Recognizing the ESS

How can we recognize an ESS? I will start with the answer and then explain why it makes sense. First some notation. The name "Ego" will refer to the individual whose fitness is being calculated. In this section I will assume that the game is between a pair of individuals so that Ego has a single opponent. Below, this logic will be extended to the case of games between N individuals.

The symbol I will represent some strategy in which we are interested. Let $\Pi(I, J)$ denote the payoff to an individual who plays I against an opponent who plays J. There are two ways in which I can be an ESS. Strategy I is an ESS if *either* of the following conditions holds for every alternative strategy J:

$$\Pi(I, I) > \Pi(J, I) \tag{9.1}$$

or

$$\left. \begin{array}{rcll} \Pi(I,I) & = & \Pi(J,I) & \text{and} \\ \Pi(I,J) & > & \Pi(J,J) & \end{array} \right\} \tag{9.2}$$

To understand these conditions, the first step is to realize that we are only interested in what happens when I is very common. This is a matter of definition: an ESS is a strategy that resists invasion by *rare* mutants. It need not resist invasion by common ones. This restriction is reasonable because every mutant is rare to begin with.

If the mutant J is rare, then nearly all individuals (of either type) will play against I. Consequently, the fitness of I is $\Pi(I,I)$ and that of J is $\Pi(J,I)$. Thus, strategy I has the higher fitness if inequality 9.1 is satisfied.

Inequality 9.1 is just another way of saying that strategy I has higher fitness when strategy J is rare. We have seen this condition already—it is a paraphrase of result 8.1 from chapter 8. Refer back to panel 1 of Figure 8.4 on page 51. There, strategies A and B are both ESSes because each has a higher fitness when common than its alternative. These strategies each satisfy inequality 9.1. But this inequality cannot help us with panels 2 and 3. In panel 3, $w_A = w_B$ when $p_A = 1$, so $\Pi(I,I) = \Pi(J,I)$ and inequality 9.1 does not hold. Yet we already know that strategy A is an ESS. Condition 9.2 is designed to deal with cases such as this.

To justify condition 9.2, write the fitnesses of the two types as

$$\begin{array}{rcl} w_I & = & Q_0\Pi(I,I) + Q_1\Pi(I,J) \\ w_J & = & Q_0\Pi(J,I) + Q_1\Pi(J,J) \end{array}$$

where Q_0 is the probability that Ego's opponent is of type I and Q_1 is the probability that this opponent is of type J. (The subscript counts the number of opponents of type J.) Since we are concerned here with the case in which $\Pi(I,I) = \Pi(J,I)$, the difference between these two fitnesses is

$$w_I - w_J = Q_1[\Pi(I,J) - \Pi(J,J)]$$

Thus, $w_I > w_J$ whenever (9.2) holds and $Q_1 > 0$.

9.1.2   The Hawk-Dove game

We used graphical methods to study the Hawk-Dove game in section 8.4. Here, we re-visit that game in order to illustrate the algebraic approach described above. For ease of reference, I re-write the payoff matrix (table 8.2, page 52) using the new notation:

$$\begin{array}{rcl} \Pi(H,H) & = & (V-C)/2 \\ \Pi(H,D) & = & V \\ \Pi(D,H) & = & 0 \\ \Pi(D,D) & = & V/2 \end{array}$$

Recall from the graphical analysis that Dove is never an ESS. This conclusion also follows from the observation that since, $\Pi(D,D) < \Pi(H,D)$, Dove satisfies neither condition 9.1 nor 9.2. The graphical analysis also showed that Hawk is an ESS when $V > C$. To verify this using condition 9.1, note that $\Pi(H,H) > \Pi(D,H)$ when $V > C$.

But what if $V < C$? In that case no pure strategy can be stable, because neither pure strategy satisfies either of the conditions for an ESS. Let us therefore consider a mixed strategy. Let I denote

a strategy that plays Hawk with probability p and Dove with probability $1 - p$. In a population playing I, the fitnesses of hawks and doves are

$$
\begin{aligned}
\Pi(H, I) &= p\Pi(H, H) + (1 - p)\Pi(H, D) \\
&= p(V - C)/2 + (1 - p)V \\
\Pi(D, I) &= p\Pi(D, H) + (1 - p)\Pi(D, D) \\
&= (1 - p)V/2
\end{aligned}
$$

These fitnesses must be equal if I is an ESS.[1] This requires that $p = V/C$.

⋆ EXERCISE **9–1** Verify that $\Pi(H, I) = \Pi(D, I)$ when $p = V/C$.

The algebra above shows that, when the population is playing I, a mutant playing hawk would have exactly the same fitness as a mutant playing dove. What about a mutant playing some mixed strategy? It turns out that their fitnesses are identical. Thus, all strategies—pure or mixed—receive the same payoff when the population is dominated by strategy I, so strategy I satisfies the first half of condition 9.2.

⋆ EXERCISE **9–2** Explain why any mixed strategy would receive the same payoffs as Hawks and Doves in a population dominated by strategy I.

The second half of condition 9.2 requires that $\Pi(I, J) > \Pi(J, J)$ for any J. We cannot guarantee that I is an ESS until we have checked that this condition holds for all conceivable invaders. It must resist invasion by hawks, by doves, and by other invading mixed strategies. Each of these potential invaders requires its own paragraph of algebra. I illustrate this rather tedious process in the appendix to this chapter. Here, I take a short cut that avoids a lot of algebra.

9.1.3 Shortcut for pairwise games with only two strategies

You can avoid a lot of algebra if you are dealing with a symmetric pairwise game in which there are two pure strategies. Recall that, in strategy I of the Hawk-Dove game, p was chosen so as to make the fitness of Hawk equal to that of Dove. These fitnesses would also be equal in a population comprising a fraction p of Hawks and a fraction $1 - p$ of Doves. Thus, p is a point where the lines cross in a graph like those in panels 1 and 2 of Figure 8.4. The question is, which of these graphs is relevant? p is not an ESS in panel 1 but is an ESS in panel 2. There is an easy way to distinguish the two cases: in panel 1 the two pure strategies ($p = 0$ and $p = 1$) are stable, whereas in panel 2 they are unstable. In the Hawk-Dove game the two pure strateies are unstable, so panel 2 is relevant and the mixed equilibrium must be stable. This reasoning is fair because we are dealing with a pairwise interaction, where the lines in the graph necessarily are straight. In N-person games this reasoning would not apply, and you would need to do the sort of analysis illustrated in the appendix to this chapter.

---

[1]If Hawk had a higher fitness than Dove, then it would also have a higher fitness than I since I is a mixture of Hawk and Dove. Consequently I cannot be an ESS unless the two pure strategies have equal fitness.

Figure 9.1: *Perissodus eccentricus*, a scale-eating fish.

9.2   Asymmetric games

In Lake Tanganyika, in the African Rift Valley, there is a fish (*Perissodus eccentricus*) that lives by eating the scales of larger fish [39]. It attacks from behind, snatching a scale or two from the side of its prey. *P. eccentricus* is aided in this effort by an interesting adaptation: its mouth is twisted to one side (figure 9.1). Some individuals have mouths twisted to the right (so-called right-handed fish) whereas others are left-handed. Right-handers attack from the left, left-handers from the right. Prey try to avoid attacks, but they cannot watch their left and right at the same time. If they are watching left, then attacks from the left seldom succeed but attacks from the right usually do.

This is an example of an asymmetric game—one in which the two parties to the contest play different roles. The scale-predator chooses between two strategies: right-handed (RH) and left-handed (LH). These strategies are not relevent to the prey, whose choice is between: watch left (WL), and watch right (WR). In an asymmetric game, the row and column player need not even have the same number of strategies to choose from.

For asymmetric games, the payoff matrix must include not only the payoffs to row but also those to column. This is done in Table 9.1 by writing each entry in the matrix as a pair of numbers (j, k), where j is the payoff to row and k is the payoff to column. The payoff matrix shown here embodies the following assumptions: (1) When the predator attacks the side that is defended, the attack fails and neither party gains or loses anything. (2) When the predator attacks the undefended side, the predator gains a units of fitness and the prey loses b units.

9.2.1   Response functions

We can study this game using the method of response functions. In this model, x is the probability that a random predator is right-handed, and y is the probability that a random prey watches left (guarding against right-handed attacks). The method of response curves asks first how the optimal y depends on x, and then how the optimal x depends on y.

Let $R_X(y)$ denote the value of x that maximizes the fitness of predators given y. It is called the

		Prey	
		WL	WR
Scale-Eater	RH	$(0,0)$	$(a,-b)$
	LH	$(a,-b)$	$(0,0)$

Table 9.1: Payoffs to scale-eating fish and their prey. Within each pair of parentheses, the first number is the payoff to row and the second is the payoff to column.

response function of the predator. Similarly, let $R_Y(x)$ (the response function of the prey) denote the value of y that maximizes the fitness of prey given x. We will use these functions to search for (x, y) values such that x is the best reply to y, and y is also the best reply to x.

If the value of y were fixed, then fitnesses would be density-independent and selection would increase the mean fitness of the predator population by moving x toward $R_X(y)$. Similarly, if x were fixed, then selection would move y toward $R_Y(x)$. Even when x and y are not fixed, selection tends to change the value of x whenever $x \neq R_X(y)$ and to change that of y whenever $y \neq R_Y(x)$. Thus, equilibria occur where x and y are each equal to their corresponding response functions.

To calculate the optimal response function of the scale-eating fish, we begin by writing down the fitnesses of strategies RH and LH as a function of y. From Table 9.1, these are

$$w_{RH} = (1 - y)a \qquad \text{and} \qquad w_{LH} = ya$$

The mean fitness of the predator population is

$$
\begin{aligned}
w_1 &= x w_{RH} + (1 - x) w_{LH} \\
&= x(1 - y)a + (1 - x)ya \\
&= x(1 - 2y)a + ya
\end{aligned}
$$

Notice that if $y < 1/2$, then w_1 increases with x so the optimal value of x is one. On the other hand, if $y > 1/2$ then w_1 decreases with x so the optimal x is zero. Thus, the optimal response function of the predator is

$$R_X(y) = \begin{cases} 1 & \text{if } y < 1/2 \\ 0 & \text{if } y > 1/2 \end{cases}$$

This expresses the obvious fact that left-handers have the advantage if prey are guarding against right-handers, and vice versa. An analogous argument gives the optimal response function of the prey:

$$R_Y(x) = \begin{cases} 1 & \text{if } x > 1/2 \\ 0 & \text{if } x < 1/2 \end{cases}$$

$\star$ EXERCISE **9–3** Calculate the optimal response function of the prey.

Let us now use these response functions to find the game's equilibrium. An equilibrium point is a pair of values $(\hat{x}, \hat{y})$ such that

$$\hat{x} = R_X(\hat{y}) \qquad \text{and} \qquad \hat{y} = R_Y(\hat{x})$$

An easy way to find the equilibrium is to draw a graph such as that in Figure 9.2, where the horizontal axis is x and the vertical axis is y. The prey response function, $R_Y(x)$, is graphed against the horizontal (x) axis as a solid line; the predator response function, $R_X(y)$, is graphed against the vertical (y) axis as a dashed line. The lines cross at the point where x and y both equal $1/2$, i.e. where the predators are divided equally between left-handers and right-handers and the prey spend half their time looking over each shoulder. When the system is at this point, neither species can gain any advantage by changing. Thus, this point is an equilibrium.

It is not yet clear, however, that this equilibrium is stable. Suppose that $x > 1/2$ and $y > 1/2$ so that the system is in the upper-right quadrant. Because most predators are right-handed, it is advantageous for the prey to defend against right-handers, so y will increase. But because most prey defend against right-handers left-handed predators flourish, and x will decrease. As a result, the system moves in the direction of decreasing x but increasing y, as indicated by the arrow in

Fish Response Curves
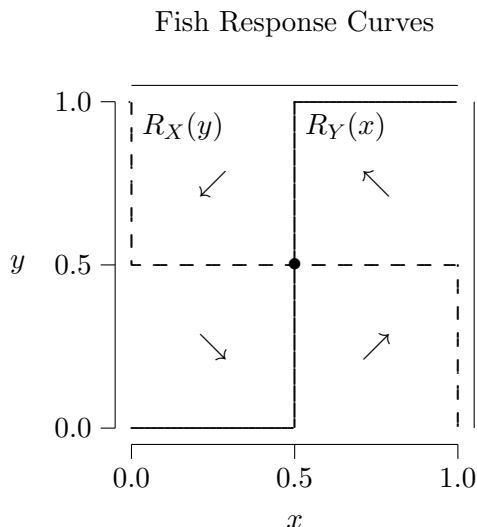
Figure 9.2: Response curves for scale-predators and their prey. Solid line shows the response curve $R_Y(x)$ of prey; dashed line shows the response curve $R_X(y)$ of predator. Arrows show the direction of motion. (I apologize for the form of the graph. No political message is intended.)

the upper-right quadrant of the figure. Similar reasoning justifies the arrows shown in the other quadrants of the figure. Taken together, these arrows suggest that the system will tend to cycle in a counter-clockwise direction. This is exactly what is shown by the field data.

There are several possibilities that we have not resolved. The cycles may be damped, growing smaller and smaller in magnitude until they eventually converge to the equilibrium in the middle. On the other hand, they may be undamped, continuing to swirl around forever. In order to find out, we would need to turn the model into a system of differential equations [38]. These methods are beyond our scope here.

9.3   Games with continuous strategies: The clutch size game

[2]Wasps lay eggs in fly pupae. Let us assume that the survival probability of young wasps is a linear function of crowding,

$$s = a - bc$$

where a and b are positive constants and c is the number of eggs laid in the pupa. Assume further that two wasps lay eggs in each pupa, so that

$$c = c_1 + c_2$$

where c_1 and c_2 are the eggs laid by wasps 1 and 2. The fitnesses of the two wasps are

$$\begin{aligned} w_1 &= c_1 s = c_1(a - b(c_1 + c_2)) \\ w_2 &= c_2 s = c_2(a - b(c_1 + c_2)) \end{aligned}$$

The optimal value of c_1 depends on the value of c_2 and vice versa. To find the equilibrium, we need game theory. I will treat c_1 and c_2 continuous variables so that I can use this example to illustrate the methods that are used for games with continuous strategy spaces.

There are two equivalent ways to attack this problem, which I will illustrate one at a time.

---

[2]I thank Eric Charnov for this example.
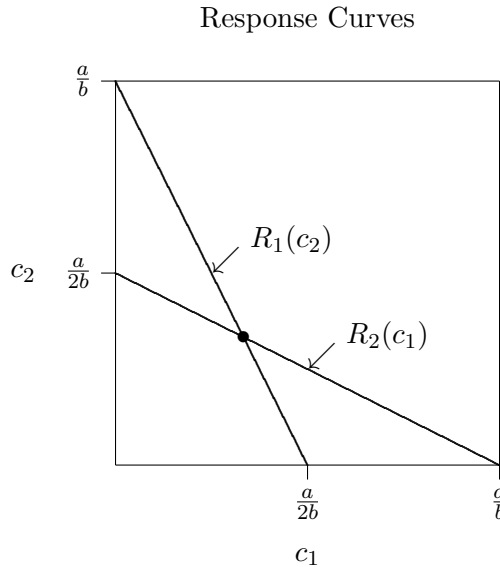
Response Curves

Figure 9.3: Response curves for wasp clutch size game

9.3.1 Response curves

To find a response curve for individual 2, we need to find the value of c_2 that maximizes fitness, taking c_1 as given. This requires that $\partial w_2 / \partial c_2 = 0$. The solution to this equation is the response curve for individual 2:

$$R_2(c_1) = (a/b - c_1)/2 \tag{9.3}$$

Similarly, the response curve for individual 1 is

$$R_1(c_2) = (a/b - c_2)/2 \tag{9.4}$$

These can be plotted on the same graph as shown in figure 9.3.

Let (c_1^*, c_2^*) denote the value of (c_1, c_2) at equilibrium. Graphically, (c_1^*, c_2^*) is the point where the two lines cross in figure 9.3. We can find this point algebraically as follows. First note that since, c_1^* and c_2^* are optimal responses to each other,

$$\begin{aligned} c_1^* &= R_1(c_2^*) \\ c_2^* &= R_2(c_1^*) \end{aligned}$$

Substituting each of these into the other gives

$$\begin{aligned} c_1^* &= R_1(R_2(c_1^*)) \\ c_2^* &= R_2(R_1(c_2^*)) \end{aligned}$$

These give

$$c_1^* = c_2^* = \frac{a}{3b} \tag{9.5}$$

The stronger the effect, b, of crowding, the fewer eggs should be laid in each pupa.

9.3.2 An alternative approach for symmetric games

In this game, wasps 1 and 2 face an identical decision problem so at equilibrium, their behavior must be the same. The method of response functions, shown in the previous section, works even

when this sort of symmetry does not exist. But where it does exist the problem can be solved more simply. Imagine that most wasps lay c_1 eggs, but a very few (rare mutants) lay c_2 eggs. If c_1 is an ESS, then the optimal value for the mutant is $c_2 = c_1$. We can find the ESS in two steps: The first step finds the optimal value of c_2 as a function of c_1; then the second asserts that this optimal value is the same as c_1.

First step   The mutant's fitness is

$$w_2 = c_2(a - b(c_1 + c_2)),$$

just as above. The question is, what value of c_2 will maximize w_2? Let us call this value c_2^*. The value of c_2^* is given above by Eqn. 9.3:

$$c_2^* = (a/b - c_1)/2$$

Second step   If c_1 is an ESS, then the most invasive mutant must be c_1 itself. The final step is thus to set $c_2^* = c_1$ and simplify. This yields

$$c_1 = c_2^* = (a/b - c_1)/2$$

This is an equation in one variable, c_1. Solving it gives

$$c_1 = \frac{a}{3b},$$

just as we obtained using response curves.

Note that if only one female parasitized each pupa, the optimal number of eggs would be larger: $R_1(0) = a/(2b)$. Parasitism reduces the optimal number of eggs laid by each female.

Appendix: Is the mixed strategy stable against invaders in the Hawk-Dove game?

This section completes the analysis of section 9.1.2 above by demonstrating that the mixed strategy, I, resists invasion by rare individuals playing either hawk, dove, or some mixed strategy that differs from I.

Resistance to invading Hawks   When J is Hawk, the second half of condition 9.2 becomes $\Pi(I, H) > \Pi(H, H)$, or
$$p\Pi(H, H) + (1 - p)\Pi(D, H) > \Pi(H, H)$$

After substituting values from the payoff matrix, this becomes $p(V - C)/2 > (V - C)/2$, which is true because $0 < p < 1$ and $V - C$ is negative. Thus, I resists invasion by Hawk.

Resistance to invading Doves   When J is Dove, the second half of condition 9.2 becomes

$$pV + (1 - p)V/2 > V/2$$

which is always true because the left-hand side is an average of $V/2$ and something larger than $V/2$. Thus, I resists invasion by Dove.

Resistance to invading mixed strategies   Suppose that mutants play a mixed strategy, J, which plays H with probability u and D with probability $1 - u$. When I is pitted against J, the payoff to I is

$$\Pi(I, J) = pu(V - C)/2 + p(1 - u)V + (1 - p)(1 - u)V/2$$

Here, pu is the probability that both contestants play H, $p(1 - u)$ is the probability that I plays H and and J plays D, and so on. When J is pitted against J, the payoff to each contestant is

$$\Pi(J, J) = u^2(V - C)/2 + u(1 - u)V + (1 - u)^2 V/2$$

We already know that $\Pi(I, I) = \Pi(J, I)$, because *every* strategy gets the same payoff against I. To ensure that I is an ESS using Eqn. 9.2, we must show that $\Pi(I, J) - \Pi(J, J) > 0$. The left side of this inequality reduces, after some algebra, to

$$\Pi(I, J) - \Pi(J, J) = (C/2)(V/C - u)^2, \tag{9.6}$$

which is clearly greater than zero whenever $u \neq V/C$. Thus, both of the conditions in Eqn. 9.2 are satisfied and strategy I is an ESS.

Chapter 10

N-Person Games with Discrete Strategies

10.1 Recognizing the ESS

It adds hardly any difficulty to generalize conditions 9.1 and 9.2 to the case of a game played between N competitors. Once again, I start with the answer. I assume now that Ego competes against K opponents that are drawn at random from some large population. Let $\Pi(I, J^x I^{K-x})$ denote the payoff to an individual who plays I against K opponents of whom x play J, and $K - x$ play I.[1]

As in the case of pairwise games, there are two conditions to consider: I is an ESS either if

$$\Pi(I, I^K) > \Pi(J, I^K) \tag{10.1}$$

or

$$\left.\begin{array}{rcl} \Pi(I, I^K) & = & \Pi(J, I^K) \quad \text{and} \\ \Pi(I, J^1 I^{K-1}) & > & \Pi(J, J^1 I^{K-1}) \end{array}\right\} \tag{10.2}$$

These conditions are perfectly analogous to conditions 9.1 and 9.2 and can be justified in the same way. For details, see the appendix to this chapter.

10.2 Hunting group size

The optimal size of a hunting group is $N \equiv K + 1$ and gives fitness w to each group member. If another member joins, fitness drops to $w - c$. There are two strategies:

Strategy C (Cooperate) Fights to keep others from joining. Fighting reduces the fitness of each C-strategist by k.

Strategy D (Defect) Doesn't fight.

The group will remain at optimal size if it contains at least one C.

Payoffs to row are

---

[1]This notation is *not* meant to imply that some number J is raised to the xth power and multiplied by the $(K - x)$th power of I. I and J are not numbers at all; they are simply symbols that indicate different strategies.
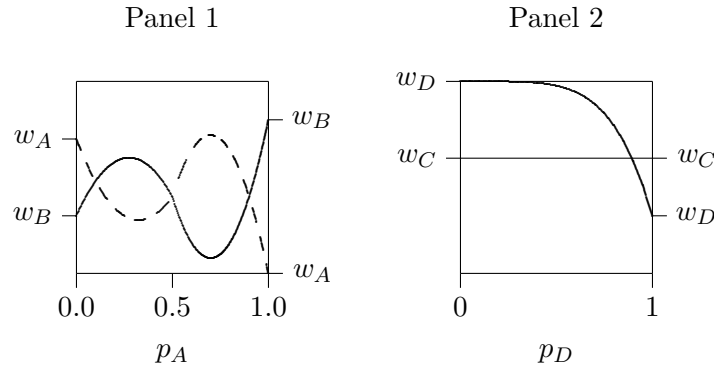
Panel 1   Panel 2

Figure 10.1: Equilibria in non-linear games
In panel 1 the, the two pure strategies ($p_A = 0$ and $p_A = 1$) are unstable. The mixed equilibrium at $p = 0.5$ is unstable. The other two mixed equilibria are stable. Panel 2 shows fitnesses in the game of hunting group size. The two pure strategies are unstable; the single mixed equilibrium is stable.

	At least one opponent plays C	All opponents play D
C	$w - k$	$w - k$
D	w	$w - c$

If $c < k$, D is an ESS. If $c > k$ there is no pure ESS, so consider a mixed strategy:

Strategy I (Mixed) play D with probability p

Payoffs to C and D in a population of Is are

$$\Pi(C, I^K) = w - k$$
$$\Pi(D, I^K) = p^K(w - c) + (1 - p^K)w = w - cp^K$$

since p^K is the probability that none of Ego's opponents play C. If I is an ESS, then the two pure strategies must have equal fitness, so $\Pi(C, I^K) = \Pi(D, I^K)$. This requires that

$$p = (k/c)^{1/K} \tag{10.3}$$

This makes all fitnesses equal, and is a necessary condition for an ESS. To guarantee that I is an ESS, we must also satisfy the inequality in condition 10.2 for all possible alternatives.

The straightforward approach would take the possibilities one at a time. Does I resist invasion by D? By C? By some different mixed strategy J? But this is tedious, and there is a quicker way. We know that the two pure strategies are unstable. If the fitnesses were linear functions of p_D, then we could reason as in section 9.1.3: In a pairwise game, the fitnesses are straight lines and can therefore cross only once. If the two pure strategies are unstable, the mixed strategy in the middle is bound to be stable. But here the fitnesses are not linear functions of p_D. It is conceivable that the graph of this game may look like panel 1 of figure 10.1. The game depicted there has unstable pure strategies, yet the mixed equilibrium at $p = 0.5$ is unstable. This is a point worth remembering: *In nonlinear games, mixed equilibria may be unstable even when all pure strategies are unstable.*

So far the news is bad, but there is better news in sight. The unstable mixed equilibrium in panel 1 of figure 10.1 occurs because the fitness functions cross several times. Algebraically, this means that there are several solutions to the equation $w_A(p_A) = w_B(p_A)$. Yet in the game of hunting group size, there is only one solution to this equation. This means that the fitness functions cross at only one value of p_D. This is another point worth remembering: *If a game has just two pure strategies, both of which are unstable, and the game has a single mixed equilibrium, then that equilibrium must be stable.* Thus, the mixed equilibrium in the game of hunting-group size is an ESS.

Appendix

Derivation of inequalities 10.1 and 10.2

If the mutant strategy, J, is rare, then most copies of the non-mutant strategy, I, will be in groups that contain no Js at all, and most copies of J will be in groups that contain no other J. Thus, the fitness of I is $\Pi(I, I^K)$ and that of J is $\Pi(J, I^K)$. Strategy I has the higher fitness if inequality 10.1 is satisfied.

