Evolutionary Games in Natural, Social and Virtual Worlds[1]

Daniel Friedman[2] Barry Sinervo[3]

January 10, 2012

[1]acknowledgements go here.

[2]Economics Department, University of California, Santa Cruz, CA, 95064. *dan@ucsc.edu*

[3]Biology Dept, University of California, Santa Cruz, CA, 95064. *sinervo@biology.ucsc.edu*

Contents

Part I

Basics

Chapter 1

Population Dynamics

Evolutionary game theory draws on two different branches of mathematical theory — dynamical systems theory, and traditional game theory. This first chapter introduces basic ideas from the dynamical systems branch. It focuses on fitness and the discrete replicator equation, and looks at applications and extensions of these basic ideas when there is no strategic interaction.

Chapter 2 begins to introduce ideas from game theory, and takes a first look at dynamics when there is strategic interaction or, in biological jargon, frequency dependence. Chapter 3 looks at more complex forms of interaction, and Chapter 4 rounds out the basics by using traditional game theory to describe the equilibrium states. Chapter 5 takes a broader look at social interactions, and Chapter 6 develops a useful way to consider spatial aspects of interactions.

The rest of the book develops the ideas in the context of specific applications.

1.1 Fitness

This book is mainly concerned with how populations change over time. Often we focus on a single population and a set of two or more alternative strategies or traits $i = 1, 2, ...$ We denote by s_i the share (or fraction of the population) using strategy i, and ask how it changes over time.

Fitness is the key. Denoted by w_i, it refers to the ability of a strategy (or trait) to increase its share. Later we will see examples where fitness can be measured directly, e.g., via by tracking gene frequencies from one cohort to the next. In other cases, we can't observe fitness directly, but can find a good observable proxy, such as foraging success.

1.1.1 Examples

In a population of *Escherichia coli* bacteria, some fraction $s_1 > 0$ is resistant to penicillin, and the remaining fraction, $s_2 = 1 - s_1 > 0$ is not resistant. If there is no penicillin around, then the non resistant trait has higher fitness, $w_1 < w_2$, because it involves a more efficient metabolism. Over time, s_1 decreases and s_2 increases. On the other hand, when there is a high concentration of penicillin in the neighborhood, $w_1 >> w_2$, and s_1 quickly rises towards 1.0.

Suppose that there are only three types of memory chips that can be used in a computer. In some environments, their fitnesses w_1, w_2, w_3 depend only on the relative speeds. In other environments, price or capacity or some combination might determine fitness. The chip with highest fitness will gain share, and the share will dwindle for the chip with lowest fitness. Of course, according to Moore's famous law, after a few months a new chip will be introduced that has yet higher speed and greater capacity at no higher price. It will gain share at the expense of existing chips, until a still fitter chip appears.

The idea of fitness also applies to economic questions. For example, an organization can use an old fashioned hierarchy ($i = 1$: workers report to foremen who report to floor managers who report to assistant vice presidents who report to VPs who report to EVPs who report to the CEO) or a flexible governance structure ($i = 2$: teams form for specific tasks and dissolve when it is completed). In a new industry like smart phone software, the flexible stucture is likely to generate greater profits, so $w_1 < w_2$ and s_2 will increase. In a mature industry like coal mining, it is more likely that $w_1 > w_2$, and s_1 approaches 1.0.

Social customs can be analyzed from the same perspective. A taboo against eating pork products ($i = 1$, say) kept its adherents healthier in historical times and places when trichonosis and other diseases were prevalent. Thus it achieved high fitness and in some places became universally adopted, so there $s_1 = 1.0$.

The custom of primogeniture ($i = 1$: the oldest surviving son inherits the entire family estate) in much of the world 1000 years ago was rare compared to "fair" division among all surviving sons ($i = 2$). However, in medieval Europe, a single large estate could better protect itself than a collection of smaller estates, so once again $w_1 > w_2$, and s_1 rose towards 1.0.

In more recent times, one can find times and places in which investors like Warren Buffet who follow "fundamentalist" strategies in the stock market (e.g., you buy a stock whose price is below your estimate of its fundamental value, the present value of its future earnings) are more profitable and increase their share relative to "chartist" strategies (e.g., you buy a stock whose price recently exceeded 1.3 times its average over the last 6 months). And one can find other times and places in which the chartists do better and increase their share.

One last example. Since around 1800, economies oriented towards market exchange have typically generated more wealth than centrally controlled economies or feudal economies. Wealth is a good proxy for fitness, since greater wealth gives a country important military and diplomatic advantages. Hence the share of the world's economy dominated by the market has increased over time and now is, by most definitions, at least $s_1 = 0.8$. On the other hand, when Russia attempted a quick transition to a market economy in the early 1990s, the results were disastrous, arguably because Russia at that time lacked the necessary legal and political infrastructure. The fitness of a strategy is not absolute; it depends on the environment.

1.2 Tradeoffs

We defined the fitness of a trait as the ability to increase its share over time. But where does that ability come from, and how does it depend on environment? Traits often bring some advantages and some disadvantages relative to alternative traits, and then estimating fitness comes down to analyzing the tradeoffs.

1.2.1 Biological tradeoffs

The fitness of any biological trait has two components, survival and fecundity. Individuals with a given trait have to survive to maturity. The greater the survival rate, the greater the fitness. Second and equally important, those individuals have to reproduce, and bring into the world new individuals carrying the trait.

Often there is a tradeoff between these components. That trade-off is spotlighted in the distinction between r-strategists versus K-strategists in a given population. The r-strategy trait produces many progeny but these progeny recruit poorly into a dense population. For example, certain plant species produce vast numbers tiny seeds that don't germinate well except on open ground. By contrast, the K-strategy trait produces a few well-provisioned progeny that recruit well at high density. For example, other sorts of plants produce relatively few seeds, but with devices like shells or hooks that help them germinate and get established even with a thick ground cover.

Population growth is governed by the logistic equation

$$\frac{dN}{dt} = rN(1 - \frac{N}{K}). \tag{1.1}$$

The instantaneous change in N, the population size, is a function of the maximal growth rate r and carrying capacity K, which are fixed parameters in equation (1.1). Now suppose that there are r-strategists with a high maximal growth rate r_1 but low carrying capacity K_1, and a K-strategists with a high carrying capacity $K_2 > K_1$ but low maximal growth rate $r_2 < r_1$:

$$\frac{dN_1}{dt} = r_1 N_1(1 - \frac{N_1}{K_1}) \tag{1.2}$$

$$\frac{dN_2}{dt} = r_2 N_2(1 - \frac{N_2}{K_2}) \tag{1.3}$$

where $r_1 > r_2$ and $K_1 < K_2$.

The dynamics are simple if each type resides in a separate population. The r-strategists will increase more rapidly and rapidly approach the carrying capacity, K_1. The K-strategists will increase more slowly and but eventually will approach a larger population size $K_2 > K_1$.

