

# Elements of Evolutionary Anthropology

*A Theoretical Minimum for  
Studying Human Evolution*





# *Contents*

Preface	ix
<b>Part 1. Light Elements</b>	<b>1</b>
Chapter 1. What the Fitness	3
Chapter 2. Probability	13
Chapter 3. Optimality	17
Chapter 4. The Evolution of Conflict	19
Chapter 5. The Evolution of Cooperation	31
Chapter 6. The Evolution of Relationships	37
Chapter 7. The Evolution of Families	51
<b>Part 2. Transition Elements</b>	<b>59</b>
Chapter 7. Causal Inference	61
Chapter 11. Senescence	63
Chapter 18. Parental Investment	75
Chapter 19. Social Learning	77



## *Preface*

The big problems in human evolution are all woven together. A typical research project is threaded with demography, development, cognition, evolution, and probability theory. Every specific empirical case is colored by details of cultural institutions like kinship and politics.

Anthropology is inherently undisciplined.

The goal of this text is to present and support one view of the theoretical minimum a researcher requires to conduct responsible research in human evolution and ecology.

The goal is not to engender mastery in all aspects. Mastery comes with practice and application. Besides, for many research skills, mastery means only recognizing contexts where some technique or fact is useful. You can look up the rest afterwards.

The elements metaphor is only playful. There is no natural or unique typology of topics. The structure I've chosen is designed to reduce complexity into chunks that can be individually studied and referenced, as well as to indicate which chunks are foundational to others. For example, a student wishing to research the evolutionary reasons for human menopause will need to become acquainted with age-structured life history theory, among other things, and that itself requires first studying some relevant mathematics.



## Part 1

# Light Elements





# 1 F

## *What the Fitness*

No one reads Herbert Spencer anymore. Spencer was the most authoritative philosopher of his generation, selling over a million copies of his various works. Along the way he invented the phrase **survival of the fittest**. This loathsome phrase is the only part of his work that is remembered. It too should be forgotten.

The concept of evolutionary or Darwinian **fitness** is central to the study of evolution. It is a beguilingly simple concept, nearly always taught incorrectly. The problem is that evolution does not in general favor the survival of the fittest, under common definitions of “fitness.” And what evolution does care about has no simple, universal definition other than perhaps “the ability to persist in the long run.” This sounds like a platitude, and we should aspire to more than platitude.

Before getting any deeper into evolutionary theory, you need to better understand fitness. Otherwise you might misunderstand how to evaluate evolutionary explanations. Let’s tour some common misunderstandings. In this chapter, I’ll present two realistic factors that complicate real fitness: timing and variation. In later chapters, these factors will be important for understand theories of the evolution of aging and learning.

### *Reproductive success*

*Different genotypes with the same lifetime reproductive success can differ dramatically in number of descendants.*

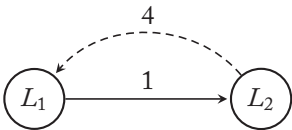
Fitness is most often defined as the **average reproductive success** of a certain genotype. Reproductive success means the number of offspring produced by an individual in its lifetime. Sometimes this is called “counting babies.” Counting babies isn’t a sound guide to which genotypes persist. You can dress this definition up with decorations like “offspring who themselves survive to reproduce,” but a costume doesn’t fix it. Counting babies is in principle wrong, even if it sometimes gives the right answer. Understanding the principle matters, unless you are happy to be right only by accident.

Consider the following example. Imagine a species in which individuals live for exactly two years. There are two genotypes. The **early** genotype starts breeding

in the first year and produces **2** offspring each year. The **late** genotype does not reproduce in its first year, but it produces **4** offspring in the second year. The early type has a lifetime reproductive success of 4. The late type also has a lifetime reproductive success of 4.

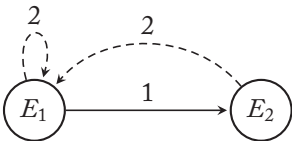
5 It will be useful to draw these alternative genotypes as life history diagrams. Let's start with the late-reproducing genotype, since it is simpler.

Figure 1.1



10 In this diagram, the circles are the age classes, with one-year-olds on the left and two-year-olds on the right. The arrow from  $L_1$  to  $L_2$  represents aging. In this model, there is no mortality during aging, so the 1 on this path represents 1 two-year-old for each one-year-old. Then there is reproduction, represented by the dashed arrow from 2 to 1. Each two-year-old contributes four offspring to the one-year-olds in the next year—that's the dashed arrow with the 4 on it represents. We can draw a similar diagram for the early-reproducing genotype:

Figure 1.2



15 This diagram is very similar to the previous, but now there is a second dashed reproduction arrow, looping from 1 back to 1. This represents reproduction of one-year-olds.

20 These two genotypes have the same reproductive success, but one of them will very quickly replace the other. To see this, you have to leave heuristics and metaphors behind and just calculate. To make the calculation easier, let's start with just one individual of each genotype of each age. So the population, to begin, has just 4 individuals: an early type of age 1, an early type of age 2, a late type of age 1, and a late type of age 2. Let's abbreviate the numbers of these types as  $E_1$ ,  $E_2$ ,  $L_1$ , and  $L_2$ . So this is the state of the population in the first year:

25

Year	$E_1$	$E_2$	$L_1$	$L_2$
1	1	1	1	1

30 When a year passes, we must update the **state variables** (here  $E_1$ ,  $E_2$ ,  $L_1$ ,  $L_2$ ) according to the rules expressed in the diagrams above. There is a simple algorithm for translating a diagram into a set of equations for updating the numbers of each type. Here it is:

- (1) Pick a variable.

- (2) Find each arrow entering the variable. Each arrow contributes individuals in the next year. The number it contributes is equal to the the variable at the source of the arrow multiplied by the rate on the arrow itself.
- (3) The value of the variable in the next year is the sum of the arrow products.

Let's apply this to  $E_1$ . There are two arrows entering  $E_1$ . The first is the loop from  $E_1$  back to itself. And the rate on the arrow is 2. So this arrow contributes  $2 \times E_1$  individuals in the next year. The second arrow is the one from  $E_2$ . The rate is again 2. So this arrow contributes  $2 \times E_2$ . Adding the contributions of the two arrows:

$$E_{1,t+1} = 2E_{1,t} + 2E_{2,t}$$

Applying the same procedure to the other state variables, we get four equations that let us update the state of the population each year:

$$\begin{aligned} E_{1,t+1} &= 2E_{1,t} + 2E_{2,t} & L_{1,t+1} &= 4L_{2,t} \\ E_{2,t+1} &= E_{1,t} & L_{2,t+1} &= L_{1,t} \end{aligned}$$

Now we plug in the values from year 1,  $t = 1$ , to update.

$$\begin{aligned} E_{1,t+1} &= 2(1) + 2(1) = 4 & L_{1,t+1} &= 4(1) = 4 \\ E_{2,t+1} &= 1 & L_{2,t+1} &= 1 \end{aligned}$$

The genotypes are still tied. But let's do another year, and then update the table:

Year	$E_1$	$E_2$	$L_1$	$L_2$
1	1	1	1	1
2	4	1	4	1
3	10	4	4	4

The early type has now pulled ahead, with  $10 + 4$  individuals compared to late's  $4 + 4$ . If you keep turning the crank on this calculation, you'll see that early explodes compared to late. Here's some simple code to turn the crank for you.<sup>1</sup> Let's do 6 years of population growth.

```
let E1=1, E2=1, L1=1, L2=1
println("\tE1", "\tE2", "\tL1", "\tL2")
for t in 1:6
    E1 , E2 = 2*(E1+E2) , E1
    L1 , L2 = 4*L2 , L1
    println(t, "\t", E1, "\t", E2, "\t", L1, "\t", L2)
end
end
```

  
1.1

	$E_1$	$E_2$	$L_1$	$L_2$
1	4	1	4	1
2	10	4	4	4
3	28	10	16	4

20

<sup>1</sup>This is Julia code. Julia is a high-performance scientific computing language. It's faster and better designed than R or Python. Download it at <https://julialang.org/>.

4	76	28	16	16
5	208	76	64	16
6	568	208	64	64

The early type has a huge advantage, despite having the same lifetime reproductive success. By year 10, there are more than 10 times as many  $E_1$  as  $L_1$ . The advantage of the early type is so big, that you can give the late type a higher reproductive success, and early can still win. Repeat the calculation above, but modify the code so that the late type produces 6 offspring instead of 4. You should get:

	E1	E2	L1	L2
10	1	4	1	6
	2	10	4	6
	3	28	10	36
	4	76	28	36
	5	208	76	216
15	6	568	208	216

So the early type grows faster, even though it has lower reproductive success.

Of course if the late type’s reproductive success is large enough, it will beat the early type. It is possible to find a general condition for the early type to beat the late type, and the box on the next page shows you how. These calculations will be emphasized much more in later chapters. But for now, I’ll smuggle them into optional boxes. You don’t have to understand everything at once.

The early type grows faster because it reproduces earlier. There is no risk in this example of dying before reproduction, so that isn’t the early type’s advantage. Instead the advantage is that early’s offspring start reproducing earlier themselves. Late in contrast takes a year off, and then its offspring take a year off. One way to summarize this result is that early has a shorter **generation time**, and shorter generation times can help a population grow.

But the broadest lesson is that we should not use heuristics like lifetime reproductive success to understand the genotypes and phenotypes that appear in nature. Adding another heuristic like “shorter generations are better” is no good either. Sometimes natural selection favors a longer generation time.

If you really want a heuristic to walk away with, I suggest “**timing matters**”.

*Average reproductive success*

*Even when reproductive success is what matters, more than the average may matter.*

Another problem with defining fitness as **average reproductive success** is the average part. When reproductive success varies from year to year or individual to individual, the entire distribution can matter, not just the average.

Again consider a simple example, a species that lives and reproduces for only one year. There are two genotypes. The **constant** (C) genotype always produces 2 offspring and then dies. The **variable** (V) genotype produces 1 offspring in each even year and 3 offspring in each odd year, for an average of 2. The average reproductive success of both genotypes is the same.

**Condition for early reproduction to beat late reproduction.** We can write the population dynamics as a system of **recursions**, equations that update the numbers of each type. I'll replace the fixed values in the example with variables, so we can find which values allow the early type to beat the late type.

$$\begin{aligned} E_{1,t+1} &= b_1 E_{1,t} + b_2 E_{2,t} & L_{1,t+1} &= B L_{2,t} \\ E_{2,t+1} &= E_{1,t} & L_{2,t+1} &= L_{1,t} \end{aligned}$$

The variables  $b_1$  and  $b_2$  are the fertilities of age 1 and 2 early types, and  $B$  is the fertility of age 2 late types. Assume  $B > b_1 + b_2$ . When can the early type beat the late type, even though the late type has higher lifetime reproductive success?

To find the answer, we calculate the **long-term growth rate** of each type. The long-term growth rate is the per-year growth rate of the population in the long run. For the late type, we can find it first with intuition. The late type grows by a factor of  $B$  every two years. Look at the table on the previous page. In that example  $B = 4$ . The late type population starts at 5, then grows to 20 two years later, then to 80 two years after that. This means that after a large number of years  $x$ , the population will be approximately  $B^{x/2}$  times larger. The growth rate per year is then the  $x$ -th root,  $(B^{x/2})^{1/x} = \sqrt{B}$ .

Intuition is nice, but isn't always reliable. We can also prove this in a principled way. The long term growth rate  $\lambda_L$  of the late type satisfies these equations:

$$L_{1,t+1} = \lambda_L L_{1,t} \qquad L_{2,t+1} = \lambda_L L_{2,t}$$

Why? Because the long-term growth rate is just the relative size of population from one year to the next. So  $\lambda_L = L_{1,t+1}/L_{1,t} = L_{2,t+1}/L_{2,t}$ . Substituting in the recursions and removing the  $t$  subscripts:

$$B L_2 = \lambda_L L_1 \qquad L_1 = \lambda_L L_2$$

Solve for  $\lambda_L$  and you'll get two solutions:  $\lambda_L = \pm\sqrt{B}$ . Only the positive solution makes biological sense here. So  $\lambda_L = \sqrt{B}$  is the long-term growth rate of the late type. The same procedure for the early type produces these two equations:

$$b_1 E_1 + b_2 E_2 = \lambda_E E_1 \qquad E_1 = \lambda_E E_2$$

Solve for the growth rate  $\lambda_E$ . Again you'll find two roots, but only one is positive. All of this work gives us a long-term growth rate for each type:

$$\lambda_E = \frac{1}{2} \left( b_1 + \sqrt{b_1^2 + 4b_2} \right) \qquad \lambda_L = \sqrt{B}$$

Before moving on, note that neither of these expressions resembles **lifetime reproductive success**. Lifetime reproductive success of the early type is  $b_1 + b_2$ . And of the late type is  $B$ . Don't count babies.

Now we ask when  $\lambda_E > \lambda_L$ . After some algebra, the condition simplifies to:

$$b_1 > \frac{B - b_2}{\sqrt{B}}$$

Arranged this way, we can see how much more important reproduction in the first year is for long-term growth. Suppose  $b_2 = 0$  for example. Then the above simplifies to  $b_1 > \sqrt{B}$ . So a species that produces 11 offspring in its first year and dies can out-compete (in the long run) one that waits one year and then produces 100 offspring. This is the extraordinary advantage of early reproduction, at least in a population that grows without bound. In a later chapter, you'll see that this advantage may depend upon **population regulation**, how population density influences survival and reproduction. Different mechanisms of population regulation influence natural selection in different ways.

However the growth rates of these genotypes are not the same. Suppose there are 10 individuals of each type. Now let's compute the population size of each genotype over the next 7 years.

julia  
1.2

```
let C=10, V=10
println("\tC\tV")
for t in 1:7
    C = 2*C
    V = (t%2==0) ? 1*V : 3*V
    println(t, "\t", C, "\t", V)
end
end
```

	C	V
5 1	20	30
2	40	30
3	80	90
4	160	90
5	320	270
10 6	640	270
7	1280	810

The constant type races ahead, even though both genotypes have the same average reproductive success.

The average isn't what matters, because population growth is not an additive process. The size of a population is not the sum of the offspring produced in previous years. Instead, population growth is multiplicative. If ever there are zero offspring, then the genotype goes extinct. One zero is enough for extinction.

Variation in reproductive success interacts with the multiplicative nature of population growth in very powerful ways. To figure out why, and see how this explains why the constant type wins in our example, we need to get formal. Let the population size of the constant type in a given year  $t$  be  $C_t$ . In the next year,  $t + 1$ , the population size is:

$$C_{t+1} = 2C_t$$

The population of constant individuals doubles every year. After another year, the population size is:

$$C_{t+2} = 2C_{t+1} = 2 \times 2C_t = 2^2 C_t$$

After yet another year:

$$C_{t+3} = 2C_{t+2} = 2 \times 2^2 C_t = 2^3 C_t$$

You can probably see now that in any year  $t + x$ , the population size will be:

$$C_{t+x} = 2^x C_t$$

But what about the other type? Recall that the variable type V produces 1 offspring in an even year and 3 in an odd year. Suppose year  $t$  is even. Then in the next year, the population size of the variable type will be:

$$V_{t+1} = 3V_t$$

And in the next year, which is even again:

$$V_{t+2} = 1V_{t+1} = (1)(3)V_t$$

And the next, which is odd again:

$$V_{t+3} = 3V_{t+2} = (1)(3^2)V_t$$

And one more year to draw out the implications:

$$V_{t+4} = 1V_{t+3} = (1^2)(3^2)V_t$$

Yes, that  $1^2$  just equals 1, but if that 1 were a different fertility, the term would make a difference. So it is worth keeping it there.

This is obviously more complicated than the constant type. Note first that we do not average over years, because in no year does the V genotype get 2 offspring. It alternates between getting 1 and 3. And the population size in any year is the product of those previous fertilities—one 3 for each odd year and one 1 for each even year—and the initial population size  $V_t$ .

In the long run, over very many years  $x$  after year  $t$ , V will experience nearly the same number of even and odd years,  $x/2$  even years and  $x/2$  odd years. So its population size will be approximately:

$$V_{t+x} = (1^{x/2})(3^{x/2})V_t$$

We can compare this growth to the growth of C. If the initial populations are the same size,  $C_t = V_t$ , then C will be larger than V when:

$$2^x > (1^{x/2})(3^{x/2})$$

To compare the left and right sides, we need to get rid of  $x$ . To get  $x$  out of this, we can take the  $x$ -th root. This means raising each side to the power  $1/x$ , like this:

$$(2^x)^{1/x} > ((1^{x/2})(3^{x/2}))^{1/x}$$

$$2 > 1^{1/2}3^{1/2}$$

Now square both sides:

$$2^2 > (1)(3)$$

This condition is true. The constant type has a higher growth rate.

The average isn't what matters, in general. In this example, what matters is actually the **geometric mean** of the reproductive success over years. Without noticing, we calculated the geometric mean growth of each genotype. I provide the details in the math box below. The geometric mean isn't what matters in general.

It just happens to be what matters in simple models like this one. If the species lives more than one year, something else will matter.

Also in the math box, I show that a constant type with lower average reproductive success can beat a variable type with higher average reproductive success provided the variance is large enough. As an intuitive extreme case, if the variable type ever experiences a year with zero reproductive success, it will go extinct. One zero is sufficient.

*Geometric mean fitness.* To abstract the derivation in the main text, notice that the growth rate of the variable type is the product of two terms,  $1^{1/2}$  and  $3^{1/2}$ . Each of these terms is a possible fertility, 1 or 3, raised to the probability that it happens in a long sequence of years. In many years, half the years will result in a growth factor 1 and half instead in 3. This applies to more complex models, with more possible growth factors with different probabilities, as long as there is no age structure. If there is age structure or stage structure or any number of other complications, then this result doesn't perfectly generalize. But the intuition you get from it remains valuable in more complicated models, and understanding the general derivation can help you compute what matters in more complicated models. So it's worth understanding.

Suppose for example that there are  $m$  possible growth factors  $f_i$ . In a long sequence of  $x$  years, each factor occurs with proportion  $p_i$ . Then the growth rate over these  $x$  years is given by:

$$f_1^{p_1 x} f_2^{p_2 x} f_3^{p_3 x} \cdots f_m^{p_m x} = \prod_{i=1}^m f_i^{p_i x}$$

which you can read as “the product of all growth factors, each raised to the number of times it happened in the sequence.” If we wish to compare the growth record to some other lineage over the same number of years  $x$ , it will be convenient to remove  $x$  from the expression. The only way to do that is to take the  $x$ -th root:

$$\left( \prod_{i=1}^m f_i^{p_i x} \right)^{1/x} = \prod_{i=1}^m f_i^{p_i}$$

This expression is a **geometric mean** growth rate, because the geometric mean is defined as the  $x$ -th root of the product of  $x$  values. This proves that what selection cares about in this model is not **average lifetime reproductive success**, but rather **geometric mean lifetime reproductive success**. In other models, it cares about other quantities.

It is typical to see this expression on the log scale, in which case it becomes:

$$\log \prod_{i=1}^m f_i^{p_i} = \sum_{i=1}^m p_i \log(f_i)$$

The summation is often easier to manipulate, and you can interpret it as the average log lifetime reproductive success. So if you really want to salvage baby counting as a measure of fitness, you could try with this expression, as long as you are willing to count logarithmic babies instead of whole babies.

There are important consequences of this fitness concept. For example, if an organism can do something to increase a small  $f$  value, that will have a bigger impact on its long-term growth than increasing a large  $f$  value by the same amount. We can prove this by asking how quickly the growth rate (call it  $R$ ) changes as a function of any specific  $f_i$ , and the answer is:

$$\frac{\partial R}{\partial f_i} = \frac{p_i}{f_i}$$



This is a consequence of the fact that the derivative of  $\log(x)$  with respect to  $x$  is  $1/x$ , a useful fact to remember. The consequences of this mathematical fact are biologically enormous. When  $f_i$  is small, the rate above is large. So a small change to a small growth factor will make a big difference in growth. There is a fundamental asymmetry to growth processes in that it is harder to recover from losses than it is to lose gains, and this arises from the multiplicative nature of growth.

Another consequence is that variation in fitness, keeping the same average fitness, is usually bad for growth. To see this, suppose the long-term growth rate is given again by  $R = \sum_i p_i \log(f_i) = E \log(f)$ . We don't know the distribution of  $f$ . But we can still say something useful about it through an approximation called a **Taylor expansion**. In this case, we'll approximate  $R$  assuming that individual  $f_i$  values tend to be close to the mean,  $E(f)$ . How close? Close enough that  $(f_i - E(f))^3 \approx 0$ . This gives us:

$$R \approx \log E(f) - \frac{E((f - E(f))^2)}{2 E(f)^2}$$

You may know that the **variance** of a random variable is defined as the average squared deviation from the mean. In this case that's  $\text{var}(f) = E((f - E(f))^2)$ , which appears on top of the second term above. So we can write it as:

$$R \approx \log(E(f)) - \frac{\text{var}(f)}{2 E(f)^2}$$

This says that the growth rate is approximately the (log) average reproductive success minus a term that is proportional to the variance in reproductive success. So the larger the variance, the lower the growth rate. And if the variance is small enough, then average reproductive success is a good approximation of the growth rate. Many evolutionary analyses either explicitly or implicitly assume that the variance is small and use arithmetic mean reproductive success as a measure of the growth rate. And a lot can be learned that way.