To justify condition 10.2, write

$$
\begin{aligned}
w_I &= Q_0\Pi(I, I^K) + Q_1\Pi(I, J^1 I^{K-1}) + Q_2\Pi(I, J^2 I^{K-2}) + \cdots \\
w_J &= Q_0\Pi(J, I^K) + Q_1\Pi(J, J^1 I^{K-1}) + Q_2\Pi(J, J^2 I^{K-2}) + \cdots
\end{aligned}
$$

Q_i is now the probability that an individual plays against i opponents of type J and $K-i$ opponents of type I. Since we are concerned only with the case in which $\Pi(I, I^K) = \Pi(J, I^K)$, the difference between these two fitnesses is

$$
\begin{aligned}
w_I - w_J &= Q_1[\Pi(I, J^1 I^{K-1}) - \Pi(J, J^1 I^{K-1})] \\
&\quad + Q_2[\Pi(I, J^2 I^{K-2}) - \Pi(J, J^2 I^{K-2})] \\
&\quad + \cdots
\end{aligned}
$$

If J is sufficiently rare, then Q_1 is much larger than any of the other Q_i, and this difference is approximately

$$
w_I - w_J \approx Q_1(\Pi(I, J^1 I^{K-1}) - \Pi(J, J^1 I^{K-1}))
$$

Thus, if condition 10.2 is satisfied, then $w_I > w_J$, at least when J rare.

Resistance to invading strategies in the hunting group size game

This section illustrates the use of condition 10.2. In each of the two sections below, I use it to ask whether strategy I resists some category of invaders.

Invading mixed strategies   For invading mixed strategies, condition 10.2 becomes

$$
\Pi(I, J^1 I^{K-1}) > \Pi(J, J^1, I^{K-1})
$$

where J is some mixed strategy. If J plays D with probability q, then

$$
\begin{aligned}
\Pi(I, J^1 I^{K-1}) &= w - cqp^K - k(1-p) \\
\Pi(J, J^1 I^{K-1}) &= w - cq^2 p^{K-1} - k(1-q)
\end{aligned}
$$

The difference between the two is

$$\delta \;\equiv\; \Pi(I, J^1 I^{K-1}) - \Pi(J, J^1 I^{K-1})$$
$$= \; (q-p)(cqp^{K-1} - k)$$

If I is an ESS, then δ must exceed 0 whenever $q \neq p$. To see that this is so, consider first the case in which $q - p > 0$. Then $cqp^{K-1} > cp^K = k$, so both of the factors in δ are positive. Next consider the case in which $q - p < 0$. Then $cqp^{K-1} < cp^K = k$, so both of the factors in δ are negative. Either way, $\delta > 0$ when $p \neq q$. This shows that I is indeed an ESS. We have

Condition	ESS
$k > c$	D
$k < c$	I

Resistance to invading Cs We already know that $\Pi(C, I^K) = \Pi(D, I^K)$ (see the argument leading to equation 10.3 on page 65). This implies that

$$\Pi(I, I^K) = \Pi(C, I^K)$$

as required by the first part of condition 10.2. (To see why, note that $\Pi(I, I^K) = p\Pi(D, I^K) + (1 - p)\Pi(C, I^K)$. The right-hand is a weighted average of two numbers both of which equal $\Pi(C, I^K)$. Consequently, the average is also equal to $\Pi(C, I^K)$.) The second part of condition 10.2 becomes

$$\Pi(I, C^1 I^{K-1}) > \Pi(C, C^1 I^{K-1})$$

or

$$pw + (1-p)(w - k) > w - k$$

The left-hand side is an average of $w - k$ and w, so it must be greater than $w - k$. Consequently, the inequality is satisfied and I resists invasion by C.

Chapter 11

Does Biology Constrain Culture?

This chapter is a slightly revised version of an article of mine that appeared in 1988 [60].

Abstract Most social scientists would agree that the capacity for human culture was probably fashioned by natural selection, but they disagree about the implications of this supposition. Some believe that natural selection imposes important constraints on the ways in which culture can vary, while others believe that any such constraints must be negligible. This paper employs a "thought experiment" to demonstrate that neither of these positions can be justified by appeal to general properties of culture or of evolution. Natural selection can produce mechanisms of cultural transmission that are neither adaptive nor consistent with the predictions of *acultural* evolutionary models (those ignoring cultural evolution). On the other hand, natural selection can also produce mechanisms of cultural transmission that are highly consistent with acultural models. Thus, neither side of the sociobiology debate is justified in dismissing the arguments of the other. Natural selection may impose significant constraints on some human behaviors, but negligible constraints on others. Models of simultaneous genetic/cultural evolution will be useful in identifying domains in which acultural evolutionary models are, and are not, likely to be useful.

11.1 Introduction

The term *culture* means different things to different people [43]. At least within anthropology, however, there is consensus on one point: culture is viewed as something that we learn from each other. When individuals learn from each other, each generation may inherit the knowledge, attitudes, and behaviors of its predecessors even if these are not transmitted genetically. This constitutes a mechanism of *cultural inheritance*. I use a definition of culture that is broad enough to encompass most others; the term will refer here to any pattern of behavior that is influenced by cultural inheritance.

This definition emphasizes that social learning is the basis of culture. Most social scientists would agree that human learning mechanisms were, in all likelihood, crafted by natural selection. They are in profound disagreement, however, about the implications of this supposition. Some scientists believe that the action of natural selection on learning mechanisms constrains culture in important ways, while others are convinced that any such constraints are negligible.

This disagreement is most clearly displayed in the debate over human sociobiology. Critics of sociobiology often argue as follows:

Argument for Weak Constraints

    1.1 Any biological constraints on culture are too weak to be significant.

    1.2 Therefore, models of genetic evolution will be of little use in understanding variation in human behavior.

Thus, evolutionary biology has little to offer the social sciences, and further progress in that field will require attention to the dynamics of cultural transmission [71, 24, 33, 25, 42, 64].

    Against this position, sociobiologists offer what Boyd and Richerson [4] have termed the "argument from natural origins" [2, 40, 44, 19, 20, 67]:

    Argument from Natural Origins (i.e., For Strong Constraints)

    2.1 If culturally acquired behaviors tended to reduce Darwinian fitness, then the capacity for culture would be altered or destroyed by natural selection.

    2.2 Therefore, behavior should tend to enhance fitness whether transmission is cultural or not.

This argument implies that biological constraints on culture are so strong that it is reasonable for social scientists to ignore cultural transmission. It provides a rationale for the use of what I shall call *acultural* evolutionary models—those that ignore the dynamics of cultural evolution—even in species where cultural transmission is known to exist.

    The argument for weak constraints has been strengthened by recent demonstrations [7, 4] that, under some models of cultural transmission, cultural evolution leads to patterns of behavior that are inconsistent with acultural models. However, neither of these demonstrations is conclusive. Cavalli-Sforza and Feldman [7] do not deal with the simultaneous evolution of genes and culture, and their work therefore tells us little about the extent to which genetic evolution constrains the evolution of culture. Boyd and Richerson [4] do study simultaneous genetic/cultural evolution, but their conclusions concerning the argument from natural origins are based on other models in which genes and culture evolve independently. Thus, the extent to which culture is constrained by the natural selection is still an open question.

    A comprehensive theory of behavior, of course, would incorporate both cultural and genetic evolution. The dynamics of cultural evolution, however, are complex, and the comprehensive theory is still beyond our grasp. For the present, all theories of behavior rely on some form of simplification. This article considers the merits of two such simplifications: purely cultural models, which ignore genetic evolution, and purely acultural models, which ignore cultural evolution. The first are justified only when biological constraints on culture are weak, the second only when they are strong.

11.2   Methods

This article will evaluate the premises of the two arguments above, a much easier task than evaluating the truth of their conclusions. This task is especially easy because these premises (propositions 1.1 and 2.1) have been advanced, not as peculiar features of human culture, but as general properties of culture or of evolution. If they hold in general, then they must also hold in any particular case that we care to examine. If either fails in *any* particular case, then we can conclude that it does not hold in general.

    Rather than search for particular cases in nature, it will be convenient to conduct a "thought experiment". I shall consider the evolution of a hypothetical species, the "snerdwump", which is characterized by a rudimentary form of culture, described below. By analysis of a graphical and a

mathematical model of simultaneous genetic and cultural evolution, it will be shown that neither of the premises above is justified for snerdwumps. I will conclude that, since they fail for snerdwumps, they cannot be general properties of culture or of evolution.

The proposition that species having culture will behave as predicted by acultural models is often confused with another—that such species will behave "adaptively". The latter proposition is not justified by the argument from natural origins since natural selection does not necessarily lead to adaptation. Apparently, this fact is appreciated mainly by population geneticists and behavioral ecologists [28, p. 119][79, 68], for the distinction between adaptation and agreement with acultural models is often ignored. Thus, it will be useful evaluate what might be called the strong form of the argument from natural origins—the proposition that, if the capacity for culture evolved by natural selection, then culture must be adaptive.

11.3   Culture and Adaptation

Naturalists have long been impressed with the many ways in which plants and animals seem suited to their environments, and refer to the characteristics that make them so as *adaptations*. The importance of Darwin's theory stems in large part from its success in explaining these adaptations. This usage of "adaptation" is widespread in the literature of evolutionary biology [23, p. 41][28], and is also found in standard dictionaries. Unfortunately, another usage is also common. For some evolutionists, natural selection produces adaptation *by definition* [77, p. 25] [78, pp. 47–48] [63, p. 49]. This, however, confuses what is being explained (adaptation) with the theory used to explain it (natural selection). If adaptation is defined to be the product of natural selection, then some other word is needed to refer to the observation that individuals seem well suited to their environments, and no such word has been suggested. Use of the same word for both purposes is confusing, and has probably contributed to the misconception that Darwin's theory is circular (see Dunbar [18]). To avoid confusion, I use the word "adaptation" only to refer to the observation that organisms are somehow suited to their environments. In any particular case, the relationship of adaptation to natural selection is a matter of hypothesis, not of definition.

Two hypotheses are involved: selection can account for adaptation only when (1) those individuals with highest fitness are also those that are "best adapted", by the criteria being used to recognize adaptation, and (2) selection tends to increase the mean fitness of the population. Fisher's [23] "fundamental theorem of natural selection" specifies circumstances under which selection will increase mean fitness. This theorem, however, is not nearly as general as was once thought [79, 21]. When selection does not maximize mean fitness, there is no reason to think that it will produce adaptation. Thus, the capacity for culture is expected to be adaptive only if the mean fitness of a population with culture is higher than that of one without.

Most of us would agree that human culture is an adaptation [75, p. 8]. It is widely assumed that this is to be expected if the capacity for culture evolved through natural selection [19]. Thus, no special explanation has been sought for the adaptiveness of human culture. As we shall see, however, it is far from obvious that the forms of cultural transmission produced by natural selection will be adaptive.

Cultural transmission occurs when individuals learn from each other as well as from the environment. Thus, to understand the evolution of the capacity for culture, we must study the evolutionary forces that affect mechanisms of learning. The graphical argument presented below is adapted from that of Harpending, Rogers, and Draper [32], and incorporates ideas developed by Boyd and Richerson [4]. Following Boyd and Richerson, I distinguish *individual learning* (i.e., learning directly from the environment) from *social learning* (i.e., learning from others). For example, consider the
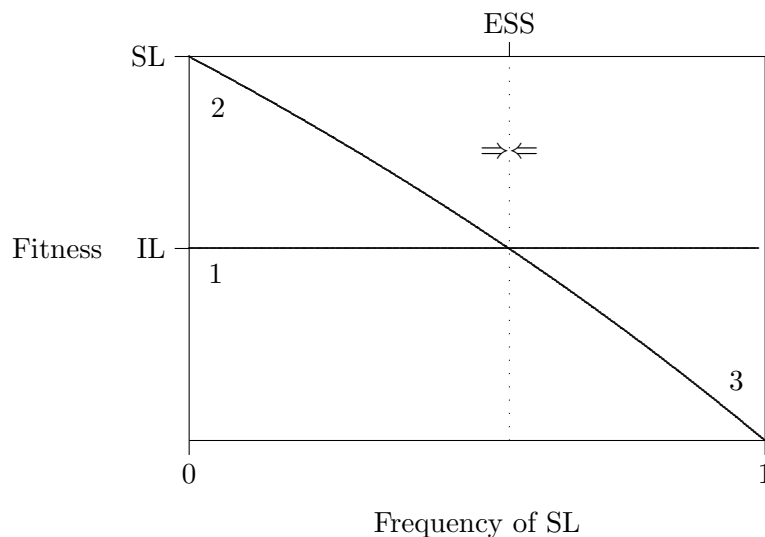
Figure 11.1: The fitnesses of both learning strategies and also the mean fitness of the population are plotted against the frequency of social learning, assuming that $w = b = 1$, $c = 0.9$, $s = 0$, and $u = 0.8$. The numbers refer to the following assumptions: (1) the costs and benefits of individual learning are independent of its frequency; (2) when social learning is rare, its fitness value exceeds that of individual learning because it provides relatively up-to-date information at low cost; and (3) when social learning is very common, its fitness value falls below that of individual learning because the information it provides is out of date.

problem of deciding what foods to eat. One way to do this is by experiment—to eat things that you encounter, and wait to see what happens. If you get sick, you would be wise not to try that food again. Otherwise, you might include it in your diet. This is individual learning, and rats are very good at it. Because of their willingness to experiment with novel foods, they have been able to expand over the much of the world, occupying many new habitats.

However, individual learning entails costs as well as benefits. Rats, for example, undoubtedly ingest poison from time to time. Thus, the value of individual learning reflects a tradeoff between the benefit of a broad diet and the cost of occasional poisoning. A rat that learned his diet socially, by copying an elder, would be less often poisoned, but would be unable to use novel foods. Thus, cultural transmission of diet reduces the risk of being poisoned, at the cost of a narrower diet. Similar remarks have been made by Pulliam and Dunford [58] and by Boyd and Richerson [4].

To clarify the implications of these assumptions, let us perform a "thought experiment". Consider the evolution of culture in a hypothetical species, the snerdwump, which inhabits a varying environment. Snerdwumps may adopt various behaviors, and the fitness consequences of each behavior depend on the state of the environment. Some snerdwumps cope with this situation by individual learning—they obtain, at some cost, information about the environment, and then exhibit the behavior most appropriate to that environment. Flexible responses such as this are often favored by selection in varying environments [6, 73, 4]. Other snerdwumps employ social learning; they simply adopt the behavior of some individual—their "cultural parent"—of the previous generation. For the moment, it will not matter much whether the cultural parent is also a biological parent or not.

The implications of these assumptions are illustrated graphically in Fig. 11.1. The numbers there refer to the numbered paragraphs below.

1. The fitness effect of individual learning depends on its costs and benefits, but not on what others are doing. Thus, its fitness is a horizontal line in Fig. 11.1.

2. If social learning is very rare, nearly all cultural parents will be individual learners. Thus,

social learners will acquire relatively recent information about the environment. They will acquire, that is, behaviors that were appropriate in the immediately preceding generation. Their fitness will therefore exceed that of individual learners, provided that social learning is sufficiently cheap and environmental change sufficiently slow.

3. When all learning is social, no one is monitoring the environment, and the information acquired will eventually be many generations old. Social learning will then have lower fitness than individual learning because, by assumption, information is worthless if sufficiently out of date.

These assumptions imply that the fitness of social learning exceeds that of individual learning when its frequency is low, but falls below as its frequency increases. The shape of the curve connecting points 2 and 3 will depend on the structure of the cultural inheritance system, i.e. on the tendency of social learners to learn more from some individuals than from others. The curve in Fig. 11.1 was drawn using the mathematical formulation presented below, but any monotonic curve connecting these points will produce a graph of the same general form. The conclusions that follow hold regardless of the structure of the cultural inheritance system.

This representation is consistent with a variety of assumptions about the genetics of social learning. It applies equally when social learning is inherited as a single Mendelian character, and when what is inherited is merely a propensity to learn socially, affected by genes at many loci. Under most (but not all) forms of genetic inheritance, natural selection will tend to increase the frequency of the learning mechanism with the highest fitness, and I assume this to be the case. From this assumption several conclusions follow.

The frequency of social learning among snerdwumps will increase under natural selection until it reaches an equilibrium at the point where the curves for individual and social learning cross, and I assume this equilibrium to be stable. At this equilibrium the fitness of social learning (and the mean fitness of the population as a whole) is exactly the same as that of individual learning; a population (or individual) with social learning has no higher mean fitness than one without. The evolution of culture among snerdwumps, therefore, does not maximize mean fitness. Selection often fails to maximize mean fitness when, as here, the fitness of a morph decreases with its frequency. This result is hardly new [28, 79]. The novelty here is the insight that the evolutionary process that produced culture is likely to have been characterized by this form of frequency dependence.

Recall that natural selection is expected to produce adaptation only if mean fitness is increased. Since mean fitness is not increased in the present model, there is no reason to expect that snerdwump culture will be adaptive. This does not, of course, mean that human culture is not adaptive. It does show, however, that if human culture is adaptive, it must be so because of some feature not included in the present model. Adaptation is neither an inevitable consequence of evolution nor of culture.

Rindos [59] agrees that evolution can produce forms of culture that are not adaptive. He also claims (p. 316), however, that the evolution of culture *is* an adaptive process when fitness depends on the relationship of behavior to the physical, rather than the social environment. The snerdwump example shows that this is not necessarily so.

These results also contradict the widely held belief that, if the capacity for culture evolved through natural selection, then culture must, on average, enhance fitness [20, p. 43]. However, it is a mistake to assume, as does Schwartz [65], that sociobiology must stand or fall on this issue, for maximization of fitness is not a general feature of evolutionary models. The relevance of acultural evolutionary models depends on how well they predict human behavior, not on whether culture maximizes fitness. It is important to ask, then, how well such models can be expected to predict behavior in species with cultural transmission.

11.4 How much does biology constrain culture?

To answer a quantitative question, a quantitative model is needed, and the one offered here is a simple formulation of the assumptions just discussed. Suppose that there are two environmental states, labeled 0 and 1, two behaviors labeled in the same way, and that individual fitness depends on behavior and environment as follows:

		Environmental State	
		0	1
	0	$w + b$	$w - b$
Behavior			
	1	$w - b$	$w + b$

Note that the fitness effect of each behavior is independent of the frequency with which it is exhibited. Thus, although the fitness effects of snerdwump learning mechanisms are frequency-dependent, the effects of the behaviors learned are not. Further suppose that, in each generation, the environment is perturbed with probability u, and that after perturbations it is equally likely to be in either state.

In the absence of any information about present or past environmental states, the environment is equally likely to be in either state. Therefore, the unconditional fitness of either behavior is $w + \frac{1}{2}b - \frac{1}{2}b = w$. Fitness can be enhanced by obtaining information about the environment. Suppose that the state of the environment can be determined accurately by individual learning, but that this effort reduces fitness by bc units due to risk of predation or expenditure of time and energy. Then the fitness of a snerdwump adopting individual learning is

$$w_I = w + b(1 - c). \tag{11.1}$$

How does this compare to the fitness of a snerdwump who learns socially?

The fitness, w_S, of a social learner, depends on the probability that social learning will provide accurate information, and this probability changes as cultural evolves. Let p denote the frequency after selection of individuals adopting social learning in the current generation. If cultural evolution is fast compared with genetic evolution, then p can be treated as a constant for the purpose of finding what I shall call the *cultural equilibrium*, $\tilde{w}_S$. This is the equilibrium value of w_S under the action of cultural evolution only. Through social learning, an individual acquires behavior that was originally acquired by individual learning in some previous generation. The probability that this occurred τ generations ago is $p^{\tau-1}(1 - p)$, if cultural parents are random members of the previous generation. The expected benefit obtained by adopting such behavior is zero if the environment has been perturbed in the past τ generations, and b if no perturbation has occurred. The probability of no perturbation is $(1 - u)^\tau$. Putting this together, we can write

$$\begin{aligned}
\tilde{w}_S &= w + b \sum_{\tau=1}^{\infty} p^{\tau-1}(1 - p)(1 - u)^\tau \\
&= w + b(1 - p)(1 - u) \sum_{\tau=1}^{\infty} p^{\tau-1}(1 - u)^{\tau-1}.
\end{aligned}$$

By the formula for the sum of a geometric series, this is

$$\tilde{w}_S = w + \frac{b(1 - p)(1 - u)}{1 - p(1 - u)}. \tag{11.2}$$
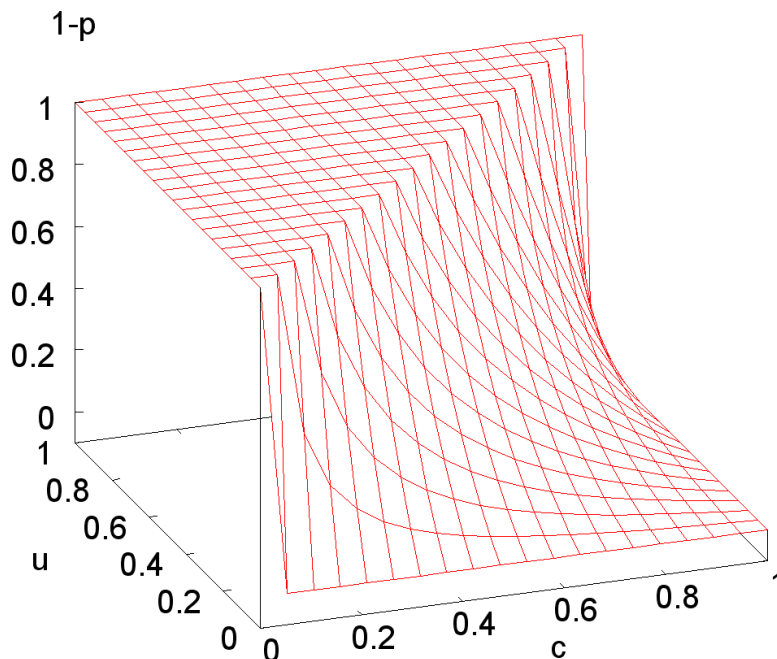
Figure 11.2: The the equilibrium frequency $1 - \hat{p}$ of individual learners, graphed against the cost c of individual learning and the rate u of environmental perturbations.

At this point, the 1988 paper suggested that social learning may be costly as well, and introduced notation to measure this cost. In retrospect, this seems like a bad idea. It introduces the complexity of an additional parameter without providing any real insight. We are dealing, after all, with a model in which social learn is advantageous because it avoids the costs of individual learning. Let us therefore study the extreme case in which social learning has no cost whatever.

It can be shown that $\tilde{w}_S$ is a stable equilibrium and that convergence toward it is rapid unless $(1 - u)p \approx 1$. As in Fig. 11.1, $\tilde{w}_S$ is a monotonically decreasing function of p that exceeds w_I when p is small, but falls below it as p increases.

Changes in p, the frequency of social learning, are produced by genetic evolution, which is assumed to be much slower than cultural evolution. In studying the evolution of p, therefore, it is reasonable to assume that convergence toward (11.2) is effectively instantaneous. Equation 11.2 implies that $\tilde{w}_S = w_I$ when p is equal to

$$\hat{p} = \frac{c - u}{c(1 - u)}.$$

This is the *evolutionary equilibrium* of p, i.e. the equilibrium under the combined effects of cultural and genetic evolution. It is between zero and one provided that $u < c$.

The fraction $1 - \hat{p}$ of individual learners is graphed in Fig. 11.2. When $u > c$, this fraction equals unity because there are no social learners. The fraction of individual learners drops off steeply, however, as c increases or u increases. You would not find many individual learners in a world where individual learning was costly and the environment changed slowly.

We are interested in measuring the consistency of behavior under this model with the behavior that would be predicted by a model ignoring cultural transmission. Because the fitness effects of the two behaviors are frequency independent, an acultural model would predict that natural selection would maximize fitness. It would, in other words, predict appropriate behavior in each environment. Thus, the consistency between behavior and the predictions of the acultural model can be measured by the correlation between behavior and environment. To define such a correlation, we must first define two random variables. Let B equal zero or unity depending on the behavior exibited by a
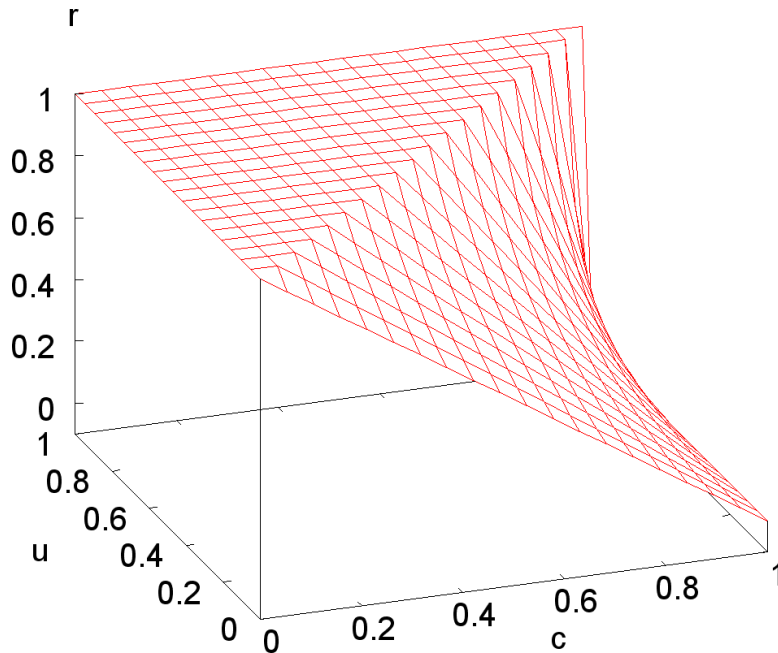
Figure 11.3: The correlation r between the predictions of cultural and acultural models as a function of the cost c of individual learning and the rate u of environmental perturbations.

random member of the population, and let S take the same values depending on the state of the environment. These variables have the same means and variances: $E\{S\} = E\{B\} = 1/2$, and $Var\{S\} = Var\{B\} = 1/4$. The correlation of S and B is

$$r = \frac{E\{SB\} - E\{S\}E\{B\}}{\sqrt{(Var\{S\}Var\{B\})}} = 4E\{SB\} - 1, \tag{11.3}$$

and measures the tendency of individuals to behave appropriately. The acultural model would predict that $r = 1$.