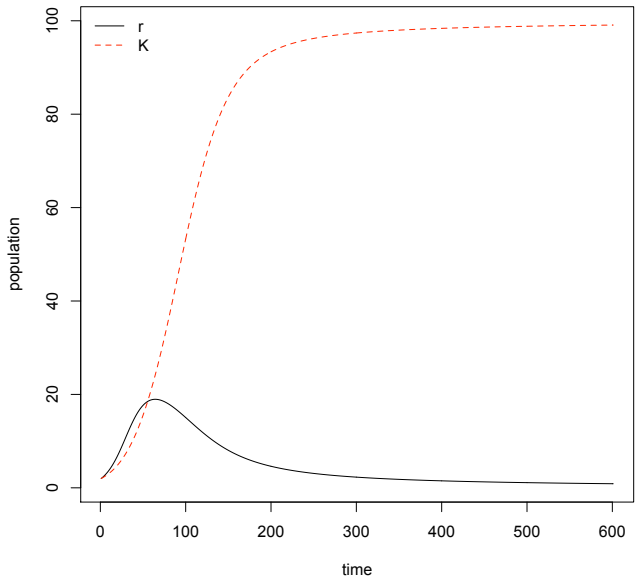

Figure 1.1: The rK-game (1.4 - 1.5) solved for the parameters $r_1 = 0.08, r_2 = .05, K = 50, c_{11} = 2, c_{12} = 0.5, c_{22}$=0.5, and c_{21}=0.5, which allows the K strategy to win.

More interesting dynamics ensue when both types are mixed in a single population. Here we must track the raw numbers in a population, and also look at the shares $s_1 = \frac{N_1}{N_1+N_2}$ and $s_2 = 1 - s_1$. The system is described by coupled differential equations similar to the Lotka competition equations used in ecology (Sinervo et al. 2000).

The rest of this subsection is not used later; it is included only for readers curious about r- vs K-strategists. The coupled differential equations are:

$$\frac{dN_1}{dt} = r_1 N_1 (1 - c_{11}\frac{N_1}{K} - c_{12}\frac{N_2}{K}) \tag{1.4}$$

$$\frac{dN_2}{dt} = r_2 N_2 (1 - c_{22}\frac{N_2}{K} - c_{21}\frac{N_1}{K}) \tag{1.5}$$

where K is the overall carrying capacity.

The parameters c_{11} and c_{22} specify the own-type competitive effects, while c_{ij} for $i \neq j$ specify the competitive effect of the other type j on type i. In these equations, a K-strategist type has lower values for c, since lowering c is equivalent to raising K in the denominator. For example, in Figure 1.1, strategy 2 has a lower c than strategy 1's value $c_{11} = 2$, so it is the K strategy (shown as a dotted line). Even though the r strategy 1 increases share faster

at first, it is eventually overtaken by the K strategy.

Equations (1.4 - 1.5) can also describe quite rich dynamics in situations where a K-strategist does well against the other type (e.g., c_{ij} for $i \neq j$) but relatively poorly against self types c_{ij} for $i = i$.

1.2.2 Investors' tradeoffs

Investors trade off risk against average return. According to standard theory (backed by a fair amount of historical evidence), some financial assets (such as US Treasury bills) reliably pay a steady income, while other assets (such as newly-listed stocks) on average pay a higher return, but are riskier in the sense that the return sometimes is low or even negative.

The tradeoff between risk and return depends on the investor's age. A young professional should target high returns and accept a fair amount of risk, because things should average out over the decades remaining before retirement. Her fitness or payoff (economists might try to measure this as the expected utility of consumption over her lifetime) is highest with a fairly risky portfolio. On the other hand, a retiree whose income mainly comes from his investments is much less able to tolerate risk, so he should accept a lower average return. His fitness is maximized with a much lower risk portfolio.

1.3 Memes

Evolution is the central organizing principle of biology. Its genetic mechanisms, some of which are quickly summarized in Section 1.7 below, are now are fairly well understood.

Many of the examples already used—stock market investors, food taboos, 1990s Russia, for example—are not governed by those biological mechanisms. There is no gene for markets or for stock market fundamentalism. Rather, individuals chose among alternative strategies, consciously or unconsciously, and over time the strategy shares change within the group of individuals.

Most of this book will focus on regularities that apply to any sort of evolution, whether or not the alternative strategies are coded in genes. When genes are not directly involved,

sometimes it is helpful to have an analogous word to describe the carriers of alternative strategies. Then, following Dawkins (1976), we will refer to *memes*.

Like genes, the fitness of memes has two components: persistence within individuals (survival), and transmission across individuals (fecundity). But these components work differently for memes than for genes.

Genes typically persist in individuals from birth to death. By contrast, an individual may switch memes many times within a single lifetime. Perhaps even more important, genes can only be transmitted "vertically," from parents to their children. Memes often transmit vertically, but they can also transmit horizontally, from peer to peer (especially among adolescents!), or diagonally. They can even transmit reverse vertically: one of the author's kids persuaded him buy and use a smart phone.

You might think that memes are more complicated to analyze because of all the possible transmission routes and individual switches. Yet genes can also be quite complicated, since in most species half of them come from each parent, and the mapping from genotype to phenotype is often unclear and complex. In any application, whether it involves genes or memes, you have to think through the components of fitness, and to identify the factors that seem most important.

Finding a useful simplification is often challenging, and it takes scientific insight and creativity. Once you have a tractable simplified model, things become much more routine because then you can then apply techniques explained in textbooks like the one in your hands.

1.4 State space geometry

An evolving system is modeled by listing the alternative strategies $i = 1, 2, ..., n$, and keeping track of the number of individuals N_i employing them in each population. Sometimes we need to study the population size $\sum_{i=1}^{n} N_i$, but usually we focus on the shares $s_i = \frac{N_i}{\sum_{i=1}^{n} N_i}$. Then the *state* of the system is the vector of shares $\mathbf{s} = (s_1, s_2, ..., s_n)$, and an evolutionary model seeks to describe how the state $\mathbf{s}$ changes over time.

The smallest number of individuals choosing a strategy is zero so $N_i \geq 0$, and any relevant

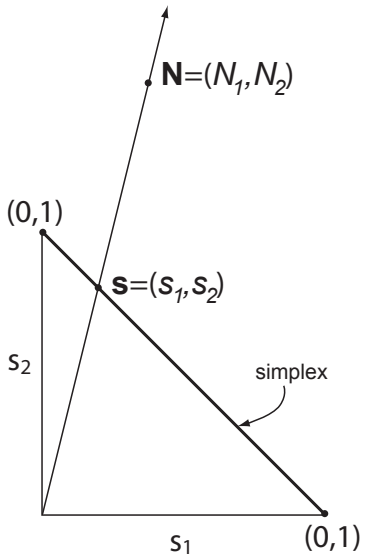

Figure 1.2: The geometry of the two-strategy simplex.

population contains some individuals so $\sum_{i=1}^{n} N_i > 0$. It follows that shares are non-negative ($s_i \geq 0$) and that they sum to $\sum_{i=1}^{n} s_i = \frac{\sum_{i=1}^{n} N_i}{\sum_{i=1}^{n} N_i} = 1$. Therefore any possible state of a given population is represented by a point in the *n-simplex*, the set of vectors whose components are non-negative and sum to 1. The evolutionary process is visualized as the state moving over time within its simplex.