But a lot can be missed as well. Notably, behavior that reduces variance in reproductive success can be beneficial. A lineage with lower average reproductive success can spread, provided its variance is also sufficiently lower. Suppose a constant genotype has mean reproductive success  $E(k)$ . A variable type has mean reproductive success  $E(v) > E(k)$ . The constant type will nevertheless grow faster, if:

$$\frac{\text{var}(v)}{E(v)^2} > 2(\log E(v) - \log E(k))$$

This is just an approximation of course, and it depends upon the simple population structure and life history of the example. But it is a symptom of the general fact that natural selection does not inherently care only about average reproductive success.

## Lessons

The previous sections presented two simple evolutionary scenarios in which average reproductive success, the most common definition of fitness, gives the wrong answer. There are many core problems in the study of evolution, including human evolution, where this matters. 5

If you study enough models, you get a sense for general principles, like for example that variation is usually bad for a species. Or that shorter generation times are better. But sometimes variation is good. And sometimes a longer generation time is better. In a novel scenario, there is just no getting around making a formal model and doing the work. Intuition is not enough. 10

### *Invasion Fitness*

The previous sections focus on the long-term growth rate of an individual genotype, or lineage. They give the impression that the correct way to compare genotypes is to compute their long-term growth rates in isolation. The genotype  
5 with the highest growth rate will win, in the long run.

But this is not correct. A genotype does not spread and become common because of its growth rate in isolation. Instead it spreads because of its growth rate in the presence of other genotypes. Once common, a genotype could actually have a lower growth rate than the genotypes it replaced. There are published,  
10 peer-reviewed papers that make this mistake.<sup>2</sup> These papers ask which genotype maximizes population growth rate and then assume that genotype will replace the others. Sometimes this approach gives the right answer, by accident. But it is never the correct approach.

Many aspects of social behavior contain some tradeoff between the ability of a  
15 phenotype to spread when rare and to persist when common. Kin selection, cooperation, conflict, parental care, sexual selection, and learning all present powerful examples. We'll work through these examples in detail in later chapters.

Let's consider a very simple example, to illustrate the basic procedure.

### *Empirical fitness*

20 [how to sensibly estimate lineage growth rate]

Age at first reproduction can be the most important thing. If fitness varies with temporal environmental variation, then cross-sectional data may mislead.

---

<sup>2</sup>Cite that evolution of menopause paper

# <sup>2</sup> P

## Probability

The greatest confusion with probability is believing it exists.<sup>1</sup>

Consider a coin toss. An early Roman coin was made of bronze and simply called “a bronze,” *aes*. An *aes* had a head on one side and a ship on the other. The head belonged to Janus, the Roman god of duality, among other things. Janus’ head was usually displayed with two faces, one looking to the future and the other into the past.

Janus was supposed to have been the first Italian to make coins, his faces embossed on one side. The ship on the other side commemorated Saturn, who arrived in Italy by ship and taught Janus how to farm, among other things. When Romans decided things by the toss of a coin, asking Janus to decide the outcome, they called it “heads or ships” (*capita aut navia*) rather than “heads or tails.”



*A Roman aes from around 230 BCE*

Coin tosses are useful, because they are **random**. The outcome cannot be easily influenced nor predicted, and so both parties may trust the coin to resolve a decision fairly. But what is it about the toss of a coin—or the roll of the dice—that makes it random? Coins are physical objects, and their behavior is determined by physical law. A precise experiment, which applies the same forces every time, can repeat the outcome every time.<sup>2</sup> Coin tosses are deterministic. Neither Janus nor any other deity influences the toss.

<sup>1</sup>This paraphrases Bruno de Finetti, who wrote PROBABILITY DOES NOT EXIST. All caps in original. He felt very strongly about it.

<sup>2</sup>Diaconis, Holmes, Montgomery paper.

The random numbers that your computer generates are technically known as **pseudo-random numbers**. Computer calculations are also deterministic, and computers produce random sequences of numbers using mathematical algorithms. Some of these algorithms have flaws that allow clever individuals to predict parts of the sequence. For this reason, some people prefer to tap natural phenomena, like atmospheric noise and radio static, for their random numbers. The service RANDOM.ORG does exactly this. Since 1998, it has supplied artisanal **true random numbers**, distilled from the atmosphere. Of course atmospheric noise is also deterministic. Atmospheric dynamics may be chaotic and difficult to predict. But the initial conditions completely determine the outcome, just as with a coin toss.

Randomness is not a property of natural phenomena. It is a property of knowledge. The randomness of a coin toss is not a property of the coin. It is a property of our lack of information about the physical conditions that determine the outcome. The randomness of radio static is no more “true” than the randomness of numbers generated inside a computer. Both are produced by deterministic phenomena. In both cases, the precision of knowledge necessary to accurately predict the outcome is practically unobtainable. But the randomness in both cases is a property of us. Our lack of information makes it possible for us to use a coin or radio static as a social device for making decisions.

The strictly epistemological—belief based—nature of randomness is not controversial. But some scientists do not like it. Two clever objections immediately come up. The first is what I call “quantum begging.” The second is to claim that randomness is a property of sequences of numbers. Both objections are mistaken.

**Quantum begging** is appealing to the hypothetically indeterminate nature of quantum phenomena. At some level of description, quantum phenomena are currently only describable by probability distributions. There has been a century of debate about whether this indeterminacy is merely epistemological—like all other randomness—or rather fundamentally real. Einstein famously sided with determinism. This debate is not over, although many books and films present it as such. However, this debate is entirely irrelevant to coins and dice, because coins and dice are not quantum objects. Every model agrees that physics at the scale of human lives is deterministic.

The second objection is to say that only **sequences** can be random. A sequence of coin tosses may be defined as random, if heads and tails appear nearly as often in the sequences and there is no useful information in previous outcomes to help you predict future outcomes. This definition is fine, but it doesn’t solve the problem. A single coin toss can be usefully random, provided neither party knows whether heads or tails will turn up. The coin doesn’t even have to be fair, have to present heads and tails with equal frequency in a very long sequence of tosses.

Why is a book about human evolution getting into the philosophy of randomness? Because all evolutionary theory is built from probability theory, and the conventional view of probability is wrong. It’s not that all evolutionary theory is

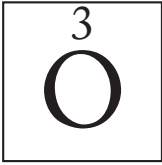
wrong—although some of it is. Rather to understand how to read and construct evolutionary theory, you must stop believing that probability exists.

*The conventional view*

The conventional view is that probability is defined as the proportion of some event in an infinite sequence of repeated trials. But all randomness, probability, is epistemological. It is defined by the observer's state of information.

5





## *Optimality*

Nulla in ipsum. Praesent eros nulla, congrue vitae, euismod ut, commodo a, wisi. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Aenean nonummy magna non leo. Sed felis erat, ullamcorper in, dictum non, ultricies ut, lectus. Proin vel arcu a odio lobortis euismod. Vestibulum ante ipsum primis in faucibus orci luctus et ultrices posuere cubilia Curae; Proin ut est. Aliquam odio. Pellentesque massa turpis, cursus eu, euismod nec, tempor congrue, nulla. Duis viverra gravida mauris. Cras tincidunt. Curabitur eros ligula, varius ut, pulvinar in, cursus faucibus, augue.

Nulla mattis luctus nulla. Duis commodo velit at leo. Aliquam vulputate magna et leo. Nam vestibulum ullamcorper leo. Vestibulum condimentum rutrum mauris. Donec id mauris. Morbi molestie justo et pede. Vivamus eget turpis sed nisl cursus tempor. Curabitur mollis sapien condimentum nunc. In wisi nisl, malesuada at, dignissim sit amet, lobortis in, odio. Aenean consequat arcu a ante. Pellentesque porta elit sit amet orci. Etiam at turpis nec elit ultricies imperdiet. Nulla facilisi. In hac habitasse platea dictumst. Suspendisse viverra aliquam risus. Nullam pede justo, molestie nonummy, scelerisque eu, facilisis vel, arcu.

Curabitur tellus magna, porttitor a, commodo a, commodo in, tortor. Donec interdum. Praesent scelerisque. Maecenas posuere sodales odio. Vivamus metus lacus, varius quis, imperdiet quis, rhoncus a, turpis. Etiam ligula arcu, elementum a, venenatis quis, sollicitudin sed, metus. Donec nunc pede, tincidunt in, venenatis vitae, faucibus vel, nibh. Pellentesque wisi. Nullam malesuada. Morbi ut tellus ut pede tincidunt porta. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Etiam congrue neque id dolor.





# 4 H

## *The Evolution of Conflict*

Sometimes contemporary evolutionary theory is called “Neo-Darwinian.” But it would be better to call it “Post-Darwinian.” Darwin employed no formal mathematics, got some big things wrong, and most of contemporary evolutionary theory was created from mathematical scratch afterwards, in the 20th century.

The cult of prestige surrounding Darwin is misleading. For example, Darwin argued that four postulates make evolution by natural selection adaptive. These postulates are:

- (1) Individuals vary in their characteristics
- (2) There is competition to survive and reproduce
- (3) Variation has consequences for this competition
- (4) Offspring resemble parents

If these four things are true, the argument goes, then natural selection occurs and produces adaptations. There are several problems here, as any contemporary evolutionary biologist knows. The postulates are still taught in introductory courses, but we all know they are a kind of white lie. Post-Darwinian evolutionary theory is much more complicated and much more interesting.

One important thing missing from the postulates is **frequency dependence**. When the consequence of a variable characteristic depends upon the frequency of that characteristic in the population, then natural selection may not produce adaptations at all. What this means is that natural selection need not increase the frequency of characteristics that help a species survive or reproduce in its environment. Sure, we could redefine “adaptation” so that it is only about some narrow sense of relative welfare of individuals in the same species. But the original, and I think still meaningful, concept of adaptation is the view of design—a species that is adapted to survival seems designed for it. In that sense, under frequency dependence, favored characteristics can seem designed to destroy the species.

This chapter introduces **evolutionary game theory**, a method of defining and analyzing theories of the evolution of social behavior, theories in which frequency dependence is important. I won’t spend much time talking about game theory in

general, because it is diverse and honestly not really coherent—none of its methods are distinct from general methods for modeling dynamical systems.

But evolutionary game theory does have a distinct style. It's style is that it is profoundly stupid. However, verbal theories are just as stupid. The value of game theory is that it exposes that stupidity, reveals that we still cannot intuit the consequences of simple constellations of assumptions, and provides a framework for deducing the correct consequences. Transparent stupidity is good.

This chapter begins with the topic of conflict, since it is foundational to evolutionary theory. It's right there in Darwin's postulates. But it isn't so simple. To begin, we need a very basic game theoretic model, stripped of ecological and physiological complexity. Then we can ask and answer basic questions step by step, folding in additional complexity as our questions grow. So in the sections that follow, we start with the simplest model I know. Like most simple game theoretic models, it cannot be applied directly to any realistic ecological context. It makes no quantitative predictions, even though it is a mathematical model. But it does provide qualitative insight—few people anticipate its results, and it is possible to understand how the details of the model lead to the results.

### *Hawks, doves, and chickens*

*When conflict is very costly, animals may evolve to avoid it. But hardly ever is the evolutionarily stable amount of conflict optimal for population growth or survival.*

Rattlesnakes like to eat other snakes. They are convenient to digest, I suppose. But when rattlesnakes meet one another, they don't usually try to eat one another. Instead they wrestle. They twist around one another until one of them grows tired or bored and cedes the battleground to the superior wrestler.

Darwin's postulates posit conflict as foundational: There is competition for survival and reproduction, and so animals fight over the resources that aid survival and reproduction. But many animals act like the rattlesnake. They have strong natural weapons, like fangs and venom. But when they contest a resource, they do not use these weapons. Instead they engage in ritualized combat.

How can natural selection explain this restraint? To address this question logically, we have to construct a scenario in which the resources, competitors, strategies, and costs and benefits are defined. Game theory does not provide the assumptions. Those are our responsibility and depend upon many subjective choices in moving from a vague question to a logical scenario. But game theory does give us an objective way to deduce the consequences of our assumptions. And that's a huge help, because intuition alone is terrible at deduction.

So suppose a population of animals in which pairs of adults compete for a non-divisible resource, such as a territory or food item or mate. Suppose that the value of the resource, in units of fitness, is  $v$ . Suppose also that the species

has natural weaponry—such as claws or fangs—and that injury by these weapons reduces fitness by  $c$ .

There are two possible behaviors: (1) **fight** or (2) **display**. Fight means attack with weaponry, possibly inflicting injury on the competitor. Display means to compete in some way that does not inflict injury. Natural examples include wrestling, dancing, and low-intensity fighting.

Now the theme in game theory is always that behavior has fitness consequences, but behavior isn't what evolves. Instead **strategies** evolve. A strategy is a program that an animal uses to regulate its behavior. In this model, we will consider two strategies, each mapped unconditionally to one of the two possible behaviors. The first is **Hawk**, which always fights. The second is **Dove**, which always displays. If Dove is attacked with weapons, it retreats to avoid injury.

In each contest, two individuals meet and either can be Hawk or Dove. So there are three kinds of contests: Hawk vs Hawk, Hawk vs Dove, and Dove vs Dove.

When Hawk meets Hawk, they both fight. One of them wins at random and the other is injured. The winner gains  $v$ . The loser loses  $c$ . So each Hawk has a half chance of each of these outcomes, for an expected fitness change of  $\frac{1}{2}v - \frac{1}{2}c$ .

When Hawk meets Dove, the Hawk fights, the Dove retreats. So the Hawk gains  $v$ . And the Dove gains nothing and loses nothing.

When Dove meets Dove, they both display. One of them eventually wins, gaining  $v$ , but the loser is not injured. So the average payoff is  $\frac{1}{2}v$ .

Now for Hawk-Hawk and Dove-Dove contests, I wrote above that the winner is random. What does this mean? In the chapter on probability, I explained that all that “random” ever means is that we lack information required to determine the outcome. It does not mean that the result is not deterministic. In this case, of course differences in fighting ability influence which competitor wins the contest. It's just that none of the information in the model tells us about fighting ability. That is all that “random” means. And since which individual is focal is arbitrary, the probability of winning must be symmetric and so it is one-half. In later modifications of the model, this will change.

To summarize the consequences of these contests, let's make a simple table:

	Hawk	Dove
Hawk	$\frac{1}{2}v - \frac{1}{2}c$	$v$
Dove	$0$	$\frac{1}{2}v$

The rows are a focal individual's behavior. The columns are her opponent's behavior. The cells in the table are fitness changes for the focal individual.

Our goal is to figure out what evolution does in the long term, given the assumptions above. It's easier if we break this down into some simple questions. First, can Hawk or Dove ever be **evolutionarily stable**? This means that other strategies cannot invade, when either strategy is common. Second, are there any

stable mixes of Hawk and Dove? Third, what would be best for the population growth rate, and when is it different what evolves?

**Evolutionary stability of Hawk.** Suppose the only strategy in the population in Hawk. Then a mutant Dove appears. Will the proportion of Doves tend to increase? To rigorously answer that question, we'd need to specify a lot more than we have so far—population regulation, mutation patterns, the nature of heritability, and so on. But heuristically, if we confine our question to what natural selection would do, if it got its way, we can make progress without any additional assumptions. We just assert that higher fitness will increase the representation of that strategy over time, if offspring resemble parents in strategy.

It's not that we should believe that natural selection always gets its way. We know it does not. But this is what models are for: Only in a model can we remove other forces and learn about each force individually. This is very valuable. But it will be equally valuable later to add other forces. But only slowly so that we learn something other than "things are complicated." We knew that already.

In this case, we first note that when Hawk is common, almost every Hawk fights another Hawk. So the average fitness change of a Hawk is just  $(v - c)/2$ . We are ignoring a Hawk who meets the first mutant Dove. But if the population is large, this is fine. The single Hawk who gets a different payoff ( $v$  for victory!) will have almost no impact on the average fitness of the Hawk strategy.

What about the rare mutant Dove? The Dove meets a Hawk for sure. And so it loses, but without injury. So no change in fitness (zero).

Now we ask which of these expected fitness changes is greater. If the expected fitness of Hawk is greater, it can resist invasion by rare Doves. This is the condition for Hawk to be evolutionarily stable:

$$\underbrace{\frac{v - c}{2}}_{\text{Hawk-vs-Hawk}} > \underbrace{0}_{\text{Dove-vs-Hawk}}$$

Before simplifying this condition, pause to make sure you understand what it means. For a strategy to be evolutionarily stable, it must be able to live with itself well enough that a rare alternative strategy cannot do better. So the left side asks how well Hawk lives with itself (Hawk-vs-Hawk), while the right side asks how well the rare Dove does against a Hawk. This is the basic logic of evolutionary stability.

Okay we can simplify the condition above to:

$$v > c$$

This means: If the value of the resource ( $v$ ) exceeds the cost of injury ( $c$ ), rare Doves cannot invade a population of Hawks. But if  $v < c$ , because the natural weaponry is dangerous, then Doves can increase when rare.

**Evolutionary stability of Dove.** So now we also have to ask about Doves. When can a population of Doves resist invasion by a rare Hawk? The same logic as before gives us this condition for Dove to be evolutionarily stable:

$$\underbrace{\frac{v}{2}}_{\text{Dove-vs-Dove}} > \underbrace{v}_{\text{Hawk-vs-Dove}}$$

This condition is satisfied only when  $v < 0$ . That would make the model silly, since it implies the resource reduces fitness. So we conclude that Doves are never evolutionarily stable. 5

**Hawk-Dove mixes.** So far we've learned that Dove is never evolutionarily stable. Hawk can be evolutionarily stable when  $v > c$ . But if  $v < c$ , then neither strategy is stable. So something must happen as the two strategies interact in a mixed population. Let's figure out what. 10

Now we have to imagine that the population has some Hawks and some Doves. We still want to avoid actually modeling all of the population dynamics. So we'll just let  $p$  be the proportion of the population that is Hawks. Then  $1 - p$  is the proportion that is Doves. The total number of individuals might increase or decrease each generation. But each possible value of  $p$ , between zero and one, will either tend to increase in the next generation, decrease, or stay the same. 15

Consider for example when  $p = 1$ . This is the Hawk world we already analyzed. We asked if a rare Dove would increase, because it had higher fitness than a Hawk. The answer was yes, if  $v < c$ . So when  $p = 1$ , the proportion of Hawks tends to decrease, but only if  $v < c$ . And for  $p = 0$ , a rare Hawk can always increase. So  $p$  tends to increase when  $p = 0$ . 20

What about all of the values between zero and one? To answer what happens for other values of  $p$ , we need to write expressions for the average fitness change in a mixed population. This requires nothing more than taking an average. Conceptually, the average fitness change of Hawk in a mixed population of Hawks and Doves is: 25

$$\begin{aligned} & (\text{probability meet a Hawk})(\text{Hawk-Hawk fitness change}) \\ & + (\text{probability meet a Dove})(\text{Hawk-Dove fitness change}) \end{aligned}$$

Using the definition of  $p$  and the fitness changes in the old table from before, this becomes:

$$p \frac{v - c}{2} + (1 - p)v$$

Similarly, we get an expression for the average fitness change of Doves:

$$p(0) + (1 - p)\frac{v}{2}$$

Now we could ask for any specific value of  $p$ , say  $p = 0.5$ , whether selection tends to increase or decrease the proportion of Hawks. But instead of plugging in 30

numerical values, we should just solve for all the values of  $p$  for which  $p$  tends to increase. Like this:

$$\underbrace{p \frac{v-c}{2} + (1-p)v}_{\text{Hawk fitness}} > \underbrace{p(0) + (1-p)\frac{v}{2}}_{\text{Dove fitness}}$$

- Simplifying, Hawk tends to increase when  $p < v/c$ . If we flip the question around and ask when Dove tends to increase, we get the opposite result:  $p > v/c$ . This implies that when  $p = v/c$ , natural selection does not favor either strategy. This is an evolutionarily stable mix of Hawk and Dove.

**What maximizes population growth?** We've figured out now what natural selection will do in the long run. If  $v > c$ , it's all Hawks, because injury isn't sufficient to deter aggression. If instead  $v < c$ , then the population evolves to a mix of Hawks and Doves at which  $v/c$  of the population is Hawks and  $1 - v/c$  is Doves. So the larger  $c$  is relative to  $v$ , the more Doves you get. So this seems to support to idea that non-lethal contests evolve because they

But we have one more question to address: What mix of Hawks and Doves would maximize the population growth rate? We can't answer this in a detailed sense, not until we specify more about the demography and population regulation. But we can ask which value of  $p$  maximizes average fitness. In this model, average fitness means average offspring. In a very wide range of ecologies, more offspring means more population growth (but not always). So let's figure it out.

Now the orthodox way to figure this out is to define an expression for the average fitness in the population, as a function of  $p$ . Then we can take the derivative with respect to  $p$  and solve for any maxima (and minima). But there is a more intuitive way.