The value of r is easy to find if $u > c$ since that condition implies that social learning will not evolve. Thus, all snerdwumps are individual learners, and $r = 1$. In the appendix it is shown that

$$r = \frac{1 - c}{1 - u}, \tag{11.4}$$

in the more interesting case when $u < c$. As illustrated in Fig. 11.3, the behavior of r depends on the relationship between the cost c of individual learning and the rate u of environmental perturbations. If environmental change is slow, or individual learning is cheap, $c \leq u$, and social learning does not evolve. Under such circumstances, $r = 1$.

At larger values of c/u, social learning does evolve, and r declines linearly with increasing c. This decline is especially steep where u is large. If c is appreciably greater than u, r is low and an acultural model would predict poorly. In such cases, the biological constraints on culture weak. In other words, a biologist who used genetic models to study snerdwump behavior would get the wrong answer much of the time. On the other hand, if individual learning is cheap, or the environment changes rapidly, c will be small relative to u, and $r = 1$. In such cases, acultural models would work well.

When social learning does evolve (that is when $c > u$), Eqn 11.4 can also be written as

$$r = (1 - \hat{p})c/u > 1 - \hat{p}$$

In other words, r is always greater than the frequency of individual learners. If we knew that 80% of snerdwumps were individual learners, we could conclude that $r > 0.8$ even if we knew nothing

about c and u. The predictive value of an acultural model is at least as large the fraction of individual learners. This does not, however, prevent r from dropping to low values when c/u is large. It does so because, in these cases, individual learning is rare indeed.

11.5 Discussion

The study of culture is in a position similar to that of biology before Darwin. Most of us would agree that human culture is adaptive—culturally acquired behaviors are essential to humans, and have enabled them to colonize habitats far outside the ecological niche of their primate ancestors. Yet, we have no satisfactory explanation of this adaptation. It is possible to devise models in which natural selection produces adaptive forms of culture (Robert Boyd, personal communication, 1987). As we have seen, however, selection can also produce forms of culture that are not adaptive. This leaves us with an important unanswered question: What accounts for the adaptiveness of human culture?

In answering this question, some theoretical work will be necessary. We must identify ways in which adaptive mechanisms of cultural inheritance can evolve. The model studied here ignores a variety of factors that may tend to make culture adaptive. For example,

1. There is no cultural selection—individuals with high fitness are no more likely than others to serve as cultural parents.

2. There are no learning biases favoring transmission of behaviors that enhance fitness.

3. Individual learning does not modify or improve behaviors acquired culturally. Therefore, the value of culturally transmitted information does not accumulate; it decays.

4. Decisions about which form of learning to adopt are not contingent on the quality of available environmental information.

Departures from any or all of these assumptions may contribute toward the apparently adaptive nature of human culture, but their relative importance is unclear. Models of simultaneous genetic and cultural evolution should help clarify this question, and useful starts in this direction have already been made by several authors [47, 4, 5].

Many of us believe that acultural evolutionary models (those that ignore cultural evolution) provide useful insights into the evolution of behavior—even that of species in which cultural transmission is important. One way to justify this position is what Boyd and Richerson [4] have called the argument from natural origins. In its simplest form, this argument rests on the premise that a mechanism of cultural inheritance can evolve by natural selection only if it is adaptive. As we have seen, this premise is false. However, this simple form of the argument from natural origins has few supporters and is of no great importance. Most modern evolutionists are well aware that natural selection can be maladaptive, especially where interactions between individuals are involved.

The weak form of the argument from natural origins assumes that mechanisms of cultural transmission can evolve only if the behaviors produced tend to be consistent with the predictions of acultural evolutionary models. The consistency between behavior and acultural models is measured here by r, a coefficient of correlation. Equation 11.4 shows that, at least within our hypothetical species, the snerdwump, there is some tendency for behavior to accord with acultural models. However, if $c \gg u$, and this tendency will be negligible. There is therefore no basis for the claim that, if the capacity for culture evolved by natural selection, then culture must be explicable in terms of ordinary, acultural evolutionary models.

On the other hand, equation 11.4 also shows that culture will be highly consistent with the predictions of acultural models if $c \leq u$. Thus, there are circumstances under which biological constraints on culture are important. There is, therefore, no basis for the claim that since humans have culture, models of genetic evolution are necessarily irrelevant to contemporary variation in human behavior. Thus, neither extreme position in the sociobiology debate can be justified by appeal to general properties of evolution or of culture. It is possible for biological constraints on culture to be negligible; it is also possible for them to be of fundamental importance. If either extreme position is correct, it must be because of some peculiar feature of human culture that has yet to be identified.

The advantages of theories of simultaneous genetic/cultural evolution are obvious, but such theories are difficult to construct and it seems likely that both purely cultural and purely acultural approaches to social science will continue in importance. These approaches, however, need not arise from contradictory philosophical positions about the causes of human behavior. The past two decades of psychological research have undermined the old notion that there are general laws of learning that apply universally to all behaviors and all species [47, 62]. Learning abilities are species-specific, and are constrained and biased in many ways. It is likely that each mammalian species possesses a constellation of learning abilities, each evolved for a specific purpose. Within a single species, biological constraints may be strong for some behaviors, but negligible for others. Purely acultural explanations would be needed for the former behaviors, and purely cultural explanations for the latter. To reconcile the two approaches to social science, what is needed is a way to distinguish behaviors whose evolutionary dynamics are primarily genetic ($r \approx 1$) from those whose dynamics are primarily cultural ($r \approx 0$). This will be possible if appropriate models of simultaneous genetic/cultural evolution can be developed.

The present model is far too simple to be advanced as a realistic description of human culture. The conclusions drawn so far, however, do not depend on the model's realism but only on its consistency with what is known about the process of natural selection. This cannot be said of the conclusions I am about to draw, and they should be regarded with some skepticism. However, experience has shown that simplistic models are often successful in capturing the main features of phenomena that are too complex to model in detail. For example, the premises of the Hardy-Weinberg theorem of population genetics almost never hold in nature, yet its predictions are approximately correct more often than not. Nontheless, this theorem is fundamental both to empirical and theoretical studies in population genetics. Only experience will tell whether the model advanced here is similarly robust. In the meantime, the model's broader implications seem worthy of discussion.

The model assumes that the fitness effects of particular behaviors are independent of their frequency. This may be true of behaviors, such as techniques of farming or foraging, with which resources are extracted from the environment, but it is unlikely to be true of social behavior: the success of particular social behaviors ordinarily does depend on what others are doing. As we have seen, such frequency dependence can profoundly alter the outcome of evolution. Thus, the model developed here is probably a poor guide to the evolutionary dynamics of social behavior.

The model's main result is equation 11.4, which relates the correlation between behavior and the environment to the cost of individual learning and the rate of environmental change. As these undoubtedly vary from behavior to behavior, it follows that some behaviors should track the environment well, others poorly. Consequently, ecology should be more useful in understanding some behaviors than others, a fact that cultural ecologists have long understood [70, 54]. This equation also implies that acultural evolutionary models are unlikely to be helpful in understanding behaviors that are usually learned socially.

Behavioral ecologists have had considerable success in using acultural models to explain aspects

of human foraging. For example, Hawkes, Hill, and O'Connell [34] showed that the resource species chosen by Aché foragers were consistent with the predictions of a simple evolutionary model. The availability of various food items may change so rapidly that tradition would be a poor source of information about the optimal choice of diet. Thus, we might expect a reliance on individual learning, leading to a strong correlation between behavior and environment. This conclusion can be obtained from equation 11.4 by noting that a high rate of environmental change (large u) tends to increase r. It may explain the predictive success of the acultural model used by Hawkes, Hill, and O'Connell. Hunting techniques, on the other hand, may be much easier to learn socially than by trial and error, and should not rapidly become obsolete ($u \ll 1$). If so, equation 11.4 indicates that $r \ll 1$, and acultural models should do a poor job of explaining variation in hunting techniques.

11.6   Conclusions

Natural selection can produce mechanisms of cultural transmission that are not adaptive and that may be only weakly consistent with the predictions of *acultural* evolutionary models (those that ignore cultural evolution). This refutes what Boyd and Richerson have called "the argument from natural origins" as a general principle, and raises the question: Why is culture adaptive? On the other hand, natural selection can also produce mechanisms of cultural transmission that are highly consistent with acultural evolutionary models. This refutes the claim that, since humans have culture, acultural models are necessarily irrelevant to human behavior. It seems likely that both purely cultural and purely acultural theories of behavior will continue in importance. These conclusions do not depend on the realism of the model studied here.

   To the extent that the model is accurate in spite of its simplicity, several additional conclusions can be drawn. The correlation between behavior and the predictions of acultural models should be better for some behaviors than others. It should be greatest where individual learning is easy and environmental change is rapid or, equivalently, where the frequency of social learning is low.

Appendix

To evaluate r when $u < c$, note that the term $E\{SB\}$ in (11.3) is

$$E\{SB\} = \Pr\{\, S = 1 \text{ and } B = 1 \,\} = 0.5 x_T,$$

where x_T is the probability that a random individual in the population is behaving appropriately with respect to the current environment. Since all individual learners adopt appropriate behavior, x_T is

$$x_T = 1 - \hat{p}(1 - x_S), \tag{11.5}$$

where x_S is the probability that a social learner behaves appropriately. Now the fitness of a social learners is $w + b$ with probability x_S, and $w - b$ with probability $1 - x_S$. Thus,

$$\hat{w}_S = w + b(2x_S - 1).$$

At equilibrium, we also have

$$\hat{w}_S = w_I = w + b(1 - c),$$

since both strategies have equal fitnesses at equilibrium. Equating these expressions leads to $1 - x_S = c/2$, and substituting this into equation 11.5 produces

$$x_T = 1 - \frac{c - u}{2(1 - u)}.$$

Finally, substituting into the expressions for and $E\{SB\}$ and then into equation 11.3 produces equation 11.4.

Chapter 12

Selection and Covariance

In the early 1970s George Price introduced a new way to write models of natural selection [56, 57]. In his approach, the rate of change in a character equals the covariance between its value and relative fitness. This has been extremely useful, largely because of the ease with which covariances can be decomposed into terms representing different effects. This chapter begins with definitions, then explains Price's main result, and uses it in a series of examples.

12.1   Definitions

Imagine a population that can be subdivided in some way. The subdivisions will be called "components." They may be smaller populations, individual organisms, or individual genes. The same algebra will apply in all cases. Much of the power of Price's algebra derives from the fact that we can be as creative as we like in deciding how the population will be subdivided.

Character value   We are interested in the evolution of some character, whose mean value within component i will be written as x_i. This character could be the frequency of some allele, or stature, or selfishness, or nearly anything else. The only restriction is that there be some resemblance between parent and offspring. This might result from genetic inheritance or from cultural inheritance. Price's method will work no matter what the mechnism of inheritance.

Absolute fitness   is defined in the usual way. The number of offspring of component i is $n_i W_i$, where n_i is the size of component i and W_i is its absolute fitness.

Expectation   In statistical jargon, an "expectation" is an average, and I will write the expectation of X as $E[X]$. Below, expectation will refer to an average taken over the components of a population, weighted by component size. For example, the mean absolute fitness of the entire population is

$$W = E[W_i] = \sum_i W_i f_i \tag{12.1}$$

where

$$f_i = n_i \Big/ \sum_i n_i$$

is the relative size of component i. Similarly, the mean character value within the population as a whole is

$$x = E[x_i] = \sum_i x_i f_i \tag{12.2}$$

80

Below, I will make use of the properties of expectations that are outlined in equations C.3–C.5, on page 125 of Appendix C.

Relative fitness   The relative fitness of the ith component is

$$w_i = W_i/W \qquad (12.3)$$

In other contexts within population genetics, relative fitnesses are obtained by dividing fitness by some arbitrary constant. But in Price's model, this constant is not arbitrary. Fitnesses must be made relative to the mean. Note that the mean relative fitness is

$$w = Ew_i = 1 \qquad (12.4)$$

Covariance   The covariance between two variables, X and Y, is a measure of their tendency to increase (and decrease) together. It can be written in any of the following equivalent forms

$$
\begin{aligned}
\mathrm{Cov}(X,Y) &= E[(X - \bar{X})(Y - \bar{Y})] & (12.5) \\
&= E[X(Y - \bar{Y})] & (12.6) \\
&= E[XY] - \bar{X}\bar{Y} & (12.7)
\end{aligned}
$$

where $\bar{X} = E[X]$ and $\bar{Y} = E[Y]$ are the expectations of X and Y. For further explanation, see section C.6 of Appendix C.

12.1.1   Example

To make these definitions clear, let us consider a concrete example. Table 12.1 describes a hypothetical population that is divided into two components.

Table 12.1: Hypothetical Population

Component	f_i	x_i	W_i
1	1/3	0	2
2	2/3	3	1

A fraction $f_1 = 1/3$ of the population is in component 1, and the remaining 2/3 is in component 2. The mean character value within the population as a whole is defined by Eqn. 12.2. With these data, it is

$$x = \frac{1}{3} \times 0 + \frac{2}{3} \times 3 = 2$$

The mean fitness (Eqn. 12.1) is

$$W = \frac{1}{3} \times 2 + \frac{2}{3} \times 1 = \frac{4}{3}$$

The relative fitnesses of the two components (Eqn. 12.3) are $w_1 = 2/(4/3) = 3/2$ and $w_2 = 1/(4/3) = 3/4$.

$\star$ EXERCISE **12**–**1** Verify that the mean relative fitness equals 1, as claimed by Eqn. 12.4.

The covariance between x_i and w_i is

$$\begin{aligned} \text{Cov}(x_i, w_i) &= E[x_i w_i] - E[x_i]E[w_i] \\ &= 6/4 - 2 = -1/2 \end{aligned}$$

These quantities are the ingredients of the Price equation, to which we now turn.

The population is divided into N components. The n_i members of component i have mean character value x_i; their n_i' offspring have mean character value is x_i'. These parents have mean fitness

$$W_i = n_i'/n_i \tag{12.8}$$

In the population as a whole, mean character value of offspring is

$$x' = \frac{\sum n_i' x_i'}{\sum n_i'} = \frac{\sum n_i W_i x_i'}{\sum n_i W_i} \qquad \text{(by Eqn. 12.8)}$$

Multiplying and dividing by $\sum n_i$ gives

$$x' = \frac{\sum n_i W_i x_i'}{\sum n_i} \frac{\sum n_i}{\sum n_i W_i} = \frac{E[W_i x_i']}{E[W_i]}$$

Next, divide numerator and denominator by W to convert from absolute to relative fitnesses:

$$x' = \frac{E[W_i x_i']/W}{E[W_i]/W} = \frac{E[w_i x_i']}{E[w_i]} = E[w_i x_i']$$

The denominator has disappeared because of Eqn. 12.4. The change in mean character value across one generation is

$$\Delta x = x' - x = E[w_i x_i'] - x$$

To re-express this in terms of covariances, we begin by adding and subtracting x_i.

$$\Delta x = E[w_i(x_i + x_i' - x_i)] - x$$

Now, make use of the fact that the expectation of a sum is a sum of expectations (see page 125).

$$\begin{aligned} \Delta x &= (E[w_i x_i] - x) + E[w_i(x_i' - x_i)] \\ &= (E[w_i x_i] - wx) + E[w_i(x_i' - x_i)] \qquad \text{(Since } w = 1\text{)} \\ &= \text{Cov}(w_i, x_i) + E[w_i(x_i' - x_i)], \end{aligned}$$

which is Price's equation (Eqn. 12.9).

Box 6: Derivation of Price's equation

12.2   Price's Equation

The change in mean character value between generations is, by definition, $\Delta x = x' - x$. George Price showed that it equals

$$\Delta x = \text{Cov}(w_i, x_i) + E[w_i(x_i' - x_i)] \tag{12.9}$$

In this equation, x_i' is the mean character value among *offspring* whose parents were in component i. This formula is derived in Box 6. Here, I concentrate on explaining what it means.

The first term on the right is the covariance between relative fitness and character value. It measures the effect of natural selection on differences between components. The second term on the right measures the effect of selection on differences within components.

It is sometimes useful to apply Price's equation once again to the quantity $x_i' - x_i$ that appears on the right-hand side of Eqn. 12.9. This makes it possible to "peel the onion," decomposing the effect of natural selection into a sum of covariances, each associated with a different level of the model. Often, this process ends when $x_i' = x_i$ and the second term in Eqn. 12.9 disappears.

12.2.1 Evolution of allele frequency in a population of randomly mating diploids

Much of population genetics is about the special case in which the components are diploid individuals who mate at random, and the character of interest is the frequency of some allele. In this case, Price's equation takes a particularly simple form[1]

$$\Delta x = \text{Cov}(w_i, x_i)/2 \tag{12.10}$$

Here, x_i is the allele frequency (0, 1/2, or 1) within the ith individual, and w_i is the fitness of that individual. The second term in Eqn. 12.9 has disappeared, and the first is half as large.

12.3 Examples

12.3.1 Hypothetical population

For the hypothetical population in Table 12.1, there is not much left to do. The covariance between x_i and w_i was calculated above, and equals –1/2. If offspring were identical to their parents, the second term in Price's equation would equal zero, and Price's equation would give

$$\Delta x = -1/2$$

In other words, mean character value among offspring should be 1/2 unit smaller that that of their parents.

12.3.2 Genic selection

Section 4.2 (page 20) derived a famous result for the change in allele frequency under natural selection. This section will rederive the main result of that section using Price's equation.

In this model there are two alleles, A_0 and A_1 (a slight change in notation from section 4.2). To apply Price's method, let us say that the population comprises two components, one consisting of all copies of allele A_0 and the other of all copies of A_1. The notation associated with this model is shown in Table 12.2. There, x_i is the character value associated with component i. Note that I have defined it to equal the relative frequency of allele A_1 within the component. Thus, $x_1 = 1$, and $x_0 = 0$. With this definition, the mean character value ($E[x_i]$) equals p, the frequency of allele A_1 within the population as a whole. The genic fitnesses of the two alleles are W_0 and W_1, and mean fitness is $W = pW_1 + qW_0$.

---

[1]Each of i's offspring gets one gene from i and one from the population at large. If mating is at random, these two genes are copies of allele A with probabilities x_i and x, respectively. Thus, $x_i' = (x_i + x)/2$, and $x_i' - x_i = -(x_i - x)/2$. The second term in (12.9) becomes $-E[w_i(x_i - x)]/2 = -\text{Cov}(w_i, x_i)/2$. (This uses Eqn. C.11, on page 126.) Plugging this into Eqn. 12.9 yields Eqn. 12.10.

Table 12.2: Notation for Model of Genic Selection

Allele	f_i	x_i	W_i	w_i
A_1	p	1	W_1	W_1/W
A_0	$q = 1 - p$	0	W_0	W_0/W

$$W = pW_1 + qW_0$$

To use Price's equation, the first step is to calculate $\text{Cov}(w_i, x_i)$. This can be done in several ways, as shown above in Eqns. 12.5–12.7. I tried several of these and settled on Eqn. 12.7. To use this approach, we need three ingredients

$$
\begin{aligned}
x &= px_1 + qx_0 = p \\
w &= 1 \quad \text{(see Eqn. 12.4)} \\
E[w_i x_i] &= pW_1/W + q \times 0 = pW_1/W
\end{aligned}
$$

Each of these is an expected value calculated using the values in Table 12.2. Substituting now into Eqn. 12.7 gives

$$
\begin{aligned}
\text{Cov}(w_i, x_i) &= E[w_i x_i] - xw \qquad\qquad (12.11)\\
&= pW_1/W - p \\
&= \frac{pq(W_1 - W_0)}{W}
\end{aligned}
$$

This is the first term in Price's equation. The second term is the average change in character value within components. In other words, it is the average difference between the character value of a component and that of the component's offspring. Absent mutation, there are no such differences, so the second term in Price's equation equals zero. Thus, the average change in character value is

$$\Delta x = \frac{pq(W_1 - W_0)}{W}$$

Apart from differences in notation, this is identical to Eqn. 4.3 (page 20). Thus, the Price equation give us the same answer in this familiar context.

12.3.3   Genic selection with mutation

Suppose now that each generation, mutation converts each A_1 to A_0 with probability u, and each A_0 to A_1 with probability v. What is the average frequency x_1' of allele A_1 among "offspring" of allele A_1? The answer is $x_1' = 1 - u$, since this is the fraction that did not mutate into copies of A_0.

⋆ EXERCISE **12–2** What is the average frequency x_0' of allele A_1 among "offspring" of allele A_0?

⋆ EXERCISE **12–3** Calculate the second term in Price's equation (Eqn. 12.9) under this model.

12.3.4 Cultural evolution of charisma

This example deals with a characteristic that I will call charisma. The idea is that people who have charisma are imitated by others, who therefore may become more charismatic themselves. However, charisma is not easy to imitate, so imitators do not always succeed in becoming charismatic themselves.

Let

$$x_i = \begin{cases} 1 & \text{if individual } i \text{ is charismatic} \\ 0 & \text{otherwise} \end{cases}$$

Suppose that each charismatic person is imitated by b others. The fitness of an individual of type i (in this cultural context) is the number of people who imitate him:

$$W_i = bx_i$$

Note that, for non-charismatic people, x_0 and W_0 are both zero. Mean fitness is

$$W = bx$$

where x is the fraction of the population that is charismatic. The relative fitness of individual i is

$$w_i = W_i/W = x_i/x$$

Imitation is imperfect, so of the bx_i imitators of i only a fraction r become charismatic themselves. Thus, the frequency of charisma among the imitators of i is

$$x_i' = rx_i.$$

To use Price's formula, we must first calculate the covariance between x_i and relative fitness. Using Eqn. 12.7,

$$\begin{aligned} \text{Cov}(w_i x_i) &= E[w_i x_i] - wx \\ &= [xw_1 x_1 + (1-x)w_0 x_0] - wx \\ &= 1 - x \end{aligned}$$

The second line above uses the fact that x and $1 - x$ are the fractions of charismatics and non-charismatics within the population. The other term in Price's formula measures the difference in charisma between the "parent" and the imitators:

$$\begin{aligned} E[w_i(x_i' - x_i)] &= xw_1(x_1' - x_1) + (1-x)w_0(x_0' - x_0) \\ &= r - 1 \end{aligned}$$

Now applying Price's formula (12.9),

$$\begin{aligned} \Delta x &= (1-x) + (r-1) \\ &= r - x \end{aligned}$$

Notice that

$$\Delta x \begin{cases} > 0 & \text{when } x \text{ is small} \\ < 0 & \text{when } x \text{ is large} \end{cases}$$

The system reaches equilibrium when $\Delta x = 0$, i.e. when

$$x = r$$

At equilibrium, the frequency of charisma within the population is equal to the ease with which charisma can be imitated.

⋆ EXERCISE **12–4** Make a table that summarizes the assumptions of the charisma model. The table should have the following headings: Component, f_i, x_i, W_i, and w_i.

⋆ EXERCISE **12–5** Use Price's equation to find Δx from the table below, assuming that $x_i' = x_i$:

Component	f_i	x_i	W_i
1	0.25	1	1
2	0.50	3	2
3	0.25	2	3

Chapter 13

Kin Selection

The chapters on clutch size, foraging, and game theory had little to say about genetics. They made use of the "phenotypic gambit," a method of finding evolutionary equilibria that ignores genetics altogether (see section 6.2, p. 32). The present chapter deals with cooperation between relatives. In this context, evolutionary change depends on genetic relatedness and, for the first time, the genetics of our model will need to be explicit.

Many forms of behavior affect the fitness of more than one individual. For example, if Jack steals food from his brother Joe, this may increase Jack's probability of survival while reducing Joe's. If theft among brothers is encouraged by some allele, the increase in Jack's survival will increase the fitness of this allele. However, since Joe is Jack's brother, he is also likely to carry the allele. The reduction in Joe's survival will therefore reduce the allele's fitness. There is a tradeoff, then, between the fitnesses of these different individuals. The theory that deals with such tradeoffs is called the theory of "kin selection," and was originally developed by W.D. Hamilton [29, 30].

Although I introduced the subject with an example about theft, it is more conventional to discuss examples of "altruism." Altruism will refer here to a behavior that reduces the actor's own fitness by an amount C while increasing by B the fitness of some other individual (called the recipient). The central question is, what forms of altruism are favored by selection? The answer will tell us not only when altruism will evolve, but when selfish behavior such as theft will evolve. Selfishness is simply the special case in which C and B are both negative.

13.1 Hamilton's Rule

If the recipient is a relative of the donor, then the recipient may have a good chance of carrying the genes that encouraged the donor's altruism. If this happens often enough, alleles favoring altruism may be favored because they enjoy the benefits of altruism as well as suffering the costs. The theory of kin selection asserts that this will happen when

$$rB - C > 0 \tag{13.1}$$

where r measures the relatedness between donor and recipient in a sense that will be made precise below. This inequality is known as "Hamilton's rule," and is the central result of the theory of kin selection.

13.2   Derivation of Hamilton's Rule

A genetic model is needed to derive this result, because the tradeoff in question is caused by the genetic relationship between donor and recipient. It will be convenient first to ask how altruism affects an arbitrary allele A. I postpone until later the question of how this allele affects behavior. This will enable us to consider various assumptions about the mode of gene action.

Let p_i denote the frequency of A in the ith individual, and let

$$\bar{p} = E[p_i] = \frac{1}{N}\sum_{i=1}^{N} p_i$$

denote the frequency of A in the population. Here, N is the size of the population. Throughout this section, I will use the expectation notation, "$E[\cdot]$, " to denote a simple average across the members of the population. To avoid clutter, I will often use a superposed bar, as in $\bar{p}$, to indicate the average or expected value.

In the absence of altruism, each individual has fitness w^*. Each act of altruism adds B to the recipient's fitness and subtracts C from the donor's. The same individual may participate in several of these interactions, may be donor in some and recipient in others. Let b_i represent the sum of the benefits that individual i receives from altruism and c_i the sum of his costs. Thus, individual i has fitness

$$W_i = w^* + b_i - c_i \tag{13.2}$$

Individual i's relative fitness is $w_i = W_i/\bar{W}$, where $\bar{W}$ is mean fitness.