Simplexes have a very nice geometry (Figure 1.2). To begin with the simplest case, consider two alternative strategies for individuals in a single population. The raw population vector is $\mathbf{N} = (N_1, N_2)$, a point in the positive quadrant of the 2-dimensional plane. Multiplying or dividing the components of $\mathbf{N}$ any positive number can be visualized as a move along the ray from the origin (0,0) through $\mathbf{N}$. Dividing by the sum of the components, in order to get the corresponding state, picks out the point where that ray intersects the line segment $s_1 + s_2 = 1$. Hence the simplex is the one-dimensional line segment that connects the point $\mathbf{s} = (s_1, s_2) = (1, 0)$, where everyone uses the first strategy, to the point $(s_1, s_2) = (0, 1)$, where everyone uses the second strategy. Alternatively, we can visualize this simplex by looking only at its first component s_1, with with s_2 understood to be $1 - s_1$, as in Panel A of Figure 1.3.

Often we will consider three possible strategies in a single population. The raw population vector $\mathbf{N} = (N_1, N_2, N_3)$ now is a point in the positive orthant of 3-dimensional space, and the

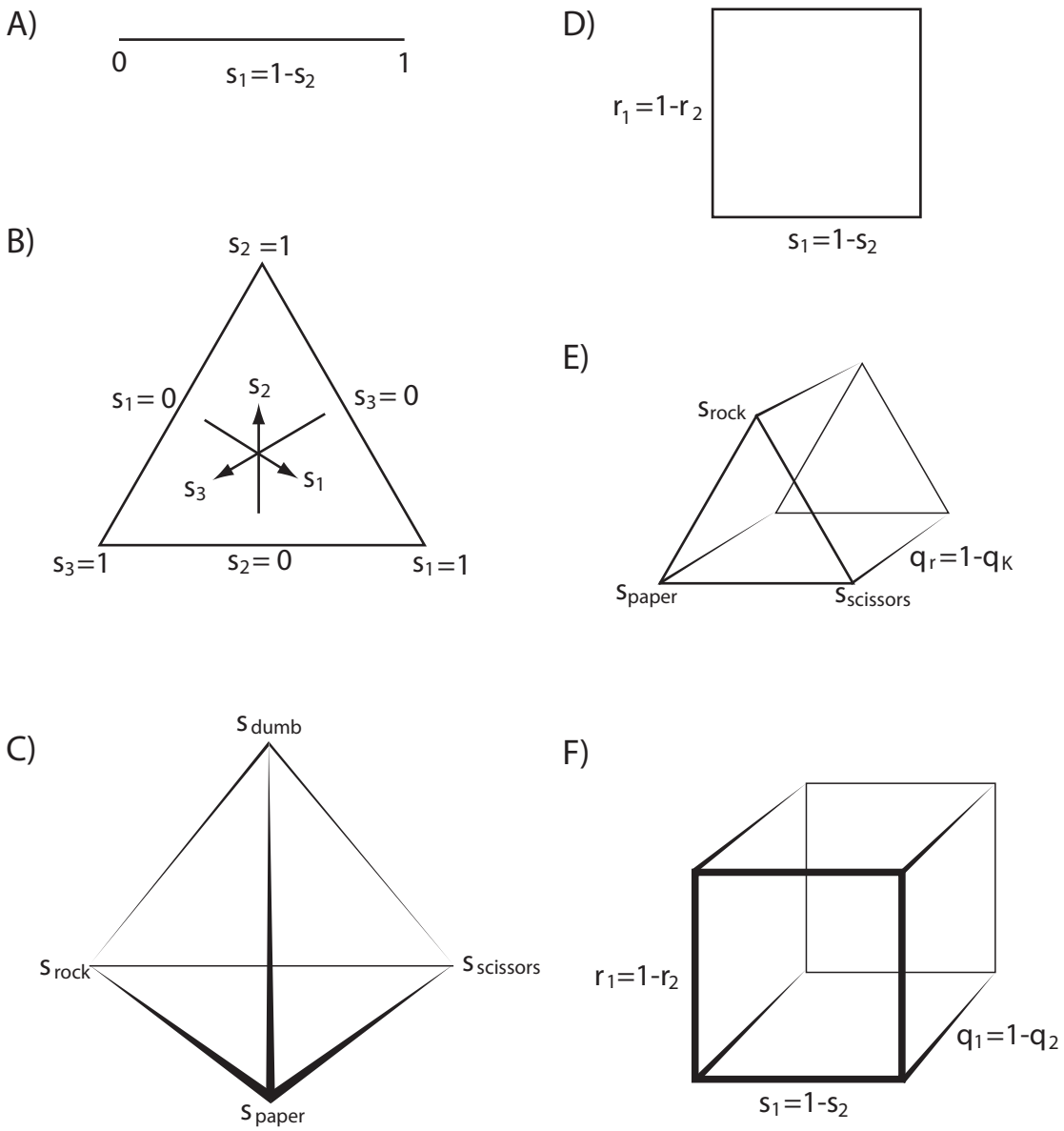

Figure 1.3: Single population state spaces for A) 2 alternative strategies, and B) 3 alternative strategies. Panel C) is for a game with 4 strategies, called rock, paper, scissors and dumb, discussed in Chapter 3. State spaces for simple two-population games are shown in D) and E), while F) is the state space for a game where each of 3 populations has 2 alternative strategies.

state $\mathbf{s} = (s_1, s_2, s_3)$ lies in a two dimensional subspace due to the constraint $s_1 + s_2 + s_3 = 1$. The simplex now is the equilateral triangle whose corners are the points $\mathbf{e}_1 = (1, 0, 0)$, $\mathbf{e}_2 = (0, 1, 0)$, and $\mathbf{e}_3 = (0, 0, 1)$. In Panel B of of Figure 1.3, this triangle is laid out flat on the page. Each corner represents a state in which everyone uses the same strategy i, and the opposite edge represents all states in which only the other two strategies are used. The share of strategy i increases steadily as we move back towards that corner via line segments parallel to the opposite edge.

The coordinates of any state $\mathbf{s}$ describe it as a weighted average of the corners of the triangle, $\mathbf{s} = (s_1, s_2, s_3) = s_1\mathbf{e}_1 + s_2\mathbf{e}_2 + s_3\mathbf{e}_3$. These sort of coordinates are known as Barycentric, which perhaps explains why they are so loved by one of the authors. Occasionally we will represent the same simplex in a different way. The projection of the simplex into the two dimensional $s_1 - s_2$ plane is the right triangle $\{(s_1, s_2) : s_i \geq 0, s_1 + s_2 \leq 1\}$. It also represents the state because, by setting $s_3 = 1 - s_1 - s_2$, we recover the omitted coordinate.