At one extreme,  $p = 1$ , the average fitness is  $(v - c)/2$ . At the other extreme,  $p = 0$ , the average fitness is  $v/2$ . Since  $v - c < v$  as long as  $c > 0$ , the Dove population has higher average fitness than the Hawk population. Okay, so let's start at  $p = 0$  and ask if adding a small number of Hawks increases the average fitness in the population. Consider the first Hawk. It meets a single Dove. The Hawk gets  $v$ , and the Dove gets 0. The average fitness for these two individuals is  $v/2$ . But that's just what two Doves get on average as well. So the first Hawk makes no difference at all to mean fitness. If there is a second Hawk, and it meets the first Hawk, there is now injury. And we know that injury reduces mean fitness. So there is no internal  $p$  that maximizes population growth. A population Doves is best for the species. In hindsight this may be intuitive, because Doves divide the resource without wasting any of it on injury.

Okay, now here's the calculus proof. The mean fitness change in a population with  $p$  proportion Hawks is:

$$\underbrace{p \left( p \frac{v-c}{2} + (1-p)v \right)}_{\text{Hawks}} + \underbrace{(1-p) \left( p(0) + (1-p)\frac{v}{2} \right)}_{\text{Doves}}$$

The derivative of this with respect to  $p$  is remarkably simple:

$$-cp$$

This is always decreasing in  $p$ , so more Hawks only decreases population growth.

**Summary.** The Hawk-Dove game is also known as the **Chicken** game. It is mathematically similar and the conclusions are qualitatively the same. But the motivating story is instead two foolish young men racing cars towards a cliff or wall. The first driver to swerve loses. But if neither swerves they both suffer some injury, presumably fatal. This version of the game makes it clear that the Hawk-Dove/Chicken game favors behaving in the opposite way from one's opponent. If you knew that your opponent would fight, you should not. If you knew that your opponent would not fight, you should. If strategies are heritable, this tends to result in mixed populations that do not maximize general welfare.

*Practice: Hawk-Dove with display costs.* It doesn't make sense that Dove's display has no fitness cost. If nothing else, it costs time and energy. Let  $d$  be the cost of display. Assume that Dove pays this cost whenever it meets another Dove, whether it wins the resource or not, but not when it retreats from a Hawk. Analyze this new version of the game.

*Practice: Escalating display costs.* Suppose the modified game with display costs  $d$ . Now consider a mutant Dove with a larger display cost  $D > d$  who always wins against a regular Dove who pays only  $d$ . Reanalyze the game.

### *Retaliation doesn't pay*

What happens if Doves fight back, if they displayed but always retaliated against aggression? The intuition is that retaliation could keep Hawks from invading and then peace would reign. But it turns out this doesn't make much difference. Let's figure out why.

Suppose the original Hawk-Dove game as described in the previous section. But now introduce a third strategy **Retaliator**. Retaliator displays but fights if its opponent fights. This means it plays like Dove with itself and with a Dove. It plays like Hawk with a Hawk.

Let's answer three questions. First, can Retaliator ever be evolutionarily stable? Second, can Retaliator invade a population of Hawks and Doves? Third, what happens in the long run?

**Retaliator evolutionary stability.** When Retaliator is common, its average fitness change is  $v/2$ , because it plays like Dove with itself. A rare Hawk will always make a fight, getting  $(v - c)/2$ . This is always worse than  $v/2$ . So Hawk cannot invade.

The tricky element is Doves. A rare Dove also gets  $v/2$ . So natural selection on behavior doesn't keep the Dove out. But it doesn't favor it either. So other forces—mutation, developmental costs—will determine what happens.

Retaliator is never evolutionarily stable.

**Retaliator invasion.** There are two scenarios to consider. If  $v > c$ , then the first Retaliator encounters a population of Hawks. If  $v < c$ , then the first Retaliator encounters instead a mixture of Hawks and Doves. Let's analyze each.

In a population of Hawks, the first Retaliator also gets into a fight, earning the same  $(v - c)/2$  as every Hawk. So Hawk is not stable against Retaliator. And if a small number of Retaliators is present, whenever they interact with one another they do better. So Retaliator can increase when rare, in a world of Hawks.

In a mixed population of Hawks and Doves, the average fitness change is:

$$\frac{v}{c}(0) + \left(1 - \frac{v}{c}\right) \frac{v}{2}$$

This is the average fitness of Doves, who encounter another Dove  $1 - v/c$  of the time and get zero otherwise. But since the fitness of a Hawk must be the same at evolutionary stability, the expression above is correct for Hawks too.

A rare Retaliator needs to do better than this. A rare Retaliator earns:

$$\frac{v}{c} \frac{v - c}{2} + \left(1 - \frac{v}{c}\right) \frac{v}{2}$$

Simplifying, this is greater than  $(1 - v/c)v/2$  only when  $v > c$ . That's the condition for Hawk to exclude Dove. So if Doves are present, Retaliator cannot invade.

If Doves are absent, Retaliator can invade and it will increase, gradually replacing Hawk, because in a mixed population of  $p$  Hawk and  $1 - p$  Retaliator, Hawk always gets in to a fight so earns  $(v - c)/2$  still. But Retaliator gets:

$$p \frac{v - c}{2} + (1 - p) \frac{v}{2}$$

Retaliator will replace Hawk.

**The long run.** In the long run, what happens will still depend upon Doves. There are two scenarios to consider. First  $v < c$ , in which case Retaliator cannot invade at all. Second  $v > c$ , in which case Retaliator invades a pure Hawk population and replaces Hawk. But since Dove can invade a pure population of Retaliator, we need to think hard now about what happens next.

Other forces that we haven't modeled will determine whether or not Dove increases in a mixed population of Retaliators and Doves. In the model as we have it, the proportion of Dove will just drift around. If Dove ever becomes sufficiently common, it will allow Hawks to invade. Let  $q$  be the proportion the population



that is Retaliator. The average fitness in a mix of Retaliator and Dove is always  $v/2$ , because neither strategy fights. But the average fitness of a rare Hawk will be:

$$q \frac{v-c}{2} + (1-q)v$$

Hawk can invade if this is greater than  $v/2$ , which simplifies to:

$$q < \frac{v}{v+c}$$

So in the long run, if other forces don't prevent this threshold being crossed, Hawks will invade, the population will evolve again towards a mix of Hawk and Retaliator, and the pure Retaliator, and then a mix of Doves and Retaliator, and so on until the end of time.

Retaliator is a poor police officer, because it tolerates lazy Doves.

*Practice: Retaliator errors.* Suppose Retaliator sometimes makes an implementation error and judges a display as an attack. This results in a fight, when paired with Hawk or another Retaliator, or a flawless victory, when paired with a Dove. Let  $x$  be the probability of an error. When can Retaliator be evolutionarily stable?

## Assessment

Typically there are cues that an animal can use to assess the fighting ability of its opponent. Suppose for example that the competitors vary in size and that size influences fighting ability. Since which individual is focal is arbitrary, there is a half chance that the focal individual is larger than its opponent. Let  $x > 0.5$  be the probability that the larger individual wins a fight.

Now consider a strategy **Assessor** that uses size asymmetry to decide whether or not it fights or displays. If Assessor is larger, it fights. If it is smaller, it displays. When Assessor is common, its average fitness change is  $v/2$ , because half the time it will be larger and get  $v$ , because its opponent will retreat. The other half of the time, it will be smaller, so it retreats and earns zero. So Assessor manages to get the socially optimal payoff when common. That's a good start.

Can Hawk invade Assessor? A rare Hawk gets:

$$\underbrace{\frac{1}{2}v}_{\text{Hawk larger}} + \underbrace{\frac{1}{2}((1-x)v - xc)}_{\text{Hawk smaller}}$$

Assessor's  $v/2$  is larger than this when  $x > v/(v+c)$ . So for example if  $v = 2$  and  $c = 1$ , the fighting advantage of body size needs to be greater than  $2/3$ . But Assessor can potentially be evolutionarily stable against Hawk, even if  $v > c$ . It pays to choose your fights, if there is a reliable cue of fighting ability.

Can Dove invade Assessor? A rare Dove gets:

$$\underbrace{\frac{1}{2} \frac{v}{2}}_{\text{Dove larger}} + \underbrace{\frac{1}{2}(0)}_{\text{Dove smaller}}$$

This is also always smaller than  $v/2$ . Assessor is evolutionarily stable against Dove.

Can Assessor invade a mixed population of Hawks and Doves? As before, at the mixed equilibrium, Hawks and Doves earn average fitness  $(1 - v/c)v/2$ . A

5 rare Assessor earns:

$$\underbrace{\frac{1}{2} \left( \frac{v}{c}(xv - (1-x)c) + (1 - v/c)v \right)}_{\text{Assessor larger}} + \underbrace{\frac{1}{2} \left( \frac{v}{c}(0) + (1 - v/c)\frac{v}{2} \right)}_{\text{Assessor smaller}}$$

Invasion requires that this is larger than  $(1 - v/c)v/2$ . You can start reducing and find that this is true whenever  $x > 1/2$ .

In reality, size is a continuous cue and so the animal needs some function that maps a perceived size difference to  $x$ . For competitors of similar sizes,  $x$  may not be  
 10 large enough (size may not be predictive of fighting ability), and so other strategies might take over.

### *Conventions in conflict*

A **convention** is a cue that is not associated with fighting ability but that is used to decide contests. A famous convention is which animal arrives first. Suppose  
 15 there a strategy that fights when it arrives first at the resource and displays when it arrives second. When common, this strategy earns  $v/2$ , because half the time the focal individual arrives first and the opponent lets it have the resource. So the convention is just a way of randomly allocating the resource without conflict.

Can such a strategy be evolutionarily stable? A rare Hawk gets:

$$\underbrace{\frac{1}{2}v}_{\text{Hawk first}} + \underbrace{\frac{1}{2} \frac{v-c}{2}}_{\text{Hawk second}}$$

20 This is always worse than  $v/2$ . Hawk cannot invade. A rare Dove gets:

$$\underbrace{\frac{1}{2} \frac{v}{2}}_{\text{Dove first}} + \underbrace{\frac{1}{2}v}_{\text{Dove second}}$$

This is also always less than  $v/2$ . So the conventional strategy can be evolutionarily stable, at least against simple Hawk and Dove strategies.

Can it also invade? At the mixed Hawk-Dove equilibrium, the convention of fighting when first earns:

$$\underbrace{\frac{1}{2} \left( \frac{v}{c} \frac{v-c}{2} + (1-v/c)v \right)}_{\text{Focal first}} + \underbrace{\frac{1}{2} \left( \frac{v}{c}(0) + (1-v/c)\frac{v}{2} \right)}_{\text{Focal second}}$$

If you reduce this, you'll find it is equal to  $(1-v/c)v/2$ . So it gets the same payoff as a Hawk or Dove. The convention can enter the population, and if it becomes more common, it does better against itself than the Hawks and Doves do against it, so it will increase. The convention can evolve and be stable, despite the fact that it has nothing to do with fighting ability. Any other convention would work just as well.

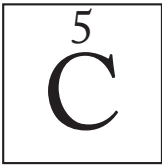
5

### *Summary*

x

10





## *The Evolution of Cooperation*

Hiroo Onoda (小野田寛郎) was born in 1922 in the Empire of Japan. As a boy, he herded cattle. In 1940 he joined the army. In 1944, he traveled to a small island in the Philippines with orders to defend the island against the United States. The Empire of Japan surrendered the next year, but no one could find Onoda to tell him the bad news.

5

Onoda surrendered 29 years later. His 29 years of persistence had no effect on the war. No effect aside from his own long deprivation. But no one saw it that way. He returned to Japan a hero, a symbol of dedication, honor, and altruism.

Human societies make role models of people like Onoda, because they recognize the conflict between the interests of individuals and groups. The groups are not always so large as the Empire of Japan. And the individuals are not always so capable as Hiroo Onoda. But conflict is the context in which cooperation happens. Typically each individual also has ways to achieve its goals that do not promote the goals of others. This creates conflict between the interests of individuals and the interests of the groups and societies they are members of.

10

15

In this chapter, you'll meet simple game theoretic models of cooperation. You'll learn how to analyze these models to understand precisely how to express the conflict between individuals and groups. And then you'll study a general mechanism that helps cooperation to evolve. In later chapters, the problem of cooperation continues. But it gets much more complex.

20

### *Pure conflict: The prisoner's dilemma*

*Individual costs can undermine the production of common benefits.*

Cooperation is a messy category. So it's useful to study a simple, extreme example. Then we'll consider less simple, less extreme cases.

25

Suppose pairs of individuals share a common territory. Each individual can choose to patrol the territory. If either individual patrols, a common benefit  $b$  is produced for both individuals. If only one individual patrols, it costs that individual  $c$  units of fitness. If both patrol, they split the cost. Each pays only  $c/2$ . This is the payoff matrix for this game:

30

	Rest	Patrol
Rest	0	$b$
Patrol	$b - c$	$b - c/2$

Now analyze the game's evolutionary dynamics. Patrol is evolutionarily stable if  $b - c/2 > b$ , which is never true when  $c > 0$ . A rare Rest can always invade. Rest is evolutionarily stable if  $0 > b - c$ , which implies  $c > b$ . Suppose  $b > c$ . Then this game has dynamics like the Hawk-Dove game: there will be a stable mix of Rest and Patrol at equilibrium. To be precise, let  $p$  be the proportion of Patrol. Then the equilibrium is found where:

$$p(b - c/2) + (1 - p)(b - c) = pb + (1 - p)0$$

Solve for  $p$ :

$$p = \frac{b - c}{b - c/2}$$

As long as  $b > c$ , this point implies a mix of Rest (Hawk) and Patrol (Dove).

But now suppose  $c > b$ . If the cost of patrolling alone is greater than the benefit, no one will patrol. The population does best when someone patrols. But evolution won't sustain it.

When  $c > b$ , this game is an example of the **prisoner's dilemma**. In a prisoner's dilemma, non-cooperation is favored, no matter how common cooperation is in the population. Biologists and social scientists study the prisoner's dilemma a lot. Maybe too much. But they study it because it is an extreme example of the basic puzzle of cooperation: How can animals sustain cooperation when costs are private while benefits are shared? In a prisoner's dilemma, the interests of the individual and group are completely opposed. Each individual does better by resting. But the group does better when someone cooperates (patrols).

This dilemma may seem obvious to you. But there were many scholars in both biology and the social sciences who did not realize it. In the early 20th century, it was common for biologists to think that animals were designed for the good of their social groups and for social scientists to believe that people cooperate just because cooperation produces shared benefits.<sup>1</sup> Simple games like the prisoner's dilemma taught us that it isn't so simple.

### *Partial conflict: Coordination*

*The interests of individuals and groups are not always perfectly in opposition. But risk and mis-coordination are other reasons it can be hard to achieve cooperation.*

Suppose a pair of animals who can choose to search for food that is easy but low quality or instead risky but high quality. The easy option always produces a

<sup>1</sup>For example, the framing of Mancur Olson's *The Logic of Collective Action* (1965) seems weird now. But Olson was writing for an audience that had not thought deeply about the conflict between individual and group interests.

payoff  $b$ , regardless of what the other individual does. But the risky option only produces  $B > b$  when both individuals choose it. Otherwise it returns on average zero. This is the payoff matrix:

	Safe	Risky
Safe	$b$	$b$
Risky	0	$B$

The population will be better off if everyone pursues the risky option. Now what happens when Safe is common? Can Risky invade? No, because  $b > 0$ . Can Risky be stable when common? Yes, because  $B > b$ . So Risky is evolutionarily stable, but it cannot invade. This is a different kind of conflict than in the prisoner's dilemma, but just as serious. Everyone would be better off, if they could agree to choose the risky option. But there is no path for evolution to take, in this simple example. And since Safe always produces  $b$ , it is hard to evolve away from. The game above is a classic, and it is usually called the **stag hunt**.

Another kind of cooperation dilemma is **coordination**. Suppose there are two behaviors, and to receive a non-zero payoff, an individual must match the behavior of its partner. But the different behaviors produce different payoffs. Like this:

	A	B
A	$b$	0
B	0	$B$

Again assume  $B > b$ . The most important thing in this game is to match your partner. But both behaviors are evolutionarily stable. So if a population is fixed on A, evolution cannot evolve to B, even though it would make everyone better off. This is another type of cooperation conflict, one in which the dilemma is how to get everyone to coordinate to choose the better option.

In both of the games above, both options are evolutionarily stable. This implies that there is some proportion of each strategy at which the fitness of each is equal, an equilibrium. But this equilibrium will be unstable. If the population moves to one side or the other, natural selection will push towards an extreme of one behavior or the other dominating the population. This is the opposite dynamic of what we saw in the previous chapter, with Hawk-Dove. In Hawk-Dove, natural selection pushed the population to a stable mix. In these games, natural selection will instead push the population to the extremes.

Still the equilibrium is interesting, because it defines the **basin of attraction** for each strategy. A basin of attraction is the range of proportions that favor a strategy. For the stag hunt, we find the unstable equilibrium that defines the basins of attraction by setting the average fitness of both strategies equal and solving for the proportions that make the equality true. Let  $p$  be the proportion choosing Risky. This implies the equilibrium is found where:

$$pB + (1 - p)0 = b$$

This is true only when  $p = b/B$ . When  $p < b/B$ , natural selection favors the safe option. When  $p > b/B$ , natural selection favors the risky option. So if  $b$  is small enough, then it might not be so hard to for other forces to push a population over the equilibrium so it can reach the fitness-enhancing behavior. But if for example  
 5  $b = 1$  and  $B = 2$ , then the unstable point is right in the middle and populations likely get stuck at the inferior behavior, even though the alternative is twice as good.

Can you find the unstable equilibrium in the coordination game?

### *Positive assortment*

10 *A general solution to the problem of evolving cooperation is when cooperative strategies tend to encounter one another more often than chance.*

Think again of the patrol game. Suppose pairs were always comprised of the same strategy. So pairs of Rest and pairs of Patrol, but no mixed pairs. So Rest always gets zero, and Patrol always gets  $b - c/2$ . Cooperation would always evolve.  
 15 If evolution can find a way for cooperative strategies to **positively assort**, this would help those strategies to evolve.

There are several mechanisms that biologists and social scientists think can produce positive assortment of cooperative strategies. We'll model some of them in later chapters. For now, let's think of assortment in purely statistical terms, without  
 20 specifying the mechanism that produces it. Specifically, suppose the probability that an individual with the strategy Patrol is paired with another Patrol is:

$$\Pr(\text{Patrol}|\text{Patrol}) = r + (1 - r)p$$

where  $r$  is the probability of assortment and  $p$  is the proportion of the population that is Patrol. So when  $r = 1$ , the above evaluates to  $\Pr(\text{Patrol}|\text{Patrol}) = 1$ . Perfect assortment. But for  $r < 1$ , the amount of assortment depends upon how common  
 25 the strategy is in the population. The corresponding chance Patrol meets Rest is:

$$\Pr(\text{Rest}|\text{Patrol}) = (1 - r)(1 - p)$$

Patrol meets rest when assortment fails, so no  $r$  term, and then meets Rest with chance  $1 - p$ . These two probabilities are the statistical assortment model. They let us explore the general implications of assortment, without worrying yet about the mechanisms that produce it. That comes later.

30 So let's reconsider when Patrol can be evolutionarily stable. When common, the fitness of Patrol is  $b - c/2$ , because it meets itself, assortment or not. What is the fitness of a rare Rest? Rest will experience assortment, even when rare. It's fitness is:

$$\underbrace{(r + (1 - r)(1 - p))0}_{\text{Rest}|\text{Rest}} + \underbrace{(1 - r)p b}_{\text{Patrol}|\text{Rest}}$$



When Patrol is common,  $p \approx 1$ , so the above simplifies to:

$$(1 - r)b$$

And so the condition for Patrol to be evolutionarily stable is:

$$b - c/2 > (1 - r)b$$

Let's rearrange this so it is a condition in terms of  $r$ :

$$r > \frac{c}{2b}$$

This can be satisfied, if  $r$  is large enough, even when  $c > b$  (a prisoner's dilemma). For example, if  $c = 3$  and  $b = 2$ , then this requires  $r > 3/4$ . 5

Assortment also changes the condition for Patrol to invade. When Rest is common, it always gets zero. A rare Patrol gets:

$$\underbrace{(r + (1 - r)p)(b - c/2)}_{\text{Patrol|Patrol}} + \underbrace{(1 - r)(1 - p)(b - c)}_{\text{Patrol|Rest}}$$

Now  $p \approx 0$ , so this simplifies to:

$$r(b - c/2) + (1 - r)(b - c)$$

For Patrol to invade a population, we require:

$$r(b - c/2) + (1 - r)(b - c) > 0$$

This condition simplifies to  $r > 2(1 - b/c)$ . This is not the same condition that Patrol needs for stability. But again, this can be satisfied when  $c > b$ . But it might require implausible amounts of assortment, so don't start celebrating. 10

Assortment behaves in different ways in different games. It changes stability conditions, as well as basins of attraction. Then details of the evolutionary process, such as the genetic architecture of the strategies, or whether strategies are learned rather than genetic, can influence whether cooperative strategies succeed in the long run. 15

### *Cooperation in large groups*

[need a section on how cooperation in groups larger than pairs influences the different games] 20



# 6 R

## *The Evolution of Relationships*

The Portuguese man o' war (*Physalia physalis*) is neither Portuguese nor a warship. It looks like a jellyfish, but it's not. It's not even an animal. Instead it's a colony of individual animals. Each individual has a specialized job, fusing together in cooperation with the others to sail and sting their way through the ocean. Like a pirate ship made of jelly.