Let us use the Price model (chapter 12) to find out how $\bar{p}$ changes in response to natural selection during a single generation. As discussed in chapter 12, the Price model imagines a population that is divided into components. For present purposes, we can take the components to be individual diploid organisms. If these mate at random, then the change in allele frequency over a single generation is (see Eqn. 12.10)

$$\begin{aligned}
\Delta\bar{p} &= \mathrm{Cov}(w_i, p_i)/2 = \mathrm{Cov}(W_i, p_i)/2\bar{W} \\
&= \mathrm{Cov}(w^* + b_i - c_i, p_i)/2\bar{W}
\end{aligned}$$

Covariances can be decomposed like variances, so

$$\Delta\bar{p} = \mathrm{Cov}(w^*, p_i)/2\bar{W} + \mathrm{Cov}(b_i, p_i)/2\bar{W} - \mathrm{Cov}(c_i, p_i)/2\bar{W}$$

The first covariance is zero because w^* is a constant. After dropping it, we have

$$\Delta\bar{p} = \mathrm{Cov}(b_i, p_i)/2\bar{W} - \mathrm{Cov}(c_i, p_i)/2\bar{W} \tag{13.3}$$

Box 7 re-expresses these covariances and allows us to write

$$\Delta\bar{p} = (u - \bar{p})(r\bar{b} - \bar{c})/2\bar{W}$$

where u is the mean frequency of allele A among donors and r is the regression of the allele frequency in recipients on that in donors.

This is almost (but not quite) the answer we need. The trouble is that $\bar{c}$ and $\bar{b}$ are the mean effects of altruism per individual, not the effects (C and B) per act of altruism. Fortunately, it is easy to translate. If m is the mean number of altruistic interactions per person, then $\bar{c} = mC$ and $\bar{b} = mB$. Thus,

$$\Delta\bar{p} = K(u - \bar{p})(rB - C), \tag{13.4}$$

The two covariances in Eqn. 13.3 are

$$\text{Cov}(b_i, p_i) = E[b_i p_i] - \bar{b}\bar{p}, \qquad \text{and} \qquad \text{Cov}(c_i, p_i) = E[c_i p_i] - \bar{c}\bar{p}$$

where $\bar{b} = E[b_i]$ and $\bar{c} = E[c_i]$ are the mean benefit and cost of altruism per individual. Let us rewrite the covariances in terms of two new variables,

$$u = \sum c_i p_i \Big/ \sum c_i, \qquad \text{and} \qquad v = \sum b_i p_i \Big/ \sum b_i.$$

The first equation says that u is an average of p_i weighted by c_i. Since c_i is proportional to the number of times i has served as donor, u is the average frequency of allele A among donors. Similarly, v is the average frequency among recipients. Rearranging, $\sum c_i p_i / N = u \sum c_i / N$ and $\sum b_i p_i / N = v \sum b_i / N$, or $E[c_i p_i] = u\bar{c}$ and $E[b_i p_i] = v\bar{b}$. The two covariances become

$$\text{Cov}(b_i, p_i) = (v - \bar{p})\bar{b} \qquad \text{and} \qquad \text{Cov}(c_i, p_i) = (u - \bar{p})\bar{c},$$

Now set $v - \bar{p} = r(u - \bar{p})$. This defines r as the regression coefficient of the allele frequency of recipients on that in donors. The first covariance becomes

$$\text{Cov}(b_i, p_i) = r(u - \bar{p})\bar{b}$$

Box 7: Re-expressing the covariances in Eqn. 13.3

where $K = m/2\bar{W}$ is a positive constant.

Allele A will increase in frequency when $\Delta\bar{p} > 0$. When is this so? The answer does not depend on K, for that is always positive. It may, however, depend on $u - \bar{p}$. This quantity is positive if carriers of A perform more altruism than random members of the population. Formally, it is proportional to the covariance between the frequency p_i of the allele within individuals and the rate at which they perform altruism, as shown in Box 7. It seems reasonable to define an altruist allele as one for which $u - \bar{p}$ is positive. Given this definition, Eqn. 13.4 tells us that altruist alleles increase in frequency if $rB - C > 0$. This is Hamilton's Rule.

We turn now to a series of examples that will make this discussion a little less abstract. In the process, it will turn out that our definition of "altruist allele" has a surprising implication.

13.2.1 Example 1

Suppose that the three genotypes perform altruism at the following rates.

Genotype	AA	Aa	aa
Frequency	p^2	$2pq$	q^2
Rate of altruism	3	2	1
p_i	1	$1/2$	0

What do these data imply about the frequency u of allele A in a random donor? The ratio of AA to Aa to aa in the population at large is $p^2 : 2pq : q^2$. To get the corresponding ratio among donors, weight by rates of altruism to get $3p^2 : 4pq : q^2$. Finally, use these as weights to calculate the weighted average of p_i:

$$u = \frac{3p^2 \times 1 + 4pq \times 0.5 + q^2 \times 0}{3p^2 + 4pq + q^2} = \frac{p(p+2)}{2p+1}$$

To state this recipe differently: (1) multiply the three numbers in each column of the table and sum the resulting products, (2) repeat this process, using only the values in rows 1 and 2, and (3) divide the result of step 1 by that of step 2. The result is u, the frequency of allele A among donors.

⋆EXERCISE **13–1** Enter this formula into Maple, and graph $u - \bar{p}$ as a function of $\bar{p}$. Is $u - \bar{p}$ always positive, always negative, or sometimes positive and sometimes negative? Is there any point where $u - \bar{p} = 0$? Use Eqn. 13.4 to interpret your results.

13.2.2   Example 2

Suppose that the three genotypes perform altruism at the following rates.

Genotype	AA	Aa	aa
Frequency	p^2	$2pq$	q^2
Rate of altruism	2	3	1
p_i	1	1/2	0

Then, if the population is close to Hardy-Weinberg equilibrium (an assumption that requires selection to be weak), the frequency u of A among donors will be approximately

$$u = \frac{2p^2 + 3pq}{2p^2 + 6pq + q^2}$$

⋆EXERCISE **13–2** Enter this formula into Maple, and graph $u - p$ as a function of p. Is $u - p$ always positive, always negative, or sometimes positive and sometimes negative? Is there any point where $u - p = 0$?

You should discover that $u - p$ is positive when p is small, but negative when p is large. This happens because the heterozygote performs altruism at a higher rate than any other genotype. For such alleles, Hamilton's rule gives the condition for the allele to increase when rare and to decrease when common.

This may seem to reduce the generality of Hamilton's rule. However, when p is common, most individuals are either of genotypes AA or Aa, and the latter genotype provides more altruism than the first. In other words, A is an altruist allele only when it is rare. When A is common, a is the altruist allele. Hamilton's rule gives the condition under which selection will tend to increase the provision of altruism, regardless of the underlying genetic system. In this example, the result is a balanced polymorphism at which both alleles remain in the population.

13.3   Inclusive Fitness

Hamilton's rule refers to the case in which an individual's actions affect just two individuals, himself and a relative. What if his behavior affects several individuals? An allele that predisposes an individual toward such an action is favored by selection when a quantity called "inclusive fitness" is positive. Inclusive fitness, I, is defined as

$$I = \delta W_0 + \sum_i r_i \delta W_i \tag{13.5}$$

where δW_0 is the effect of the behavior on the individual who performs the behavior (let us call this individual Ego), δW_i the behavior's effect on Ego's ith relative, and r_i the coefficient of relationship between this relative and Ego.

In the case of altruism directed toward a single recipient, the effect on Ego is $\delta W_0 = -C$. The effect on the single recipient is $\delta W_1 = B$. Thus, $I = -C + rB$, and Hamilton's rule gives the condition under which $I > 0$. But inclusive fitness is more general than this. The effect on Ego can be negative as well as positive, and there can be multiple recipients.

If I could travel back in time to 1963, I would advise Hamilton to choose some name other than "inclusive fitness." People are used to thinking that selection maximizes ordinary Darwinian fitness. This gets them into trouble, for selection does *not* maximize inclusive fitness. The altruist allele continues to spread so long as $I > 0$. It does not stop spreading when I reaches some maximal value.

To compound the confusion, some people apply the term "inclusive fitness" to a different quantity,

$$Y(x) = W_0(x) + \sum_i r_i W_i(x),$$

where x is the rate at which the donor (individual 0) exhibits altruism. Y is the sum of the donor's Darwinian fitness plus that of other individuals weighted by their relatedness (r_i) to the donor. If some mutation causes a small increases the donor's rate of altruism, the corresponding change in Y is equal to I. Thus, Hamilton's inclusive fitness is not Y; it is more like dY/dx. Had Hamilton defined Y as inclusive fitness, it would have avoided a lot of confusion. We could talk sensibly about selection maximizing inclusive fitness, and we could search for optima by setting the derivative equal to zero. Instead, we must all deal with a literature in which different authors use the term differently. For further discussion, see Dawkins [17, pp. 185–186].

This section will use the term Hamilton's way: inclusive fitness will refer to I rather than Y. I will use it to study yet another variant of the Smith-Fretwell model of clutch size and parental investment.

13.4 Conflict between Mother and Fetus

How much resource should a mother allocate to each fetus? The answer depends on your point of view. Trivers [72] pointed out that genes that are expressed in the mother may favor a different allocation than genes expressed by the offspring. The resulting conflict of interest has apparently led to an evolutionary arms race between mothers and fetuses. Fetuses evolve strategies that extract more resources from mothers, while mothers evolve strategies that limit what fetuses are able to extract [27]. To explore this subject, I will consider three models. The first assumes that the mother is in charge; the others deal with two forms of control by offspring, assuming either that (a) the offspring's behavior increases its own fitness only, or that (b) the offspring's behavior increases the fitness of his entire litter.

13.4.1 Maternal control

Recall from chapter 6 that Smith and Fretwell [66] wrote the parent's fitness as

$$W = bs(x) = Ts(x)/x$$

where $b = T/x$ is the initial number of offspring, $s(x)$ the probability that an offspring receiving x units of resource will survive to maturity, and T the total quantity of resource that is available to the parent. Smith and Fretwell made the reasonable assumption that the investment, x, per offspring is determined by the parent. They showed that, under this assumption, evolutionary equilibrium requires that
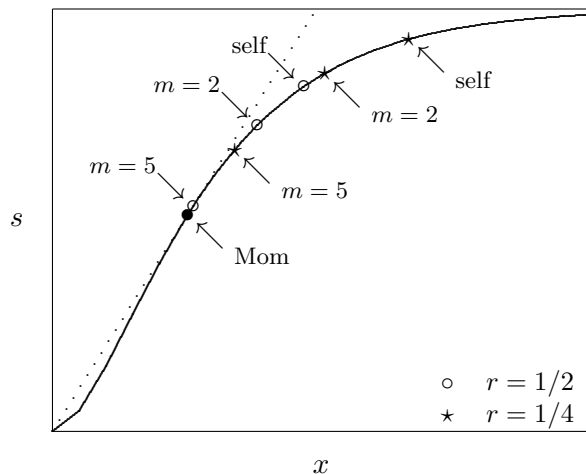
$$s'(x)/s(x) = 1/x \tag{13.6}$$

Figure 13.1: Optimal allocation to offspring from various points of view
The solid line shows the fitness s of an offspring who receives x units of parental resource. The filled circle shows the optimal allocation from the mother's point of view, as given by the model of Smith and Fretwell. The other symbols show optima according to various models in which allocation is controlled by offspring. Open circles refer to cases in which all offspring are full siblings, stars to cases in which they are half siblings. The points labeled "self" refer to optima for manipulations that benefit only the bearer of the gene. The remaining points refer to manipulations that benefit an entire litter of size m.

This equilibrium is shown as a filled circle in figure 13.1.

13.4.2   Fetal manipulations on behalf of the individual fetus

Consider now a mutant allele, expressed by the fetus, which extracts more maternal nutrients for itself. To model such phenomena, let us suppose that a fetal mutation changes the resource allocated to its bearer from x to y, but does not affect the allocation to littermates, who still receive x. Under what circumstances will this allele invade?

I assume that most individuals in the population are A_1A_1 homozygotes, but that there are a few A_1A_2 heterozygotes. These heterozygotes are reproduced via two types of matings:

$$\overbrace{A_1A_2}^{\text{mother}} \times \overbrace{A_1A_1}^{\text{father}} \qquad \text{and} \qquad \overbrace{A_1A_1}^{\text{mother}} \times \overbrace{A_1A_2}^{\text{father}}.$$

(We can ignore matings involving more than one A_2 if that allele is sufficiently rare.) Denote by r the fraction of the mother's offspring that are A_1A_2. If the heterozygous parent is the mother, then $r = 1/2$. If the heterozygous parent is the father, then $r = 1/2$ if the female mates with just one male, and $r < 1/2$ if she mates with more than one. Each heterozygous fetus extracts y units of resource from the mother, whereas each A_1A_1 homozygote extracts x units.

Fitness, W, is the number of A_1A_2 offspring produced by the mother, regardless of whether she is the homozygous or the heterozygous parent.[1] W will be the product of the number b of offspring

---

[1]Why, you may ask, should we count the mother's offspring even when it is the father who is the heterozygote? The reproductive success of a copy of the A_2 allele in a male is the product of the number of females with whom he mates times the number of heterozygous offspring produced by each mated female. I am assuming that the A_2 allele has no effect on male mating success; it affects only the number of offspring produced per mating. Thus, it is appropriate to count the heterozygous offspring of the female even when the A_2 gene resides in the father.

produced, the fraction r of these that are $A_1 A_2$, and the fraction $s(y)$ of these births that survive to adulthood. Thus,

$$W = rbs(y)$$

where b satisfies the constraint equation

$$T = b((1-r)x + ry).$$

Here, T is the total resource available to the mother, and $(1-r)x + ry$ is the amount of resource used by an average offspring. Solving the constraint equation for b and substituting the result into the expression for W gives

$$
\begin{aligned}
W &= \frac{rTs(y)}{(1-r)x + ry}, \quad \text{and} \\
\ln W &= \text{const.} + \ln s(y) - \ln((1-r)x + ry)
\end{aligned}
$$

To find the evolutionary equilibrium, I use the method described in section 9.3.2. The first step is to find the value of y that maximizes the mutant's fitness. The second is to set $x = y$, since x is a stable equilibrium if it is the best reply to itself.

The derivative of $\ln W$ with respect to y is

$$\frac{\partial \ln W}{\partial y} = \frac{s'(y)}{s(y)} - \frac{r}{(1-r)x + ry}$$

Setting this equal to zero and then setting $y = x$ gives the condition for equilibrium:

$$s'(x)/s(x) = r/x \tag{13.7}$$

The right-hand side of this equation is smaller than the corresponding part of Eqn. 13.6. If $s(x)$ increases at a decreasing rate, then smaller values of $s'(x)/s(x)$ imply a greater allocation to each offspring. This is illustrated in Figure 13.1, where the points labeled "self" refer to equilibria defined by Eqn. 13.7. Offspring control leads to smaller litters of better nourished offspring. The difference between the optima for mothers and for offspring implies a conflict of interest. The intensity of conflict depends only on r, the mean coefficient of relationship between siblings. As the figure shows, conflict is most pronounced where r is small because each female mates with multiple males. This might lead, for example, to the evolution of placentas that invade maternal tissue deeply in an effort to extract nutrients. The level of conflict should be independent of litter size.

$\star$ EXERCISE **13–3** Suppose that

$$s(x) = 1 - e^{-(x-1)}$$

(Ignore the negative values that this formula gives when x is small. They will not affect the answer.) What is the equilibrium under parental control (the standard Smith-Fretwell model), under monogamy (where $r = 1/2$), and under offspring control with promiscuous mating (where $r = 1/4$)?

13.4.3 Fetal manipulations on behalf of the entire litter

Other forms of fetal manipulation require a different model. For example, fetuses also manipulate maternal allocations by secreting hormones into the mother's bloodstream. The effects are enjoyed by the entire litter. To model such manipulations, I assume that a mutant fetal gene changes the

allocation to itself and all its litter mates from x to y. A random fetus uses x units of resource if neither it nor any of its littermates is a heterozygote; it uses y units of resource otherwise. The first of these possibilities has probability $(1 - r)^m$, where r is the probability that a random fetus is a heterozygote and m is the number of fetuses in a litter. Thus, a random fetus uses $(1 - r)^m x + (1 - (1 - r)^m)y$ units of resource. The constraint equation is

$$T = b((1 - r)^m x + (1 - (1 - r)^m)y)$$

and the mutant's fitness and log fitness are

$$
\begin{aligned}
W &= \frac{rTs(y)}{(1 - r)^m x + (1 - (1 - r)^m)y} \\
\ln W &= \text{const.} + \ln s(y) - \ln((1 - r)^m x + (1 - (1 - r)^m)y)
\end{aligned}
$$

To find the evolutionary equilibrium, the first step is to differentiate:

$$\frac{\partial \ln W}{\partial y} = \frac{s'(y)}{s(y)} - \frac{1 - (1 - r)^m}{(1 - r)^m x + (1 - (1 - r)^m)y}$$

Next, set the derivative equal to zero and set $y = x$ to obtain the condition for equilibrium:

$$\frac{s'(x)}{s(x)} = \frac{1 - (1 - r)^m}{x} \tag{13.8}$$

This result produced the equilibria labeled $m = 2$ and $m = 5$ in figure 13.1. Conflict decreases both with relatedness (r) and with litter size (m). When $m > 1$ the right side of Eqn. 13.8 is greater than that of Eqn. 13.7. Thus, the equilibrium value of $s'(x)$ is larger for manipulations benefitting the litter than for those benefitting the individual fetus. Because of the shape of the function s, large values of $s'(x)$ imply small values of x. Consequently, if fetal manipulations benefit the entire litter, then each fetus will be less selfish than if manipulations benefit only a single fetus.

This effect is especially pronounced when litters are large. In the limit, as litter size becomes infinite, $(1 - r)^m$ goes to zero and Eqn. 13.8 becomes identical to Eqn. 13.6—the equation for maternal control. When litters are large, there is little conflict between mother and fetus because the optima for the two parties are roughly the same. When litters are small, the opposite is true and conflict should be pronounced. Thus, we should expect to see most evidence of conflict in species with small litters.

There is some supportive evidence. Mammalian pregnancies are maintained by progesterone, which is secreted by the corpus luteum (a maternal tissue). One can terminate a pregnancy by removing the corpus luteum in such mammals as hamsters, goats, pigs, rabbits [27, p. 507], and dogs [15]. In humans however, this is true only until the eighth week of gestation. After that time, the placenta (a fetal tissue) produces enough progesterone to maintain the pregnancy. David Haig interprets this as an adaptation by the fetus to maintain pregnancies that the mother would like to abort. Since such abortions affect all litter mates equally, a fetus that prevents them is increasing the allocation of resources not only to itself but also to its litter mates. Thus, Eqn. 13.8 is most nearly appropriate and conflict is most likely where litter-sizes are small. This agrees with the observation that human placentas produce enough progesterone to maintain pregnancy whereas the placentals of hampsters, pigs, and rabbits do not. The model does not explain why goats, whose litters are also small, have not evolved similar counter-measures.

⋆ EXERCISE **13**–**4** Suppose once again that

$$s(x) = 1 - e^{-(x-1)}$$

and also assume that sibs are full sibs. If manipulations affect an entire litter of size m, what is the equilibrium if $m = 1$, if $m = 5$, if $m = 15$, and if $m = 100$.

Chapter 14

Growth of Populations with Age Structure

Many questions in evolutionary ecology involve trade-offs across time. For example, which strategy would selection favor: one that has many children late in life, or one that has fewer but does so early? The winner need not be the one with the most children. For such problems, we need a different currency. The relevant currency is related to the rate of population growth, so let us begin there.

The models below ignore sex. There is no distinction between the vital rates of males and females. In fact, there are no males at all. In demography, this is generally justified by appeal to the principle of *female demographic dominance*, which maintains that males don't matter much—fertility is determined by the number of females. The idea is that it doesn't take many males to fertilize a large number of females. Whatever the merit of this argument, it is a useful place to start if only because it makes things simple.

There are several ways to formulate the problem of population growth in an age structured population. Of these, the renewal equation and the Leslie matrix are used far more often than any other. We begin with the renewal equation.

14.1 The Renewal Equation

The renewal equation provides a method for calculating a population's rate r of growth from its vital rates. The equation that makes this possible is called the *characteristic equation* of the renewal process [22]:

$$1 = \sum_x e^{-rx} l_x m_x, \tag{14.1}$$

and is derived in Box 8. In this equation, l_x is the probability that a female will survive from birth to age x, and m_x is the age-specific fertility: the mean number of girl babies born per woman between the ages of x and $x + 1$.

Let $n_x(t)$ represent the number of females of age x alive at time t. The number of births, $n_0(t)$, is the sum of the fertilities of individuals of all ages:

$$n_0(t) = \sum_x n_x(t) m_x \tag{14.2}$$

But the number of x-year-old individuals at time t is just the number of births at time $t - x$ times the fraction of those births that survive to age x. In other words, $n_x(t) = n_0(t - x) l_x$, and

$$n_0(t) = \sum_x n_0(t - x) l_x m_x. \tag{14.3}$$

This is called the *renewal equation* [22]. If the population has reached a stable age distribution (explained in the text), then the number in each age class is increasing (or decreasing) exponentially. In other words,

$$n_x(t) = e^r n_x(t - 1) \tag{14.4}$$

where r is the rate of exponential growth. Substituting this into Eqn. 14.3 gives

$$n_0(0) e^{rt} = n_0(0) e^{rt} \sum_x e^{-rx} l_x m_x$$

Dividing through by $n_0(0) e^{rt}$ gives Eqn. 14.1.

Box 8: Derivation of the characteristic equation of the renewal process

---

Eqn 14.1 relies on the central result of "stable population theory," which holds that if the age-specific birth and death rates are unchanging, then population will converge to a state in which although the number of individuals in each age class may change, their relative frequencies do not. This array of unchanging relative frequencies is called the *stable age distribution*.[1]

It may be hard to imagine that the numbers could change without affecting the relative frequencies, so consider an example. Suppose that in one generation some population has 100 individuals of age 1 and 200 of age 2. The relative frequencies within these two age classes are thus 1/3 and 2/3. A few generations later, we revisit the population and discover that the number in each age class has doubled. There are now 200 of age 1 and 400 of age 2. The population has grown, but the relative frequencies (1/3 and 2/3) have not changed.

Once the population has reached a stable age distribution, the number in each class grows (or shrinks) by a factor of e^r in each time period, and so does the number in the population as a whole. After t time periods, the population is e^{rt} times as large as it was to begin with. r is called the rate of exponential growth and appears as a parameter in Eqn. 14.1.

$\star$ EXERCISE **14–1** If a population is growing at rate r, how long does it take for it to double in size?

---

[1]Actually, there is a little more to it than this. Some schedules of age-specific fertility and mortality lead not to exponential growth, but to various kinds of oscillations. However, if there is positive fertility in two adjacent age classes, then the population always converges to a stable age distribution.

Table 14.1: Vital statistics for female grey squirrels. Source: Table 1.3 of [8].

x	l_x	m_x
0	1.000	0.00
1	0.253	1.28
2	0.116	2.28
3	0.089	2.28
4	0.058	2.28
5	0.039	2.28
6	0.025	2.28
7	0.022	2.28

Eqn 14.1 is important because it provides information about r. If we know the numerical values of l_x and m_x for all ages, then r is the only unknown in the equation. We can find r simply by experimenting with different values until we find one that satisfies the equation. This would be tedious to do by hand, but it is easy with a computer.

On the other hand, the numerical approach only works when we have numerical values for l_x and m_x. Theoreticians often prefer to work with symbols because this makes their answers general. There is no way to rearrange Eqn. 14.1 so that r is on one side and everything else is on the other. We cannot, in other words, solve for r. Nonetheless, the equation does define r as an *implicit* function of l_x and m_x. This makes it possible (as explained in Appendix E) to derive formulas for such quantities as $\partial r / \partial m_x$ and $\partial r / \partial l_x$. I will make use of both approaches—numerical and symbolic—in what follows.

14.1.1   Numerical Example

Table 14.1 shows the demographic parameters of a population of gray squirrels. To calculate r from these data using Maple, I begin by using the data in table 14.1 to define arrays for l_x and m_x:

```
l := <1, 0.253, 0.116, 0.089, 0.058,
    0.039, 0.025, 0.022>;
m := <0, 1.28, 2.28, 2.28, 2.28, 2.28,
    2.28, 2.28>;
```

There is a source of potential confusion here. In the math, the first age is age 0, so our initial survivorship and fertility are l_0 and m_0. But the first index of a Maple array is 1, not 0, so l_0 corresponds to Maple's `l[1]`, and m_0 to Maple's `m[1]`.

Since the data exhibit positive fertility in several adjacent age classes, we are justified in using Eqn. 14.1 (see footnote 1). I therefore define a variable that is equal to the right hand side of Eqn. 14.1.

```
rh := sum('exp(-r*x)* l[x+1]*m[x+1]', x=0..7);
```

Here `rh` stands for "right-hand side."

Since `rh` is a function of a single variable, it can be plotted in Maple by typing `plot(rh, r=-0.2..0.2)`. This produces a graph similar to figure 14.1. The solution to Eqn. 14.1 is the value of r that makes `rh` equal to unity. In figure 14.1, this is the point where the solid line crosses the horizontal dotted line drawn at `rh` = 1. The figure shows that the solution is somewhere between 0 and 0.05.