What if there are 4 alternative strategies for a single population? It is hard to visualize the 4 dimensional space in which the W_i's reside, but the simplex is 3 dimensional because it satisfies the linear constraint $s_1 + s_2 + s_3 + s_4 = 1$. The corners of this simplex are the points $\mathbf{e}_1 = (1, 0, 0, 0)$, $\mathbf{e}_2 = (0, 1, 0, 0)$, $\mathbf{e}_3 = (0, 0, 1, 0)$ and $\mathbf{e}_4 = (0, 0, 0, 1)$, and they can be arranged in 3 dimensional space to form the equilateral tetrahedron shown in Panel C of Figure 1.3. Again, each corner represents a pure state where everyone uses the indicated strategy, and the opposite face (a triangle) represents a mixed state where only the other three strategies are used. Edges represent states in which two strategies are used and the other two are not. Barycentric coordinates are natural here as well.

It is hard to draw on the page the $(n-1)$ dimensional simplex corresponding to $n = 5$ or more alternative strategies in a single population. Some readers may be able to able to visualize the corners, edges, etc. of these higher dimensional simplexes, but if you can't, don't worry! We will seldom need them in this book.

In later chapters we will consider evolutionary games involving two or more distinct populations. For example, in the biological realm, a predator population may co-evolve with a prey population, or a bumblebee population with a flower population. In a digital domain, the population of content providers may co-evolve with the population of end-users (consumers). Panel D illustrates the simplest such state space — one population has two

alternative strategies with shares s_1 and $s_2 = 1-s_1$, and it interacts with a second population also with two alternative strategies with shares denoted r_1 and $r_2 = 1 - r_1$. The state space is the cartesian product of the state spaces for each population, in this case, the cartesian product of two line segments, i.e., the unit square.

Panels E and F show two other possibilities. One population has 3 alternative strategies and the other has two in Panel E, so the result is a triangular prism, the cartesian product of a triangle and a line segment. In Panel F there are three distinct populations, each with two alternative strategies. The state space is the cube, the cartesian product of three line segments.

1.4.1 Finite vs infinite population models

With a finite population $\sum_{i=1}^{n} N_i < \infty$, the possible states are lattice points in the simplex, that is, each share in each population belongs to the set $\{0, \frac{1}{N_i}, \frac{2}{N_i}, ..., \frac{N_i-1}{N_i}, 1\}$. Even in continuous time, dynamics consist of *jumps* from one lattice point to another, and will generally be *stochastic*, i.e., some randomness in the jumps. The natural way to describe dynamics is via a Markov process.

For many purposes, we can get a very good approximation of dynamic behavior in large finite populations by using a model that assumes an infinite population. In continuous time, the changes are then usually *smooth*, with no jumps, and *deterministic*, with no randomness. This simplifies the mathematics considerably.

Most of the first part of the book works with dynamic models that describe changes in shares in continuous time using systems of ordinary differential equations (ODEs), or that describe shares in discrete time using systems of difference equations. We occasionally will deal with applications in which randomness is important (for example, genetic drift), and then we mention what happens when we model finite populations in more realistic detail.

1.5 Replicator dynamics

Consider now a simple model with two alternative strategies, $s = (s_1, s_2)$. Of course, shares are nonnegative ($s_i \geq 0$) and sum to unity ($s_1 + s_2 = 1$.) The initial state $s(0) = (s_1(0), s_2(0))$ is given. Suppose that the population is observed periodically (e.g., every summer) and the fitnesses W_1, W_2 are constant.

The discrete time replicator equation then describes the population over time. At times $t = 0, 1, 2, ...$ the state is given by

$$s_i(t+1) = \frac{W_i}{\overline{W}} s_i(t), \quad i = 1, 2, \tag{1.6}$$

where $\overline{W} = s_1 W_1 + s_2 W_2$ is the average fitness in the population.

The equation says that the share increases or decreases according to relative fitness. That is, the current share $s_i(t)$ is multiplied by its relative fitness $\frac{W_i}{\overline{W}}$. If relative fitness is greater than 1.0 — i.e., if fitness W_i is greater than average fitness $\overline{W}$ — then i's share will increase. It will decrease if relative fitness is less than 1.0.

For example, suppose $W_1 = 4$ and $W_2 = 3$ and the initial state is $s(0) = (s_1(0), s_2(0)) = (0.6, 0.4)$. Then $\overline{W} = s_1 W_1 + s_2 W_2 = 4(0.6) + 3(0.4) = 3.6$ and the next state has shares $s_1(1) = \frac{W_i}{\overline{W}} s_1(0) = \frac{4}{3.6} 0.6 = 2/3 > 0.6$, and $s_2(1) = \frac{3}{3.6} 0.4 = 1/3 < 0.4$.

Equation (1.6) is written for the case with only two alternative strategies. But the replicator equation looks exactly the same for any finite number $n > 2$ of alternatives: just replace $i = 1, 2$ by $i = 1, 2, ..., n$, and set $\overline{W} = \sum_{i=1}^{n} s_i W_i$.

Do replicator dynamics keep the state in the simplex? That is, at all times the shares must be nonnegative and sum to 1.0. If we have such a state at time t and apply (1.6), then we are asking: (a) are the shares still positive at time $t+1$ and (b) do they still sum to 1.0? Both answers are Yes in the simple example above, but what assurance do we have more generally?

Recall that fitness W_i is defined as the expected number of descendents in the next generation. This can't be a negative number. Unless the entire species immediately vanishes, at least one trait (or stategy) must have positive fitness. Therefore average fitness is positive and the ratio $\frac{W_i}{\overline{W}}$ is guaranteed to be nonnegative. Thus we have an affirmative answer to

(a): share i at time $t+1$ is the product of a nonnegative share at time t and a nonnegative ratio, and so must be nonnegative.

Question (b) also has an affirmative answer, as shown by the following algebra.

$$\sum_{i=1}^{n} s_i(t+1) = \sum_{i=1}^{n} \frac{W_i}{\overline{W}} s_i(t) = \frac{1}{\overline{W}} \sum_{i=1}^{n} W_i s_i(t) = \frac{1}{\overline{W}} \overline{W} = 1.0 \tag{1.7}$$

Replicator dynamics are very straightforward, and can be taken as the definition of fitness. Yet there are alternative dynamics worth considering, which we will do in Chapter 3.

1.5.1 Continuous time versions

One limitation of (1.6) is that it only gives the shares at discrete time intervals. In some cases that is all you would want to know. For example, adult Mayflies live only for a few weeks each year, and are in dormant forms most of the year. So it is reasonable to focus on shares of alternative Mayfly traits only once each year.