Each man o' war colony can last for a year or more. So the evolutionary biologist must ask, how did this cooperation evolve and how is it stable against cheaters? It is the same kind of question we have to ask about more familiar relationships among animals, like cooperating lions or breeding penguins. Or actual pirates.

In all of these examples, the evolutionary game to analyze is unlike those in the previous chapters. In the previous chapters, every example was a group of individuals who met, interacted once, and then never met again. There are animal societies like that. But primate groups and man o' war colonies are not like that. They persist and the same individuals interact many times in their lifetimes.

To model the evolution of these relationships, we need to learn how to model **repeated games**. In a repeated game, individuals form groups, interact potentially many times, and use past experience to modify group membership or adjust their own behavior towards other members of the group. The question we want to ask in this framework is: Which strategies lead to the evolution of lasting cooperative relationships?

Nearly all important game theory, as applied to humans and other animal societies, concerns repeated games. And unfortunately simple games, once repeated, are no longer so simple. The longer individuals interact, the larger the possible number of strategies. And even rare errors in the implementation or the interpretation of behavior can have huge effects. But we'll begin as always with something simple and understandable, before layering on the necessary complexity.

### *The iterated prisoner's dilemma*

*In a repeated game with uncertain end, reciprocity can be evolutionarily stable against non-cooperation. But it is never stable against other cooperative strategies.*

The most studied repeated game is the **iterated prisoner's dilemma**. The simple game at its heart an additive prisoner's dilemma. Suppose in a pair of individuals each has the opportunity to help the other. Helping costs the helper  $c$ . But it produces  $b > c$  for the helped. This is the resulting payoff matrix:

5

	Cooperate	Not
Cooperate	$b - c$	$-c$
Not	$b$	0

In this matrix, mutual cooperation is beneficial. But each individual still does better by switching to non-cooperate, because  $b > b - c$  (left column) and  $0 > -c$  (right column).

10

Now we embed this simple game in some repeating structure, so that the same pair of individuals interact more than once. There are many ways to assume repeat interaction. The structure of repetition itself influences the evolution of behavior. Some structures make it harder for cooperation to evolve. For example, a defined end point to interaction is unhelpful. If for example the pair interacts exactly 10 times, it can be as if the game was not repeated at all. Whatever the pair does in the first 9 plays, the 10th time neither individual has any (evolutionary) incentive to cooperate. It's just like a one-shot game at that point. Then it follows that in the 9th play, it is also a one-shot game, because neither player expects (evolutionarily) the other to cooperate in the 10th play. So there's no incentive to cooperate in the 9th either. And so on, all the way back to the first play.

20

If there is uncertainty about the end, things can be different. So to repeat this game, we allow a pair of individuals to play once and then interact again with probability  $w$ . And then if they do interact again, they get the same chance  $w$  again of interacting again. And so on. This  $w$  is a constant chance of interacting again after each interaction, and it does not depend upon whether the individuals cooperate or not. Of course you are ready to object. It makes sense that the persistence of the relationship depends upon how the individuals treat one another. But we start here, analyze the consequences, and then consider something more dynamic in a later section.

30

When the pair has a constant chance  $w$  of interacting again after each game, it means that some pairs will interact only once, while others will interact many times. The probability of interacting only once is  $1 - w$ . The probability of interacting twice is  $w(1 - w)$ . The chance of interacting three times is  $w^2(1 - w)$ . And so on out to infinity. Yes, the individuals will die of natural causes before that. But unless  $w$  is very close to 1, or the time between interactions is very long, that isn't an event we need to worry about. If selection is strong, then the entire distribution of pair durations could matter (as explained in **1F**). But if selection is weak, then the average duration will give us a useful approximation. And this is what nearly everyone does, use the average. Later we'll consider what this misses, however.

The average number of interactions is  $1/(1-w)$ . There are several ways to calculate this average. You could proceed in the orthodox manner of just defining the expected duration using the probability distribution above. Let  $R$  be the average length of a relationship. Then  $R$  must satisfy:

$$R = \sum_{i=1}^{\infty} w^{i-1}(1-w)(i)$$

The probability  $w^{i-1}(1-w)$  is the chance of persisting exactly  $i$  interactions. So the above just takes each  $i$  value, multiplies it (weights it) by the chance it happens, and then adds up all these values. That's the average. The expression above is an infinite series. But it's geometric, so it closes easily. If you don't know how to do that, it just means you are a normal and healthy person. I show you how in the box below. Don't worry. It's safe.

A much easier way to derive the mean  $R = 1/(1-w)$  is to realize that since the chance the interaction continues is always the same, no matter how long a pair persists, the expected number of future interactions never changes. It seems weird, but it's true. A pair just meeting for the first time expects  $1/(1-w)$  interactions. And so does a pair that has already interacted 100 times. They expect  $1/(1-w)$  future interactions. This implies:

$$R = 1 + wR$$

*Closing the geometric series.* We want to take this infinite series and find a finite expression:

$$R = \sum_{i=1}^{\infty} w^{i-1}(1-w)(i) = (1-w)(1) + w(1-w)2 + w^2(1-w)3 + w^3(1-w)4 + \dots$$

The trick with a **geometric infinite series** is to recognize that it is always embedded in itself, because every step of the series has the same future expected value. So what we do to close it is factor it so we reveal this fact. Then it closes almost instantly. Let's factor  $w$  out of every term after the first:

$$R = (1-w)(1) + w \underbrace{(1-w)2 + w(1-w)3 + w^2(1-w)4 + \dots}_{Q}$$

Now the term labeled  $Q$  is not equal to  $R$ , but it is close. Each term just has 1 added to it. Like this:

$$R = (1-w)(1) + w \underbrace{((1-w)(1+1) + w(1-w)(2+1) + w^2(1-w)(3+1) + \dots)}_{Q}$$

All of the those  $+1$  terms just sum to 1. Why? Because the series  $(1-w)(1) + w(1-w)(1) + w^2(1-w)(1) + \dots$  is the total probability distribution. It must equal 1. This gives us:

$$R = (1-w)(1) + w \underbrace{(1 + (1-w)1 + w(1-w)2 + w^2(1-w)3 + \dots)}_R$$

And now the sequence in parentheses is equal to  $R$ . So we have  $R = (1-w)(1) + w(1+R) = 1 + wR$ . Solving yields  $R = 1/(1-w)$ .

Why? Because whatever value  $R$  takes, it must satisfy the equation above so that it has the property that the expectation never changes as the pair persists. It says that the pair interacts once (1) and then has chance  $w$  of having the same expectation that they just had. So solving the above for  $R$  and get  $R = 1/(1 - w)$ .

5 Unless we also introduce strategies that use information about past behavior, repeating the game makes no qualitative difference. For example, let ALLC be a strategy that always cooperates. In a pair of ALLC individuals, each will earn  $b - c$  each time they interact. So the expected fitness change is  $R(b - c)$ . But a rare individual who never cooperates, call this NO-C, gets  $b$  every time it interacts with ALLC, so it earns  $Rb$ . This is still a prisoner's dilemma and cooperation is  
10 not stable.

But consider a simple contingent strategy that uses past behavior. Suppose a strategy that begins by cooperating and then copies its opponent's behavior in every interaction afterwards. This strategy is usually called Tit-for-Tat (TFT). Can  
15 TFT be evolutionarily stable against NO-C? A pair of TFT individuals always cooperate with one another, because they start by cooperating and then copy one another in each interaction afterwards. So each earns  $R(b - c)$ . A rare NO-C earns  $b$  on the first interaction with TFT, but then TFT copies the non-cooperation on the next interaction. So both individuals earn zero until the pair ends. So TFT  
20 can be evolutionarily stable against NO-C if:

$$R(b - c) > b$$

This is easy to satisfy, if  $w$  is large enough. Consider  $w = 0.9$ . Then  $R = 10$ . Even if  $b - c$  is small, multiplying it by a lasting relationship can make up for the costs of cooperation. We can also express the condition above after substituting  $R = 1/(1 - w)$ , as:

$$wb > c$$

25 So if  $b/c = 2$ , then  $w$  just needs to be more than 0.5. That doesn't seem so demanding. Relationships don't even have to last that long, for reciprocity to pay.

But hang on. Tit-for-Tat is rather famous in game theory.<sup>1</sup> It reveals the basic logic of reciprocity and how it can maintain cooperation. But Tit-for-Tat itself is a bad strategy and it is never evolutionary stable. Consider a rare ALLC in  
30 a population of TFT. The rare ALLC also earns  $R(b - c)$ . So ALLC can increase. If there are enough of them, NO-C can increase as well. It's like the problem with retaliation in the chapter on conflict (4H). Tit-for-Tat tolerates strategies that do not guard the wall.

However it seems odd that TFT and ALLC receive exactly the same payoff,  
35 in the absence of NO-C. Surely these strategies differ in other ways?

---

<sup>1</sup>Robert Axelrod wrote a book talking about how great it is.

### *Little mistakes, large consequences*

*Errors matter in repeated games. If games are repeated enough, even rare errors can influence the dynamics of relationships.*

Suppose an individual who intends to cooperate sometimes fails to do so. This is because cooperation requires action, and sometimes action goes wrong, no matter our intentions. Let  $x$  be the probability that an individual who intends to cooperation ends up non-cooperating instead. In game theory, this is called an **implementation error**. Implementation errors differentiate TFT and ALLC. Errors hurt both of these strategies. But TFT retaliates against an error by its opponent—it doesn't know it is an error. ALLC never retaliates.

To compute the expected payoffs in the presence of errors, we need to account for all the possible sequences of mistakes in every possible duration of the relationship. That sounds hard. But it is made easier by representing the problem as transitions between **intention states**. For a pair of TFT individuals, there are four possible pairs of intentions for each play of the prisoner's dilemma: (1) They both intend to cooperate (CC); (2) Partner intends to cooperate but focal does not (NC); (3) Focal intends to cooperate but partner does not (CN); (4) Neither intends to cooperate (NN). Implementation errors influence the transitions between these states. Let's represent it all with a diagram, which I explain below it.

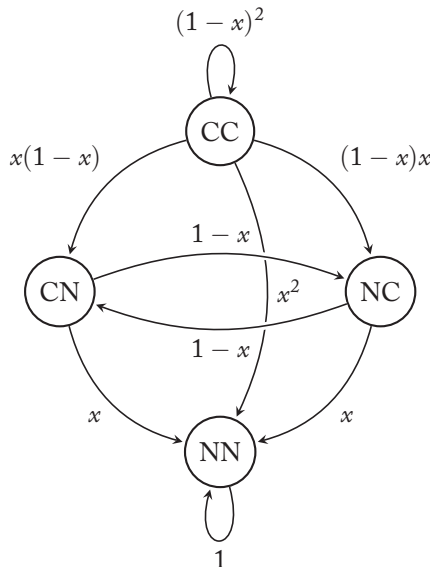


Figure  
6.1

*Two Tit-for-Tats, in a game tree, F-E-U-D-I-N-G*

The circles are the intention states and the arrows are possible transitions, labeled with probabilities. Start at the top. A pair in the CC state remains there

only if neither commits an error. The probability neither commits an error is  $(1 - x)^2$ . If both commit an error in the same play, the relationship moves to NN. This happens with probability  $x^2$ . For each state of the relationship (CC, NC, CN, NN), the same logic implies the probability of moving to any state (or staying in the same state) in the next play.

We can use this diagram to derive the expected payoff to a TFT paired with another TFT. The intuition is that we write an expression for the expected payoff at each state. We represent future payoffs by adding the other state expressions as needed. This is weird, I know. Consider the payoff for CC. Let  $V_{CC}$  be the expected payoff of a pair at the CC intention state. If neither individual commits an error, the focal TFT receives  $b - c$  and then with chance  $w$  the pair interacts again at the same CC state. So the expected future payoff is  $wV_{CC}$ . So the payoff is  $b - c + wV_{CC}$ , but only with chance  $(1 - x)^2$ , the chance neither commits an error. With the same logic, we can build the full expectation at CC and all of the other states. The box on the next page has the details.

It turns out that the expected payoff for TFT against TFT is:

$$V(\text{TFT}|\text{TFT}) = \frac{(1 - x)(b - c)}{1 - w(1 - x)}$$

Errors reduce benefits (the numerator) and the effective duration of the relationship (denominator), because once the pair enters mutual non-cooperation (NN), there is no way out. But notice that when  $x = 0$ , we recover the original  $(b - c)/(1 - w)$  payoff to sustained cooperation.

How does a rare ALLC do against TFT? Shown in the same box, ALLC receives:

$$V(\text{ALLC}|\text{TFT}) = (1 - x) \frac{(1 - xw)b - c}{1 - w}$$

This looks more complicated, but it has the same structure. Errors reduce benefits. But in this case, they do not reduce the duration of the relationship (denominator), because ALLC never intentionally non-cooperates. So ALLC never gets trapped in NN like TFT does. Sometimes people say that ALLC is forgiving. But really it has no concept of a grudge in the first place.

TFT is now evolutionarily stable against ALLC when  $(1 - x)wb < c$ . This is easier to satisfy when  $x$  is large. Errors hurt both TFT and ALLC. But TFT retaliates against ALLC's errors, while ALLC never retaliates for TFT's errors.

Errors help TFT against ALLC. But unfortunately they do the opposite with NO-C. Also in the box, I show that the condition for TFT to be stable against NO-C is  $(1 - x)wb > c$ . This is the opposite of the previous condition. Errors hurt TFT relative to NO-C, because pairs of TFT end up feuding, when errors are common. So TFT is either stable against ALLC or NO-C, but never both.

Before you read too much into this analysis, notice that we assumed that implementation errors apply only to cooperation. Things can look quite different if



**Calculating payoffs when there are implementation errors.** To calculate the expected payoff for TFT when it is common,  $V(\text{TFT}|\text{TFT})$ , we need to write an expression for the expected payoff for each pair of *intended* actions. The first is mutual cooperation, CC. The payoff is:

$$V_{\text{CC}} = \underbrace{(1-x)^2(b-c+wV_{\text{CC}})}_{\text{neither makes a mistake}} + \underbrace{x(1-x)(b+wV_{\text{CN}})}_{\text{focal fails to cooperate}} + \underbrace{(1-x)x(-c+wV_{\text{NC}})}_{\text{partner fails to cooperate}} + x^2(0)$$

On the far right, the last term is just zero, because if ever both TFT simultaneously fail to cooperate, with probability  $x^2$ , they will non-cooperate until the relationship ends. We need similar expressions for what happens at  $V_{\text{CN}}$  and  $V_{\text{NC}}$ . These are easier, because if there is no mistake the pair just moves back-and-forth from CN to NC forever. If there is a mistake, then there is mutual non-cooperation and the relationship is essentially over.

$$V_{\text{CN}} = (1-x)(-c+wV_{\text{NC}}) + x(0) \quad V_{\text{NC}} = (1-x)(b+wV_{\text{CN}}) + x(0)$$

Now we have three simultaneous equations and three unknowns,  $V_{\text{CC}}$ ,  $V_{\text{CN}}$  and  $V_{\text{NC}}$ . If you aren't going to use some software like Mathematica to do this, I suggest solving for  $V_{\text{CN}}$  and  $V_{\text{NC}}$  first, they are like mirrors of one another and not so complicated. Then solve for  $V_{\text{CC}}$ . You should make sure you know how to solve such systems in software like Mathematica however, because eventually systems of equations get too big for a sane person to waste time on doing algebra. Of course you still need to be confident with algebra, since you need to manipulate the solutions and understand them. So practicing on this three equation system might be useful. The solutions are:

$$V_{\text{CN}} = (1-x) \frac{w(1-x)b-c}{(1-w(1-x))(1+w(1-x))} \quad V_{\text{NC}} = (1-x) \frac{b-w(1-x)c}{(1-w(1-x))(1+w(1-x))}$$

$$V_{\text{CC}} = \frac{(1-x)(b-c)}{1-w(1-x)}$$

Since a pair of TFT individuals begin at CC,  $V(\text{TFT}|\text{TFT}) = V_{\text{CC}}$  and we are done. Notice that in the limit of  $x \rightarrow 0$ , we recover the original mutual cooperation payoff,  $(b-c)/(1-w)$ . To figure out when TFT is evolutionarily stable, we now need  $V(\text{NO-C}|\text{TFT})$  and  $V(\text{ALLC}|\text{TFT})$ . First:

$$V(\text{NO-C}|\text{TFT}) = (1-x)b + x(0) + w(0) = (1-x)b$$

We require  $V(\text{TFT}|\text{TFT}) > V(\text{NO-C}|\text{TFT})$ . This simplifies to  $(1-x)wb > c$ . For ALLC, we just need  $V_{\text{CC}}$  and  $V_{\text{CN}}$ , because ALLC always intends to cooperate.

$$V_{\text{CC}} = \underbrace{(1-x)^2(b-c+wV_{\text{CC}})}_{\text{no mistakes}} + \underbrace{(1-x)x(-c+wV_{\text{CC}})}_{\text{partner (TFT) error}} + \underbrace{x(1-x)(b+wV_{\text{CN}})}_{\text{focal (ALLC) error}} + \underbrace{x^2(0+wV_{\text{CN}})}_{\text{both error}}$$

$$V_{\text{CN}} = (1-x)(-c+wV_{\text{CC}}) + x(0+wV_{\text{CN}})$$

Now two equations and two unknowns. Solving for  $V_{\text{CC}}$ :

$$V_{\text{CC}} = (1-x) \frac{(1-xw)b-c}{1-w} = V(\text{ALLC}|\text{TFT})$$

We require  $V(\text{TFT}|\text{TFT}) > V(\text{ALLC}|\text{TFT})$  for TFT to be stable against ALLC. This simplifies to  $(1-x)wb < c$ , which is the *opposite* of the condition for TFT to be stable against NO-C. We can reexpress these conditions in terms of  $x$ . For TFT to resist ALLC, we require  $x > 1-c/(wb)$ . So if  $x$  is large enough, then rare ALLC cannot invade. Nice! But if  $x$  is large enough to keep out ALLC, then it is also large enough to allow rare NO-C to invade.

we also allow errors for non-cooperation. How can a non-cooperator end up accidentally cooperating? Suppose cooperation means “stinting,” limiting harvest of a scarce resource, to allow it to regenerate. If both individuals stint, then both can harvest more in the future. Non-cooperation means harvesting. But harvesting can fail, because it requires action. Maybe the resource is a pond with fish, and the non-cooperator is bad at fishing. In the practice box below, I ask you to analyze this situation.

The details of the payoffs can change things as well. The default prisoner’s dilemma is *additive*. There are no synergies when multiple individuals cooperate. This can also change things. In the second practice problem, I ask you to analyze one example.

*Practice: When cooperation means not doing.* Suppose implementation errors happen instead to non-cooperation intentions—an individual who intends to non-cooperate (N) instead cooperates (C) with chance  $y$ . Rederive the stability conditions for TFT against ALLC and NO-C. If you are feeling lucky, allow at the same time the previous chance  $x$  that intent to cooperate (C) results in non-cooperation (N). What qualitative conclusions change?

*Practice: Synergistic cooperation.* Suppose the payoff when both individuals cooperate is now  $B - c$  instead of  $b - c$ , where  $B > b > c$ . Analyze the iterated game again. Does this change any of the qualitative conclusions?

### *Learning and mistakes*

*Some strategies try to learn which behavior, cooperation or non-cooperation, results in higher rewards. These strategies can do better than TFT in the presence of implementation errors.*

Another way to think about the iterated prisoner’s dilemma, especially for animals like primates, is that individuals do not arrive with fixed strategies for how to behave. Rather they learn dynamically how to make a relationship work or whether to abandon it. Algorithmically, learning strategies are still strategies, so we don’t need any new tricks (yet) to study a simple example.

Consider a learning strategy named **Pavlov**. Pavlov has an aspiration level for the benefit it derives from an interaction. If it receives a payoff at or above that level, it keeps its behavior (cooperation or non-cooperate). Otherwise it changes behavior. Suppose the aspiration level is set to  $b - c$ , the payoff to mutual cooperation (CC). So Pavlov doesn’t change when it receives  $b - c$  or greater, which includes  $b$ , the payoff to tricking an opponent (NC). If it receives instead 0 (NN) or  $-c$  (CN), it switches behavior. So you see Pavlov is not a nice, moralistic strategy like TFT. It is just learning how to get a satisfactory payoff.

First consider how Pavlov behaves in the absence of implementation errors. If two Pavlov individuals begin by cooperating, they will receive  $b - c$  and stay with

cooperation. If they begin instead with mutual defection, they will receive 0 and switch to CC, where they'll stay. So Pavlov does well when common. That's a good start.

Now let's figure out how they do when there are implementation errors. As in the previous section, suppose  $x$  is the probability that an individual who intends to cooperate instead non-cooperates. Let's draw a finite-state diagram again, to help us understand how this calculation is built.