To find the solution more exactly, I used Maple's `fsolve` function. Since, the solution appears to be near zero, I instructed `fsolve` to search the interval `0..0.05`.
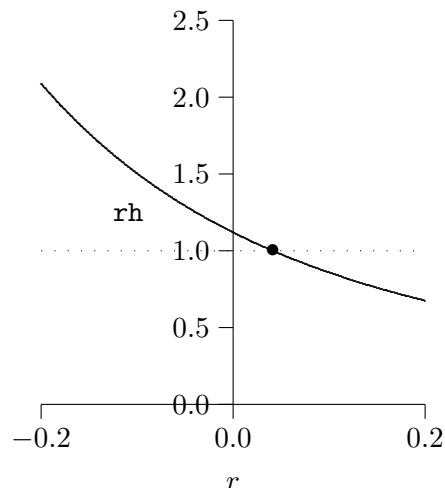
Figure 14.1: Right-hand side of characteristic equation for data in table 14.1

```
> rhat := fsolve(rh=1, r, 0..0.05);
      rhat := .04130450880
```

Thus, the intrinsic rate of increase is about $r = 0.04$. To check that this solution is correct, substitute it back into `rh`.

```
> evalf(subs(r=rhat, rh));
      .9999999999
```

The answer is equal to 1 to within the limits of Maple's precision. Thus, `rhat` is the value that makes `rh` $= 1$, just as required by Eqn. 14.1.

Figure 14.1 suggests that `rh` is a function that always decreases. If so, then there can be no other values of r that satisfy 14.1. Thus, the figure suggests that $r = 0.04$ is the unique solution: the demographic parameters are consistent with only this single rate of exponential growth. This conclusion must be qualified in two ways. First, the demographic parameters of some populations— for example those with fertility in only a single age class—imply that the population will never grow at a constant exponential rate, as we have assumed. Second, the intrinsic growth rate only tells us the rate of growth after a stable age distribution has been reached. It tells us neither how rapidly the population will approach its stable age distribution nor how it will fluctuate along the way. Both questions are more easily addressed using another formulation of the mathematics of population growth—the Leslie matrix.

14.2   The Leslie Matrix

The growth of an age-structured population can also be described using what is called the *Leslie matrix* [45, 46, 55]. Whereas the renewal equation keeps track of births only, the Leslie matrix keeps track of the number of individuals in each age class. Let us begin by asking what determines the number of 3-year olds at time $t + 1$.

The number of 3-year-olds this year equals the number of 2-year-olds in the population last year times the fraction of those 2-year-olds that survived. Thus, if P_x is the probability that a female of age x will survive to age $x + 1$, then

$$n_{x+1}(t + 1) = P_x n_x(t). \tag{14.5}$$

Given the age distribution at time t, this equation predicts the number of individuals at time $t+1$ in age categories $1, 2, \ldots, K$ where K is the maximum age attainable. But it does not predict the number in the youngest age class.

The youngest age class in the data will depend on when during the year the population is censused and on how the population distributes births across the year. In discussing the renewal equation, I assumed without saying so that births were seasonal and that the population was censused just after the season in which births occur. This allowed me to equate the number of births with the number of observed births. In practice, some births will not be observed because the newborns will die before they are counted. The Leslie matrix makes the opposite assumption: that the population is censused just before births occur, so that the youngest age class is nearly one year old.[2] The new assumption demands a new notation. Let

$$f_x \equiv m_x P_0$$

denote what is called the *net fecundity* of females of age x: the mean number of 1-year-old young in year $t+1$ produced per x-year-old mother in year t. The number of newborns at time $t+1$ is

$$n_1(t+1) = \sum_{x=1}^{K} n_x(t) f_x \tag{14.6}$$

Compare this with Eqn. 14.2 above.

If the maximum age is (for example) $K = 3$, equations 14.5 and 14.6 can be written as

$$\left.\begin{array}{rcl} n_1(t+1) &=& n_1(t) \times f_1 + n_2(t) \times f_2 + n_3(t) \times f_3 \\ n_2(t+1) &=& n_1(t) \times P_1 + n_2(t) \times 0 + n_3(t) \times 0 \\ n_3(t+1) &=& n_1(t) \times 0 + n_2(t) \times P_2 + n_3(t) \times 0 \end{array}\right\} \tag{14.7}$$

Before proceding, make sure that you see that equations 14.7 are simply another way of writing equations 14.5 and 14.6.

In matrix notation, these same equations are written as

$$\begin{pmatrix} n_1(t+1) \\ n_2(t+1) \\ n_3(t+1) \end{pmatrix} = \begin{pmatrix} f_1 & f_2 & f_3 \\ P_1 & 0 & 0 \\ 0 & P_2 & 0 \end{pmatrix} \begin{pmatrix} n_1(t) \\ n_2(t) \\ n_3(t) \end{pmatrix} \tag{14.8}$$

Here, the vertical columns containing $n_x(t)$ and $n_x(t+1)$ are called *vectors*, and the square array is called a *matrix*. Eqn. 14.8 says that the vector on the left is the product of the matrix and vector on the right. For a synopsis of matrix multiplication, see Box 9.

Let us now give names to the three quantities in Eqn. 14.8. Let $n(t+1)$ denote the vector on the left, $n(t)$ the vector on the right, and let L (for Leslie) denote the matrix. Eqn. 14.8 becomes

$$n(t+1) = Ln(t) \tag{14.10}$$

This equation makes it easy to project a population forward in time. Given any initial vector $n(0)$ of numbers of individuals in each age category, one can calculate $n(1) = Ln(0)$, $n(2) = Ln(1) = L^2 n(0)$, $n(3) = Ln(2) = L^3 n(0)$, and so forth. In general,

$$n(t) = L^t n(0)$$

---

[2]With human data, neither assumption is appropriate. Consequently, demographers have to modify their formulas to avoid introducing bias into estimates. I will ignore these issues in this chapter, but you should *not* ignore them if you are analyzing human data.

What does it mean to multiply a matrix by a vector, as is done in Eqn. 14.8? The simplest answer is that it is *defined* to mean the same thing as equations 14.7. Here is a 2×2 example:

$$\begin{pmatrix} am + bn \\ cm + dn \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} m \\ n \end{pmatrix} \qquad (14.9)$$

The vector on the left is the product of the matrix and vector on the right. The first entry of the left-hand vector is obtained by multiplying the elements of the first row of the matrix times the right-hand vector, and adding the results. The second entry of the left-hand vector is calculated in the same way using the matrix's second row. When this is clear to you, re-examine equations 14.7 and 14.8. They illustrate matrix multiplication on a larger scale.

● EXAMPLE **14–1**

Calculate the following matrix product.

$$\begin{pmatrix} 1 & 0 \\ 2 & 3 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

○ ANSWER

$$\begin{pmatrix} 1 \\ 5 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 2 & 3 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

● EXAMPLE **14–2**

Multiply the matrix in Eqn. 14.9 by the vector on the left side of the equals sign there.

○ ANSWER

$$\begin{pmatrix} a^2 m + abn + bcm + bdn \\ cam + cbn + dcm + d^2 n \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} am + bn \\ cm + dn \end{pmatrix}$$

Box 9: Matrix multiplication

14.2.1 Numerical example

In this example, I use the squirrel data in table 14.1 to calculate a Leslie matrix, use this matrix to project the population forward in time, and to calculate the intrinsic rate of increase. This will show that analysis of the Leslie matrix yields exactly the same answer as analysis of the renewal equation.

The table provides the probability l_x of surviving to age x, but not the probability P_x of surviving from x to $x + 1$. These latter probabilities are easily calculated from the former, since

$$\begin{aligned} l_x P_x &= l_{x+1}, \qquad \text{or} \\ P_x &= l_{x+1}/l_x \end{aligned}$$

I did this calculation in Maple as shown to the right. The first statement allocates an array, P, to hold the answers. Note that Maple represents P as a column of zeroes: a "column-vector." To save space, I will often re-write vectors as rows.

```
>P := Vector(7, 0)


                [0]
                [ ]
                [0]
                [ ]
                [0]
                [ ]
       P  :=    [0]
                [ ]
                [0]
                [ ]
                [0]
                [ ]
                [0]
```

Next, the `for` statement calculates `P[i]` for values of `i` ranging from 1 through 7.

```
>for i from 1 to 7 do
>  P[i] := l[i+1]/l[i];
>  end do;
    P[1] := .253
    P[2] := .4584980237
    P[3] := .7672413793
    P[4] := .6516853933
    P[5] := .6724137931
    P[6] := .6410256410
    P[7] := .8800000000
```

To keep the answers consistent with the calculations of section 14.1, we have to assume that the youngest age category in the squirrel data refers to individuals just after birth. This means that `P[1]` really refers to P_0, `P[2]` to P_1, and so forth. Until Maple implements vectors whose indices begin at zero, we will have to live with such inconsistencies.[3]

```
leslie := proc(P,m)                       end do;
   local Lmat, i, j, len;              for i from 1 to (len-1) do
   len := Dimension(m) - 1;              Lmat[i+1,i] := P[i+1];
   # len X len matrix of zeroes          end do;
   Lmat := Matrix(len,len,0);         Lmat;
   for j from 1 to len do             end;
     Lmat[1,j] := P[1]*m[j+1]
```

Box 10: Maple procedure for making a Leslie matrix.

Before using matrix routines, tell Maple to read the LinearAlgebra package

```
with(LinearAlgebra):
```

Box 10 shows a Maple procedure that automates the task of constructing a Leslie matrix from `P` and `m`. With the two vectors just defined, this procedure gives

```
> L := leslie(P,m);

L := [.32384, .57684, .57684, .57684,
            .57684, .57684, .57684]
      [.4584980237, 0, 0, 0, 0, 0, 0]
      [0, .7672413793, 0, 0, 0, 0, 0]
      [0, 0, .6516853933, 0, 0, 0, 0]
      [0, 0, 0, .6724137931, 0, 0, 0]
      [0, 0, 0, 0, .6410256410, 0, 0]
      [0, 0, 0, 0, 0, .8800000000, 0]
```

---

[3]Maple 10 does have such vectors, but many of my students still use Maple 7.

This is the Leslie matrix for the squirrel population. The structure is easier to see when it is rewritten as:

$$L = \begin{pmatrix} 0.324 & 0.577 & 0.577 & 0.577 & 0.577 & 0.577 & 0.577 \\ 0.457 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.767 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.652 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.672 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.641 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.880 & 0 \end{pmatrix}$$

Note that this matrix is identical in form to that in Eqn. 14.8. The net fecundities are listed across the first row, the age-specific survival probabilities are listed on the first subdiagonal. Matrices of this form are called Leslie matrices.

To show how Leslie matrices can be used, I begin by using Eqn. 14.10 to project the squirrel population forward in time. First, I define an arbitrary initial vector in which there are 10 females at each age:

```
n0 := <10,10,10,10,10,10,10>;
```

Next, the following line calculates the population at time 1:

```
> nn := L . n0;
nn := [ 37.84880, 4.584980237, 7.672413793,
  6.516853933, 6.724137931, 6.410256410,
  8.800000000 ]
```

Here, the "." operator stands for matrix multiplication. Thus, the command `nn := L . n0` has the same meaning as Eqn. 14.10. It says to perform the matrix multiplication $Ln(0)$ and assign the result to `nn`. Thus, `nn` is now equal to $n(1)$. Notice that there are 37.8 1-year-olds in year 1, about 4 times the number in year 0. On the other hand, all of the other age categories have decreased in size. Thus, the age distribution is changing rapidly.

We now have $n(0)$ and $n(1)$. To calculate $n(2), n(3), \ldots, n(15)$, use Maple's `for` command:

```
#calc age dist for t=2...20:
for i from 2 by 1 to 20 do
  nn := L . nn;
  end do;
```

At each pass through the loop, the value of `nn` is replaced by the product of L with the old `nn`. Maple prints out each of these products, and the last three lines of output look like this:

```
nn := [ 63.00238786, 27.71749776, 20.40644546,
  12.76274474, 8.232158038, 5.062908857,
  4.274427622 ]
nn := [ 65.65935759, 28.88647032, 21.26601121,
  13.29858244, 8.581845601, 5.277024383,
  4.455359794 ]
nn := [ 68.42861842, 30.10468569, 22.16289533,
  13.85874888, 8.942150261, 5.501183077,
  4.643781457 ]
```
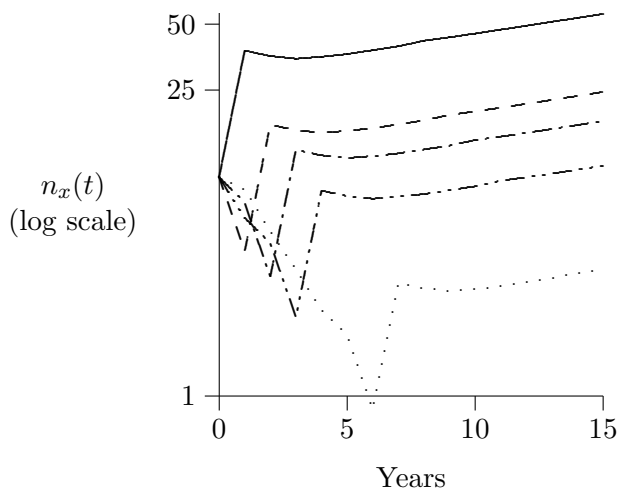
Figure 14.2: Projection of gray squirrel population

Shown here are $n(18)$, $n(19)$, and $n(20)$. I captured all this output, and edited it to produce figure 14.2.

To keep the figure readable, I show only a few age classes. Notice that after a few generations of wild oscillations, the time path of the age classes become parallel straight lines. Since the vertical axis is on log scale, straight lines mean that the age classes are increasing exponentially. The fact that the lines become parallel is also important: it means that the relative frequencies of the age classes have stopped changing. In other words, the population has reached a stable age distribution. The speed of this process is remarkable; it appears that the stable age distribution is reached in less than ten years.

Let us now use these data to calculate the rate of increase. Eqn. 14.4 asserts that, once the stable age distribution has been reached, the number in each age class increases by a constant ratio e^r. For convenience, I rewrite this equation as

$$n_x(t+1) = \lambda n_x(t) \tag{14.11}$$

where $\lambda \equiv e^r$ is called the ratio of increase, and is constant at the stable age distribution. Using the data generated above, I calculate that

$$\begin{aligned} n_1(14)/n_1(13) &= 1.04229 \\ n_1(15)/n_1(14) &= 1.04248 \\ n_7(15)/n_7(14) &= 1.04189 \end{aligned}$$

At the stable age distribution, these numbers should be identical. They are nearly so, and this indicates that after 15 years the squirrel population has almost, but not quite, reached its stable age distribution.

Are these numbers consistent with the value of $r = 0.04130450880$ that we calculated from the renewal equation? With this value, $\lambda = e^r = 1.042169407$, which is certainly close to the estimates above. To get a more accurate estimate of λ from the Leslie matrix, I calculate its *eigenvalues*. Without attempting to explain what eigenvalues are, I will simply assert that λ is equal to the eigenvalue of largest magnitude. To calculate them in Maple, type:

```
> evals := Eigenvalues(L);
```

Table 14.2: Female demographic data for 1906 human population of Taiwan

x (years)	l_x	m_x
0	1.0000	0.00000
5	0.6498	0.00000
10	0.5945	0.00000
15	0.5680	0.25600
20	0.5353	0.80235
25	0.4948	0.82955
30	0.4514	0.75690
35	0.4061	0.64001
40	0.3615	0.27713
45	0.3193	0.02500
50	0.2753	0.00000

Note: Ages after 50 contribute no
births, and are therefore omitted.

```
evals :=
  1.042169410,
    .3409946041 + .5071133337 I,
    .3409946041 - .5071133337 I,
 - .5388684430 + .2549467576 I,
 - .5388684430 - .2549467576 I,
 - .1612908651 + .5802170652 I,
 - .1612908651 - .5802170652 I
```

The dominant eigenvalue here is $\lambda = 1.042169410$, which is essentially identical to the answer provided by the renewal equation.

This exercise has shown that the renewal equation and the Leslie matrix are different ways to arrive at the same answer. The Leslie matrix tells us what is happening in greater detail, but is somewhat more cumbersome.

14.3 Exercises with human demographic data

Table 14.2 shows the female demographic data for the 1906 population of Taiwan.

⋆ EXERCISE **14–2** Use the renewal equation to calculate r from the Taiwan data.

⋆ EXERCISE **14–3** Calculate the Leslie matrix for the data in table 14.2.

⋆ EXERCISE **14–4** What value of λ is implied by the Taiwan Leslie matrix? How does this answer compare to the result from the renewal equation?

⋆ EXERCISE **14–5** These exercises tell us about the rate of population growth per time period, where each time period is 5 years long. What is the annual growth rate (i.e. the rate of growth per year).

Chapter 15

Life History Evolution

15.1   When can a rare allele invade?

The previous chapter developed models for the growth of age-structured populations. This one applies these models to a different issue: that of selection between alternate alleles within a single population. A comprehensive theory of selection with age-structure is hard, and we deal here with a small piece of the problem—we ask whether an allele can increase in frequency when rare.

To simplify the exposition, I will assume that the population has a monoecious sexual system of reproduction. This means that there is sex but not sexes: There is no distinction between male and female, and everyone can mate with everyone else. Adding sexes to the model is not hard, but the argument is simpler without them.

Consider the dynamics of a rare allele B in a population dominated by a common allele A. Since B is rare, we need consider only two genotypes (AA and AB), and two types of mating ($AA \times AA$ and $AA \times AB$). The other genotypes and mating types are negligibly rare. Prior to the introduction of B, all individuals were of genotype AA, and the population increased by the Leslie matrix equation,

$$n^{AA}(t+1) = \frac{1}{2}L^{AA}n^{AA}(t)$$

where $n^{AA}(t)$ is a vector whose xth entry is the number of individuals of age x and genotype AA at time t, and where L^{AA} is the Leslie matrix describing the pattern of reproduction and survival of individuals in $AA \times AA$ matings. The factor $1/2$ appears because each $AA \times AA$ mating requires two AA individuals, so there are only $1/2$ as many matings as there are individuals. I will assume that this Leslie matrix has a stable age distribution (see footnote 1 on page 97). Once the AA component of the population has reached its stable age distribution, the pattern of increase is even simpler:

$$n^{AA}(t+1) = \frac{1}{2}\lambda^{AA}n^{AA}(t)$$

where λ^{AA} is a number rather than a matrix. This number is referred to as the "dominant eigenvalue" of the matrix L^{AA}. This equation says that each age category is multiplied by $\lambda^{AA}/2$ each time period. $\lambda^{AA}/2$ is often called the "ratio of increase," since it is the ratio of the number of individuals of age x to the number in the preceding time period. Once the stable age distribution has been reached, each age category has the same ratio of increase. Furthermore this same ratio of increase also holds for the population of AA individuals taken as a whole.

Since the B allele is rare, the AB component of the population reproduces via matings of type $AA \times AB$, whose offspring are evenly divided between genotypes AA and AB. Thus, individuals

of genotype AB follow the Leslie matrix equation

$$n^{AB}(t+1) = \frac{1}{2}L^{AB}n^{AB}(t)$$

where the vector $n^{AB}(t)$ and the matrix L^{AB} are analogous to $n^{AA}(t)$ and L^{AA}. The factor $1/2$ appears because only half the offspring of a $AB \times AA$ mating are of type AB.[1] I assume once again that this Leslie matrix has a stable age distribution. Once this age distribution is reached, the number of AB individuals in each age category will increase according to

$$n^{AB}(t+1) = \frac{1}{2}\lambda^{AB}n^{AB}(t)$$

where λ^{AB} is the dominant eigenvalue of matrix L^{AB}.

At this point, the *number* of each genotype is increasing (or decreasing) exponentially. Genotype AA is increasing with ratio $\lambda^{AA}/2$ and genotype AB is increasing with ratio $\lambda^{AB}/2$. If

$$\lambda^{AB} > \lambda^{AA}$$

then the number of AB individuals will grow faster than the number of AA individuals. Consequently, the *relative frequency* of AB will increase and that of AA will decrease. This also implies that the relative frequency of allele B will increase. However, the entire argument here is based on the premise that B is rare, so the inequality above does not guarantee that B will increase to fixation. It only implies that B will spread when rare.

If allele A represents an ESS, then no alternative allele can increase, even when rare. Thus, A can be an ESS only if

$$\lambda^{AB} \leq \lambda^{AA}$$

for every alternative allele B. This condition is bound to hold if AA has the largest possible ratio of increase. For this reason, we search for strategies that maximize λ.

It also works to search for strategies that maximize the rate, r, of exponential growth (see page 97). Since r is equal to $\ln \lambda$, it is an increasing function of λ. Consequently, any strategy that maximizes λ will also maximize r.

15.2 Selection on age-specific modifiers of fertility and mortality

The optimal life history is the one whose vital rates (l_x and m_x) maximize r. To find an optimum, we need to know how r responds to changes in vital rates. Let us attack this problem using the renewal equation, which was introduced in section 14.1. As in that section,

$$l_x = \text{Pr[surviving to age } x].$$
$$m_x = \text{fertility of individuals of age } x.$$

Suppose that a mutation has just one effect: it increases fertility at age x. This mutation will be favored by selection, since fertility is beneficial. But how strong is this effect? To find out, we must calculate $\partial r/\partial m_x$. But there is no obvious way to calculate this derivative since we have no explicit formula for r. Instead, r is defined implicitly by the characteristic equation

$$1 = \sum_x e^{-rx}l_x m_x,$$

---

[1]The remaining $1/2$ contribute to the reproduction of type AA, but the $AB \times AA$ mating type is so much rarer than $AA \times AA$ that these AA offspring can be ignored in calculating the rate of reproduction of AA.
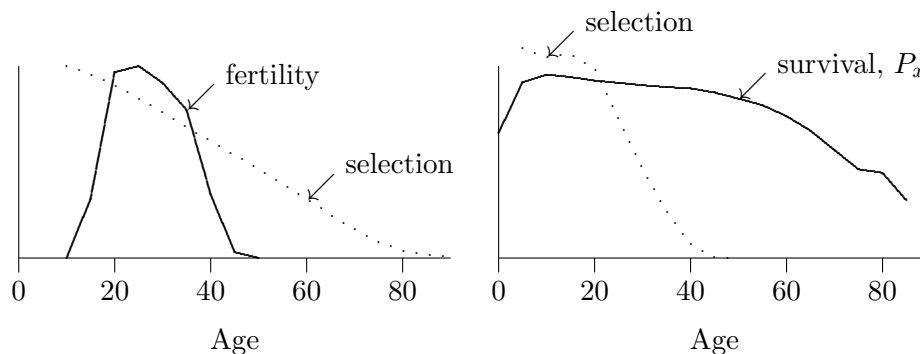
Figure 15.1: The Problem. *Left*: Age-specific fertility declines far faster than the decline in the force of selection on fertility. *Right*: Survival from age x to $x+5$ declines far slower than the force of selection on survival. Data are from 1906 Taiwan [31]. The strength of selection is calculated assuming a stationary population and using Eqns. 15.2 and 15.3.

which was introduced on p. 97. To calculate the derivative of r, we can use the rule for implicit functions (Appendix E). If g is the right-hand side of the equation above, then

$$\frac{\partial r}{\partial m_z} = -\frac{\partial g/\partial m_z}{\partial g/\partial r} \tag{15.1}$$

To evaluate this expression, we need two partial derivatives:

$$\partial g/\partial m_z \;=\; e^{-rz}l_z$$

$$\partial g/\partial r \;=\; -\sum_{x=0}^{\infty} xe^{-rx}l_x m_x \equiv -T$$

where T can be interpreted as the average age of mothers at reproduction. Substituting into Eqn. 15.1 gives

$$\partial r/\partial m_z = e^{-rz}l_z/T \tag{15.2}$$

In stable populations, $e^{-rz}l_z$ is proportional to the fraction of the population at age z. Thus, the force of selection on the fertility of x-year-olds is proportional to the number of such individuals in the population. If twice as many have age 20 as age 40, then selection for fertility is twice as strong at 20 as at 40.

The same method can also be used for the force of selection on age-specific mortality. It turns out that

$$\frac{\partial r}{\partial P_z} = \frac{\sum_{x=z+1}^{\infty} e^{-rx}l_x m_x}{P_z T} \tag{15.3}$$

where P_z is the probability of survival from age z to age $z+1$. Both results were originally derived by Hamilton [31], using exactly the argument that I have presented here. Charlesworth [8] extends the argument in various ways.

These results were used to draw the graphs in figure 15.1. The left panel shows that the fertility of human females declines faster with age than it ought to, according to Eqn. 15.2. The right panel shows that survival declines more slowly than it ought to. Why should this be? Is the discrepancy mere noise or has something been left out of the model?

15.3 Evolution of aging

These results are at the heart of the evolutionary theory of aging [31, 8]. They contribute to two less-mathematical theories, which I summarize below. Both theories are based on the fact that the force of selection is strong early in life but declines with increasing age. This is implied both by Eqn. 15.2 and by Eqn. 15.3, and is apparent in both panels of Fig. 15.1. Thus, it is true of mutations that affect fertility and also of those that affect mortality.

Both theories assume the existence of many genetic loci that act at specific ages. No one really thinks that any locus turns on precisely at (say) age 34, only to turn off again at 35. But there are age differences in the rates at which loci are expressed. (One obvious example is fetal hemoglobin, which is expressed prenatally but disappears within a few weeks after birth.) Once we understand how selection affects fictional loci that act at single ages, it will be easy to infer how it affects loci with more realistic trajectories of expression.

15.3.1 The theory of mutation-selection balance

Suppose that a deleterious mutation affects individuals of particular age. If it acts late in life, the opposing force of selection will be weak. Consequently, the mutant may increase in frequency until its damaging effects are common.

15.3.2 Antagonistic pleiotropy

Suppose now that a single mutation has effects at two different ages, one early in life and the other late. For example, a mutation that reduces the amount of DNA repair that occurs at age 20 will increase survival and fertility at age 20 because it frees the resources that are needed to repair DNA. But this mutant also has an adverse effect at later ages because, when the DNA is not repaired cancer is likely to occur later on. The same mutation thus has a beneficial effect at age 20 and an adverse effect later on.