For many situations, however, we are interested in how shares change in continuous time. For example, advertisers want to know brand shares every month (or week or day), not just once a year, and you might want to know the how the state of the stockmarket (e.g., the fundamentalist share) changes minute by minute when you are about to make a big trade.

To see how to deal with shorter time intervals, first subtract $s_i(t)$ from both sides of equation (1.6) to obtain

$$\Delta s_i \equiv s_i(t+1) - s_i(t) = \frac{W_i}{\overline{W}} s_i(t) - s_i(t) = \left[\frac{W_i - \overline{W}}{\overline{W}} \right] s_i(t). \tag{1.8}$$

This is fine for $\Delta t = 1$, but what if $\Delta t < 1$? It makes sense to put an adjustment factor $c(\Delta t)$ in front of the expression, where $c(\Delta t) \to 0$ as $\Delta t \to 0$ — roughly speaking, the shorter the elapsed time, the smaller the change in the share. This gives us the more flexible expression

$$\Delta s_i = c(\Delta t) \left[\frac{W_i - \overline{W}}{\overline{W}} \right] s_i(t). \tag{1.9}$$

Divide both sides of (1.9) by Δt to obtain

$$\frac{\Delta s_i}{\Delta t} = \frac{c(\Delta t)}{\Delta t} \left[\frac{W_i - \overline{W}}{\overline{W}} \right] s_i(t). \tag{1.10}$$

Taking the limit as $\Delta t \to 0$ we get

$$\frac{ds_i}{dt} = \frac{c'(0)}{\overline{W}} \left[W_i - \overline{W}\right] s_i(t). \tag{1.11}$$

This equation looks even nicer if we chose the time scale (hours or weeks or half-days or whatever) so that $c'(0) = \overline{W}$ and use the notation $\frac{ds_i}{dt} = \dot{s}_i$. We then obtain the standard continuous time replicator equation

$$\dot{s}_i(t) = (W_i - \overline{W}) s_i(t). \tag{1.12}$$

This derivation has a caveat. Since the shares change over time, so does the (share-weighted) average fitness $\overline{W}$. Hence the time scale in (1.12) may not be quite proportional to the time scale in (1.6).

There is an alternative derivation of the continuous replicator equation that has a slightly different implementation. It goes as follows. Rather than putting in a flexible adjustment rate expression $c(\Delta t)$, instead generalize (1.6) directly by writing

$$s_i(t + \Delta t) = \left[\frac{W_i}{\overline{W}}\right]^{\Delta t} s_i(t). \tag{1.13}$$

This equation agrees with (1.6) when $\Delta t = 0$ or 1, but it also makes sense for intermediate values of Δt. Now take the natural logarithm of both sides to obtain

$$\ln s_i(t + \Delta t) = \Delta t \left[\ln W_i - \ln \overline{W}\right] + \ln s_i(t). \tag{1.14}$$

Subtract $\ln s_i(t)$ from both sides and divide both sides by Δt to obtain

$$\frac{\ln s_i(t + \Delta t) - \ln s_i(t)}{\Delta t} = \left[\ln W_i - \ln \overline{W}\right]. \tag{1.15}$$

Take the limit $\Delta t \to 0$ and the left hand side of (1.15) becomes $\frac{d \ln s_i(t)}{dt} = \frac{1}{s_i}\frac{ds_i}{dt} = \frac{\dot{s}_i}{s_i}$. Using the notation $w_i = \ln W_i$ and $\bar{w} = \ln \overline{W}$, the right hand side of (1.15) becomes $[w_i - \bar{w}]$. Multiply through by s_i and we get the very nice version of continuous replicator equation:

$$\dot{s}_i(t) = (w_i - \bar{w}) s_i(t), \quad i = 1, ..., n. \tag{1.16}$$

The main difference from (1.12) is that fitnesses are logarithmic. Negative w's correspond to W's less than 1, and positive w's correspond to W's above 1. (It turns out that logs, and

meaningful negative fitnesses, are convenient in data estimation.) Either way, relative fitness is the difference (not the ratio) between a strategy's fitness and the population average. A subtle difference is that average fitness of the w's corresponds to the geometric mean fitness of the W's, since $\sum_{i=1}^{n} s_i w_i$ is the natural log of $W_1^{s_1} W_2^{s_2} \cdots W_n^{s_n}$.

It is easy to check that (1.12) or (1.16) keeps the state in the simplex. The growth rate $\frac{\dot{s}_i}{s_i}$ can be positive or negative, but no growth rate can turn a positive number into a negative number. Summing (1.12) over all strategies, we get $\sum_{i=1}^{n} \dot{s}_i(t) = \sum_{i=1}^{n} (W_i - \overline{W}) s_i(t) = \sum_{i=1}^{n} W_i s_i(t) - \overline{W} \sum_{i=1}^{n} s_i(t) = \overline{W} - \overline{W} * 1 = 0$, that is, there is no change over time in the sum of the shares, so it remains at 1.0. The same algebra works for (1.16) as long as $\bar{w}$ is the arithmetic average of the w's (corresponding to the geometric average of the W's).

Consider again the simple example $W_1 = 4$, $W_2 = 3$ and $s(0) = (0.6, 0.4)$. Then $w_1 = \ln 4 \approx 1.4$ and $w_2 = \ln 3 \approx 1.1$. We get $\bar{w} = 0.6 w_1 + 0.4 w_2 \approx 1.27$. Then equation (1.16) says that $\dot{s}_1(0) = (w_i - \bar{w}) s_i(t) \approx (1.4 - 1.27) 0.6 = 0.078$. Thus $s_1(1) \approx s_1(0) + 1 * \dot{s}_1(0) \approx 0.678$, which is fairly close to the answer using discrete replicator, $s_1(1) = 2/3 \approx 0.667$.

1.6 Steady states and stability

Of special interest are situations where the system has settled down. These situations are easier to observe, and often tell us much of what we want to know about the system. This section will introduce the main ideas from the dynamic perspective: steady states, stability, and basins of attraction. Chapter 4 will link this dynamic perspective to the traditional static perspective of game theory, and will develop the ideas in much more detail.

A system is in **steady state** if the shares of all strategies are constant over time. That is, $\Delta s_i(t) = 0$ or $\dot{s}_i(t) = 0$ at all times $t \geq 0$. Inspection of (1.6) or (1.16) reveals that, under replicator dynamics, a state $s^* = (s_1^*, ..., s_n^*)$ is steady iff (shorthand for "if and only if") for each strategy i, either $s_i^* = 0$, or $W_i = \overline{W}$ (or both). That is, all surviving strategies achieve the same average fitness.

In the cases considered so far in this chapter, fitness is constant for each strategy. Except for a very unlikely coincidence in which two different strategies had exactly the same fitness, the steady states must be in pure strategies, at a corner or vertex of the simplex. (In the

next section we'll see how diploid genetics can lead to an interior steady state.)