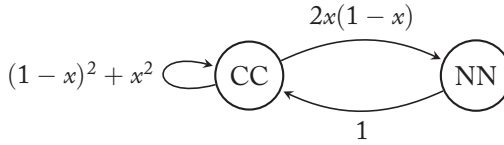


Figure  
6.2

*A pair of Pavlov individuals: Meet, play, love*

Begin with the CC node. If no mistake it made, the pair remains at CC. But they also remain at CC if *both* make a mistake. Why? These are implementation errors. To the players know when they make a mistake. So when both make a mistake, they play NN, which results in zero for both. So Pavlov decides zero is bad, and then both play C on the next turn. When only one makes a mistake, they play CN or NC. Either way, one of the Pavlov individuals receives  $b$  and the other  $-c$ . The one who received  $b$  stays at N. The one who received  $-c$  changes to N. So the next turn, they play NN. But then they immediately revert back to CC. Pavlov recovers from mistakes very quickly, because it is trying to find a good payoff. In the box on the next page, I show that when Pavlov is common, it receives on average:

$$V(\text{Pavlov}|\text{Pavlov}) = \frac{(1-x)(b-c)}{(1-w)(1+2wx(1-x))}$$

This is a bit of a mess. But notice that when  $x = 0$ , we recover  $(b-c)/(1-w)$ , the payoff to sustained cooperation. Errors reduce benefits (top), as well as the effective duration of cooperation (bottom).

We want to compare this payoff to a rare ALLC. Why? Because failing to keep out rare nice strategies was the basic flaw of TFT, remember. And errors didn't help. ALLC will never intend to non-cooperate. So we only have to consider the CC and CN nodes. In the box, I show that rare ALLC cannot invade common Pavlov. The reason is that ALLC lets Pavlov exploit at the CN node. Pavlov likes getting  $b$ , and ALLC will just keep paying  $-c$ , until it makes a mistake. Then both get zero, and Pavlov decides to cooperate again. This phenomenon highlights that Pavlov is not really a "cooperative" strategy in the typical sense. It just wants a good enough payoff,  $b-c$  or more. And it is happy to exploit a partner to get it.

Pavlov's weakness is NO-C. It will keep trying to cooperate with NO-C, bouncing from CN to NN and back, until the relationship ends. If errors are

not too common, Pavlov can be stable against rare NO-C, but it is not as easy as with TFT, which just stops cooperating with NO-C right away. See the box for the details.

### *Standing and contrition*

5 [section on standing-based strategies and Contrite-TFT]

### *Partner choice*

*Strategies that leave a non-cooperative partner to search for a new one must cope with the cost of search and the market for partners.*

*Pavlovian payoffs.* The calculations are analogous to the TFT example in the previous section. We need only two equations for this calculation:

$$\begin{aligned} V_{CC} &= (1-x)^2(b-c+wV_{CC}) + x(1-x)(b+wV_{NN}) \\ &\quad + (1-x)x(-c+wV_{NN}) + x^2(0+wV_{CC}) \\ V_{NN} &= 0 + wV_{CC} \end{aligned}$$

Solving for  $V_{CC}$  gives us:

$$V_{CC} = \frac{(1-x)(b-c)}{(1-w)(1+2wx(1-x))} = V(\text{Pavlov}|\text{Pavlov})$$

This doesn't look so nice. But we want to compare it to the payoffs for rare invading strategies. A rare ALLC visits CC and CN nodes. These are the implied expressions for each:

$$\begin{aligned} V_{CC} &= (1-x)^2(b-c+wV_{CC}) + x(1-x)(b+wV_{CN}) \\ &\quad + (1-x)x(-c+wV_{CN}) + x^2(0+wV_{CC}) \\ V_{CN} &= (1-x)(-c+wV_{CN}) + x(0+wV_{CC}) \end{aligned}$$

Solving for  $V_{CC}$ , we find:

$$V_{CC} = \frac{(1-x)(b(1-w(1-x)) - c(1-w(1-(3-x)x)))}{(1-w)(1-w(1-(3-x)x))} = V(\text{ALLC}|\text{Pavlov})$$

This expression is even worse. We ask when  $V(\text{Pavlov}|\text{Pavlov}) > V(\text{ALLC}|\text{Pavlov})$ , and the answer is whenever  $b > c$ . ALLC cannot invade. What about are NO-C? We need only consider the NC and NN nodes.

$$\begin{aligned} V_{NC} &= (1-x)(b+wV_{NN}) + x(0+wV_{NC}) \\ V_{NN} &= 0 + wV_{NC} \end{aligned}$$

Solving for  $V_{NC}$ :

$$V_{NC} = \frac{(1-x)b}{(1-w)(1+w(1-x))} = V(\text{NO-C}|\text{Pavlov})$$

These expressions are not so nice to work with, so I suggest plotting  $V(\text{NO-C}|\text{Pavlov})$  and  $V(\text{Pavlov}|\text{Pavlov})$  as functions of  $x$ . As long as errors aren't too common, and  $b$  is large enough, Pavlov can be stable against both NO-C and ALLC.

The iterated prisoner's dilemma can be interpreted in a variety of ways. We can see it as each individual having a single partner and sacking with for  $R = 1/(1-w)$  plays. Or we can see it as each individual having many partners all at once. For example, in a group of baboons, every adult interacts repeatedly with every other adult. Some pairs sustain cooperation. Others do not. When non-cooperation means not interacting, as in the default prisoner's dilemma game, the result will look like **partner choice**: An individual will cooperate with only a subset of the other adults.

As a general model of partner choice, the iterated prisoner's dilemma is badly flawed. First, often individuals cannot partner with everyone. Relationships are special and help is limited, and so animals must choose who to help. Second, pairs who sustain cooperation may want to interact more often, so to maximize the benefits of the relationship. To explore these ideas, we need some way to model dynamic partner choice.

Imagine now that individuals interact an iterated prisoner's dilemma with only one partner at a time. Each time step, each individual can decide to interact or to dissolve the pair and search for a new partner. Relationships can end for other reasons, and the  $w$  probability of persistence remains. However the relationship ends, search takes time. There is a chance  $s$  each time step that a new partner is found. So on average it takes  $S = 1/(1-s)$  periods to find a new partner. Yes, a constant chance  $s$  is odd—what if every other individual in the local community is already in a relationship? But we start easy.

Lifespan is not infinite, and this makes search costly. Time searching is time not interacting, and time is limited. Let  $\lambda$  be the probability an individual survives at each time step. So the average lifespan is  $L = 1/(1-\lambda)$ .

The next step is to calculate fitness. What is the lifetime fitness contribution of these pairwise relationships? Suppose TFT is common. Any pair of TFT individuals still earns  $R(b-c)$ . And then each TFT spends on average  $S$  turns searching for another partner. So an individual's lifespan is divided up into segments of length (on average)  $R+S$ . In a lifespan of length  $L$ , there will be (on average)  $L/(R+S)$  of these segments. In each segment, the individual receives (on average)  $R(b-c)$  increments of fitness. Whew. So in all the lifetime fitness contribution is:

$$V(\text{TFT}|\text{TFT}) = \frac{L}{R+S} R(b-c)$$

Does partner choice help TFT resist invasion by NO-C? A rare NO-C individual will interact once with each TFT partner, earning  $b$ , before the relationship is terminated by the TFT partner. Then  $S$  turns of search before the next victim is found. So NO-C earns:

$$V(\text{NO-C}|\text{TFT}) = \frac{L}{1+S} b$$

So TFT is evolutionarily stable when:

$$\frac{L}{R+S}R(b-c) > \frac{L}{1+S}b$$

This condition is not so easy to interpret, because it depends strongly upon both  $R$  and  $S$ . I'm going to plot it, which will help. But first I want to manipulate it so you can see how we can extract insight from these results.

- 5 Recall that in the absence of partner choice TFT is evolutionarily stable when  $R(b-c) > b$ . This implies that the critical value of  $R$  is  $R > b/(b-c)$ . Let's compare this condition to our new one. Expressing the partner choice stability condition in terms of  $R$ :

$$R > \frac{Sb}{(1+S)(b-c) - b}$$

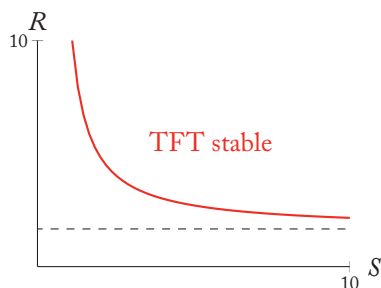
- 10 Is this ever easier to satisfy than the previous condition? We can ask the oracle of algebra:

$$\frac{b}{b-c} > \frac{Sb}{(1+S)(b-c) - b}$$

This simplifies to  $c < 0$ , which is never true. Partner choice never improves the stability of TFT.

- 15 It will help to plot the critical value of  $R$  as a function of  $S$ . This means to plot  $R = Sb/((1+S)(b-c) - b)$ . Above this curve, TFT is stable against rare NO-C. Below it, NO-C invades. Here's what it looks like:

Figure  
6.3



- The above is drawn for  $b = 2$  and  $c = 1.2$ . The dashed horizontal line is  $R = b/(b-c)$ , the value required for TFT to be stable in the absence of partner choice. Above the red curve, TFT is stable. So you see that partner choice, in this model, never helps. The reason is because partner choice sets rare un-cooperative individuals free to victimize new partners. When a TFT sticks with NO-C, the NO-C can't extract benefits from anyone else.

- 25 Don't jump to the conclusion that partner choice never helps. There are lots of dynamics missing from this basic model. For example, when TFT is common the only process that will free up new victims for NO-C is the exogenous  $w$  end

of relationships. So the expected search time should be a function of  $w$ . But this model serves as another example of how careful, logical statements of assumptions help us explore social causes in strategic contexts.

*Indirect reciprocity*

x

5

*Variance reduction and risk pooling*

x





# 7 K

## *The Evolution of Families*

From an evolutionary perspective, the family is a hostage situation. Offspring are gametes (eggs and sperm) who run away together with stolen copies of parental genes. Their stolen goods put offspring in a position of power. A genetic variant in a parent that hurts offspring very likely hurts a copy of itself. A genetic variant that helps offspring very likely helps a copy of itself. This is not an entirely cooperative scenario, because offspring are not their parents. They do not have exactly the same interests. They may use their power to extract more help than they really need. “Give me another worm,” the baby bird cries, “or your lineage gets it!”

Animal societies scaffold on this hostage scenario. Everywhere in nature, common descent is an important source of **positive assortment**. But this engenders always a mix of conflict and cooperation. If there is cooperation within the family, there is also competition between families. In many mammals for example, social groups are sisterhoods. And these sisterhoods compete intensely with one another for space, food, and a safe future for their offspring. Sisters also compete with one another. And the balance between conflict and cooperation can tip suddenly to either side. The baboon says: *I against my sister. I and my sister against my cousin. I, my sister, and my cousin against the world.*<sup>1</sup>

Human societies build absurd and amazing institutions on top of hostages. In medieval Europe, royalty dominated political power. And royalty is a family. But royal families routinely raised one another’s children, literal royal hostages, to ensure that they could remain at peace. “Don’t invade my lands,” the princess cries, “or your princess gets it!”<sup>2</sup>

This chapter is about **kin selection**.

### *Hamilton’s rule*

*When relatives interact, it can promote cooperation. Under strong assumptions, there is a simple rule.*

<sup>1</sup>This is a slightly modified proverb (Nafisa Haji, The Sweetness of Tears).

<sup>2</sup>Peacemaking in Middle Ages citation

Suppose two relatives interact in a one-shot **prisoner's dilemma**. When will selection favor cooperation instead of non-cooperation? Countless students have been taught in introductory evolutionary biology and anthropology courses that there is a simple answer to this question:

$$rb > c$$

- 5 where  $r$  is the **coefficient of relatedness** between the pair, and  $b$  and  $c$  have their usual meanings in the prisoner's dilemma. This condition is usually called **Hamilton's rule**.<sup>3</sup> This rule is a valuable qualitative result. It tells us that recent common descent, which increases  $r$ , can increase the range of conditions under which natural selection favors cooperation. The model it is derived from also helps us  
10 understand why.

Here is a very simple and sloppy derivation. The rest of this section derives the rule in a more careful and educational way. Recall the statistical positive assortment model from **5C**. In this model, the probability an individual of any strategy X is paired with another individual with the same strategy is:

$$\Pr(X|X) = r + (1 - r)p$$

- 15 where  $r$  is a coefficient of assortment and  $p$  the proportion of the population with strategy X. Now imagine pairs of individuals interact once in the additive prisoner's dilemma from the previous chapter. The condition for rare cooperative (C) individuals to invade a population of non-cooperators (N) is:

$$\Pr(C|C)(b - c) + \Pr(N|C)(-c) > 0$$

The left side of the condition above is the expected fitness of the rare C individual.

- 20 The right side is the average fitness of a common N. Now we fill in the abstract interaction probabilities with the statistical assortment model:

$$(r + (1 - r)p)(b - c) + (1 - r)(1 - p)(-c) > 0$$

When C is rare,  $p \approx 0$ , so this simplifies to  $rb - c > 0$ . This is Hamilton's rule.

- Like most results in evolutionary game theory, this rule is a result of very specific set of assumptions. So it is not completely general. But there is a tendency  
25 to exaggerate its generality, especially in the human evolutionary sciences. And this leads to mistakes. For example, the founders of Evolutionary Psychology wrote that it "should be true of any species on any planet at any time."<sup>4</sup> But consider the following two puzzles.

- 30 In many species of bird, offspring do not sexually mature in the first year but can instead stay behind and help their parents at the nest. Suppose a situation arises in which the helper can sacrifice her own life to save four siblings. So  $b = 4$ ,

<sup>3</sup>cite Hamilton

<sup>4</sup>Page 168 of Berkow, Cosmides, Tooby (1992) The Adapted Mind. It is tempting to dismiss this, because few people expect evolutionary psychologists to have a strong grasp of evolutionary theory. But many biologists are equally confused. They are just less likely to put it in writing.

$c = 1$ , and  $r = 0.5$ . Hamilton's rule is satisfied, but any individual in which a gene arises that leads to this self-sacrifice would die and remove the allele from the population. Why does Hamilton's rule arrive at the wrong answer?

The coefficient of relatedness  $r$  is  $1/2$  for full siblings,  $1/4$  for cousins, and so on, declining with genealogical distance. This coefficient is often interpreted as the probability a pair of individuals share an allele. But this can't be right, because at many sites in the genome, in most species, there are only a few alleles. So many distantly related individuals will share alleles. Instead  $r$  is the probability the individuals share an allele by common descent. They must have inherited it from a common ancestor. But again, this can't be right. If mutations are rare, and a population is not very large, then most individuals who share an allele must share it by common descent. In such populations, why doesn't natural selection favor universal cooperation? Or why doesn't Hamilton's rule predict that it should, if indeed  $r$  is the probability of sharing an allele by common descent?

**The Price equation.** To sort out Hamilton's rule, we need to back up and start over. We need a more general statement of natural selection. Then we can introduce assumptions, until we get the rule back. The place to start is the **Price equation**:

$$\bar{w}\Delta p = \text{cov}(w_i, p_i)$$

This equation states a simple fact about evolution: The change in the proportion of the population  $p$  that has any particular allele is equal to the covariance between individual fitness  $w_i$  and individual allele frequency  $p_i$ . I show you how to derive this equation in the box on the next page.

The Price equation is a very general statement about evolutionary change. It isn't about natural selection, specifically. If  $w_i$  is random, then it's a neutral model of evolutionary change. It is also perfectly general to a wide variety of inheritance systems, which is important when modeling some social insects for example, where females are diploid and males haploid.

Now what we'll do is make one assumption at a time until we recover Hamilton's rule.<sup>5</sup>

**Additive fitness.** To start, we assume an additive prisoner's dilemma as the cooperation game. Remember, there is nothing necessarily additive about cooperation payoffs. Assume:

$$w_i = \omega + h_i(-c) + y_i b$$

The symbol  $\omega$  is baseline fitness unrelated to the cooperation game. The variables  $b$  and  $c$  have their usual meanings. Now  $h_i$  is the probability that individual  $i$  provides help, and  $y_i$  is the probability that same individual receives help from its partner.

<sup>5</sup>This derivation is a simplified version of Grafen's geometric view of relatedness article.

**Deriving the Price equation.** We focus on a particular allele “S” and derive an expression for its expected change in frequency in the population. Let  $n_i$  be the number of copies of any allele that individual  $i$  could possess, its *ploidy*. So for example, humans are  $n = 2$  for most chromosomes. Let  $N$  be the size of the allele population,  $N = \sum_i n_i$ . Let  $p_i$  be the proportion of individual  $i$ ’s alleles that are S. Then just by definition, the proportion of S in the population is:

$$p = \sum_i \frac{n_i p_i}{N}$$

Now we use a prime symbol  $'$  to indicate the value of a variable in the next generation. So the proportion of S in the next generation is just  $p' = \sum_i (n'_i p'_i) / N'$ . Now let  $w_i$  be the “fitness” of individual  $i$ , the number of offspring it produces. Let  $\bar{w} = \sum_i (n_i / N) w_i$  be the average fitness. Then we can substitute  $n'_i = w_i n_i$  and  $N' = \bar{w} N$ .

$$p' = \sum_i \frac{w_i n_i p'_i}{\bar{w} N}$$

So far, nothing but definitions of random variables. Let’s make the above into a difference equation, instead of a recursion,  $\Delta p = p' - p$ :

$$\Delta p = \sum_i \frac{w_i n_i p'_i}{\bar{w} N} - p$$

Multiplying both sides by  $\bar{w}$ :

$$\bar{w} \Delta p = \sum_i \frac{n_i}{N} w_i p'_i - p \bar{w} = \sum_i \frac{n_i}{N} w_i p'_i - p \sum_i \frac{n_i}{N} w_i$$

Now just for fun, let’s add and subtract the term  $\sum_i (n_i / N) w_i p_i$  to the right side. This doesn’t change the value, because we added it and subtracted it at the same time. But then we can arrange and factor to make a simpler expression. Like this:

$$\bar{w} \Delta p = \underbrace{\sum_i \frac{n_i}{N} w_i p'_i}_A - \underbrace{p \sum_i \frac{n_i}{N} w_i}_B + \underbrace{\sum_i \frac{n_i}{N} w_i p_i}_C - \underbrace{\sum_i \frac{n_i}{N} w_i p_i}_D$$

The terms A and B are the originals. We added C and subtracted D. Now let’s rearrange:

$$\begin{aligned} \bar{w} \Delta p &= \left( \underbrace{\sum_i \frac{n_i}{N} w_i p_i}_C - \underbrace{\sum_i \frac{n_i}{N} w_i}_B \right) + \left( \underbrace{\sum_i \frac{n_i}{N} w_i p'_i}_A - \underbrace{\sum_i \frac{n_i}{N} w_i p_i}_D \right) \\ &= \sum_i \frac{n_i}{N} w_i (p_i - p) + \sum_i \frac{n_i}{N} w_i (p'_i - p_i) \end{aligned}$$

Now the last step. The definition of covariance is  $\text{cov}(x, y) = E(x(y - \bar{y}))$ . So the term on the left above is the covariance between  $w_i$  and  $p_i$ , since  $p$  is the average  $p_i$ . And the term on the right above is an average. Using these definitions, we arrive at the Price equation:

$$\bar{w} \Delta p = \text{cov}(w_i, p_i) + E(w_i \Delta p_i)$$

The term  $\Delta p_i = p'_i - p_i$  is the change within individual  $i$ . In most cases, this is zero, because the frequency of an allele does not change within an individual within its lifetime. But it’s not impossible. Setting  $\Delta p_i = 0$  gives us the simplified form in the text.

Additive fitness is a very special case that disallows synergies and diminishing returns. For most resources, if you give me twice as much, my benefit is not twice as large.

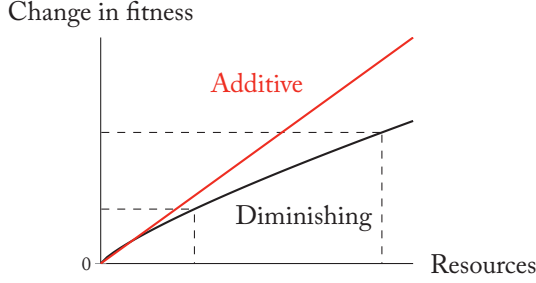


Figure  
7.1

Now insert our definition for  $w_i$  into the Price equation:

$$\bar{w}\Delta p = \text{cov}(\omega + h_i(-c) + y_i b, p_i)$$

Now since a covariance is an expectation, and an expectation is a sum, we can decompose the above across addition, giving us a sum of three covariances:

$$\bar{w}\Delta p = \text{cov}(\omega, p_i) + \text{cov}(h_i(-c), p_i) + \text{cov}(y_i b, p_i)$$

The first covariance,  $\text{cov}(\omega, p_i)$ , is zero. The reason is that  $\omega$  is a constant, and constants don't vary, so they also don't covary. This leaves the second and third covariances. For these, we factor out the constants:

$$\bar{w}\Delta p = (-c) \text{cov}(h_i, p_i) + b \text{cov}(y_i, p_i)$$

And this is greater than zero, meaning cooperation increases, when:

$$b \frac{\text{cov}(y_i, p_i)}{\text{cov}(h_i, p_i)} > c$$

This is not yet Hamilton's rule. Not even close. But it does resemble it. Instead of  $r$ , we have the ratio of the covariance between receiving help and genotype to the covariance between giving help and genotype.