What will selection do with such a mutant? Because the force of selection declines with age, selection is more sensitive to the early benefit than to the cost that arives late in life. Consequently, mutations that do damage late in life may be favored by selection.

The distinction between these models is important for several reasons. If aging occurs primarily by mutation-selection balance, then the forces affecting aging are weak: selection on these genes is no stronger than mutation, and mutation is a very weak force. Because these directional forces are weak, genetic drift is likely to be important. We should expect great and unpredictable variation in rates of senescence among characters and among populations.

On the other hand, if aging is due mainly to antagonistic pleiotropy then the force that produces it is strong. It results from changes in allele frequencies that are driven by selection on effects early in life, when selection is a strong force. Consequently, mutation and genetic drift are likely to be less important. Characters may age at different rates because of different trade-offs across ages, but where these trade-offs are similar, characters should senesce at similar rates. Differences between populations should reflect different selective environments and should therefore be predictable.

15.4 Evolution of menopause

In human females, fertility declines to low levels during the decade between ages 40 and 50; it stops altogether a few years later. Does this make evolutionary sense? If the human fertility schedule represents a fitness optimum, then any mutation that changes it must reduce fitness. To

find out whether this is so, we investigate the effect of a rare allele that causes a small increase in the fertility of its bearer between ages 45 and 50. If this increased fertility were its only effect, it would necessarily increase fitness. If menopause is an adaptation, there must also be some indirect disadvantage. Two possibilities are investigated in the sections below. The first discusses a simplified version of a model that I introduced a few years ago [61]. The second discusses the "Grandmother hypothesis" of Kristen Hawkes and her colleagues [37, 35].

15.4.1   The childbirth mortality model

Long ago, George Williams [76] suggested that the negative effect of continued fertility reflects the risk of death in childbirth. This risk rises with age both because an older woman is more likely to have children who need her and also because the probability of child-birth mortality increases with age.

   Suppose that some mutation increases the probability that a mother (Ego) will give birth in the interval between ages 45 and 50. If she dies giving birth to this additional child, then she cannot care for the children produced before age 45. For simplicity, I assume that Ego's death would have no effect on the children produced before she was 40, but that all the children produced between ages 40 and 45 would die. In effect, this *reduces* Ego's birth rate in the interval between ages 40 and 45.

   To measure these effects, I introduce two parameters: q is the risk of death during child-birth, and δ is the expected number of extra daughters that Ego produces between ages 40 and 50. (She also produces an extra δ sons, but this will not be important until later.) The mutation changes Ego's birth rate (for female births) between ages 45 and 50 to

$$m_{45} = m_{45}^* + \delta(1 - q).$$

where m_{45}^* is the fertility of normal mothers. This model is not intended to imply that normal women suffer no childbirth mortality. Those deaths do not appear explicitly because they are already worked into the m_x^* values. We only need to account for the excess mortality that is suffered by the mutants.

   I assume that children over five years of age are unaffected by their mother's death but that the effect on children under five is profound—I assume that all such children die when their mother dies. This assumption is certainly too extreme and exaggerates the cost of increased fertility in 45-year-old women. The implications of this extreme assumption are discussed below. For the moment, bear with me.

   When the mother dies between ages 45 and 50, she loses all the children that were born between ages 40 and 45. We can model this by pretending that these children were never born, i.e. by modifying m_{40}. With probability $2\delta q$ the mother has an additional birth and dies. (Here, the factor of 2 reflects the fact that Ego can die when giving birth either to a son or a daughter.) If Ego does die, then $m_{40} = 0$. With probability $1 - 2\delta q$, either no additional birth occurs or else the mother gives birth successfully; in this event $m_{40} = m_{40}^*$, where m_{40}^* is the fertility of non-mutant mothers. Thus,

$$m_{40} = 0 \times 2\delta q + m_{40}^*(1 - 2\delta q) = m_{40}^*(1 - 2\delta q).$$

   A death between ages 45 and 50 also deprives the mother of any children she might have produced after the age of 50. But this effect does not need to be considered here because, in the Taiwan data, fertility is zero after the age of 50.

   To determine whether this mutant is favored or opposed by natural selection, we use Eqn. 15.2 to calculate the derivative of r with respect to δ. If this model describes the mechanism by which

natural selection maintains menopause, then $dr/d\delta$ must be negative. (Otherwise, a mutation increasing the fertility of older women would be favored.) The derivative of r with respect to δ is the sum of effects in the two age intervals:

$$\frac{dr}{d\delta} = \frac{\partial r}{\partial m_{40}}\frac{\partial m_{40}}{\partial \delta} + \frac{\partial r}{\partial m_{45}}\frac{\partial m_{45}}{\partial \delta} \tag{15.4}$$

Here, the derivatives of r with respect to m_x are given by Eqn. 15.2. The other two derivatives are

$$\partial m_{40}/\partial \delta = -2qm_{40}$$
$$\partial m_{45}/\partial \delta = 1-q$$

Substituting all this into Eqn. 15.4 gives

$$dr/d\delta = -2qm_{40}e^{-r40}l_{40}/T + (1-q)e^{-r45}l_{45}/T \tag{15.5}$$

Menopause is maintained by natural selection only if this quantity is negative. This requires that

$$\frac{q}{1-q} > \frac{e^{-5r}l_{45}}{2m_{40}l_{40}}. \tag{15.6}$$

Here, r is the growth rate per year, which was approximately 0.02 in the Taiwan data (see exercise 14–5, page 105). From table 14.2, $m_{40} = 0.2771$, $l_{45} = 0.3193$ and $l_{40} = 0.3615$. Thus, the condition becomes

$$\frac{q}{1-q} > \frac{\exp[-5 \times 0.02] \times 0.3193}{2 \times 0.2771 \times 0.3615} = 1.44$$

or

$$q > 1.44/(1+1.44) = 0.59 \tag{15.7}$$

Menopause is maintained under this model only if mothers die in child-birth more than 59% of the time. This is wildly unrealistic—in human data, $q \approx 1/200$. This is especially impressive in view of the fact, discussed above, that our model *exaggerates* the cost of increased fertility by 45-year-old women. Had we assumed (more realistically) that only some fraction of a women's young children die when she dies, the cost of increasing fertility would have been less, and the right side of inequality 15.7 would have been even larger. Before abandoning the model altogether, however, we should try making it a little more realistic. The section that follows does just this, by incorporating density-dependent population growth.

Density-dependent population growth

The model above requires estimates of several demographic parameters. We used estimates taken from a particular human population, but it would have been better to use estimates that reflect some kind of average of the last several hundred thousand years of the human species. We don't know much about this ancient population, but we can be sure of one thing—it did not grow at 2% per year throughout human history. A population growing at that rate doubles every 35 years, and would increase by 868 orders of magnitude in 100,000 years. Over the long span of human history, the average rate of growth must have been very nearly zero. Thus, it makes more sense to assume that $r = 0$ than that $r = 0.02$. We can do this by making either P_x or m_x depend on the size of the population.

Here, I take the latter approach: m_x will decrease population size N, but l_x does not respond. If $\tilde{m}_x$ is the equilibrium value of m_x, then $m_x = \tilde{m}_x$ when N equals its equilibrium value $\tilde{N}$. More generally, $m_x = \tilde{m}_x A(N)$, where $A(N)$ is a decreasing function of N, and $A(\tilde{N}) = 1$.

⋆ EXERCISE **15–1** Describe the meaning of this assumption in words. Is it reasonable? What alternative assumptions might be worth considering? Suppose you wanted to figure out how sensitive the model is to this choice. What would you do?

With these assumptions, the characteristic equation of the renewal process (Eqn. 14.1) becomes

$$1 = A(N) \sum_x e^{-rx} l_x \tilde{m}_x$$

At equilibrium, $A(\hat{N}) = 1$, and this becomes

$$1 = \sum_x l_x \tilde{m}_x \tag{15.8}$$

In demography, the quantity $R_0 = \sum l_x m_x$ is called the "net reproductive rate." When N is at equilibrium, $R_0 = 1$. In our data, however, $R_0 > 1$ because the population is growing. The observed value of R_0 estimates of the current value of $A(N)$:

$$\begin{aligned} R_0 &= \sum_x l_x m_x \\ &= A(N) \sum_x l_x \tilde{m}_x \\ &= A(N). \end{aligned}$$

Once we know R_0, we can convert m_x into $\tilde{m}_x$:

$$m_x/R_0 = A(N)\tilde{m}_x/A(N) = \tilde{m}_x \tag{15.9}$$

Thus, we have a recipe for calculating the equilibrium fertility, $\tilde{m}_x$:

First calculate the right-hand side of the characteristic equation with r set to zero. In this code, vectors `l` and `m` have length `k` and contain the Taiwan data.

```
> R0 := sum(l[x+1]*m[x+1],
                x=0..(k-1));
    R0 := 1.695105011
```

Age-category 40 corresponds to the 9th entry of our m_x vector. We are interested in m_{40}/R_0

```
> m[9];
    .27713

> m[9]/R0;
    .1634883964
```

Thus, our model of density dependence implies that in a stationary population, the fertility between ages 40 and 45 would be 0.1634. Inequality 15.6 becomes

$$\frac{q}{1-q} > \frac{0.3193}{2 \times 0.1634 \times 0.3615} = 2.701$$

or

$$q > 0.72 \tag{15.10}$$

Density dependence seems not to have made a large difference, for this condition is not very different from the one we started with. Either way, this model requires an exorbitant level of childbirth mortality.

15.4.2 The Grandmother Hypothesis

The Grandmother Hypothesis [37, 35] holds that women can increase the fertility of their daughters by feeding their grandchildren. The daughters can get pregnant again sooner because they do not have to work so hard to feed their weaned children.

In modeling this effect, we need to account for the fact that the grandmother is less closely related to her grandchildren than to her own children. The straightforward approach would involve incorporating the ideas of kin-selection into the life-history theory introduced above [9, 61]. But there is an easier way.

The sections below introduce two variants of this easier approach. The first variant considers a mutant allele that increases the fertility of a woman at about the age of menopause. If menopause is evolutionarily stable, then such alleles must be selected against. The second variant (which is more consistent with the idea of Hawkes and her colleagues [36]) considers a mutant allele that increases mortality among post-menopausal women. If menopause is evolutionarily stable, then these alleles must also be selected against.

Mutant alleles that modify fertility

Suppose that Ego carries a rare allele that inflates her fertility between ages 45 and 50. Since the allele is rare, Ego is probably a heterozygote and is equally likely to have inherited the allele from her father or mother. In the latter case, when Ego was young she would have received less help from her own mother, who would have been busy caring for her own extra children. To incorporate these effects, we need expressions for Ego's fertility between ages 20 and 25 and between 45 and 50. Consider

$$
\begin{aligned}
m_{45} &= m_{45}^* + \delta \\
m_{20} &= m_{20}^* - \frac{1}{2} \cdot \frac{l_{45}}{l_{20}} \cdot 2\delta \cdot \gamma \\
&= m_{20}^* - \frac{l_{45}}{l_{20}} \delta\gamma
\end{aligned}
$$

Here, m_{20} and m_{45} are Ego's fertility between ages 20 and 25 and between 45 and 50. On the right-hand sides of the two equations, m_{20}^* and m_{45}^* represent the fertility that Ego would enjoy if she did not carry the rare allele. Her fertility (counting daughters only) between 45 and 50 has been augmented by the quantity δ. Her fertility between 20 and 25 has been decreased by the product of $1/2$ (the probability that she inherited the rare allele from her mother), l_{45}/l_{20} (the probability that her mother survived from Ego's birth until age 45), 2δ (the increase in Ego's mother's both-sex fertility at age 45), and γ (the resulting decrease in Ego's fertility).

The next step is to calculate the derivative of r with respect to δ.

$$
\frac{dr}{d\delta} = \frac{\partial r}{\partial m_{20}} \frac{\partial m_{20}}{\partial \delta} + \frac{\partial r}{\partial m_{45}} \frac{\partial m_{45}}{\partial \delta} \tag{15.11}
$$

Here, $\partial r/\partial m_{20}$ and $\partial r/\partial m_{45}$ are as given by Eqn. 15.2 and

$$
\begin{aligned}
\partial m_{20}/\partial \delta &= -\gamma l_{45}/l_{20} \\
\partial m_{45}/\partial \delta &= 1
\end{aligned}
$$

Thus,

$$\begin{aligned}
\frac{dr}{d\delta} &= -\frac{\gamma l_{45}}{l_{20}} \cdot \frac{e^{-20r} l_{20}}{T} + \frac{e^{-45r} l_{45}}{T} \\
&= \left(\frac{e^{-20r} l_{45}}{T} \right) (-\gamma + e^{-25r})
\end{aligned} \tag{15.12}$$

The allele increasing fertility late in life is selected against if this quantity is negative. This requires that

$$\gamma > e^{-25r} \tag{15.13}$$

As before, it makes sense to set $r = 0$. In the model of density-dependence developed above, growth rate is maintained at zero by changes in fertility. Thus, changing r to zero required a compensatory change in fertility. But fertility does not enter into the inequality above. If we stick to the same model of density dependence, we can simply set $r = 0$ without further adjustments. The inequality becomes

$$\gamma > 1 \tag{15.14}$$

This means that menopause is maintained by natural selection if a new birth when Ego is 45 deprives her 20-year-old daughter of at least 1 birth. We would need to estimate γ in order to find out whether this condition holds in humans.

This model glosses over an important point—it assumes that Ego was 20–25 years old when her mother was 45–50. A more realistic model would treat the age difference between mother and daughter as a random variable.

Mutant alleles that modify mortality

The previous section considered what is required for the human female reproductive pattern to be stable against mutations that modify fertility. But this is only half the problem. Human females are unusual in that their fertility ends early *relative to their lifespan*. This pattern could be destroyed by mutations of two types: (1) by mutations that increase the fertility of older females, or (2) by mutations that increased the mortality of these females. Mutations of the first type were considered in the section above; mutations of the second type will be considered here.

Consider a rare allele that decreases the probability that its (heterozygous) female bearer will live from age 45 to age 46. Specifically, suppose that Ego (a carrier of this allele) survives this interval with probability

$$P_{45} = P_{45}^* - \delta$$

where P_{45}^* is the probability that a normal female will survive this interval. Females who carry this allele are more likely to die at age 45, and when they do they are unable to help their daughters raise children. Consequently, the daughters of a carrier have, on average, lower fertility in their youth. Let us say that the daughter's fertility at age 20 is diminished by $\delta\beta$. If Ego is a carrier, she is almost certainly a heterozygote (because the allele is rare) and received the allele from her mother with probability 1/2. Thus Ego's fertility at age 20 was

$$m_{20} = m_{20}^* - \delta\beta/2$$

So far, the mutant gene in this model has only negative effects, so it would certainly be selected against. If this model is correct, then it is easy to understand the maintenance of human female longevity. Once the fertility problem is solved, longevity takes care of itself.

But there may be trade-offs that are not captured by the model. In order to achieve low mortality at age 45, the mother may have to sacrifice something earlier in life. Let us suppose, therefore, that the mutant gene *increases* Ego's own fertility at age 20. The increased fertility early in life is paid for by decreased survivorship at age 45 via some physiological trade-off. If this were true, then Ego's fertility at age 20 would become

$$m_{20} = m_{20}^* - \delta\beta/2 + \delta\phi$$

where ϕ measures Ego's increased fertility at age 20 relative to her decreased survival at age 45. Now the mutant allele has positive effects as well as negative ones. To find out whether the allele will increase in frequency under natural selection, calculate

$$\frac{dr}{d\delta} = \frac{\partial r}{\partial m_{20}}\frac{\partial m_{20}}{\partial \delta} + \frac{\partial r}{\partial P_{45}}\frac{\partial P_{45}}{\partial \delta}$$

where

$$\begin{aligned} \partial m_{20}/\partial\delta &= \phi - \beta/2 \\ \partial P_{45}/\partial\delta &= -1 \end{aligned}$$

These results, together with Eqns. 15.2 and 15.3, give

$$\frac{dr}{d\delta} = \frac{e^{-20r}l_{20}}{T}(\phi - \beta/2) - \frac{\sum_{x=46}^{\infty} e^{-rx}l_x m_x}{P_{45}T}$$

The final term on the right-hand side contains a sum that adds up Ego's expected future fertility after age 45. But this sum will be approximately zero, since human female fertility is very low after this age. (If it is not zero at age 45, then we can repeat the analysis at age 50, or 55.) To maintain the human female pattern of survivorship, the mutant allele must be selected against, which requires that $dr/d\delta < 0$. This in turn requires that

$$\beta > 2\phi \tag{15.15}$$

The current mortality schedule resists invasion only if the benefit that older women provide to a daughter is twice as large as the benefit that they could provide for themselves (via the physiological trade-off) by emphasizing early fertility at the expense of late survival. This is simply Hamilton's rule. It is more easily satisfied if the physiological trade-off is weak than if it is strong. In order to decide whether this model is realistic, we need to estimate the tradeoff between late survival and early fertility.

15.5 Exercises

With chimpanzee data, the question is not why menopause occurs, but why it does not [51]. To attempt an answer to this question, consider the chimpanzee data shown in table 15.1. Assume that the direct effect of some gene is to *reduce* fertility at age 45. This model can be studied using the same machinery as above. The only difference is that δ is now a negative rather than a positive quantity. (Hint: changing the sign of δ changes the direction of inequalities 15.7 and 15.14.)

$\star$ EXERCISE **15–2** Use the chimpanzee data to repeat the calculation in section 15.4.1, using the model of childbirth mortality *without* density-dependent population regulation.

Table 15.1: Chimpanzee life-history table

a (years)	l_a	m_a
0	1.0000	0.0000
5	0.8280	0.0000
10	0.7922	0.1568
15	0.7654	0.2368
20	0.7331	0.2273
25	0.6760	0.3793
30	0.5128	0.2903
35	0.3805	0.2500
40	0.2114	0.5000
45	0.1510	0.2500
50	0.0000	0.0000

Source: [41]

⋆ EXERCISE **15–3** Use the chimpanzee data to repeat the calculation in section 15.4.1, using the model of childbirth mortality *with* density-dependent population regulation.

⋆ EXERCISE **15–4** Use the chimpanzee data to repeat the calculation in section 15.4.2, using the Grandmother hypothesis with a mutation affecting fertility.

⋆ EXERCISE **15–5** Use the chimpanzee data to repeat the calculation in section 15.4.2, using the Grandmother hypothesis with a mutation affecting mortality.

Appendix A

The Command Line Interpreter (Windows version)

A.1 The current directory

The command line interpreter (CLI) always looks for files in a single place, called the **current directory**. When it writes a new file, it puts it into the current directory. This makes it easy for you when you write programs: your programs will know where to find data and where to put output.

A.2 Commands

In the commands listed below, expressions like `<old>` represent file names. Do not type the `<` and `>`, just the name of the file. For example, if I had a file called foo and wanted to copy it, I would type `copy foo bar`. This would create a file named `bar` with the same contents as `foo`.

`cd mydir`	make "mydir" the current directory	`dir`	list current directory
`cd`	display current directory	`mkdir <mydir>`	make new directory
`copy <a> <b>`	make copy of file "a"	`ren <a> <b>`	rename file
`del <mydir>`	delete all files in directory	`rmdir <mydir>`	remove empty directory
`del <file>`	delete file	`type <file>`	display file on screen
		`move <a> <b>`	move file **a** to directory **b**

Examples

1. `copy old new` : copy a file named "old" into a new file named "new."

2. `del homework` : delete a directory called "homework", which must be empty.

3. `ren homework.txt hw.txt` : change the name of file "homework.txt." The new name is "hw.txt".

4. `type hw.txt` : display file "hw.txt" on the screen. It should be a text file.

A.3   History

The interpreter keeps track of the commands you execute, and you can scroll through this history using the up and down arrows. Once you have found a command that you want to execute a second time, you can edit it or just hit `<Enter>` to execute it again.

A.4   Redirecting input and output

The `>` symbol redirects output. For example,

```
dir > foo
```

would list the contents of the current directory and put the results into file foo. No output would appear on the screen.

Some programs are interactive: they ask for input from the user. You can tell such programs to read from a file instead using the `<` symbol. For example

```
myprogram < myfile
```

would run a program called "myprogram." When that program needs user input, it will read it from file "myfile."

A.5   Standard output and error output

The interpreter sends two separate output streams to your screen. One (the "standard output") consists of the program's normal output. The other (the "error output") contains any error messages. When you redirect the output into a file, it is only the standard output that goes into the file. The error output still goes to the screen.

For example, suppose that you run the program "Maple" like this:

```
cmaple > myfile
```

The standard output goes into "myfile," but any error messages will still go to the screen. This is usually what you want. But what if you want to send a copy of the output to me, including the error messages?

You need to combine the standard and error output into a single stream. Here's how

```
cmaple > myfile 2>&1
```

Now myfile will contain all the output, errors and all.

A.6   Pipes

It is also possible to link programs together in chains, using the `|` symbol. For example,

```
dir | myprogram
```

would run the dir command. The output of dir would become the input of myprogram. This is called a "pipe."

A.7 Path names

Every directory and file on the computer has a unique name—one that distinguishes it from all others. To find the full name of the current directory, type `cd` at the command prompt, and hit return. When I did this a moment ago, my computer responded by printing `Z:\home\rogers` on the screen. This is the "path name" of my current directory. Let us unpack it.

The path name `Z:\home\rogers` says that `rogers` is a subdirectory of `home`, which is itself a subdirectory of `\`, which is the root of the file system on drive `Z`. On any given drive, all the files and directories are organized in a tree, and `Z:\` is the root of the tree on drive `Z`. You can examine the contents of any of these directories by typing (for example) "`dir Z:\home`." Each drive corresponds to a device such as a diskette drive, a CDROM, or a partition of your computer's hard disk.

There is seldom a need to type out the full path name of a file. The command `mkdir txt` will make a new directory within my current directory. The full path name of this new directory will be `Z:\home\rogers\txt`, but I don't have to care. I can simply refer to it as "txt." The command interpreter assumes that all files and directories are found within the current directory unless I specify otherwise.

Suppose I type `cd txt` to move into this new directory, but then I want to go back. I could type `cd Z:\home\rogers`, but that is a lot of typing. The easy way is "`cd ..`". Here, "`..`" refers to the parental directory—the one that contains the current directory.

A.8 Text editors

Once you have got the command line going, you need to get familiar with a text editor. The best text editor to use depends on your operating system. If you want something simple, try "Notepad" or "Wordpad" under Windows, "Joe" under Linux, or "SimpleText" under Mac. More powerful alternatives are "emacs," "vi," and "jed." Whatever editor you use, make sure the output is plain text, with line breaks. There is a full discussion at `http://en.wikipedia.org/wiki/Text_editor`. Avoid word processors such as Microsoft Word and Open Office—they make things harder.

⋆EXERCISE **A–1** Open a CLI window. Make a new directory, and make it your current directory. Use a text editor to create a file within this directory called `hello.txt`, which contains the following line:

`"Hello world!"`

Back in the CLI, enter the command `dir`. The computer will type a listing of your current directory, which should include the file you just created and nothing else. Now type `type hello.txt`. The words "Hello world!" should appear on your screen.

Appendix B

The Command Line Interpreter (Mac and Linux version)

B.1   The current working directory

The command line interpreter (CLI) always looks for files in a single place, called the **working directory**. When it writes a new file, it puts it into the working directory. This makes it easy for you when you write programs: your programs will know where to find data and where to put output.

B.2   Commands

In the commands listed below, expressions like `<old>` represent file names. Do not type the `<` and `>`, just the name of the file. For example, if I had a file called foo and wanted to copy it, I would type `cp foo bar`. This would create a file named `bar` with the same contents as `foo`.

`cd mydir`	make "mydir" the current directory	`ls`	list current directory
		`mkdir <mydir>`	make new directory
`pwd`	print working directory	`mv <a> <b>`	rename file
`cp <a> <b>`	make copy of file "a"	`rmdir <mydir>`	remove empty directory
`rm *`	delete all files in current directory	`cat <file>`	display file on screen
		`mv <a> <b>`	move file **a** to directory **b**
`rm <file>`	delete file		

Examples

1. `cp old new` : copy a file named "old" into a new file named "new."

2. `rmdir homework` : delete a directory called "homework", which must be empty.

3. `mv homework.txt hw.txt` : change the name of file "homework.txt." The new name is "hw.txt".

4. `cat hw.txt` : display file "hw.txt" on the screen. It should be a text file.

B.3   History

The interpreter keeps track of the commands you execute, and you can scroll through this history using the up and down arrows. Once you have found a command that you want to execute a second time, you can edit it or just hit `<Enter>` to execute it again.

B.4   Redirecting input and output

The `>` symbol redirects output. For example,

```
ls > foo
```

would list the contents of the current directory and put the results into file foo. No output would appear on the screen.

Some programs are interactive: they ask for input from the user. You can tell such programs to read from a file instead using the `<` symbol. For example

```
myprog.py < myfile
```

would run a program called "myprog.py." When that program needs user input, it will read it from file "myfile."

B.5   Standard output and error output

The interpreter sends two separate output streams to your screen. One (the "standard output") consists of the program's normal output. The other (the "error output") contains any error messages. When you redirect the output into a file, it is only the standard output that goes into the file. The error output still goes to the screen.

For example, suppose that you run the program "Maple" like this: it like this:

```
maple > myfile
```

The standard output goes into "myfile," but any error messages will still go to the screen. This is usually what you want. But what if you want to send a copy of the output to me, including the error messages?

You need to combine the standard and error output into a single stream. Here's how

```
maple > myfile 2>&1
```

Now myfile will contain all the output, errors and all.

B.6   Pipes

It is also possible to link programs together in chains, using the `|` symbol. For example,

```
ls | myprogram
```

would run the ls command. The output of ls would become the input of myprogram. This is called a "pipe."

B.7   Path names

Every directory and file on the computer has a unique name—one that distinguishes it from all others. To find the full name of the current directory, type `pwd` at the command prompt, and hit return. When I did this a moment ago, my computer responded by printing `/home/rogers` on the screen. This is the "path name" of my current directory. Let us unpack it.

