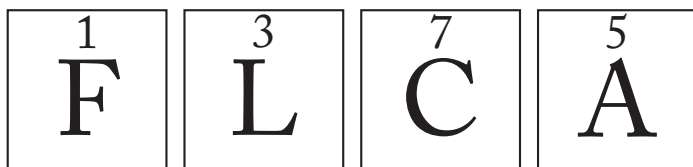


Elements of Evolutionary Anthropology

*A Theoretical Minimum for
Studying Human Evolution*



Richard McElreath

This version compiled April 17, 2021

Contents

Preface	ix
Part 1. Light Elements	1
Chapter 1. What the Fitness	3
Chapter 2. Probability	13
Chapter 3. Optimality	17
Chapter 4. Cooperation	19
Chapter 5. Kinship	21
Part 2. Transition Elements	23
Chapter 11. Senescence	25
Chapter 18. Parental Investment	37
Chapter 19. Social Learning	39
Bibliography	51

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.

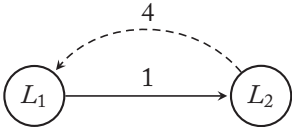
Fitness is most often defined as the **average reproductive success** of a certain genotype. Reproductive success means the number of offspring produced by an individual in its lifetime. Sometimes this is called “counting babies.” Counting babies isn’t a sound guide to which genotypes persist. You can 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.

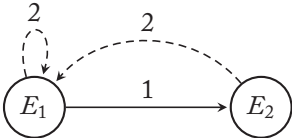
5 It will be useful to draw these alternative genotypes as life history diagrams. Let's start with the late-reproducing genotype, since it is simpler.

Figure 1.1



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



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

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:

- 30 (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.¹ 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

¹This 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 λ_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 λ_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 λ_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.² 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

[how to sensibly estimate lineage growth rate]

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

²Cite that evolution of menopause paper

² P

Probability

The greatest confusion with probability is believing it exists.¹

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.² Coin tosses are deterministic. Neither Janus nor any other deity influences the toss.

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

²Diaconis, 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
5 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
10 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
15 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
25 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
30 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 ships appear nearly as often in
35 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 ships will turn up. The coin doesn’t even have to be fair, have to present heads and ships with equal frequency in a very long sequence of tosses.

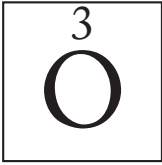
Why is a book about human evolution getting into the philosophy of randomness? Because all evolutionary theory is built from probability theory, and the
40 conventional view of probability is wrong. It’s not that all evolutionary theory is

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

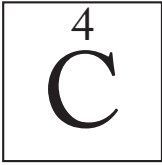


Optimality

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.

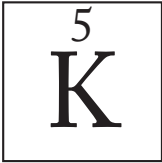


Cooperation

Etiam euismod. Fusce facilis 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 facilis 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.

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 facilis 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.



Kinship

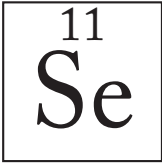
Etiam euismod. Fusce facilisis lacinia dui. Suspendisse potenti. In mi erat, cursus id, nonummy sed, ullamcorper eget, sapien. Praesent pretium, magna in eleifend egestas, pede pede pretium lorem, quis consectetuer tortor sapien facilisis magna. Mauris quis magna varius nulla scelerisque imperdiet. Aliquam 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.

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.

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.

Part 2

Transition Elements



Senescence

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.

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.

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.

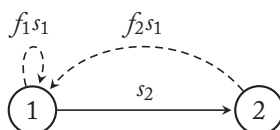
Extrinsic mortality

Imagine a species that lives for a maximum of two years.¹ 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

¹This example is based on the model in the appendix to [Day, T. and Abrams, P. A. (2020). Density dependence, senescence, and Williams' hypothesis. *Trends in Ecology and Evolution*, 35:300–302].

order to reproduce, an individual must first survive through the winter. Let s_i be the 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



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 $f_2 s_1$.

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.

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.