Not all steady states are equally important. In the simple example with $W_1 = 4$ and $W_2 = 3$, we saw that from the initial state $s(0) = (0.6, 0.4)$ the share s_1 increased over time and s_2 decreased. Using spreadsheets or analytics, the reader can show that over time, shares converge to the steady state $s^* = (s_1^*, s_2^*) = (1, 0)$, and the same is true from almost all other initial states. The system never approaches the other steady state $s^* = (0, 1)$ unless it started there.

In some sense, then, $s^* = (1, 0)$ is a stable steady state and $s^* = (0, 1)$ is unstable. Dynamical systems researchers have settled on the following definitions to formalize these ideas. They apply to any dynamical process (replicator or otherwise) and any number of alternative strategies.

A steady state s^* is **stable** if, for any initial state $s(0)$ sufficiently near s^*, the entire trajectory $\{s(t) : t > 0\}$ remains near s^*. The full definition involves epsilons and deltas to pin down what is meant by "near." It will not enlighten many readers so we relegate it to the chapter endnotes. A steady state s^* is **locally asymptotically stable** if, for any initial state $s(0)$ sufficiently near s^*, the trajectory $\{s(t) : t > 0\}$ actually converges to s^*, denoted $s(t) \to s^*$ as $t \to \infty$. The open set of initial conditions from which the trajectory converges to a particular steady state s^* is called its **basin of attraction**. Thus an unstable steady state has an empty basin of attraction. If the entire interior of the simplex is included in the basin of attraction, then the steady state is **globally asymptotically stable**.

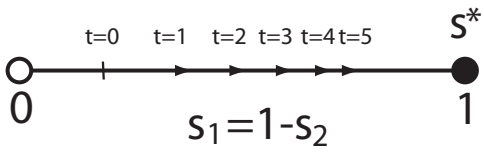

Figure 1.4: Phase portrait for continuous replicator (1.16). Fitnesses are $w_1 = 4$ and $w_2 = 3$. The length and the direction of arrows indicate the absolute value and the sign of the time derivative $\dot{s}$. Stable steady states are denoted by solid dots and unstable steady states by open circles.

Figure 1.4 illustrates the simple numerical example. You can see that the steady state

$s^* = (1, 0)$ is stable, locally asymptotically stable and, indeed, globally asymptotically stable, while the other steady state $s^* = (0, 1)$ is unstable in all three senses.

A warning is in order. Like many other authors, when we casually say "stable" we usually mean locally asymptotically stable, and we usually say "neutrally stable" when we mean stable but not asymptotically stable. We will use the jargon more carefully when it is important to do so.

1.7 An introduction to diploid dynamics

You, like most plants and animals and textbook authors, are a diploid creature and carry two sets of every gene. This section explains the formal structure of a classic diploid model. It can be skimmed quickly on first reading, because most of Part I of this book assumes a haploid form of inheritance and considers only phenotypic models. The haploid simplification is extremely useful and it captures many salient features (e.g., when a strategy can invade), but sometimes biological applications need to account for diploidy. Those applications build on the material presented here.

Some background facts and terminology. Humans are thought to carry approximately 23,000 individual genes. A fruitfly has half the number genes (13,500), but some "simple" organisms like the water flea *Daphnia* carry 31,000 genes, far more than humans. Genes reside on what are called autosomal chromosomes. Humans have 22 pairs of autosomal chromosomes (as well as 1 pair of sex chromosomes, the X and Y). Each chromosome is a very long double helical coil of the now-famous molecule called DNA.

Locus (plural loci) is the physical location of a gene on the chromosome. Each gene can come in alternative versions, called **alleles**. The **genotype** of an individual specifies which allele it carries in each locus under consideration. For example, Figure 1.5 tracks two loci labelled p and t, and each of them has two alternative alleles, labelled with the suffixes 1 and 2. The figure shows how sexual reproduction by a haploid organisms leads to new combinations via a structured random process called **segregation**. In the figure of a haploid organism with sex (a diploid organism with sex is only slightly different but a little more complicated), the female parent has genotype p1t2 — that is, both pairs of

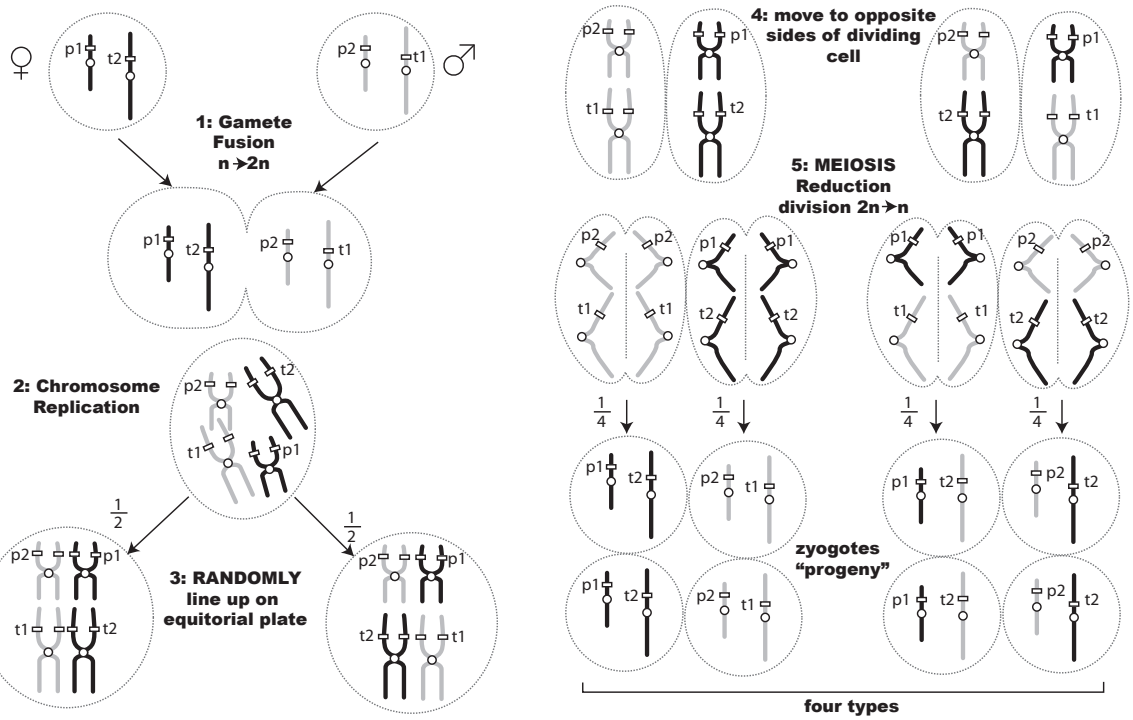

Figure 1.5: Random segregation in a haploid organism and tracking the genes on two of its chromosomes. In this example, alleles labeled t1 and t2 are available at one locus and alleles p1 and p2 at a second locus. Gamete union or sex (Step 1) doubles the number of chromosomes ($n \to 2n$) in the cell, after which each chromosome replicates (Step 2). Random segregation of chromosomes (or "crossover") in Step 3 induces variability in the progeny created in Step 4 (division of $2n$ cells) and Step 5 (meiosis, reducing the ploidy back to n).