The path name `/home/rogers` says that `rogers` is a subdirectory of `home`, which is itself a subdirectory of `/`, which is the root of the file system. All path names begin with `/`, because the files and directories are organized in a tree, and `/` is the root. You can examine the contents of any of these directories by typing (for example) "`ls /home`."

There is seldom a need to type out the full path name of a file. The command `mkdir txt` will make a new directory within my current directory. The full path name of this new directory will be `/home/rogers/txt`, but I don't have to care. I can simply refer to it as "`txt`." The command interpreter assumes that all files and directories are found within the current directory unless I specify otherwise.

Suppose I type `cd txt` to move into this new directory, but then I want to go back. I could type `cd /home/rogers`, but that is a lot of typing. The easy way is "`cd ..`". Here, "`..`" refers to the parental directory—the one that contains the current directory.

B.8   Text editors

Once you have got the command line going, you need to get familiar with a text editor. The best text editor to use depends on your operating system. If you want something simple, try "Joe" under Linux, or "SimpleText" or "TextEdit" under Mac. More powerful alternatives are "emacs," "vi," and "jed." Whatever editor you use, make sure the output is plain text, with line breaks. There is a full discussion at `http://en.wikipedia.org/wiki/Text_editor`. Avoid word processors such as Microsoft Word and Open Office—they make things harder.

⋆ EXERCISE **B–1** Open a CLI window. Make a new directory, and make it your current directory. Use a text editor to create a file within this directory called `hello.txt`, which contains the following line:

`"Hello world!"`

Back in the CLI, enter the command `ls`. The computer will type a listing of your current directory, which should include the file you just created and nothing else. Now type `cat hello.txt`. The words "Hello world!" should appear on your screen.

Appendix C

Mean, Expectation, Variance and Covariance

C.1   Sums and sigma notation

If you already understand sigma notation, please skip this section.

As we learned in grade school, the sum of 10, 12, 10, and 8 is

$$10 + 12 + 10 + 8 = 40$$

To generalize this calculation, suppose we have 4 arbitrary numbers, x_1, x_2, x_3, and x_4. Their sum is

$$x_1 + x_2 + x_3 + x_4$$

This sum can also be written using "sigma notation," as

$$\sum_{i=1}^{4} x_i$$

The "$\sum$" symbol is a Greek sigma and indicates summation. The subscript "$i = 1$" indicates that the sum begins with x_1, and the superscript "4" indicates that the sum ends with x_4.

More generally, if the number of numbers is an unknown value, N, then their sum is

$$\sum_{i=1}^{N} x_i = x_1 + x_2 + \cdots + x_N$$

Sometimes sums are written without limits, as in

$$\sum_{i} x_i.$$

This means the sum over all terms, however many there may be.

When sums are written within the text of a paragraph, the limits look like subscripts and superscripts, as in $\sum_{i=1}^{N} x_i$.

C.2 Means

The mean of N numbers, $x_1, x_2, \ldots, x_N$, is

$$\bar{x} = \frac{1}{N} \sum_{i=1}^{N} x_i$$

It is often more convenient to express the mean in terms of the counts of elements with particular values. For example, suppose as that we are dealing with data consisting of three numbers 2, 1, and 1. With these data, $N = 3$ and $\bar{x} = (2+1+1)/3 = 4/3$.

To re-express this calculation, let n_x represent the number of times the value x appears in the data. There is one 2 in these data, so $n_2 = 1$. There are two 1s, so $n_1 = 2$. With this notation, the size of the data set is

$$N = \sum_x n_x = 3$$

and

$$\bar{x} = \frac{1}{N} \sum_x x n_x = 4/3 \tag{C.1}$$

$\star$ EXERCISE **C–1** Using our three-value data set and equation C.1, verify that $\bar{x} = 4/3$ by substituting values for N, x, and n_x.

C.3 Relative frequencies

Equation C.1 can be written more compactly as

$$\bar{x} = \sum_x x p_x \tag{C.2}$$

where $p_x = n_x/N$ is the *relative frequency* within the data of elements with value x.

$\star$ EXERCISE **C–2** Given the data consisting of three numbers 2, 1, and 1, what are p_0, p_1, and p_2?

$\star$ EXERCISE **C–3** Suppose that the $p_0 = 1/10$, $p_1 = 2/10$, $p_3 = 1/10$, $p_4 = 4/10$, and all other p_i are zero. What is the mean?

C.4 Expectations

The same formula also turns up in probability theory. Suppose that some variable, X, takes values that cannot be predicted exactly, but which occur with specified probabilities, $p_1, p_2, \ldots$. Then X is called a "random variable" and its expectation, $E[x]$, is defined as

$$E[X] = \sum_x x p_x$$

This is identical to equation C.2, except that p_x is a probability rather than a relative frequencies. (This is not a coincidence, for a probability *is* a relative frequency.) An expectation is an average in which the weights are probabilities instead of relative frequencies.

Finally, suppose that X is a continuous variable, such as stature, that takes values over the range from a to b. Then you cannot enumerate the possible values of X. Its expectation is written as

$$E[X] = \int_a^b x f(x) dx$$

where $f(x)dx$ can be thought of as the probability that X takes a value within the (very small) interval from x to $x + dx$, and is thus analagous to p_x in the discrete formulation. The "$\int$" symbol stands for "integration," and is usually covered in the second course on calculus. If you have not had that course, just think of integration as summation.

C.4.1 Properties of expectations

If a and b are constants, and X and Y are random variables, then

$$
\begin{align}
E[a] &= a \tag{C.3}\\
E[aX] &= aE[X] \tag{C.4}\\
E[X + Y] &= E[X] + E[Y] \tag{C.5}
\end{align}
$$

Rather than proving these properties, I will illustrate them using averages. They hold for averages too, since expectations are just a special kind of average.

The average of a constant The average of 4, 4, and 4 is 4. This is why $E[a] = a$ when a is a constant.

The average aX Start with the numbers 1, 3, and 5. The sum of these numbers is 9, and their average is 3. Now multiply each number by a constant a. The resulting numbers sum to

$$a + 3a + 5a = a \times (1 + 3 + 5) = 9a$$

The average is $9a/3 = 3a$, which is a times the original average. This is why $E[aX] = aE[X]$ when a is a constant.

The average of $X + Y$ Consider the following table

	X	Y	$X + Y$
	0	2	2
	2	3	5
	7	7	14
Sum	9	12	21

The average of X is 3, that of Y is 4, and that of $X + Y$ is 7. This illustrates the fact that $E[X + Y] = E[X] + E[Y]$.

⋆ EXERCISE **C–4** What is $E[3]$?

⋆ EXERCISE **C–5** If $E[X] = 5$, then what is $E[2X]$?

⋆ EXERCISE **C–6** If $E[X] = 5$ and $E[Y] = 6$, then what is $E[2X + 3Y]$?

⋆ EXERCISE **C–7** If the values of a, $E[X]$, and $E[Y^2]$ are unknown, then what is $E[aX + bY^2]$?

C.5   Variance

There are many ways to measure variation, of which the *variance* is the most common. The variance is the average squared difference from the mean. For example, suppose we are dealing with data consisting of the numbers 10, 12, 10, and 8. The mean is 10, so the variance is

$$V = ((10 - 10)^2 + (12 - 10)^2 + (10 - 10)^2 + (8 - 10)^2)/4 = 2$$

There is a variety of ways to represent the variance, including

$$
\begin{aligned}
V &= N^{-1}\sum_i (x_i - \bar{x})^2 \\
&= \sum_x (x - \bar{x})^2 p_x \\
&= \sum_x x^2 p_x - \bar{x}^2 \qquad\qquad (C.6)
\end{aligned}
$$

⋆ EXERCISE **C–8** Verify that the formulas Equation C.6 are equivalent.

⋆ EXERCISE **C–9** What are the mean and variance of the numbers 3, 9, 15, and 8?

If X is a random variable (rather than data), its variance is

$$V[X] = E\left[(X - E[X])^2\right] \qquad\qquad (C.7)$$

Note the similarity between this expression and Equation C.6. The variance can also be written in either of the following ways:

$$
\begin{aligned}
V[X] &= E[X^2] - E[X]^2 \qquad\qquad (C.8) \\
&= E[X(X - E[X])] \qquad\qquad (C.9)
\end{aligned}
$$

These expressions hold irrespective of whether the random variable is continuous or discrete.

⋆ EXERCISE **C–10** Prove that if a is a constant and X a random variable, then $V[aX] = a^2 V[X]$.

C.6   Covariances

The covariance of two sets of numbers, $x_1, \ldots, x_N$, and $y_1, \ldots, y_N$, is

$$\text{Cov}(x, y) = \frac{1}{N}\sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y})$$

where $\bar{x}$ and $\bar{y}$ are the means of the two sets of numbers. Alternate expressions for the covariance include

$$
\begin{aligned}
\text{Cov}(x, y) &= \sum_{xy} (x - \bar{x})(y - \bar{y}) p_{x,y} \qquad\qquad (C.10) \\
&= \sum_{xy} x(y - \bar{y}) p_{x,y} \qquad\qquad (C.11) \\
&= \sum_{xy} xy\, p_{x,y} - \bar{x}\bar{y} \qquad\qquad (C.12)
\end{aligned}
$$

Here $p_{x,y}$ is the relative frequency of pairs of numbers with values (x, y) and the sum is taken over all possible of pairs of numbers. The notation $\sum_{xy}$ means the same thing as $\sum_x \sum_y$.

If X and Y are random variables (rather than data), their covariance is

$$\text{Cov}(X, Y) = E[(X - E[X])(Y - E[Y])] \tag{C.13}$$

Note the similarity between this expression and Equation C.12. The covariance can also be written in either of the following ways:

$$\begin{aligned} \text{Cov}(X, Y) &= E[X(Y - E[Y])] && \text{(C.14)} \\ &= E[XY] - E[X]E[Y] && \text{(C.15)} \end{aligned}$$

$\star$ EXERCISE **C–11** Calculate the covariance of x and y using the following data:

x	y
2	3
1	2
1	1

Appendix D

Using straight lines to approximate curves

It is often helpful to use a straight line to approximate a curve. Figure D.1 illustrates this approach along with its major limitation. Suppose we are interested in the curved line $f(x)$ in Panel A. The straight line $\tilde{f}(x)$ is tangent to $f(x)$ at $x = 1$, and the two lines are very close in the vicinity of this point. Thus, the straight line approximates the curve. Panel B shows a broader view of the same functions. (Note the difference in scales.) In this broader view, the two functions are generally far apart, and no one would think of using one to approximate the other. Near the point of tangency, however, the straight line is a useful approximation.

To find a formula for the straight line, we begin with two arbitrary values, x_0 and x, on the horizontal axis. The corresponding function values, $f(x_0)$ and $f(x)$, lie on the vertical axis. The straight line connecting these points has slope

$$\frac{\Delta f}{\Delta x} = \frac{f(x) - f(x_0)}{x - x_0}$$

This can be rearranged into $f(x) = f(x_0) + (x - x_0)\Delta f/\Delta x$. If x is close to x_0, the slope $\Delta f/\Delta x$ is nearly equal to the derivative, $f'(x_0)$. Thus,

$$\tilde{f}(x) = f(x_0) + (x - x_0)f'(x_0) \tag{D.1}$$

approximates $f(x)$ in the neighborhood of $x = x_0$.

Example Suppose that $w(x)$ is the fitness of individuals with character value x, and that most individuals in the population have $x = x_0$. A rare mutant arises whose character value is increased
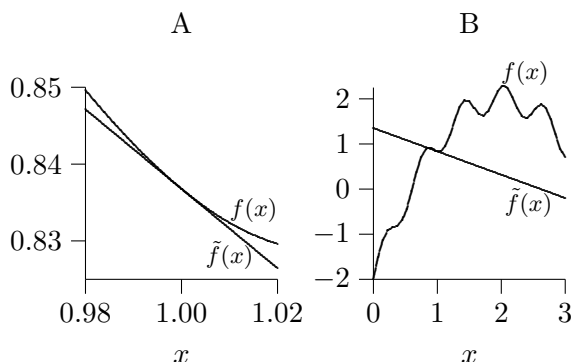
Figure D.1: The straight line $\tilde{f}(x)$ approximates the curved line $f(x)$ well near their point of tangency (see Panel A), but poorly over a broader range of values (see Panel B.)

by a small quantity dx. (In other words, the mutant's character value is $x_0 + dx$.) What is the resulting change in the mutant's fitness?

Because dx is small, we can use the linear approximation. The mutant's fitness is approximately

$$w(x_0 + dx) \approx w(x_0) + dx\, w'(x_0)$$

The change in fitness, $dw = w(x_0 + dx) - w(x_0)$, is approximately

$$dw \approx dx\, w'(x_0)$$

Appendix E

Implicit Functions

Fig. E.1 shows a contour plot of $w(x, y)$, a hypothetical function expressing fitness as a function of two variables, x, and y. Along each contour line, $w(x, y)$ is constant. Presumably, we have a function expressing fitness (w) as a function of x and y, but it may be impossible to write y explicitly as a function of x. All we know is that, along a given contour line, $w(x, y)$ is equal to some constant, k. Nonetheless, along this line y does vary in a predictable way in response to changes in x. In other words, the equation $w(x, y) = k$ defines y as an *implicit function* of x.

In evolutionary ecology, we are often interested in the derivatives of implicit functions. But how can one calculate dy/dx without an explicit formula for $y(x)$? Questions of this sort can all be answered using the implicit function theorem of calculus.

Suppose that both variables change in value, the one from x to $x + dx$ and the other from y to $y + dy$. If these changes are small, the resulting change in w will be approximately (see Appendix D)

$$dw = \frac{\partial w}{\partial x} dx + \frac{\partial w}{\partial y} dy$$

If these changes move us along a contour line, then the change in y is not arbitrary. It is a compensatory change of size $dy = \frac{dy}{dx} dx$. Thus, dw can be re-written as

$$dw = \frac{\partial w}{\partial x} dx + \frac{\partial w}{\partial y} \frac{dy}{dx} dx$$
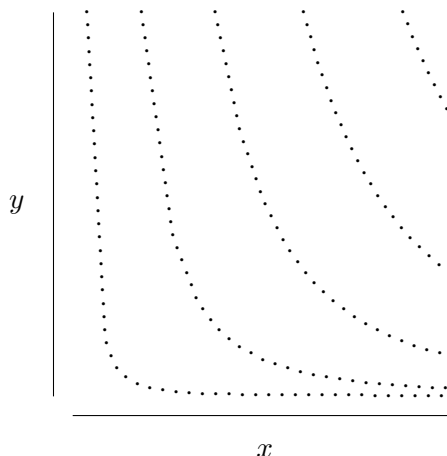
Figure E.1: Contour plot of a fitness function, $w(x, y)$. Dotted lines connect points of equal fitness.

Now along any contour line, w is a constant so $dw = 0$. This implies that

$$\frac{dy}{dx} = -\frac{\partial w/\partial x}{\partial w/\partial y} \tag{E.1}$$

Example   Suppose that

$$w(x, y) = x^{1/3}y^{2/3}$$

We can use Eqn. E.1 to find dy/dx along any contour line. We need two partial derivatives: $\partial w/\partial x = (1/3)x^{-2/3}y^{2/3}$, and $\partial w/\partial y = (2/3)x^{1/3}y^{-1/3}$. Substituting these into Eqn. E.1 gives

$$\begin{aligned}
\frac{dy}{dx} &= -\frac{(1/3)x^{-2/3}y^{2/3}}{(2/3)x^{1/3}y^{-1/3}} \\
&= -y/(2x)
\end{aligned}$$

Here is the same analysis in Maple:

```
> w := x^(1/3) * y^(2/3);
          (1/3)  (2/3)
   w := x       y


> -diff(w, x)/diff(w,y);
       y
    - ---
      2 x
```

Appendix F

Answers to Exercises

⋆ EXERCISE **2–1** 2701
⋆ EXERCISE **2–2** –36
⋆ EXERCISE **2–3** 37/39
⋆ EXERCISE **2–4** 0.94871794871795
⋆ EXERCISE **2–5** 2889
⋆ EXERCISE **2–6** π, 4π, 9π, 16π, and 25π.
⋆ EXERCISE **2–7**

```
> expand((x-2)*(x+4)*(x-5));
     3      2
    x  - 3 x  - 18 x + 40
```

⋆ EXERCISE **2–8**

```
> factor(24 + x^4 - 10*x^3 + 35*x^2 - 50*x);
    (x - 1) (x - 2) (x - 3) (x - 4)
```

⋆ EXERCISE **2–10** The Maple command is `plot(x^(1/3), x=0..1)`. I'll leave the description of
this function to you.
⋆ EXERCISE **2–11** The Maple command is `plot(1 - exp(-x), x=0..1)`.
⋆ EXERCISE **2–12**

```
> solve(1-exp(-x) = 1/2, x);
    ln(2)
```

⋆ EXERCISE **2–13** There are several ways to write the procedure, one of which is

```
g := proc(x)
  1 - exp(-x);
  end proc;
```

⋆ EXERCISE **3–1**

```
> simplify(diff(k*x^n, x));
       (n - 1)
    k x        n
```

132

```
> simplify(diff(log(x), x));
    1/x


> simplify(diff(exp(k*x), x));
    k exp(k x)


> simplify(diff(f(x) + g(x), x));
    /d        \   /d        \
    |-- f(x)| + |-- g(x)|
    \dx      /   \dx      /


> simplify(diff(f(x) * g(x), x));
    /d        \                /d        \
    |-- f(x)| g(x) + f(x) |-- g(x)|
    \dx      /                \dx      /


> simplify(diff(f(x) / g(x), x));
    /d        \                /d        \
    |-- f(x)| g(x) - f(x) |-- g(x)|
    \dx      /                \dx      /
    ------------------------------
                  2
               g(x)


> simplify(diff(f(g(x)), x));
    D(f)(g(x)) D(g)(x)
```

$\star$ EXERCISE **3–2** $df/dx = 0$

$\star$ EXERCISE **3–3** $df/dx = a$

$\star$ EXERCISE **3–4** $df/dx = 2ax$

$\star$ EXERCISE **3–5** $df/dx = 3ax^2$

$\star$ EXERCISE **3–6** $df/dx = 0 + b + 2cx + 3dx^2$

$\star$ EXERCISE **3–7** 0

$\star$ EXERCISE **3–8** 3

$\star$ EXERCISE **3–9** $3 - 6x$

$\star$ EXERCISE **3–10** $3 - 6x$ (Can you see why this derivative is the same as the last one?)

$\star$ EXERCISE **3–11** $9e^{3x}$

$\star$ EXERCISE **3–12** $1/x$

$\star$ EXERCISE **3–13**

```
> simplify(diff( (3 + 4*x - 5*x^2)/(x^2 + exp(x)), x));
         2                              2
    -4 x  + exp(x) - 14 x exp(x) - 6 x + 5 x  exp(x)
    ------------------------------------------------
                      2        2
                  (x  + exp(x))
```

$\star$ EXERCISE **4–1** The population contains NP_{11} individuals of genotype A_1A_1 and NP_{12} of genotype A_1A_2. Each of the former carries two copies of A_1 and each of the latter carries one copy. The

total number of copies of A_1 is thus $2NP_{11} + NP_{12}$. Dividing by $2N$ (the total number of genes) gives the relative frequency of A_1 within the population: $p = P_{11} + \frac{1}{2}P_{12}$.

$\star$ EXERCISE **4–2** When $w_{11} = w_{12} = 1$ and $w_{22} = 0.8$, so that A_1 is dominant, $\Delta p(0.1) = .01933174224$ whereas $\Delta p(0.9) = .001803607214$. Thus, the rate of change is faster when the dominant allele is rare, or (equivalently) when the recessive allele is common.

$\star$ EXERCISE **4–3** After using equation 4.1 to define p2, and equation 4.2 to define dp, I typed `simplify(p2 - p - dp);` and Maple responded by typing 0.

$\star$ EXERCISE **4–4** The calculation is very simple:

$$
\begin{aligned}
w_1 - w_2 &= pw_{11} + qw_{12} - qw_{22} - pw_{12} \\
&= p(w_{11} - w_{12}) + q(w_{12} - w_{22})
\end{aligned}
$$

$\star$ EXERCISE **4–5** First enter the formulas for w_1 and w_2, along with the three genotypic fitnesses. Then calculate

```
> subs(p=0.1, w1 - w2);
      .85
```

Since the answer is positive, allele 1 is favored.

$\star$ EXERCISE **4–6**

```
> subs(p=0.1, w1 - w2);
     -.35
```

Now the difference is negative, so allele 2 is favored.

$\star$ EXERCISE **4–7**

```
> solve(w1=w2, p);
     .6666666667
```

$\star$ EXERCISE **4–9** After defining `wbar` as above,

```
> f := ((w11 - w12)*p + (w12 - w22)*q);
    f := (w11 - w12) p + (w12 - w22) (1 - p)
> simplify(f - diff(wbar,p)/2);
    0
```

$\star$ EXERCISE **4–10** There are unstable equilibria at 0 and 1, and a stable equilibrium at $3/4$.

$\star$ EXERCISE **4–13** This yields only the equilibrium at $3/4$.

$\star$ EXERCISE **5–1**

```
> f := x - 2*x*x;
                2
   f := x - 2 x


# One-liner sets derivative equal to 0 and solves
> solve(diff(f, x)=0, x);
    1/4


# Substitute solution into 2nd derivative
> subs(x=1/4, diff(f, x, x));
    -4
```

The second derivative is negative, so $x = 1/4$ is a local maximum.

$\star$ EXERCISE **5–2**

```
> f := x*exp(-x);
    f := x exp(-x)


> solve(diff(f,x)=0, x);
    1


> subs(x=1, diff(f,x,x));
    -exp(-1)
```

The second derivative is negative, so $x = 1$ is a local maximum.

$\star$ EXERCISE **5–3** 0.1281250000, 0.1624310220, 0.2033491664, 0.6119042892, and 0.9999999989

$\star$ EXERCISE **5–4** 0.1657303371, 0.2489153132, 0.3371758950, 0.6574744675, and 0.7499999312

$\star$ EXERCISE **5–5** Here are two ways to sum $1/i^2$:

```
> p := 1;

                                   p := 1


> for i from 1 to 10 do p := p*1/(i*i) end do:
> p;
                            1/13168189440000


> # At this point, Maple thinks i=11.  You must unassign it
> # before using it in the product statement below.

> i := 'i';

                                   i := i


> product(1/(i*i), i=1..10);
                            1/13168189440000
```

$\star$ EXERCISE **5–6**

```
1    interface(echo=4);
2    # Using a for loop to calculate means, mean squares, and
3    # the mean cross product.
4
5    # Data
6    x := [3,2,12,12];
7    y := [6,8,4,8];
8
9    # Initialization
10   mx  := 0;
11   my  := 0;
12   mxx := 0;
13   myy := 0;
14   mxy := 0;
15
```

```
16    # n is the number of observations.
17    n := nops(x);
18
19    # Loop over data accumulating sums
20    for i from 1 to n do
21      mx := mx + x[i];
22      my := my + y[i];
23      mxx := mxx + x[i]*x[i];
24      myy := myy + y[i]*y[i];
25      mxy := mxy + x[i]*y[i];
26      end do:
27
28    # Turn sums into means
29    mx := mx/n;
30    my := my/n;
31    mxx := mxx/n;
32    myy := myy/n;
33    mxy := mxy/n;
34
35    print(mx,my,mxx,myy,mxy);
```

$\star$ EXERCISE **6–1** The stable equilibrium is at $x = 2$.

Incidentally, there is also an unstable equilibrium at $x = 0$, which you will find only if you attack the problem by solving $dw/dx = 0$ (using Eqn. 6.1). It is probably easier to solve $ds(x)/dx = s(x)/x$, or $d\log w/dx = 0$. I don't care which approach you use.

To show that the equilibrium at $x = 2$ is stable, you will need to calculate the second derivative of w (or $\log w$), and evaluate it at $x = 2$. This derivative is negative, so the equilibrium is stable. In Maple, you calculate d^2w/dx^2 as `diff(w,x,x)`.

$\star$ EXERCISE **6–2** The only positive equilibrium is stable and lies at $x = 2.1462$.

No matter how you attack this problem, Maple will probably give you an answer that involves something called "LambertW." Don't worry. Just use the evalf function to convert the answers into floating point numbers. You may get several answers, but only one should be positive. Ignore the negative answers.

$\star$ EXERCISE **6–3** Here is a Maple program that does the exercise:

```
interface(echo=4);

# Definitions
w := b*s(x);
s := proc(x) (x-1)^(1/2) end proc;
T := 20;

# xsf is solution for standard Smith-Fretwell model
xsf := solve(diff(s(x), x) = s(x)/x, x);

# Definitions for sib rivalry model
constr := c*b^2 + (x-c)*b - T = 0;
c := 1;
```

```
# Solve constraint for b using positive root of quadratic formula
b := (-(x-c) + sqrt((x-c)^2 + 4*c*T))/(2*c);


# Solve for fitness maximum in sib rivalry model
xsr := solve(diff(log(w), x)=0, x);


# Print the two solutions side by side
evalf(xsf), evalf(xsr);
```

If you run it, you will find that $\hat{x} = 2$ under Smith-Fretwell, but equals 6.16 under the sibling rivalry model.

★EXERCISE **7–1** The first-order condition gives $\hat{x} = -1 + \sqrt{2} \approx 0.414$. At this point, the second-order condition gives

$$\left.\frac{d^2 \ln w}{dx^2}\right|_{x=\hat{x}} = 1.707$$

Since this is positive, $\hat{x}$ does *not* maximize fitness. In fact, it minimizes fitness.