**Additive genetics.** Next we need to specify how genotype  $p_i$  relates to phenotypes  $h_i$  and  $y_i$ . We make each phenotype an additive function of allele frequency  $p_i$ .

$$h_i = a + kp_i$$

$$y_i = a + kp_j$$

where  $p_j$  is the allele frequency in the partner  $j$ . There are other possibilities. Many alleles are dominant, and so two copies and one copy produce the same phenotype. This model doesn't allow that. Also no recessive alleles. Each additional cooperation allele produces a dose  $k$  of helping.

We insert these definitions into our rule and simplify:

$$b \frac{\text{cov}(a + kp_j, p_i)}{\text{cov}(a + kp_i, p_i)} > c$$

As before, constants don't vary and so they don't covary, so the  $a$  terms vanish. This leaves us:

$$b \frac{k \text{cov}(p_j, p_i)}{k \text{cov}(p_i, p_i)} > c$$

The covariance of a variable with itself is its variance, and so we arrive at:

$$b \frac{\text{cov}(p_j, p_i)}{\text{var}(p_i)} > c$$

5 This is usually expressed as:

$$b\beta(p_j, p_i) > c$$

where  $\beta(p_j, p_i)$  is the regression coefficient for the line regressing the partner's genotype  $p_j$  on the focal individual's  $p_i$ . In other words, it expresses how the focal's genotype predicts the partner's genotype.

This implies there is linear equation that satisfies:

$$\begin{aligned} E(p_j|p_i) &= E(p_j) + \beta(p_j, p_i)(p_i - E(p_i)) \\ &= p + \beta(p_j, p_i)(p_i - p) \end{aligned}$$

10 Solving for  $\beta(p_j, p_i)$ :

$$\beta(p_j, p_i) = \frac{E(p_j|p_i) - p}{p_i - p}$$

Okay, this still isn't  $r$ . But it is a function of the genotypes of the two individuals. Any process that makes  $E(p_j|p_i) > p$  would support the evolution of cooperation in this model. But what does relatedness have to do with it?

15 **Very weak selection.** We need a way to make  $\beta(p_j, p_i) = r$ . And it turns out that as we let the strength of natural selection become weak,  $\beta(p_j, p_i) \approx r$ . We'll proceed by constructing an expression for  $E(p_j|p_i)$ , the expected genotype of  $j$  given  $i$ , and relating it to shared ancestry. Then we'll see why our expression requires weak selection to be valid.

20 Define  $r$  as the proportion of genes that are **identical by descent** (IBD) in individuals  $i$  and  $j$ . When a gene is IBD, it means  $i$  and  $j$  have the same allele and they got it from a common ancestor. Now when an allele at the cooperation locus is IBD, it may or may not also be an allele that increases cooperation. Suppose for example individual  $i$  has  $p_i = 0.5$ , which means perhaps a diploid species and one copy of the cooperation allele. So while  $i$  and  $j$  share overall  $r$  alleles by common  
25 descent, there's only a half chance it's a cooperation allele, because  $i$  only has half

cooperation alleles. This means the chance  $j$  shared a cooperation allele with  $i$  through common descent is  $rp_i$ .

The pair may also share cooperation alleles through the non-IBD part of their genomes, which is of size  $1 - r$ . For that part of the genome, the expected proportion of cooperation alleles is the population average,  $p$ . So the chance  $j$  shares a cooperation allele with  $i$  that is not IBD is  $(1 - r)p$ .

Together we get:

$$E(p_j|p_i) = \underbrace{rp_i}_{\text{IBD}} + \underbrace{(1-r)p}_{\text{not-IBD}}$$

Now solve for  $r$ :

$$r = \frac{E(p_j|p_i) - p}{p_i - p}$$

And this is equal to  $\beta(p_j, p_i)$ , the regression from the previous section.

Okay, so where's the weak selection? It's already in there, smuggled in under shadows. When I asserted that the proportion of cooperation alleles that are IBD is  $rp_i$ , I implicitly assumed that cooperation and non-cooperation alleles have the same chances of being passed down from a common ancestor. But if there is selection, this can't be true. The fact that you exist implies for example that you didn't inherit any lethal alleles from any ancestor. So if selection has favored cooperation (or non-cooperation) in past generations, then inheritance won't be fair in the way we usually calculate  $r$ , by just counting links in the pedigree. Technically we need to assume very weak selection (neutrality) to do that.

For close relatives, selection doesn't have to be neutral for  $r$  to be a good approximation. But for distant relatives, even weak selection will make  $r$  a terrible device for predicting shared alleles. Once an allele goes to fixation in a population, then almost everyone shares it by common descent. But that isn't what's relevant to understanding the evolution of cooperation. What's relevant is knowing what happens to a mutant allele. Close relatives tend to share recent mutations, but only if those mutations are lethal or too strongly selected against. And that creates positive assortment of phenotypes among close relatives.

Okay, if we are comfortable letting selection be weak, then we have now recovered Hamilton's rule. If you make other assumptions about the game or about genetics or about the strength of selection, you get a different rule. At least on this planet.

### *Group selection*

x [or do this in its own chapter?]

### *Neighbor-modulated fitness*

x

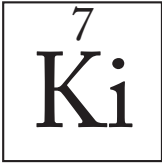




## Part 2

# Transition Elements





## *Causal Inference*

5

Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

10

Nulla malesuada porttitor diam. Donec felis erat, congue non, volutpat at, tincidunt tristique, libero. Vivamus viverra fermentum felis. Donec nonummy pellentesque ante. Phasellus adipiscing semper elit. Proin fermentum massa ac quam. Sed diam turpis, molestie vitae, placerat a, molestie nec, leo. Maecenas lacinia. Nam ipsum ligula, eleifend at, accumsan nec, suscipit a, ipsum. Morbi blandit ligula feugiat magna. Nunc eleifend consequat lorem. Sed lacinia nulla vitae enim. Pellentesque tincidunt purus vel magna. Integer non enim. Praesent euismod nunc eu purus. Donec bibendum quam in tellus. Nullam cursus pulvinar lectus. Donec et mi. Nam vulputate metus eu enim. Vestibulum pellentesque felis eu massa.

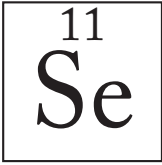
15

20

Quisque ullamcorper placerat ipsum. Cras nibh. Morbi vel justo vitae lacus tincidunt ultrices. Lorem ipsum dolor sit amet, consectetur adipiscing elit. In hac habitasse platea dictumst. Integer tempus convallis augue. Etiam facilisis. Nunc elementum fermentum wisi. Aenean placerat. Ut imperdiet, enim sed gravida sollicitudin, felis odio placerat quam, ac pulvinar elit purus eget enim. Nunc vitae tortor. Proin tempus nibh sit amet nisl. Vivamus quis tortor vitae risus porta vehicula.

25





## Senescence

5

Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

10

Nulla malesuada porttitor diam. Donec felis erat, congue non, volutpat at, tincidunt tristique, libero. Vivamus viverra fermentum felis. Donec nonummy pellentesque ante. Phasellus adipiscing semper elit. Proin fermentum massa ac quam. Sed diam turpis, molestie vitae, placerat a, molestie nec, leo. Maecenas lacinia. Nam ipsum ligula, eleifend at, accumsan nec, suscipit a, ipsum. Morbi blandit ligula feugiat magna. Nunc eleifend consequat lorem. Sed lacinia nulla vitae enim. Pellentesque tincidunt purus vel magna. Integer non enim. Praesent euismod nunc eu purus. Donec bibendum quam in tellus. Nullam cursus pulvinar lectus. Donec et mi. Nam vulputate metus eu enim. Vestibulum pellentesque felis eu massa.

15

20

Quisque ullamcorper placerat ipsum. Cras nibh. Morbi vel justo vitae lacus tincidunt ultrices. Lorem ipsum dolor sit amet, consectetur adipiscing elit. In hac habitasse platea dictumst. Integer tempus convallis augue. Etiam facilisis. Nunc elementum fermentum wisi. Aenean placerat. Ut imperdiet, enim sed gravida sollicitudin, felis odio placerat quam, ac pulvinar elit purus eget enim. Nunc vitae tortor. Proin tempus nibh sit amet nisl. Vivamus quis tortor vitae risus porta vehicula.

25

### *Extrinsic mortality*

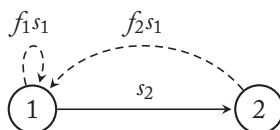
Imagine a species that lives for a maximum of two years.<sup>1</sup> Let  $n_{1,t}$  be the number of individuals of age 1 of year  $t$ , and  $n_{2,t}$  is the number of two-year-olds. In order to reproduce, an individual must first survive through the winter. Let  $s_i$  be the

<sup>1</sup>This example is based on the model in the appendix to [].

- 5 probability of surviving to reach age  $i$ . So babies are only counted if with probability  $s_1$ , and one-year-olds reach age 2 with probability  $s_2$ . Let  $f_i$  be the fertility of an individual of age  $i$ . For the moment, assume there is no population regulation—the environment is boundless and the species can grow exponentially forever. Diagrammatically, the model looks like this:

Figure  
11.1

10



15

If you aren't familiar with these state-and-arrow life history diagrams, each circle is a state. In this example, each state is an **age class**. The arrows represent contributions of each age class to another or itself in the next time step. So the arrow ① → ② represents survival to age 2, which happens with probability  $s_2$ . The other two arrows are reproduction. Each one-year-old contributes on average  $f_1 s_1$  new one-year-olds and each two-year-old contributes  $f_2 s_1$ .

20

These diagrams don't just look pretty. They also give us an algorithm for writing down a mathematical model of the population dynamics. Each node (circle) has a state variable. In this case, these are  $n_1$  and  $n_2$ . To find a mathematical rule for updating these state variables each year, we just look at all the arrows entering a node. For each arrow, we multiply the state variable of the arrow's origin by the factors on the arrow itself.

25

So to update the two-year-olds, there is just one arrow, the arrow from ① to ②. The state variable for the source ① is  $n_1$ . We multiply this by the factor on the arrow,  $s_2$ , and get an equation to update  $n_2$ :

$$n_{2,t+1} = n_{1,t} s_2$$

The number of two-year-olds is just the number of surviving one-year-olds. And the same procedure yields an update equation for  $n_1$ . But now there are two arrows.

$$n_{1,t+1} = n_{1,t} s_1 f_1 + n_{2,t} s_1 f_2$$

The number of one-year-olds is the total births from both age classes. Now we have two **recursions**—updating equations—for the two state variables of the model.

30

In a moment, we'll let these survival and fertility parameters evolve. But for now let's just figure out how the population grows, given any specific values of  $s_1$ ,  $s_2$ ,  $f_1$ , and  $f_2$ . The growth rate of this species depends upon reproduction in both age classes. So again we cannot simply count babies, the lifetime reproductive success, because timing matters (see chapter 1F). The textbook way to calculate the growth rate in this example is to calculate something called the **dominant eigenvalue**<sup>2</sup> of the transition matrix for the population. It is absolutely worth knowing how to do

<sup>2</sup>The term eigenvalue comes from German *Eigenwert*, which just means the characteristic value (of a matrix). The use of *Eigen* in English causes unnecessary angst for many students.

that, and it isn't even hard. So we'll do it. But it isn't at all intuitive why it works. So first let's also do the same calculation a possibly more intuitive way.

The long-term growth rate of the species is achieved when the population is at its **stable age distribution**. Not all models have stable age distributions—populations can fluctuate forever for example. But this one does. At the stable age distribution, the relative numbers of each age class are not changing, and every age class is growing at the same rate. Otherwise it wouldn't be stable. Call this stable rate of growth  $\lambda$ , which is the convention. Then at the stable age distribution it must be true that:

$$\lambda = \frac{N_{t+1}}{N_t} = \frac{n_{1,t+1}}{n_{1,t}} = \frac{n_{2,t+1}}{n_{2,t}}$$

Let's pull out those relations of each age class with  $\lambda$  and express them this way:

$$\lambda n_{1,t} = n_{1,t+1} \quad \lambda n_{2,t} = n_{2,t+1}$$

After substituting in the full recursions for  $n_{1,t+1}$  and  $n_{2,t+1}$ , this gives us two simultaneous equations that we can use to find  $\lambda$ . Here they are:

$$\begin{aligned} \lambda n_{1,t} &= n_{1,t} s_1 f_1 + n_{2,t} s_1 f_2 \\ \lambda n_{2,t} &= n_{1,t} s_2 \end{aligned}$$

Solving for  $\lambda$  gives two roots. If your algebra is rusty, start by solving the little equation for  $n_{2,t} = n_{1,t} s_2 / \lambda$ . Then substitute this expression for  $n_{2,t}$  into the big equation and solve for  $\lambda$ . It will be quadratic, so use the quadratic formula. Only one of the roots is positive for reasonable values of  $s_1, f_1, s_2$ , and  $f_2$ . It is:

$$\lambda^* = \frac{1}{2} \left( s_1 f_1 + \sqrt{s_1 (s_1 f_1^2 + 4 s_2 f_2)} \right)$$

This is definitely not average lifetime reproductive success, which is (for this model) just  $s_1 f_1 + s_1 s_2 f_2$ . It is however the **dominant eigenvalue**, the same as you'd get doing the convention matrix approach. To see the conventional matrix approach, dive into the shaded box on the next page. If you are curious about the solution for the stable age distribution itself, also look into the box.

Now let's make things interesting by adding a **tradeoff** in survival across age classes and asking how natural selection will manage this tradeoff. Suppose that survivorship is a function of a baseline amount of mortality that is independent of age. Call this  $x$ . This is the **extrinsic** mortality. Survivorship is also a function of an age-specific factor  $a$  that has to be allocated by a genotype across age classes. This is the **intrinsic** mortality. So let's redefine  $s_1$  and  $s_2$  this way:

$$\begin{aligned} s_1 &= (1 - x)a \\ s_2 &= (1 - x)(1 - a) \end{aligned}$$

Both  $x$  and  $a$  are between zero and one. If  $a = 1$ , then the genotype invests everything in surviving to reproduce in the first year. As  $a$  declines, the genotype

**Eigenvalues and eigenvectors.** To compute the growth rate  $\lambda^*$  using the conventional matrix approach, we first organize the system of recursions into a **transition matrix**. This is a square matrix with one row and one column for each **state variable** of the system. The state variables in this case are  $n_1$  and  $n_2$ . The entries in the matrix express the *per capita* contribution of the column variable to the row variable in the next time step. In our case, the matrix is:

$$\mathbf{L} = \begin{pmatrix} s_1 f_1 & s_1 f_2 \\ s_2 & 0 \end{pmatrix}$$

If we look at the recursions, you'll see these factors in front of each state variable on the right side of each equation. Here are the recursions again, with these factors highlighted in brackets:

$$\begin{aligned} n_{1,t+1} &= n_{1,t} [s_1 f_1] + n_{2,t} [s_1 f_2] \\ n_{2,t+1} &= n_{1,t} [s_2] \end{aligned}$$

Each one-year-old contributes on average  $s_1 f_1$  new one-year-olds to the population, as well as  $s_2$  two-year-olds (if they survive to the second year). This gives us the first column of the matrix  $\mathbf{L}$ . The second column has a zero in it, because two-year-olds contribute zero two-year-olds.

The matrix  $\mathbf{L}$  is handy, because it provides a shortcut for all that algebra we did. It turns out that we can compute the dominant eigenvalue from  $\mathbf{L}$  alone, yielding the longterm growth rate. We find the values of  $\lambda$  that satisfy:

$$\det(\mathbf{L} - \lambda \mathbf{I}) = \det \begin{pmatrix} s_1 f_1 - \lambda & s_1 f_2 \\ s_2 & 0 - \lambda \end{pmatrix} = 0$$

That “det” means the determinate of the matrix. For a two-by-two matrix  $\begin{pmatrix} A & B \\ C & D \end{pmatrix}$ , the determinate is just  $AD - BC$ . So we solve:

$$(s_1 f_1 - \lambda)(0 - \lambda) - (s_1 f_2)(s_2) = 0$$

This is quadratic in  $\lambda$ , so you get two solutions, and the largest will match  $\lambda^*$  on the previous page.

In the notebook, I show you how to use Mathematica's **Eigenvalues** function to do this. The corresponding **Eigenvectors** give the relative proportions of each age class in the stationary age distribution. The term **eigenvector** is even less helpful than eigenvalue. But it is just a list of the relative abundances of each age class. You can find it readily without the matrix, in fact. Resuming the derivation in the main text, the little equation implies that:

$$\frac{n_{2,t}}{n_{1,t}} = \frac{s_2}{\lambda}$$

The left side is the ratio of two-year-olds to one-year-olds. So once you have your solution for  $\lambda$ , just substitute it into the above to get an expression for this ratio. Call this ratio  $R_{21}$ . The proportion of two-year-olds at the stable age distribution is just  $R_{21}/(1 + R_{21})$ . Why? Let  $p_2$  be the proportion of two-year-olds and  $p_1 = 1 - p_2$  the proportion of one-year-olds. Then it must be true that  $R_{21} = p_2/p_1 = p_2/(1 - p_2)$ . Solve for  $p_2$ , and you get  $p_2 = R_{21}/(1 + R_{21})$ .

For a model with only two age classes, the matrix approach doesn't save you much work. But for larger models, it can be much easier to just find the eigenvalues and eigenvectors than to slosh around solving a large system of equations. Technically, both approaches are identical. However there are situations in which the matrix approach can trick you. For example, if the transition matrix depends upon population density, you have to be careful.<sup>a</sup>

<sup>a</sup>Cite Caswell & Takada 2004



invests more in surviving to reproduce in the second year. This function assumes a perfectly linear tradeoff in allocating  $a$ .

We already have the longterm growth rate. We can just replace  $s_1$  and  $s_2$  with the new expressions above, yielding the longterm growth rate of any genotype with a particular value of  $a$ . So which value of  $a$  has the highest longterm growth rate? We can find that by finding the value of  $a$  that maximizes  $\lambda^*$ . Why? Because if many genotypes are present in the population, then they will all be growing exponentially, but the one with the largest  $\lambda$  will come to dominate the environment, because its exponential growth rate will be larger.

I appreciate that it is weird to think about a bunch of exponentially growing lineages competing in an environment with no limits. We'll introduce population limits soon. But that will be more complicated, so it's kinder to do this exponential growth circus first.

We ask which values of  $a$  satisfy  $\frac{d\lambda^*}{da} = 0$ . The derivative is unpleasant, but if you use a symbolic system like Mathematica, it'll make quick work of it. Then when you solve for  $a$ , be careful to note the constraints on the variables:  $f_1 > 0$ ,  $f_2 > 0$ ,  $0 < a < 1$ , and  $0 < x < 1$ . In the notebook for this chapter, I show you how to do this properly with Mathematica's `Reduce` function. The answer is satisfyingly simple:

$$a^* = \frac{1}{2 - f_1/\sqrt{f_2}}$$

Notice that  $x$  is not present. This means that exogenous mortality has no impact on the evolutionary end point. This is not what happens in general. And we'll make a change next that leads to  $x$  mattering.

Before we do that however, consider how fertility matters. What matters is the ratio  $f_1/\sqrt{f_2}$ . Consider first what happens when  $f_1 = \sqrt{f_2}$ . Then  $a^* = 1$ , which means the organism invests everything in one-year-olds and no individuals survive to the second year. But as  $f_2$  increases,  $a^*$  approaches  $1/2$ , which means that the species invests equally in one-year-olds and two-year-olds. Here is a plot of  $a^*$  against  $f_2$ , with  $f_1 = 1$ .

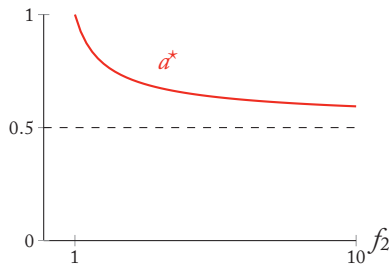


Figure  
11.2

In this model, evolution can never favor  $a^* < 1/2$ . Why  $1/2$ ? This is the value of  $a$  that maximizes the probability of attaining age 2. The probability an individual lives to the second year is not just  $s_2$ . They have to survive the first year as well. So the probability an individual lives to age 2 is  $s_1 s_2$ . Substituting in  $s_1 = (1 - x)a$  and  $s_2 = (1 - x)(1 - a)$ :

$$a(1 - a)(1 - x)^2$$

The term  $a(1 - a)$  is maximized for  $a = 1/2$ . This is a symptom of the fact that all two-year-olds were once one-year-olds. To maximize the chance of living to a later age, an organism must invest in earlier ages as well.

Before moving on, let's consider what  $a$  does to the **expected lifespan at birth**, which I'll label  $\ell$ . Which conditions favor longer life? The expected lifespan in this model is:

$$\begin{aligned}\ell &= \sum_{i=0}^2 \Pr(\text{die at age } i)(i) \\ \ell &= (1 - s_1)(0) + s_1(1 - s_2)(1) + s_1 s_2(2) = s_1(1 + s_2)\end{aligned}$$

where age 0 means a newborn. Substituting in the tradeoff functions for  $s_1$  and  $s_2$ :

$$\ell = a(2 - x)(1 - x) + a^2(1 - x)^2$$

This expression is always increasing with  $a$ , so larger values of  $a$  always increase expected lifespan. How is it possible that  $a = 1/2$  maximizes the chance of living to age 2, but  $a = 1$  maximizes lifespan? When  $a = 1$ ,  $s_1 = 1 - x$  while  $s_2 = 0$ , so the expected lifespan is just  $1 - x$ . Suppose an organism decreases  $a$  below 1 by a tiny amount. This can never increase  $\ell$ , because the reduction in  $s_1$  means additional individuals will now die before ever experiencing the better  $s_2$ . You may get a mathematical intuition for this by looking again at  $\ell = s_1(1 + s_2)$  and seeing that  $s_1$  moderates any gains in  $s_2$ .