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**² of

²The 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.

the transition matrix for the population. It is absolutely worth knowing how to do 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 λ , 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 λ 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 λ . 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 λ 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 λ . 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:

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

Eigenvalues and eigenvectors. To compute the growth rate λ^* 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 λ 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 λ , so you get two solutions, and the largest will match λ^* 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 λ , 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.^a

^aCite Caswell & Takada 2004

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 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 λ^* . Why? Because if many genotypes are present in the population, then they will all be growing exponentially, but the one with the largest λ 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 $f_1 = 1$.

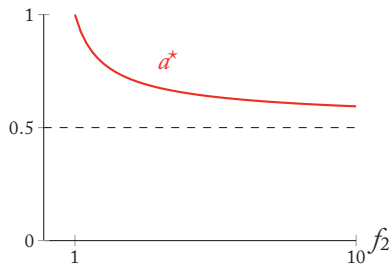


Figure
11.2

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

$$s_1 = (1 - x)a \text{ and } s_2 = (1 - x)(1 - a):$$

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

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 ℓ . Which conditions favor longer life? The expected lifespan in this model is:

$$\ell = \sum_{i=0}^2 \text{Pr}(\text{die at age } i)(i)$$

$$\ell = (1 - s_1)(0) + s_1(1 - s_2)(1) + s_1 s_2(2) = s_1(1 + s_2)$$

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 ℓ , 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 $k = 0.1$:

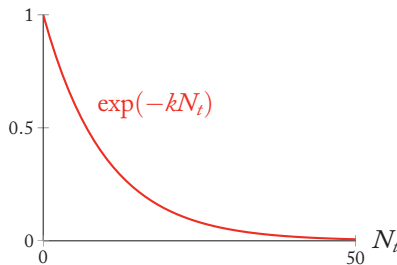


Figure
11.3

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

Substituting these new definitions into our old λ^* 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 λ^* will allow us to find the **evolutionarily stable** value of a . Like before, we want to find the value of a that maximizes λ^* . But unlike before, λ^* 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 , α , 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.

- 5 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
10 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 λ^* , while also requiring that we are at
15 demographic equilibrium. First let's find a function \hat{N} . We do this by setting $\lambda^* = 1$ and solving for N . This yields:

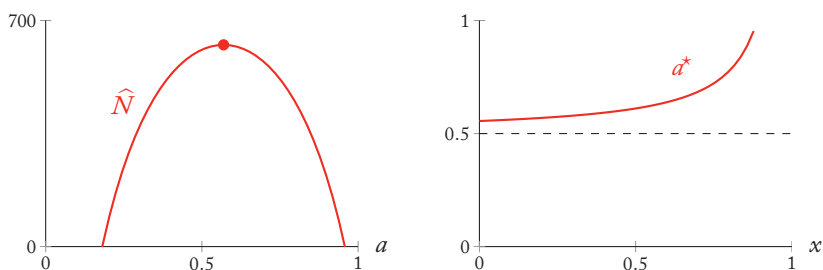
$$\hat{N} = \frac{1}{k} \log(a(1-x)(b_2(1-a)(1-x) + b_1))$$

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
20 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
25 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 ($\hat{N} = 0$).

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:

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

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:

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

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?

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 λ :

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

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

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

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 λ will adjust so that the sum of these terms is exactly 1. Likewise, if λ 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 λ 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

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.

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.

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.

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.

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 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.¹ Of course it is no paradox at all 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 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.

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:

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

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 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

¹Rogers 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:

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

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 .

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:

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

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:

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

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:

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

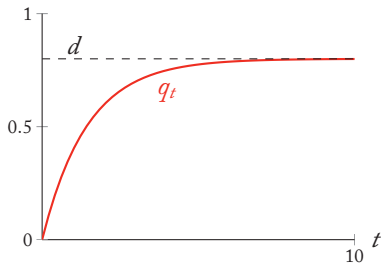
We'll plot this function in a moment. But we can learn some things about it right away. First, as t increases, q_t approaches d :

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

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.