★EXERCISE **8–1** We are interested in the left edge of both graphs, since this is where $p_B = 1$ and $p_A = 0$. In panel 1, $w_B > w_A$ at this edge, so result 8.1 implies that strategy B resists invasion by A. In panel 2, $w_A = w_B$ at this edge, and $w_B > w_A$ at points near the edge. Thus, result 8.2 implies that B resists invasion by A. Since A and B are the only feasible strategies, each graph implies that B resists invasion by all alternate strategies and is therefore an ESS.

★EXERCISE **8–3** To the left of the point $(\hat{p}_H)$ where the lines cross, $w_H > w_D$. (This follows because $V > V/2$.) To the right of $\hat{p}_H$, $w_H < w_D$ because (with these parameter values) $(V - C)/2 < 0$.

★EXERCISE **8–4** The equilibrium occurs where the two fitnesses are equal. At this point, $1/p_m = 2/(1 - p_m)$, which implies that $p_m = 1/3$. In your graph, the lines should cross at this value. The upshot is that if sons are expensive, mothers should produce fewer of them.

★EXERCISE **9–1** The difference in fitnesses is

$$\begin{aligned}
\Pi(H, I) - \Pi(D, I) &= p(V - C)/2 + (1 - p)V - (1 - p)V/2 \\
&= V(p/2 + (1 - p) - (1 - p)/2) - Cp/2 \\
&= V/2 - Cp/2
\end{aligned}$$

which equals 0 when $V = Cp$ or $p = V/C$.

★EXERCISE **9–2** Suppose that some mixed strategy J that plays Hawk with probability u and Dove with probability $1 - u$. It's payoff is

$$\Pi(J, I) = u\Pi(H, I) + (1 - u)\Pi(D, I)$$

But since H and D have the same payoff against I, we can substitute $\Pi(H, I)$ for $\Pi(D, I)$, producing

$$\Pi(J, I) = u\Pi(H, I) + (1 - u)\Pi(H, I) = \Pi(H, I)$$

Thus, J, H, and D all have the same payoff against I.

★EXERCISE **9–3** To calculate the optimal response function of the prey, we begin by writing the fitnesses of the two pure strategies as functions of x:

$$w_{WL} = -(1 - x)b \qquad \text{and} \qquad w_{WR} = -xb$$

The mean fitness of the prey is thus

$$
\begin{aligned}
w_2 &= yw_{WL} + (1-y)w_{WR} \\
&= -y(1-x)b - (1-y)xb \\
&= -y(1-2x)b - xb
\end{aligned}
$$

If $x < 1/2$, then w_2 decreases with y so the optimal value of y is zero; if $x > 1/2$, then the optimal value of y is one.

$\star$ EXERCISE **12–1** $w = (1/3) \times (3/2) + (2/3) \times (3/4) = 1$

$\star$ EXERCISE **12–2** $x_0' = v$

$\star$ EXERCISE **12–3** The average difference in character values between parent and offspring is $x_1' - x_1 = -u$ for allele A_1 and is $x_0' - x_0 = v$ for allele A_0. Thus,

$$
E[w_i(x_i' - x_i)] = -puW_1/W + qvW_0/W
$$

According to Price's equation, the mean change in character value equals this plus the covariance from the previous model (Eqn. 12.12). Thus,

$$
\Delta x = \frac{pq(W_1 - W_0)}{W} + \frac{qvW_0 - puW_1}{W}
$$

$\star$ EXERCISE **12–4** For the charisma model, the table looks like this:

Component	f_i	x_i	W_i	w_i
charismatic	x	1	b	x^{-1}
non-charismatic	$1-x$	0	0	0

$\star$ EXERCISE **12–5** With the data in the table, $x = 2.25$, $W = 2$, $E[x_i w_i] = 2.375$, and $\mathrm{Cov}(x_i, w_i) = 0.125$. Since $x_i' = x_i$, the second term in the Price equation is zero, and $\Delta x = 0.125$.

$\star$ EXERCISE **13–1** $u - \bar{p} = 0$ when $\bar{p} = 0$ or $\bar{p} = 1$. $u - \bar{p} > 0$ when $0 < \bar{p} < 1$. Consequently, in this model Hamilton's rule gives the condition for $\bar{p}$ to increase whenever $0 < \bar{p} < 1$.

$\star$ EXERCISE **13–2** In this model, u simplifies to $p(p-3)/(3p^2 - 4p - 1)$. $u - p = 0$ when p equals 0, 2/3, or 1. It is positive when $0 < p < 2/3$ and negative when $2/3 < p < 1$. When Hamilton's rule holds, p will move toward an equilibrium at $p = 2/3$.

$\star$ EXERCISE **13–3**

Survival as a function of parental investment:

```
> s := proc(x) 1 - exp(-(x-1)) end;

  s := proc(x) 1-exp(-x+1) end
```

Solve for Smith-Fretwell equilibrium—the case of maternal control). (We are interested in the positive answer only.)

```
> evalf(solve(diff(s(x), x) = s(x)/x, x));

  -.8414056605, 2.146193221
```

Now the case of offspring control, where offspring's manipulation benefits itself only.

```
> evalf(solve(diff(s(x), x)
>      = 0.5*s(x)/x, x));

  -.3733745453, 2.923938750
```

Note that $2.9 > 2.1$, so each offspring gets more.

Now assume that the sibs are half-sibs. Each offspring gets more still

```
> evalf(solve(diff(s(x), x)
>    = 0.25*s(x)/x, x));

        -.1726105583, 3.780080782
```

⋆ EXERCISE **13–4** The following continues from the answer to exercise 13–3.

Now the case in which offspring manipulations affect a litter of size m. Here is the equation that we need to solve.

```
> eq := diff(s(x),x) = s(x)*(1 - 0.5^m)/x;

    eq := exp(-x + 1)

                                    m
        (1 - exp(-x + 1)) (1 - .5 )
      = ---------------------------
                  x
```

Solve it with $m = 1$ (i.e. for litters of size 1).

```
> evalf(solve(subs(m=1,eq), x));

        -.3733745453, 2.923938750
```

We are assuming here that $r = 1/2$, so case with $m = 1$ is identical to full sib case above.

Next the case with $m = 5$.

```
> evalf(solve(subs(m=5,eq), x));

        -.8102496257, 2.178174244
```

Next the case with $m = 15$:

```
> evalf(solve(subs(m=15,eq), x));

        -.8413751429, 2.146223739
```

The more sibs are affected, the closer the optimum gets to the maternal optimum at 2.146.

With $m = 100$, the answer is nearly identical to the case of maternal control.

```
> evalf(solve(subs(m=100,eq), x));

        -.8414056604, 2.146193221
```

In summary, small litters lead to discrepancies between the fetal optimum and the maternal optimum. But when litters are large this discrepancy disappears. We should therefore expect to see more conflict in species with small litters than in species with large ones.

⋆ EXERCISE **14–1** The question can be rephrased by asking what value of t makes $e^{rt} = 2$. To solve this equation, take the natural log of both sides and then divide by r. The result

$$t_d = (\ln 2)/r \approx 0.69/r$$

is called the doubling time.

⋆ EXERCISE **14–2** The first step is to define a variable that equals the right-hand side of the renewal equation:

```
>#Set up renewal eqn:
> rh := sum('exp(-r*x)* l[x+1]*m[x+1]', x=0..10):
```

After some experimentation, I found:

```
> evalf(subs(r=0.1,rh));

    1.009415775
> evalf(subs(r=0.2,rh));

    .6123337140
```

Since these answers bracket unity, r must lie between 0.1 and 0.2. I then used fsolve to find the answer more accurately:

```
> rrenew := fsolve(rh=1,r, 0.1..0.2);

    rrenew := .1018420077
```

This is the rate of growth as determined by the renewal equation. It tells us that the population increases by roughly 10% in every unit of time. Since these demographic data were tabulated in 5-year intervals, the time unit here is five years.

⋆EXERCISE **14–3** The `leslie` procedure that I defined above can be used once again here. It yields a Leslie matrix that, after rounding to two decimal places, looks like

$$
L = \begin{pmatrix}
0 & 0 & 0.17 & 0.52 & 0.54 & 0.49 & 0.42 & 0.18 & 0.02 & 0 \\
0.91 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0.96 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0.94 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0.92 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0.91 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0.90 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0.89 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.88 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.86 & 0
\end{pmatrix}
$$

⋆EXERCISE **14–4** I used the `Eigenvalues` procedure to find λ and r:

```
> evals := Eigenvalues(L);
 evals := 0, 0, 0, 0, 0, 0, 0, 0, 0,
    1.107208535,
    .3622041335 + .7317423441 I,
    .3622041335 - .7317423441 I,
    - .08682185875 + .7249856868 I,
    - .08682185875 - .7249856868 I,
    - .4966823304 + .3155985247 I,
    - .4966823304 - .3155985247 I,
    -.5600076923, -.1046007229, 0
> lambda := %[10];

    lambda := 1.107208535

> rleslie := log(lambda);
```

```
rleslie := .1018420145
```

This answer is practically identical to the result obtained from the renewal equation.

⋆ EXERCISE **14–5** The preceding exercises show that the population is growing at an exponential rate of 0.1 per 5-year age interval. The annual rate of growth is 1/5 as large, or about 0.02. The population is growing at a rate of 2 percent per year.

⋆ EXERCISE **15–1** The assumption implies that an increase in N causes the same proportional reduction in the fertilities of all women, regardless of their age. Many other assumptions are possible. For example, as N increases the resulting decrease in fertility might fall primarily on older women. I know of no data on this issue, so it is hard to decide what assumption is most reasonable. We could however experiment with different assumptions. If all assumptions led to essentially the same answer, then it would not matter much which one we chose. This is called a *sensitivity analysis*.

⋆ EXERCISE **15–2** Changing the sign of δ changes the sign of the inequality. Thus, inequality 15.6 becomes

$$\frac{q}{1-q} < \frac{e^{-5r}l_{45}}{2m_{40}l_{40}} \tag{F.1}$$

With the chimpanzee data, $r \approx 0.0045$ so this inequality is

$$\frac{q}{1-q} < \frac{\exp[-5 \times 0.0045] \times 0.1510}{2 \times 0.5 \times 0.2114} = 0.698$$

or

$$q < 0.41$$

This condition says that the chimpanzee fertility pattern is stable against a mutation reducing fertility in old age provided that the rate of child-birth mortality among older females is less than 41%. This condition is almost certainly satisfied.

⋆ EXERCISE **15–3** The model of density dependence requires an adjustment in m_{40}, which becomes $0.5/R_0$, where $R_0 = 1.1159$ (see Eqn. 15.9). Inequality F.1 becomes

$$q/(1-q) < \frac{0.1510 \times 1.1159}{2 \times 0.5 \times 0.2114} = 0.797$$

or

$$q < 0.44$$

This says that the chimpanzee fertility pattern is stable against a mutation reducing fertility in old age provided that the rate of child-birth mortality among older females is less than 44%. Once again, it is easy to believe that this inequality is satisfied.

⋆ EXERCISE **15–4** As before, changing the sign of δ changes the sign of the inequality. Setting $r = 0$ in inequality 15.13, and reversing the direction of the inequality gives

$$\gamma < 1$$

If this condition is satisfied, then selection opposes mutations that reduce the fertility of 45-year-old females. Thus, the chimpanzee pattern is stable if the marginal effect on fertility of young females is smaller than that on fertility of older females. Recall that the human pattern was stable if the marginal effect on younger females is larger. A change from the chimpanzee pattern (of no menopause) to the human one would result from anything that increased the value of γ—from anything that increased the ability of old females to enhance the fertility of their daughters.

⋆EXERCISE **15–5** The chimpanzee pattern (no menopause) is stable if

$$\beta < 2\phi$$

It is stable, in other words, if the effect of grandmothers on the fertility of their daughters is less than twice as large as the benefit that they could provide for themselves (via the physiological trade-off) by emphasizing early fertility at the expense of late survival.

⋆EXERCISE **C–1**

$$\bar{x} = \frac{1}{3}(0 \times 0 + 1 \times 2 + 2 \times 1) = 4/3$$

⋆EXERCISE **C–2** In our data set (2, 1, and 1), there are two elements with value 1, and three elements in all, so $p_1 = 2/3$. Similarly $p_2 = 1/3$. There are no elements with value 0, so $p_0 = 0$.

⋆EXERCISE **C–3**

$$
\begin{aligned}
\bar{x} &= 0 \times 1/10 + 1 \times 2/10 + 3 \times 1/10 + 4 \times 4/10 \\
&= 2/10 + 3/10 + 16/10 \\
&= 21/10 = 2.1
\end{aligned}
$$

⋆EXERCISE **C–4** $E[3] = 3$

⋆EXERCISE **C–5** $E[2X] = 10$

⋆EXERCISE **C–6** $E[2X + 3Y] = 10 + 18 = 28$

⋆EXERCISE **C–7** $E[aX + bY^2] = aE[X] + bE[Y^2]$

⋆EXERCISE **C–9** The mean and variance are 8.75 and 18.1875.

⋆EXERCISE **C–10** First, $E[aX] = aE[X]$ by equation C.4. Next,

$$
\begin{aligned}
V[aX] &= E[(aX)^2] - E[aX]^2 \\
&= E[a^2 X^2] - a^2 E[X]^2 \\
&= a^2 \left(E[X^2] - E[X]^2 \right) \\
&= a^2 V[X]
\end{aligned}
$$

⋆EXERCISE **C–11** $\bar{x} = 4/3$, $\bar{y} = 6/3 = 2$,

$$
\begin{aligned}
\text{Cov}(x, y) &= [(2 - 4/3)(3 - 2) + (1 - 4/3)(2 - 2) + (1 - 4/3)(1 - 2)]/3 \\
&= [2/3 + 0 + 1/3]/3 \\
&= 1/3
\end{aligned}
$$

Bibliography

[1] R. McNeill Alexander. *Optima for Animals*. Edward Arnold, London, 1982.

[2] Richard D. Alexander. *Darwinism and Human Affairs*. University of Washington Press, Seattle, 1979.

[3] L. Blueweiss, H. Fox, V. Kudzma, D. Nakashima, R. Peters, and S. Sams. Relationships between body size and some life history parameters. *Oecologia*, 37:257–272, 1978.

[4] Robert Boyd and Peter J. Richerson. *Culture and the Evolutionary Process*. University of Chicago Press, Chicago, 1985.

[5] Robert Boyd and Peter J. Richerson. An evolutionary model of social learning: The effects of spatial and temporal variation. In T. Zentall and B.G. Galef, editors, *Social Learning: Psychological and Biological Perspectives*, pages 29–48. Lawrence Erlbaum, Associates, Hillsdale, NJ, 1988.

[6] Luca L. Cavalli-Sforza. The role of plasticity in biological and cultural evolution. *Annals of the New York Academy of Sciences*, 231:43–59, 1974.

[7] Luca L. Cavalli-Sforza and Marc W. Feldman. *Cultural Transmission and Evolution: A Quantitative Approach*. Princeton University Press, Princeton, 1981.

[8] Brian Charlesworth. *Evolution in Age-Structured Populations*. Cambridge University Press, Cambridge, England, 2nd edition, 1994.

[9] Brian Charlesworth and Eric L. Charnov. Kin selection in age-structured populations. *Journal of Theoretical Biology*, 88:103–119, 1981.

[10] Eric L. Charnov. Optimal foraging: The marginal value theorem. *Theoretical Population Biology*, 9:129–136, 1976.

[11] Eric L. Charnov. *The Theory of Sex Allocation*. Princeton University Press, Princeton, N.J., 1982.

[12] Alpha C. Chiang. *Fundamental Methods of Mathematical Economics*. McGraw-Hill, New York, 3rd edition, 1984.

[13] Timothy H. Clutton-Brock. Reproductive success. In Timothy H. Clutton-Brock, editor, *Reproductive Success*, pages 472–486. University of Chicago Press, Chicago, 1988.

[14] Timothy H. Clutton-Brock. *The Evolution of Parental Care*. Princeton, Princeton University Press, 1991.

[15] P. W. Concannon, J. P. McCann, and M. Temple. Biology and endocrinology of ovulation, pregnancy and parturition in the dog. *Journal of Reproductive Fertility Supplement*, 39:3–25, 1989.

[16] R. J. Cowie. Optimal foraging in Great Tits. *Nature*, 268:137–139, 1977.

[17] Richard Dawkins. *The Extended Phenotype: The Gene as the Unit of Selection*. Oxford University Press, New York, 1982.

[18] R. I. M. Dunbar. Adaptation, fitness, and evolutionary tautology. In King's College Sociobiology Group, editor, *Current Problems in Sociobiology*, pages 9–28. Cambridge University Press, 1982.

[19] William H. Durham. The adaptive significance of human social behavior. *Human Evology*, 4:89–121, 1976.

[20] William H. Durham. Title? In Napoleon A. Chagnon and William Irons, editors, *Evolutionary Biology and Human Social Behavior: An Anthropological Perspective*, page pages? Duxbury, North Scituate, 1979.

[21] Warren J. Ewens. *Mathematical Population Genetics*. Springer-Verlag, New York, 1979.

[22] William Feller. *An Introduction to Probability Theory and Its Applications*, volume I. Wiley, New York, 3rd edition, 1968.

[23] Ronald A. Fisher. *The Genetical Theory of Natural Selection*. Dover, New York, 2nd edition, 1958.

[24] S. J. Gould and R. C. Lewontin. The spandrels of San Marcos and the Panglossian paradigm: A critique of the adaptationist programme. *Proceedings of the Royal Society of London, Series B*, B205:581–598, 1979.

[25] Stephen J. Gould. Sociobiology and human nature: A postpanglossian vision. In *Sociobiology Examined*, pages 283–290. Oxford University Press, Oxford, 1980.

[26] Alan Grafen. Natural selection, kin selection, and group selection. In J. R. Krebs and N. B. Davies, editors, *Behavioural Ecology: An Evolutionary Approach*, pages 2–84. Sinaur, Sunderland, MA, 2nd edition, 1984.

[27] David Haig. Genetic conflicts in human pregnancy. *Quarterly Review of Biology*, 68(4):495–532, 1993.

[28] J. B. S. Haldane. *The Causes of Evolution*. Cornell University Press, Ithaca, 1932.

[29] William D. Hamilton. The genetical evolution of social behavior, I. *Journal of Theoretical Biology*, 7(1):1–16, July 1964.

[30] William D. Hamilton. The genetical evolution of social behavior, II. *Journal of Theoretical Biology*, 7(1):17–52, July 1964.

[31] William D. Hamilton. The moulding of senescence by natural selection. *Journal of Theoretical Biology*, 12(1):12–45, September 1966.

[32] Henry C. Harpending, Alan R. Rogers, and Patricia Draper. Human sociobiology. *Yearbook of Physical Anthropology*, 30:137–150, 1987.

[33] M. Harris. *Cultural Materialism: The Struggle for a Science of Culture*. Random House, New York, 1979.

[34] Kristen Hawkes, Kim Hill, and James F. O'Connell. Why hunters gather: Optimal foraging and the Aché of eastern paraguay. *American Ethnologist*, 9(2):379–398, May 1982.

[35] Kristen Hawkes, J. F. O'Connell, Nicholas G. Blurton Jones, Helen Alvarez, and Eric L. Charnov. Grandmothering, menopause, and the evolution of human life histories. *Proceedings of the National Academy of Sciences, USA*, 95:1336–1339, 1998.

[36] Kristen Hawkes, James F. O'Connell, and Nicholas Blurton Jones. Hardworking Hadza grandmothers. In V. Standen and Robert A. Foley, editors, *Comparative Socioecology: The Behavioural Ecology of Humans and other Mammals*, pages 341–366. Blackwell, Oxford, 1989.

[37] Kristen Hawkes, James F. O'Connell, and Nicholas G. Blurton Jones. Hadza women's time allocation, offspring provisioning, and the evolution of post-menopausal life spans. *Current Anthropology*, 38:551–578, 1997.

[38] J. Hofbauer and K. Sigmund. *Evolutionary Games and Population Dynamics*. Cambridge University Press, Cambridge, 1998.

[39] Michio Hori. Frequency-dependent natural selection in the handedness of scale-eating Cichlid fish. *Science*, 260:216–219, 1993.

[40] William Irons. Natural selection, adaptation, and human social behavior. In Napoleon A. Chagnon and William Irons, editors, *Evolutionary Biology and Human Social Behavior: An Anthropological Perspective*, pages 4–39. Duxbury, North Scituate, 1979.

[41] James Holland Jones, M. L. Wilson, and A. Pusey. Demographic analysis of the chimpanzees of Gombe National Park, Tanzania. Unpublished, 2003.

[42] P. Kitcher. *Vaulting Ambition: Sociobiology and the Quest for Human Nature*. MIT Press, Cambridge, 1985.

[43] A. L. Kroeber and C. Klukhohn. *Culture, a Critical Review of the Concepts and Definitions*, volume 47 of *Papers of the Peabody Museum of American Archeology and Ethnology*. Peabody Museum, 1952.

[44] J. A. Kurland. Paternity, mother's brother and human sociality. In N. Chagnon and W. Irons, editors, *Evolutionary Biology and Human Social Behavior: an Anthropological Perspective*, pages 145–180. Duxbury Press, North Scituate, MA, 1979.

[45] P. H. Leslie. On the use of matrices in certain population mathematics. *Biometrika*, 33:183–212, 1945.

[46] P. H. Leslie. Some further notes on the use of matrices in population mathematics. *Biometrika*, 35:213–245, 1948.

[47] C. J. Lumsden and Edward O. Wilson. *Genes, Mind and Culture*. Harvard University Press, Cambridge, 1981.

[48] John Maynard Smith. The theory of games and the evolution of animal conflicts. *Journal of Theoretical Biology*, 47:209–221, 1974.

[49] John Maynard Smith. *Evolution and the Theory of Games*. Cambridge University Press, Cambridge, 1982.

[50] John Maynard Smith and George R. Price. The logic of animal conflict. *Nature*, 246:15–18, 1973.

[51] Richard McElreath. Why menopause indeed? Unpublished, 1997.

[52] Richard McElreath and Robert Boyd. *Mathematical Models of Social Evolution*. University of Chicago Press, 2007.

[53] Douglas W. Mock and Geoffrey A. Parker. *The Evolution of Sibling Rivalry*. Oxford University Press, Oxford, 1997.

[54] S. C. Oliver. Ecology and cultural continuity as contributing factors in the social organization of the plains Indians. In Y. A. Cohen, editor, *Man in Adaptation: The Cultural Present*, pages 243–262. Aldine, Chicago, 1968.

[55] J. H. Pollard. *Mathematical Models for the Growth of Human Populations*. Cambridge University Press, New York, 1973.

[56] George R. Price. Selection and covariance. *Nature*, 227:520–521, 1970.

[57] George R. Price. Extension of covariance selection mathematics. *Annals of Human Genetics*, 35:485–490, 1972.

[58] H. Ronald Pulliam and C. Dunford. *Programmed to Learn: An Essay on the Evolution of Culture*. Columbia University Press, New York, 1980.

[59] David Rindos. The evolution of the capacity for culture: Sociobiology, structuralism, and cultural selectionism. *Current Anthropology*, 27:315–326, 1986.

[60] Alan R. Rogers. Does biology constrain culture? *American Anthropologist*, 90(4):819–831, December 1988.

[61] Alan R. Rogers. Why menopause? *Evolutionary Ecology*, 7(4):406–420, July 1993.

[62] T. J. Roper. Learning as a biological phenomenon. In T. R. Halliday and P. J. B. Slater, editors, *Genes, Development and Learning*, pages 178–212. W. H. Freeman, New York, 1983.

[63] Eugene E. Ruyle, F. T. Cloak, Jr., Lawrence B. Slobodkin, and William H. Durham. The adaptive significance of cultural behavior: Comments and reply. *Human Ecology*, 5:49–67, 1977.

[64] M. Sahlins. *The Use and Abuse of Biology: An Anthropological Critique of Sociobiology*. University of Michigan Press, Ann Arbor, 1976.

[65] B. Schwartz. *The Battle for Human Nature: Science, Morality and Modern Life*. W. W. Norton, New York, 1986.

[66] Christopher C. Smith and Stephen D. Fretwell. The optimal balance between size and number of offspring. *American Naturalist*, 108(962):499–506, 1974.

[67] Eric Alden Smith. Folk psychology *versus* pop sociobiology. *The Behavioral and Brain Sciences*, 10:85–86, 1987.

[68] Eric Alden Smith. Optimization theory in anthropology: Applications and critiques. In John Dupré, editor, *The Latest on the Best: Essays on Evolution and Optimality*, pages 201–249. MIT, Cambridge, 1987.

[69] D. W. Stephens and J. R. Krebs. *Foraging Theory*. Princeton University Press, New Jersey, 1986.

[70] Julian H. Steward. *Theory of Culture Change: The Methodology of Multilinear Evolution*. University of Illinois Press, Chicago, 1955.

[71] Science For the People. Sociobiology—another biological determinism. *BioScience*, 26:182–186, 1976.

[72] Robert L. Trivers. Parent-offspring conflict. *American Zoologist*, 14:249–264, 1974.

[73] Sara Via and Russell Lande. Genotype-environment interaction and the evolution of phenotypic plasticity. *Evolution*, 39(3):505–522, May 1985.

[74] E. E. Werner and D. J. Hall. Optimal foraging and the size selection of prey by the bluegill sunfish. *Ecology*, 55:1042–1052, 1974.

[75] Leslie A. White. *The Evolution of Culture: The Development of Civilization to the Fall of Rome*. McGraw-Hill, New York, 1959.

[76] George C. Williams. Pleiotropy, natural selection, and the evolution of senescence. *Evolution*, 11:398–411, 1957.

[77] George C. Williams. Natural selection, the costs of reproduction, and a refinement of Lack's principle. *American Naturalist*, 100:687–690, 1966.

[78] Edward O. Wilson and William H. Bossert. *A Primer of Population Biology*. Sinauer, Sunderland, Mass., 1971.

[79] Sewall Wright. *The Theory of Gene Frequencies*, volume 2 of *Evolution and the Genetics of Populations*. University of Chicago Press, Chicago, 1969.