In this simple model, evolution can favor survival in later years, but it will not result in longer expected lifespans. Any additional years of life are consumed by individuals not living to be even one year old. However, if you just census living individuals in the population, the average age will be older, for smaller values of  $a$ . This is only because you cannot census the dead.

### *Extrinsic mortality with density dependence*

*When population growth depends upon population size, extrinsic mortality can influence selection for aging.*

In the previous example, extrinsic mortality had no influence of the evolution of senescence. All this demonstrates is that extrinsic mortality, defined as a source of mortality that affects all age classes, does not always influence selection for lifespan. Now here is an example where it does.

In real populations, eventually the environment fills up. Either there aren't enough nest spots or there isn't enough food or both. As scarcity arrives, either fertility or survival or both decrease. When this happens, the population experiences **density dependent** population regulation. It will no longer grow without bound. It might level off around some stable population size or it might fluctuate, crashing down before growing again.

Let's modify the model so that the population levels off. To do this, we need to pick some biological mechanism. Suppose that in our example species the first thing that goes wrong when the population grows is that fertility declines. There are many ways that population size could impact fertility—there could be for example a fixed number of breeding sites. To keep things simple, let's suppose that fertility gradually declines with increasing population size  $N_t = n_{1,t} + n_{2,t}$ . Specifically, we'll suppose the new fertility functions are

$$\begin{aligned}f_1 &= b_1 \exp(-kN_t) \\f_2 &= b_2 \exp(-kN_t)\end{aligned}$$

Those  $b$  variables are the maximum fertilities of each age class. The term  $\exp(-kN_t)$  starts out a 1, when  $N_t = 0$ , and declines to zero as  $N_t$  increases. The variable  $k$  controls how rapidly this happens. It looks like this, for  $k = 0.1$ :

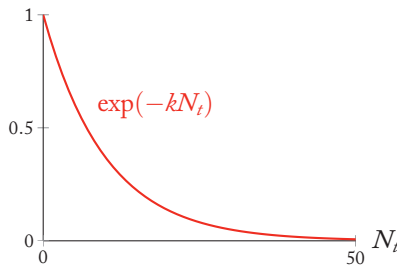


Figure  
11.3

So as  $N_t$  increases, effective fertility approaches zero, eventually halting population growth.

Substituting these new definitions into our old  $\lambda^*$  growth rate gives us a new, complicated expression for the growth rate. I won't repeat it here, since there is no intuition to draw from it. But it's in the notebook for this chapter.

The new  $\lambda^*$  will allow us to find the **evolutionarily stable** value of  $a$ . Like before, we want to find the value of  $a$  that maximizes  $\lambda^*$ . But unlike before,  $\lambda^*$  cannot in the long run be greater than 1. The reason is that the population increases only until fertility declines to match mortality. So things are more subtle now. But there turns out to be a simple solution strategy. Let me try to reason you into it, so it's not just math magic.

Consider a population with given values of  $b_1$ ,  $b_2$ ,  $\alpha$ , and  $k$ . There is one genotype with a particular value  $a$ . This genotype, after a short while, reaches

a particular **demographic equilibrium** with a particular stable population size  $\hat{N}$ . This  $\hat{N}$  is a property of the genotype. So different genotypes have different stable population sizes, even though they all regulate one another through density dependence, when they co-occur.

Now suppose there is a mutation. A new genotype appears with a slightly different value of  $a$ . The new genotype will tend to spread, if its own particular  $\hat{N}$  is larger than that of the common genotype. The reason is that even though all those common genotypes are reducing our heroic mutant's fertility, the mutant isn't suppressed to demographic equilibrium, because it has a better value of  $a$  and is better at spreading. So it will grow and eventually replace the previously common genotype, itself becoming the new common genotype.

A series of mutants like this can move the population towards values of  $a$  that are near the evolutionarily stable value of  $a$ . To solve for this special value of  $a$ , we ask again which value of  $a$  maximizes  $\lambda^*$ , while also requiring that we are at demographic equilibrium. First let's find a function  $\hat{N}$ . We do this by setting  $\lambda^* = 1$  and solving for  $N$ . This yields:

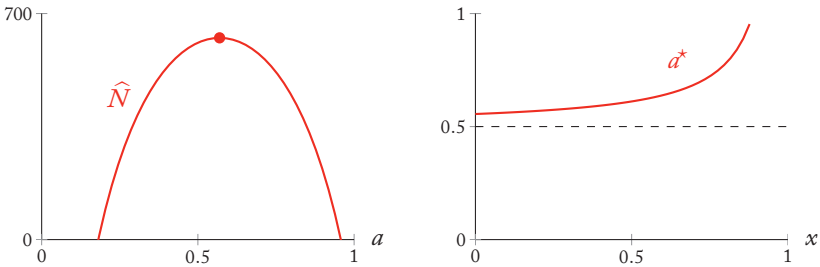
$$\hat{N} = \frac{1}{k} \log(a(1-x)(b_2(1-a)(1-x) + b_1))$$

If we plot this equilibrium population size as a function of  $a$ , you'll see that there is a clear peak. But first let's solve for that peak, the value of  $a$  that satisfies  $d\hat{N}/da = 0$ .

$$a^* = \frac{b_1 + b_2(1-x)}{2b_2(1-x)} = \frac{1}{2} \left( 1 + \frac{b_1}{b_2(1-x)} \right)$$

First note that  $x$  appears in the answer. Extrinsic mortality matters now. Let's plot both  $\hat{N}$  as a function of  $a$  (for  $x = 0.2$ ) and  $a^*$  as a function of  $x$ . Other variables are fixed at:  $b_1 = 1$ ,  $b_2 = 9$ , and  $k = 10^{-3}$ .

Figure  
11.4



The influence of  $x$  is to reduce investment in survival to age 2. You can see this in the expression above, if you first imagine that  $x = 0$  and  $b_2$  is much larger than  $b_1$ , so that  $a^* \approx 1/2$ . This just like the previous model. Now as we increase  $x$ , the amount of extrinsic mortality, it cuts away at  $b_2$  in the expression. This increases  $a^*$  until it caps out at 1. The larger the ratio  $b_2/b_1$ , the more of an impact  $x$  will have on selection. But its impact is always to reduce investment in survival to age

2. Quantitatively, the impact is rather small unless  $x$  is large. But when  $x$  is large, the species may just go extinct ( $\hat{N} = 0$ ). 5

This model shows only that there are conditions under which extrinsic mortality influences life history evolution. There are also conditions under which it does not. For example, if we modify this model so that density dependence influences survival rather than fertility, extrinsic mortality will not matter. Suppose we let fertility be simply  $f_1$  and  $f_2$  again, unaffected by density, but now we define survival as: 10

$$\begin{aligned}s_1 &= (1 - x)a \exp(-kN_t) \\ s_2 &= (1 - x)(1 - a) \exp(-kN_t)\end{aligned}$$

Repeat the same solution steps as above: find the stable population size at  $\lambda = 1$  and then solve for the  $a$  that maximizes it. If you do this, you get the same result as in the original, density-independent model: 15

$$a^* = \frac{1}{2 - f_1/\sqrt{f_2}}$$

I show the complete derivation in the chapter notebook.

### *Extrinsic confusion*

What explains these results? For something like extrinsic mortality to influence aging, it must somehow lead to differential selection across age classes. Since the extrinsic mortality in these models influence all ages the same way, why does it influence life history at all? And why only when fertility, instead of survival, is regulated by density? 20

To make sense of the different results, let's zoom out to the general model again. At demographic steady state, the model says:

$$\begin{aligned}n_{1,t+1} &= \lambda n_{1,t} = n_{1,t} s_1 f_1 + n_{2,t} s_1 f_2 \\ n_{2,t+1} &= \lambda n_{2,t} = n_{1,t} s_2\end{aligned}$$

There is a hidden implication of this model, and many other similar models, that will help us understand why selection acts as it does. To reveal it, let's eliminate  $n_{2,t}$  from the model and write down the expression for  $n_{1,t}$ . Why? Because this will give a **renewal** equation, an expression for number of births each year. To begin, divide by  $\lambda$ : 25

$$n_{1,t} = n_{1,t} \frac{s_1 f_1}{\lambda} + n_{2,t} \frac{s_1 f_2}{\lambda}$$

Now we can eliminate  $n_{2,t}$  by noting that  $n_{2,t} = n_{1,t} s_2 / \lambda$ . This gives us:

$$n_{1,t} = n_{1,t} \frac{s_1 f_1}{\lambda} + n_{1,t} \frac{s_2 s_1 f_2}{\lambda} = n_{1,t} \frac{s_1 f_1}{\lambda} + n_{1,t} \frac{s_1 s_2 f_2}{\lambda^2}$$

- 5 Now since  $n_{1,t}$  is in every term, we can divide both sides by  $n_{1,t}$  to eliminate it as well.

$$1 = \frac{s_1 f_1}{\lambda} + \frac{s_1 s_2 f_2}{\lambda^2}$$

- Maybe this doesn't look like much. But we can learn a lot by studying its structure. This equation is a special case of the **Euler-Lotka equation**, a general fact about this type of age-structured evolutionary model. See the box on the next page for a more general derivation.

The left side, the 1, means "all of the babies," 100% of them. The right side is a decomposition of all of the babies. Each term on the right is the contribution of each age class to the renewal of the population. Like this:

all of the babies = babies from one-year-olds + babies from two-year-olds

- Each term on the right side of the expression is the contribution of each age class to the renewal of the population in each year. Whatever the values of the  $s$  and  $f$  variables, the population growth  $\lambda$  will adjust so that the sum of these terms is exactly 1. Likewise, if  $\lambda$  is fixed by for example population regulation, then the equation tells us how the  $s$  and  $f$  variables must adjust.

- Let's walk through each of the three previous models and see what Euler-Lotka tells us. Recall that our model assigns  $s_1 = (1-x)a$  and  $s_2 = (1-x)(1-a)$ . Substituting:

$$1 = \frac{(1-x)a f_1}{\lambda} + \frac{(1-x)^2 a(1-a) f_2}{\lambda^2}$$

Notice that exogenous mortality  $x$  hits the second term twice. It does so because both  $s_1$  and  $s_2$  appear in the term for the second age class. The population growth  $\lambda$  also hits the second term twice, but on the bottom.

- First, consider the model without density dependence. In this model, there are two forces influencing selection on age, and they happen to cancel one another. The first force is the ordinary advantage of reproducing early. When a genotype reproduces early, and the population is growing, then offspring can take advantage of population growth. This is why the early reproducing type was favored back in chapter 1F. That same effect occurs in this model. The other force is extrinsic mortality. When extrinsic mortality increases, it has two effects. It makes it harder to reach higher age classes, because the more years go by the more chances for the asteroid to strike. But extrinsic mortality simultaneously reduces population growth, which weakens the advantage of reproducing early. These two forces cancel one another, and when populations are unregulated by density dependence—they grow exponentially forever—then extrinsic mortality does not influence selection for age-specific effects.

Now consider the model with density dependence influencing fertility. In this model, the population no longer grows exponentially forever. Most of the time,

*The Euler-Lotka equation.* General derivation

it will be near its stable population size, so selection acts at  $\lambda = 1$ , not as some  $\lambda > 1$ . This means in turn that the pure advantage of reproducing early is absent, because the population isn't growing. But the force of extrinsic mortality making it harder to reach later age classes effectively reduces selection on later age-specific effects. So extrinsic mortality does matter in this model, and it tends to reduce investment in later age classes.

The last model also has density dependence, and so the advantage of reproducing early is also canceled. So why does extrinsic mortality not matter in this case? In this model, the population is regulated through survival. When the population is at its stable size, survival has been reduced by crowding. Now imagine increasing the amount of extrinsic mortality in such a population. This will push it below the stable population size, density dependence will weaken, and survival (the part controlled by density dependence) will increase. So the impact of extrinsic mortality is canceled by the way that survival regulates population size.

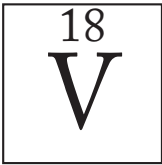
All this might sound persuasive, or not. But I haven't really proved the explanation above. To do that, we need a broader view on life history evolution. That's the goal of the next section.

## *Menopause*

Needs its own chapter







## *Parental Investment*

5

Etiam euismod. Fusce facilisis lacinia dui. Suspendisse potenti. In mi erat, cursus id, nonummy sed, ullamcorper eget, sapien. Praesent pretium, magna in eleifend egestas, pede pede pretium lorem, quis consectetuer tortor sapien facilisis magna. Mauris quis magna varius nulla scelerisque imperdiet. Aliquam non quam. Aliquam porttitor quam a lacus. Praesent vel arcu ut tortor cursus volutpat. In vitae pede quis diam bibendum placerat. Fusce elementum convallis neque. Sed dolor orci, scelerisque ac, dapibus nec, ultricies ut, mi. Duis nec dui quis leo sagittis commodo.

10

Aliquam lectus. Vivamus leo. Quisque ornare tellus ullamcorper nulla. Mauris porttitor pharetra tortor. Sed fringilla justo sed mauris. Mauris tellus. Sed non leo. Nullam elementum, magna in cursus sodales, augue est scelerisque sapien, venenatis congue nulla arcu et pede. Ut suscipit enim vel sapien. Donec congue. Maecenas urna mi, suscipit in, placerat ut, vestibulum ut, massa. Fusce ultrices nulla et nisl.

15

Etiam ac leo a risus tristique nonummy. Donec dignissim tincidunt nulla. Vestibulum rhoncus molestie odio. Sed lobortis, justo et pretium lobortis, mauris turpis condimentum augue, nec ultricies nibh arcu pretium enim. Nunc purus neque, placerat id, imperdiet sed, pellentesque nec, nisl. Vestibulum imperdiet neque non sem accumsan laoreet. In hac habitasse platea dictumst. Etiam condimentum facilisis libero. Suspendisse in elit quis nisl aliquam dapibus. Pellentesque auctor sapien. Sed egestas sapien nec lectus. Pellentesque vel dui vel neque bibendum viverra. Aliquam porttitor nisl nec pede. Proin mattis libero vel turpis. Donec rutrum mauris et libero. Proin euismod porta felis. Nam lobortis, metus quis elementum commodo, nunc lectus elementum mauris, eget vulputate ligula tellus eu neque. Vivamus eu dolor.

20

25



19

S1

Social Learning

5

In series: 1F

Etiam euismod. Fusce facilisis lacinia dui. Suspendisse potenti. In mi erat, cursus id, nonummy sed, ullamcorper eget, sapien. Praesent pretium, magna in eleifend egestas, pede pede pretium lorem, quis consectetur tortor sapien facilisis magna. Mauris quis magna varius nulla scelerisque imperdiet. Aliquam non quam. Aliquam porttitor quam a lacus. Praesent vel arcu ut tortor cursus volutpat. In vitae pede quis diam bibendum placerat. Fusce elementum convallis neque. Sed dolor orci, scelerisque ac, dapibus nec, ultricies ut, mi. Duis nec dui quis leo sagittis commodo.

Aliquam lectus. Vivamus leo. Quisque ornare tellus ullamcorper nulla. Mauris porttitor pharetra tortor. Sed fringilla justo sed mauris. Mauris tellus. Sed non leo. Nullam elementum, magna in cursus sodales, augue est scelerisque sapien, venenatis congue nulla arcu et pede. Ut suscipit enim vel sapien. Donec congue. Maecenas urna mi, suscipit in, placerat ut, vestibulum ut, massa. Fusce ultrices nulla et nisl.

## 5 *Rogers' paradox*

*Natural selection may favor social learning, but social learning may bring no evolutionary benefit for the population.*

A **paradox** is a self-contradictory statement. But in evolutionary theory, it tends to mean something that violates intuition. Evolutionary theory is full of  
 10 paradoxes. The paradox we'll consider in this section is that natural selection can easily favor social learning, at least in theory. But social learning alone brings no benefit to the population, in terms of population growth or lineage survival. In fact, it can do real damage. This paradox is known as **Rogers' paradox**, after the population geneticist who pointed it out.<sup>1</sup> Of course it is no paradox at all  
 15 that natural selection can favor traits in individuals that do not benefit the lineage. And Rogers' paradox is only another example of that fact. But it does have its own interesting features, features which persist in many models of cultural evolution. So it's worth study, even if you don't perceive any paradox.

Consider an organism with a simple life cycle. Each year, individuals are born  
 20 and live for only one year. Reproduction depends upon knowledge of the environment. When an individual possesses relevant knowledge to aid in reproduction, we'll call them **skilled**. Whether an individual is skilled or not depends upon the state of the environment, obviously. The environment changes over time, and when it does, all previously learned behavior becomes unskilled. Unskilled behavior produces  $b$  offspring, and skilled behavior produce  $B > b$  offspring.  
 25

Skilled behavior can only be learned—it is not encoded genetically. But learning strategy is influenced by genes. Suppose there are two genotypes. The first is **innovate** ( $I$ ), which attempts to innovate a new skilled behavior. An innovation attempt reduces adult fertility by a factor  $C_I > 0$  and succeeds with probability  $d$  (for “discover”). With these assumptions, the fertility of an  $I$  individual is:  
 30

$$f_I = \underbrace{d(B - C_I)}_{\text{innovation succeeds}} + \underbrace{(1 - d)(b - C_I)}_{\text{innovation fails}} = b + d(B - b) - C_I$$

Let's assume, to keep things simple, that all individuals survive to reproduce. With these assumptions, the growth rate of the  $I$  genotype is just  $f_I$ . If you are not sure why, then you should review 1F and possible Se as well.

The second genotype we'll consider is **social** ( $S$ ), which attempts to acquire a  
 35 previously innovated behavior through some combination of social and individual learning mechanisms. For example, the  $S$  individual might observe an adult acquiring some food resource and then figure out on its own how to do the same. This is less costly than innovation, carrying a cost factor  $C_S < C_I$ . Let  $q$  be the probability of acquiring skilled behavior through social learning. Then the expected fertility

<sup>1</sup>Rogers 1989. This paper focuses as well on the general question of whether we should expect behavior to be adaptive, if learning evolves by natural selection. It's a great paper, easy to read, and timelessly relevant to debates in the evolutionary sciences.

of  $S$  is:

5

$$f_s = q(B - C_s) + (1 - q)(b - C_s) = b + q(B - b) - C_s$$

This is structurally identical to  $f_t$ , but with  $q$  replacing  $s$ . However,  $q$  is very different from  $s$ , because  $q$  is dynamic. It evolves, changing each generation as a consequence of both social learning and environmental change. So  $f_s$  isn't constant, and we need to specify instead the expected fertility in a specific year  $y$ :

$$f_{s,y} = b + q_y(B - b) - C_s$$

And now we need to figure out  $q_y$ . It's not immediately obvious how to do that, so let's figure out something else instead. Let's figure out  $q_t$ , where  $t$  is the number of years since the environment most recently changed. It turns out we can write a function for this, and then we can use it to define the long-term growth rate of  $S$ .

10

Let  $p$  be the proportion of the adult population that are social learners. We need this, because the probability of acquiring skilled behavior through social learning will depend upon how many role models are themselves social learners. Let's assume that cultural dynamics are sufficiently fast, relative to gene dynamics, that  $p$  is almost constant from one year to the next. Our goal is to write a function  $q_t$  that tells us the proportion of skilled adults (who can be observed and learned from) in a population that experienced a change in the environment  $t$  years ago. When  $t = 0$ , this is straightforward:

15

$$q_0 = \underbrace{(1 - p)d}_{\text{innovators}} + \underbrace{p(0)}_{\text{social learners}} = (1 - p)d$$

The innovators are a proportion  $1 - p$  of the population and succeed  $d$  of the time. The social learners are a proportion  $p$  of the population and succeed never, because there are no skilled adults just after the environment changes. Now when the environment doesn't change, in the next year, the same logic gives us:

25

$$\begin{aligned} q_1 &= (1 - p)d + pq_0 \\ &= (1 - p)d + p(1 - p)d = d(1 - p)(1 + p) = d(1 - p^2) \end{aligned}$$

And in yet another year:

$$q_2 = (1 - p)d + pq_1 = (1 - p)d + pd(1 - p^2) = d(1 - p^3)$$

You can probably see now that  $t$  years after a change in the environment, the proportion of skilled individuals in the population will be:

$$q_t = d(1 - p^{t+1})$$

We'll plot this function in a moment. But we can learn some things about it right away. First, as  $t$  increases,  $q_t$  approaches  $d$ :

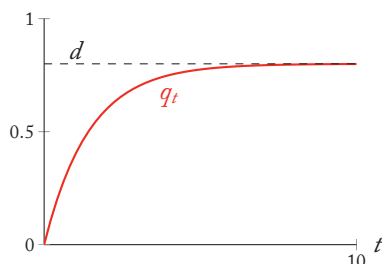
$$\lim_{t \rightarrow \infty} d(1 - p^{t+1}) = d$$

- 5 The proportion of skilled individuals in the population can never exceed the success rate of innovation. This is a hint about the overall dynamics in this model—social learning does not increase the pool of skill in the population.