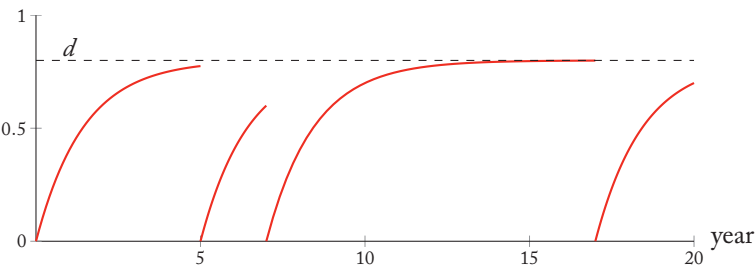
To make this more visceral, we can plot q_t as a function of t , showing how the proportion of skilled individuals increases since the last change in the environment. Here is q_t plotted for $p = 0.5$ and $d = 0.8$:

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



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. 5

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.² This implies a specific probability distribution for the probability of each t : 10

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

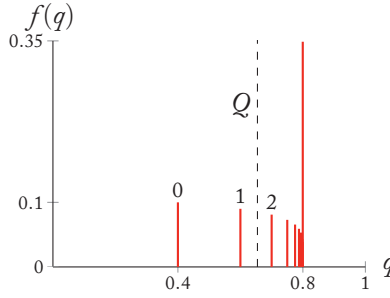
To motivate this function, consider that $v(0) = u$, because u is the chance the environment just changed. Now what is the chance of reaching $t = 1$? For that to happen, you need $u(1 - u)$. For $t = 2$, you need $u(1 - u)(1 - u)$. And so on. This is a geometric distribution. It is important to check 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. 15

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: 20

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: 25

²Real 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:

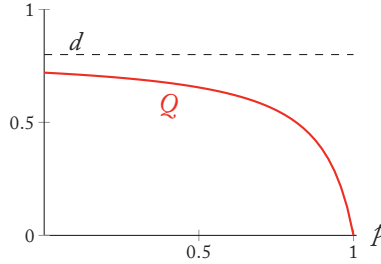
$$Q = ud \left(\frac{1-u}{u} - \frac{(1-u)p}{1-(1-u)p} \right) = \frac{d(1-u)(1-p)}{1-(1-u)p}$$

That's the result shown in the main text.

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.

- 5 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 .³ Here's a plot of Q against p , for $d = 0.8$ and $u = 0.1$.

Figure
19.4



- 10 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:

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

- 15 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 $f_S > f_I$:

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

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

- 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
25 $p \approx 0$, which means that S is rare. Can it increase in such a population? When

³To 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 $d = 0.8$.

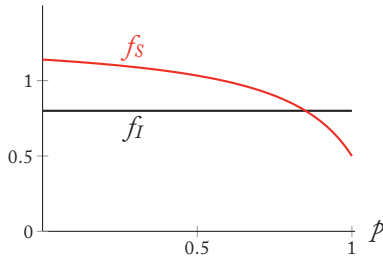


Figure
19.5

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.
- 5 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.

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

- 10 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 β . These are dimensionless parameters that map out how the system behaves. As the odds U increase, the proportion of S declines. And as β increases above 1, the proportion of S declines.
- 15

- 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.
- 20

- This result is the one called **Rogers' paradox**. A long-standing intuition was that social learning is adaptive, because it is efficient.⁴ 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.
- 30

⁴Boyd 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 σ be the probability that an individual uses social learning instead of innovation. Then there are potentially infinite genotypes with different values of σ . Which value of σ 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.

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

5 *Conditional innovation*

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
10 model is skilled. If so, the strategy invests the time to acquire the behavior through social learning. Otherwise, the strategy falls back on innovation.⁵ 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 be-
15 fore.

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
20 of the adults is skilled, the individuals pays C_I to innovate, succeeding to innovate 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
25 environment has just changed, it is:

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

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.

30 X

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

⁵Boyd 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.

Bibliography

Day, T. and Abrams, P. A. (2020). Density dependence, senescence, and Williams' hypothesis. *Trends in Ecology and Evolution*, 35:300–302.