To make this more visceral, we can plot  $q_t$  as a function of  $t$ , showing how the proportion of skilled individuals increases since the last change in the environment.

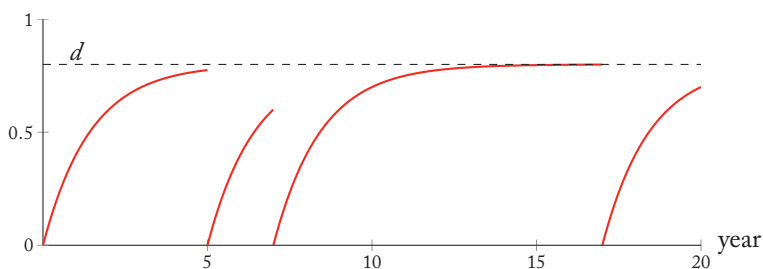
- 10 Here is  $q_t$  plotted for  $p = 0.5$  and  $d = 0.8$ :

Figure  
19.1



In any realized sequence of years, the function above gets reset to zero every time the environment changes. And changes in the environment are random, with respect to the value of  $q_t$ . Remember (2P), **random** in these models just means we aren't modeling the causes. Of course changes in the environment are deterministic. Epistemology aside, an example series of years may look like:

Figure  
19.2



- 20 The red curves are again the proportion of the skilled individuals in the population. Every time the environment changes, the curve is reset to zero. This happens in years 0, 5, and 17, in the example above.

When  $q$  is high, the growth rate of  $S$  is high. When  $q$  is low, the growth rate is low. So natural selection fluctuates as the environment fluctuates. In the long run, the growth rate of  $S$  is determined by the distribution of  $q$  values, and that distribution is determined by the rate of environmental change. We haven't specified yet how the environment changes. There are many possibilities. But whatever we choose, let  $v(t)$  be the probability that the environment goes  $t$  years without changing. Then we can write the long-term growth rate of  $S$ . First, let's

define the fertility of  $S$  using  $q_t$ :

$$f_{S,t} = b + q_t(B - b) - C_S = b + d(1 - p^{t+1})(B - b) - C_S$$

Embedding this in the same geometric mean growth expression introduced way back in **1F**, we get the long-term growth rate that averages over all the  $q$  values:

$$\log W_S = \sum_{t=0}^{\infty} v(t) \log f_{S,t} \quad (19.1)$$

A key implication of this expression is that small values of  $f_{S,t}$  will influence the growth rate more than large values.

Now supposing that gene frequencies change slowly, selection will respond to something called the **stationary distribution** of  $q_t$ . This means the distribution where each of the  $q_t$  values is weighted by its probability. So what is the probability of each  $t$ ? The simplest model is to let  $u$  be the probability the environment changes in any year.<sup>2</sup> This implies a specific probability distribution for the probability of each  $t$ :

$$v(t) = u(1 - u)^t$$

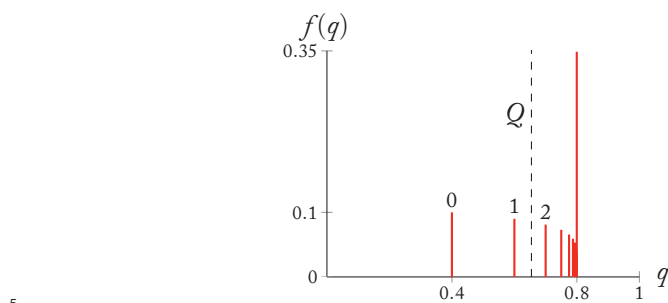
To motivate this function, consider that  $v(0) = u$ , because  $u$  is the chance the environment just changed. Now what is the chance of reaching  $t = 1$ ? For that to happen, you need  $u(1 - u)$ . For  $t = 2$ , you need  $u(1 - u)(1 - u)$ . And so on. This is a geometric distribution. It is important to check that this is a proper probability distribution, by checking that it sums to one. The fact that the sum has an infinite number of terms is no trouble. I show you the calculation on the next page.

What does the stationary distribution of  $q_t$  look like? Let's plot it, then we'll calculate with it. To visualize the distribution, imagine taking the curves in Figure 19.2 on the previous page and just plotting the histogram of the  $q$  values through the years. After enough years, the histogram stabilizes. That's the stationary distribution. Suppose for example that  $p = 0.5$ ,  $u = 0.1$ , and  $d = 0.8$ . Then the observable expected values of  $q_t$  are:

$t$	0	1	2	3	...
$q_t$	0.4	0.6	0.7	0.75	...
$v(t)$	0.1	0.09	0.081	0.0729	...

The first row are the  $t$  values, the number of years since the last change in the environment. Each of these  $t$  values implies a unique  $q_t$  and  $v(t)$ . The  $v(t)$  values tell us how often each of the unique  $q_t$  values will appear in a long time series. Let's plot these values:

<sup>2</sup>Real environmental change, at the scale people experience it at least, tends to be negatively autocorrelated. This means that change is unlikely to be quickly followed by more change. This favors more social learning than the independent-change model analyzed here.



The horizontal axis above shows the  $q$  values. The red bars show the expected proportions in a very long time series. Most  $q$  values are not realized, because the model is in discrete time. So the  $q$  values make discrete jumps. The black numbers 0, 1, 2 over the bars are  $t$  labels, the number of years after a change in the environment. The very tall red bar lies at  $q = d$ , the maximum value  $q$  can take. For  $u = 0.1$ , the population spends about 35% of its time near this maximum.

The analytical problem in this type of model is that the growth rate of  $S$  is different for each value of  $q$ . But the population doesn't spend very long at any particular value of  $q$ , as you can see above. So the long-term growth rate depends upon the distribution above, and that's what the growth rate expressed in Equation 19.1 says.

But such an expression cannot be analyzed in general—there is no way to algebraically manipulate it to close the infinite sum. So we're going to have to approximate it instead. In particular, we're going to assume **separation of time scales**. This means that we assert that the relative proportion of  $S$  is growing (or shrinking) slowly enough that the mean value of  $q_t$ , call it  $Q$ , is what matters.

To understand this, imagine a mountain stream. [need example]

Now to compute the expected value. The definition of the expected value is:

$$Q = \sum_{t=0}^{\infty} v(t)q_t = \sum_{t=0}^{\infty} u(1-u)^t d(1-p^t)$$

This is just an average, each term  $q_t$  multiplied by its probability  $v(t)$ . Then we sum up all the products to get the average. In the box on the next page, I show you how to close this infinite sum, yielding:

$$Q = \frac{d(1-u)(1-p)}{1 - (1-u)p} \quad (19.2)$$

This is the expected value of  $q$  at the stationary distribution when  $p$  changes very slowly relative to  $q$ .

We'll embed  $Q$  in our growth expression in a moment. But first, it's useful to evaluate  $Q$  at some relevant limits so that we can understand it better. Consider when  $u \rightarrow 0$ , no environmental change. In that case,  $Q = d$ . This makes sense,



*Proving  $v(t)$  sums to one.* You can confirm that  $v(t) = u(1-u)^t$  is a valid probability distribution by checking that it sums to one. Let  $Z$  be the sum of all the possible terms:

$$Z = \sum_{t=0}^{\infty} u(1-u)^t = u + u(1-u) + u(1-u)^2 + \dots$$

This is an infinite series, but it is an infinite **geometric series**. These are nearly always easy to close, by factoring the terms until you can express  $Z$  again on the right side of the equation. In this case, let's factor  $1-u$  out of every term after the first one:

$$Z = u + (1-u)(u + u(1-u) + u(1-u)^2 + \dots) = u + (1-u)Z$$

Solving for  $Z$ :

$$Z = \frac{u}{1 - (1-u)} = 1$$

All probability accounted for.

*Solving for the expected value of  $q$ .* Our goal is to close this infinite sum:

$$Q = \sum_{t=0}^{\infty} v(t)q_t = \sum_{t=0}^{\infty} u(1-u)^t d(1-p^t)$$

This is another infinite geometric series. We can start by factoring out  $ud$ , which is in every term:

$$\begin{aligned} Q &= ud \sum_{t=0}^{\infty} (1-u)^t (1-p^t) \\ &= ud(0 + (1-u)(1-p) + (1-u)^2(1-p^2) + (1-u)^3(1-p^3) + \dots) \end{aligned}$$

The infinite series inside the parentheses can be written as the difference between two different infinite geometric series. That might sound crazy. We already have one infinite series. Why double our trouble? Because we can close each separate series more easily. Let  $Z = (1-u)(1-p) + (1-u)^2(1-p^2) + \dots$ . Then:

$$\begin{aligned} Z &= ((1-u) + (1-u)^2 + (1-u)^3 + \dots) - ((1-u)p + (1-u)^2p^2 + (1-u)^3p^3 + \dots) \\ Z &= A - B \end{aligned}$$

Now we just need to close  $A$  and  $B$  separately. Both are simple geometric series and can be closed the same way. Here's how to do it for  $A$ :

$$\begin{aligned} A &= (1-u) + (1-u)^2 + (1-u)^3 + \dots = (1-u)(1+A) \\ A &= \frac{1-u}{1-(1-u)} = \frac{1-u}{u} \end{aligned}$$

The same steps applied to  $B$  yield  $B = \frac{(1-u)p}{1-(1-u)p}$ . In fact, any geometric series of the form  $\sum_{x=1}^{\infty} a^x = a/(1-a)$ . Now let's combine everything and simplify:

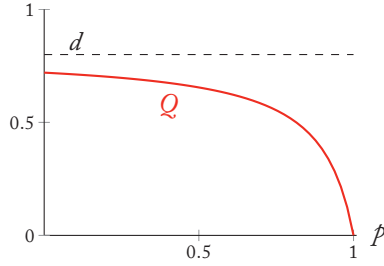
$$Q = ud \left( \frac{1-u}{u} - \frac{(1-u)p}{1-(1-u)p} \right) = \frac{d(1-u)(1-p)}{1-(1-u)p}$$

That's the result shown in the main text.

- 5 because when the environment never changes, then the proportion of skill is always as its maximum. On the reverse, when  $u \rightarrow 1$ , then  $Q = 0$ . If the environment always changes, then adults are never skilled, even though a proportion  $d$  of them become skilled in each year.

- 10 Now consider the limits of  $p$ . When  $p \rightarrow 0$ , no social learners,  $Q = d(1 - u)$ . When  $p \rightarrow 1$ , no innovation,  $Q = 0$ . It seems like increasing  $p$  reduces  $Q$ , and in fact you can prove that  $Q$  always declines with increasing  $p$ .<sup>3</sup> Here's a plot of  $Q$  against  $p$ , for  $d = 0.8$  and  $u = 0.1$ .

Figure  
19.4



- 15 Notice that  $Q$  is always less than  $d$ . An  $S$  individual will always have lower expected skill than an  $I$  individual. This doesn't mean social learning cannot evolve. Instead it means that, in this model, social learning can only evolve when it is sufficiently cheaper than innovation to make up for the less accurate learning.

So let's incorporate the learning costs now. The growth rate of  $I$  is just:

$$f_I = b + d(B - b) - C_I$$

- 20 And the expected growth rate of  $S$ , assuming selection is weak enough that it responds to the expected value  $Q$ , is:

$$f_S = b + Q(B - b) - C_S$$

$S$  will increase when  $f_S > f_I$ :

$$b + Q(B - b) - C_S > b + d(B - b) - C_I$$

$$(d - Q)(B - b) < C_I - C_S$$

- 25 The left side is the marginal cost of social learning. In this model,  $d - Q$  is always positive, because social learning is always less accurate than innovation. So the left side is essentially the skill advantage of innovation. The right side is the marginal cost of innovation. Social learning, in this model, is always cheaper than innovation. So the above states what is in hindsight obvious: social learning increases faster than innovation, when its marginal benefits outweigh its marginal costs.

What happens in the long run? Let's consider some special cases. Suppose  $p \approx 0$ , which means that  $S$  is rare. Can it increase in such a population? When

<sup>3</sup>To prove this, use calculus. Take the derivative of  $Q$  with respect to  $p$  and evaluate its sign.

$p \approx 0$ , then  $Q \approx d(1 - u)$ , so the condition becomes:

$$du(B - b) < C_I - C_S$$

This will be more interesting, if we express it as a condition on  $u$ :

$$u < \frac{C_I - C_S}{d(B - b)}$$

This says that when the rate of environmental change (the left side) is sufficiently small,  $S$  can increase when rare. More stable environments, in this model, favor social learning.

Now consider the opposite,  $p \approx 1$ . Now  $I$  is rare. When  $I$  is rare,  $Q \approx 0$ , yielding the condition for a rare  $I$  to increase faster than  $S$ :

$$d(B - b) > C_I - C_S$$

The left is the benefit of innovation and the right is the cost of innovation. Again, rather obvious. But note that this condition may not be satisfied. Suppose for example that  $d$  is small, so that on average innovation doesn't pay. Sure, once in a while, a rare innovator gets lucky and finds skilled behavior. But most innovators fail, still paying the cost  $C_I$ . So selection will not favor innovation over social learning in that case, even though social learning ends up no different than guessing.

Things are most interesting when both conditions above are satisfied. Then selection favors  $S$  when it  $I$  is common, and it favors  $I$  when  $S$  is common. What will happen in that case? There will be some mix in the population. This mix will exist at the value of  $p$  that makes  $f_S = f_I$ .

So let's find that value of  $p$ . But first let's plot the fertility expressions, so you can see what we're looking for. Here are  $f_I$  and  $f_S$  as functions of  $p$ , for  $b = B = C_I = 1$ ,  $C_S = 0.5$ ,  $u = 0.2$ , and  $d = 0.8$ .

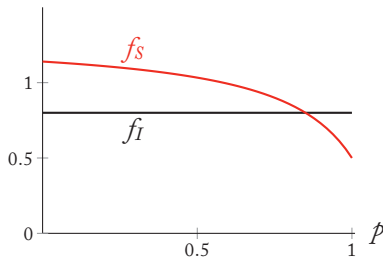


Figure  
19.5

On the left side,  $S$  grows faster than  $I$ . But as  $p$  increases,  $f_S$  declines until it intersects  $f_I$ . At the value of  $p$  where these functions meet, the two genotypes grow at the same rate. As a result, that value of  $p$  is an attractor. It isn't exactly "stable," in the traditional sense, because this model is stochastic and in a more realistic simulation, the population would fluctuate around  $p$ . But the intuition that the population is attracted to a mix of  $S$  and  $I$  holds.

And the conclusion from fitness functions shown above is that selection acts to increase social learning until its growth rate is the same as innovation. There are no population benefits to social learning.

Let's get more precise. Now let's find that point where  $f_I$  and  $f_S$  intersect. As usual in mathematics, the way to find the conditions for some situation is the assert the situation exists and then to rearrange the expression. The situation at the intersection is  $f_I = f_S$ . So we assert that:

$$b + d(B - b) - C_I = b + \frac{d(1 - u)(1 - p)}{1 - (1 - u)p}(B - b) - C_S$$

Then we solve for  $p$ . No tricks are necessary. But we'll want to rearrange the result until it has a natural grouping of the variables.

$$p = \frac{C_I - C_S + du(B - b)}{(C_I - C_S)(1 - u)} = \frac{1 + du \frac{B - b}{C_I - C_S}}{1 - u} = 1 - \frac{u}{1 - u} \left( d \frac{B - b}{C_I - C_S} - 1 \right)$$

This will be easier to interpret, if we rewrite this as:

$$p = 1 - U(\beta - 1)$$

where  $U = u/(1 - u)$  is the odds of environmental change and  $\beta = d(B - b)/(C_I - C_S)$  is the ratio of expected benefits of innovation to the expected costs of innovation. Expressed this way, you can see that this model really has only two variables that govern its dynamics,  $U$  and  $\beta$ . These are dimensionless parameters that map out how the system behaves. As the odds  $U$  increase, the proportion of  $S$  declines. And as  $\beta$  increases above 1, the proportion of  $S$  declines.

Okay, so far this model says that social learning can evolve through natural selection. But you've already seen a hint that it doesn't change the population growth rate. We just solved for the mix  $p$  of social learning and innovation where the two strategies have the same growth rate. This implies that a population with a proportion  $p$  social learners grows at the same rate as a population with zero social learners. Social learning evolves, but it doesn't provide any population benefits.

This result is the one called **Rogers' paradox**. A long-standing intuition was that social learning is adaptive, because it is efficient.<sup>4</sup> And that's true, at least in the model we just considered. But this isn't enough to produce any population benefits, implying that culture, defined as socially transmitted behavior, could be very common in nature. But it won't necessarily be any good for the species that have it. This in turn provides another example of the principle that natural selection doesn't necessarily produce adaptive behavior. Rather, if it produces any adaptation at all, it produces adaptive strategies. This in turn makes it much harder to test evolutionary hypotheses, because where behavior can be measured, strategy must always be inferred.

<sup>4</sup>Boyd and Richerson 1995 for the receipts.

This model is no reason to give up on the idea that social learning produces population benefits. It's just one model, the simplest I know. There are many alternative social learning models in which social learning does produce population benefits. We'll look at some of those in later sections and chapters.

However, the basic logic of this model holds across very many cultural evolution models. First, the amount of social learning that natural selection might favor is not simultaneously the amount of social learning that is optimal for population growth or persistence. This holds even in **cumulative cultural evolution** models, in which social learning may provide population benefits. Second, just because natural selection designs learning, that is no reason to expect learned behavior to be fitness enhancing. The **phenotypic gambit** fails in even remarkably simple models. That is no reason to reject the gambit. It's just a reminder that it is a gambit, not a principle.

### *Social learning, strong selection, and bet-hedging*

*Social learning experiences more variation in fitness, which both hurts its growth relative to innovation and favors bet-hedging.*

In the section above, I escape the trouble of evaluating the long-term growth rate of social learning

$$\log W_S = \sum_{t=0}^{\infty} v(t) \log f_{S,t}$$

by asserting that selection was weak enough that it responds only to the mean value of  $f_S$ , which is determined by the mean value of  $q = Q$ .

This approximation is okay for many purposes. The basic lesson of Rogers' paradox is not changed by a more general analysis. But there are some additional lessons to draw from the general analysis.

So how can make a more general analysis? Remember from way back in 1F that variation in growth tends to be bad for a lineage. So even without any further analysis, you can guess that social learning will suffer under stronger selection. Why? Because social learning has higher variance in growth than innovation does.

There is another important consequence of strong selection. In the analysis in the previous section, we consider competition between two genotypes, a pure  $I$  and a pure  $S$ . But consider a family of genotypes that combine these strategies in different proportions. Let  $\sigma$  be the probability that an individual uses social learning instead of innovation. Then there are potentially infinite genotypes with different values of  $\sigma$ . Which value of  $\sigma$  has the highest long-term lineage growth rate? When selection is weak, the optimal  $\sigma = p = 1 - U(\beta - 1)$ . The best mixed strategy has the same mix of social and individual learning as the population does and it has the same long-term growth rate.

- 5 But when selection is strong, it is much better to be a  $\sigma$  strategy than either pure strategy. The reason is that the  $\sigma$  strategy has lower variance in fitness—it can make it through the bottlenecks of environmental change, because some members of the lineage always innovate.

### *Conditional innovation*

- 10 *Strategies that use costly innovation only when social learning fails can benefit both individuals and populations.*

Consider a strategy that begins by deploying social learning. It tries to detect whether the socially learnable behavior would be adaptive, whether the social model is skilled. If so, the strategy invests the time to acquire the behavior through  
15 social learning. Otherwise, the strategy falls back on innovation.<sup>5</sup> This conditional innovation strategy is interesting, because it doesn't seem to require any new fancy cognition and it turns out to have rather different dynamics, both for individuals and the population, than the pure social and innovation strategies considered before.

- 20 This model has the same basic components as the previous model. Call the conditional strategy  $L$ , for “learner.” Specifically assume that  $L$  samples  $n$  adults, paying a fertility cost  $K$  to evaluate each. If any of the sampled adults is skilled, the individuals pays an additional fertility cost  $C_S$  to acquire skilled behavior. If none of the adults is skilled, the individuals pays  $C_I$  to innovate, succeeding to innovate  
25 skilled behavior with probability  $d$ .

As before, let  $u$  be the probability the environment changes. The fertility of  $L$  depends crucially upon when the environment most recently changed, because it depends upon the proportion of the adult population that is skilled. When the environment has just changed, it is:

$$(1 - d)b + dB - C_I - nK$$

- 30 which is exactly the same as innovation  $I$ , but with the extra cost  $nK$  for sampling. The reason is that when the environment just changed, none of the adults have the correct behavior, so  $L$  falls back on innovation. But it still pays the extra cost for sampling.

X

$$(1 - d)^n((1 - d)b + dB - C_I) + (1 - (1 - d)^n)(B - C_S) - Kn$$

1930

<sup>5</sup>Boyd and Richerson (1996) studied a model very similar to the presentation here. However the “Bayesian horticulturist” model in Boyd and Richerson 1985 is very similar, in that it deploys innovation conditionally. A later analysis by Enquist et al (2007) presents a more general and thorough analysis, with a focus on population benefits.