one chromosome have allele 1 at locus p and both pairs of another chromosome allele 2 at locus t. The male parent has genotype p2t1, and the Figure begins with both parents contributing gametes, one of each of the chromosome pairs. The Figure shows that progeny are equally likely to have any one of 4 genotypes, including p1t2 and p2t1, the original parental genoytpes, as well as p1t1 and p2t2, the newly recombined parental genotypes. Simply put, sex serves a powerful role in recombining the genotypes of two parents. For the number of genes (and alleles at each locus) seen in humans, the number of recombined genotypes is an astronomically large number (yes, more than the number of stars in the universe). Let us backtrack a little to the topic of memes. Consider the possibility that memes from one parent can become recombined with the memes in the other parent. The process is not unlike sexual recombination. Haploid inheritance and diploid inheritance are

not the only possibilities. The chapter endnotes mention exceptions and mixed cases.

1.7.1 The Hardy-Weinberg model

In general, diploid dynamics are complicated because the forces of selection are modulated by migration, finite population size effects, and population structure. What do we mean by population structure? Take this book's two authors, for example. Author DF's genes are drawn from a pool shared by eastern European Jews, while author BS's come from a Finnish gene pool. Historically these gene pools had little contact, but now the authors find themselves in the fluid population structure of California's central coast. DF's progeny also have genes from a variety of northern and western European pools, and they have dispersed to where they may interact with other human gene pools in China and the southern US, while BS's progeny still resides in Santa Cruz.

Our point is that human population structure is very hard to actually pin down. Later chapters will show how populations with fairly simple spatial structures can give rise to amazing emergent patterns in evolutionary games. Even without spatial structure, we will see that mate selection can dramatically affect the dynamics.

For now, however, we put such complications aside and consider only completely random gene combinations in the classic Hardy-Weinberg model. It studies a large diploid population with "panmictic" reproduction. That is, at reproduction phase of the life cycle, all surviving individuals release gametes into a common well-mixed pool of haploid egg and haploid sperm. Sperm and egg then fuse randomly to form the next generation of diploid progeny.

That assumption enables us to easily compute the shares or gene frequencies for a single locus with two alleles labelled A and a. The three genotypes are AA, Aa, and aa, whose numbers are denoted N_{AA}, N_{Aa}, and N_{aa}. The total count of A's plus a's in the gene pool is $2[N_{AA} + N_{Aa} + N_{aa}]$, so the allele shares are

$$s_A = \frac{2N_{AA} + N_{Aa}}{2[N_{AA} + N_{Aa} + N_{aa}]}, \tag{1.17}$$

$$s_a = \frac{2N_{aa} + N_{Aa}}{2[N_{AA} + N_{Aa} + N_{aa}]}. \tag{1.18}$$

Appendix 1.8 confirms that (with no migration to or from an infinite panmictic population),

Genotype:	AA	Aa	aa
shares in diploid progeny genotypes after mating:	s_A^2	$2s_As_a$	s_a^2
absolute fitness during survival phase	W_{AA}	W_{Aa}	W_{aa}
relative fitness during survival phase	1	$1-m_{Aa}$	$1-m_{aa}$

Table 1.1: Diploid shares and fitness.

we get the diploid shares simply by multiplying:

$$s_{AA} = s_A s_A = s_A^2, \tag{1.19}$$

and

$$s_{aa} = s_a s_a = s_a^2. \tag{1.20}$$

The formula for heterozygotes Aa is only slightly more complicated since each allele can come via the egg or via the sperm:

$$s_{Aa} = s_A s_a + s_a s_A = 2s_A s_a. \tag{1.21}$$

1.7.2 Fisher's fundamental theorem

We'll now see how selection works in this simple Hardy-Weinberg model. To begin, note that each genotype has its own fitness. Here fitness differences are due entirely to different survival rates, since fecundity is the same for all genotypes in this panmictic model.

As shown in Table 1.7.2, the shortfall parameters m_{Aa} and m_{aa} describe the fitness of the the Aa and aa genotypes relative to the AA genotype. Figure 1.6 shows the possible cases. The generic case is drawn for $0 < m_{Aa} < m_{aa}$, the neutral case is $0 = m_{Aa} = m_{aa}$, while the last panel shows overdominance, with $m_{Aa} < 0 < m_{aa}$.

As usual, average fitness is $\overline{W} = \Sigma_i W_i s_i$ or, after normalizing as in Table 1.7.2,

$$\overline{W} = s_A^2 + 2s_A s_a(1-m_{Aa}) + s_a^2(1-m_{aa}) = s_A^2 + 2s_A s_a + s_a^2 - 2s_A s_a m_{Aa} - s_a^2 m_{aa} \tag{1.22}$$

Since $s_A^2 + 2s_A s_a + s_a^2 = (s_A + s_a)^2 = 1^2 = 1$, equation (1.22) simplifies to:

$$\overline{W} = 1 - 2s_A s_a m_{Aa} - s_a^2 m_{aa}, \tag{1.23}$$

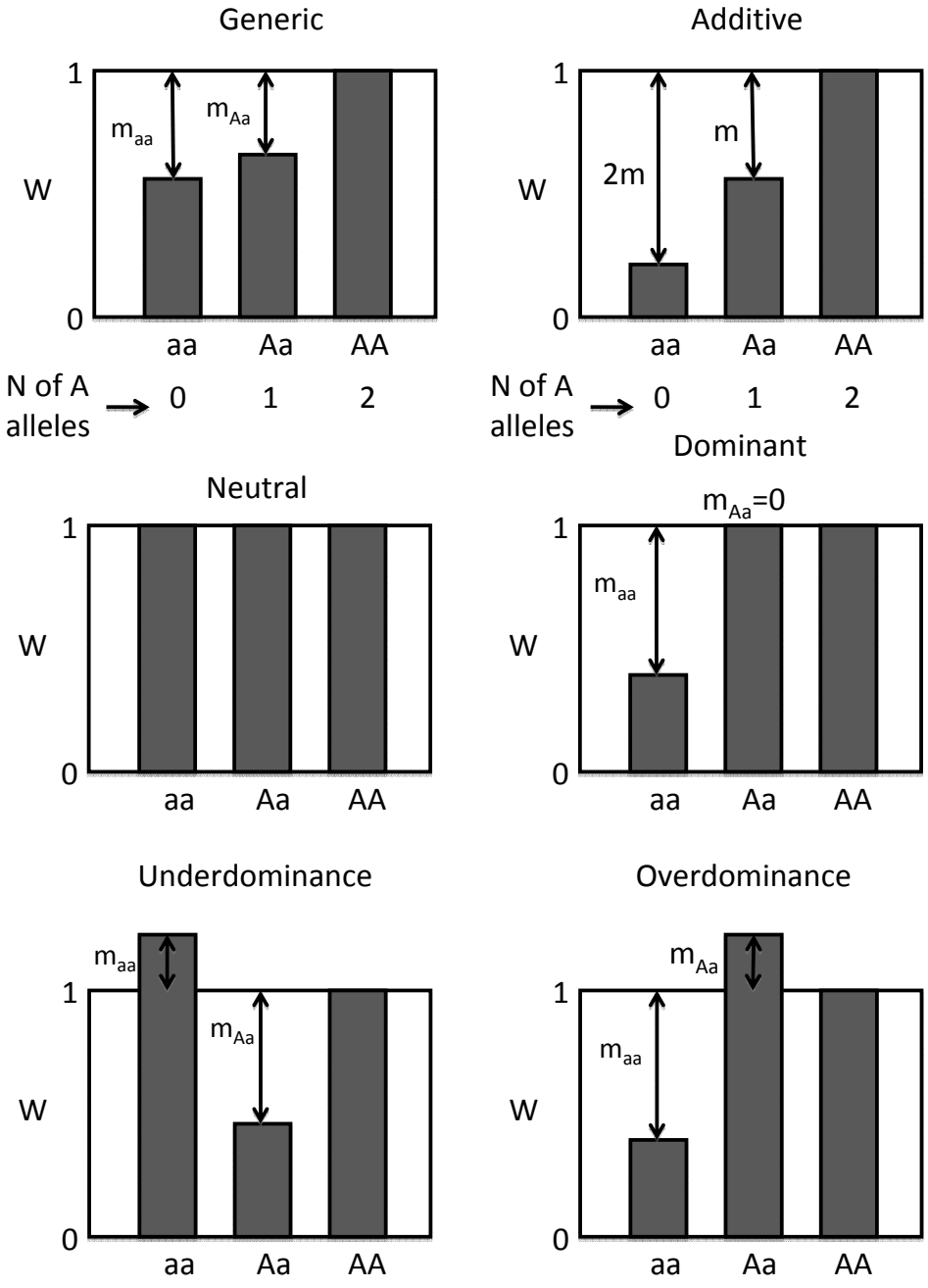

Figure 1.6: Diploid fitness for a single locus.

or, substituting out $s_A = 1 - s_a$,

$$\overline{W} = 1 - 2m_{Aa}s_a + 2m_{Aa}s_a^2 - m_{aa}s_a^2. \tag{1.24}$$

The analysis centers on the share s'_a of alleles in the next generation. Given current share s_a and fitnesses as in Table 1.7.2, we can compute $\Delta s_a = s'_a - s_a$:

$$\Delta s_a = s'_a - s_a = \frac{s_a^2 - m_{aa}s_a^2 + \frac{1}{2}(2s_A s_a - m_{Aa}2s_A s_a)}{\overline{W}} - s_a. \tag{1.25}$$

Substituting $1 - s_A$ for s_a in (1.25), we obtain:

$$\Delta s_a = \frac{s_a s_A[-m_{Aa} + 2m_{Aa}s_a - m_{aa}s_a]}{\overline{W}}. \tag{1.26}$$

This is still not all that pretty, but notice that in equation (1.22), the derivative of $\overline{W}$ with respect to s_a is

$$\frac{d\overline{W}}{ds_a} = \frac{d(1 - 2m_{Aa}s_a + 2m_{Aa}s_a^2 - m_{aa}s_a^2)}{ds_a} = 2[-m_{Aa} + 2m_{Aa}s_a - m_{aa}s_a]. \tag{1.27}$$

Thus we can rewrite Equation 1.26 more elegantly as

$$\Delta s_a = \frac{s_A s_a}{2\overline{W}} \frac{d\overline{W}}{ds_a}. \tag{1.28}$$

This last expression is a version of Fisher's Fundamental Theorem, a pillar of evolutionary biology. The factor $s_A s_a$ is the binomial variance for the haploid shares s_A, $s_a = 1 - s_A$. The factor $\frac{d\overline{W}}{ds_a}$ of course is the slope of average fitness as a function of the allele share s_a, i.e., the strength of selection. Thus equation (1.28) simply says that the rate of change in allele frequency (or share) is half the variance in allele frequency times the strength of selection. Figure 1.7 spells this out in a diagram. The variance factor captures the idea that a given strength of selection changes allele frequency more rapidly when both alleles have equal shares than when one of them is rare.

1.7.3 Overdominance and interior steady states

A numerical example will illustrate some of the implications of Fisher's theorem. Sickle cell anemia in humans is caused by a single point mutation, call it a, to the normal gene, call it A, for hemoglobin. Normal red blood cells are disk-shaped, but the alternative allele causes them to collapse into a sickle shape. In environments where malaria is prevalent, a moderate amount of sickling increases fitness because the body recycles its red blood cells faster when they sickle, and the malaria parasite (which lives on red blood cells) doesn't have as much

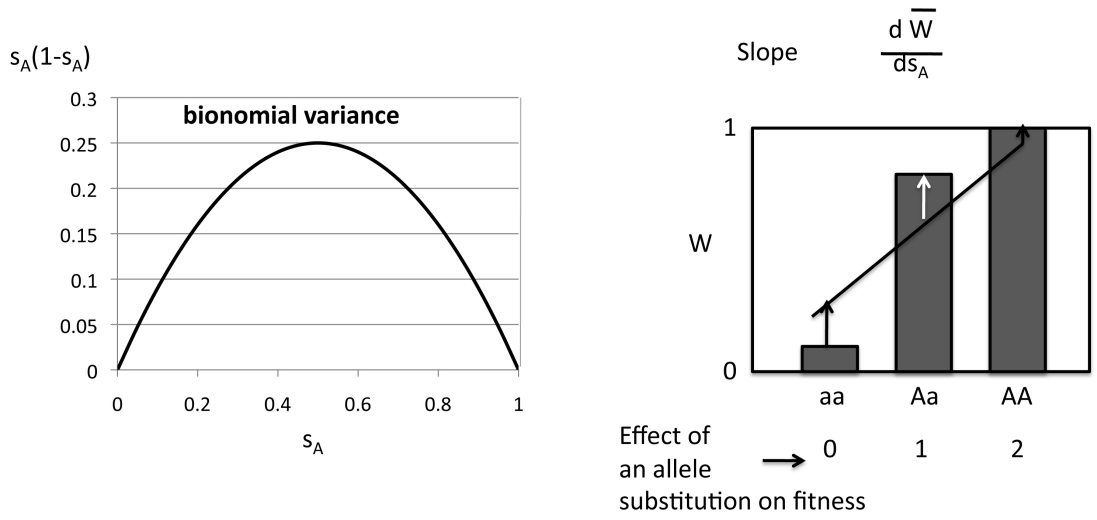

Figure 1.7: The two factors that change haploid shares.

time to grow. However, lots of sickling prevents the red blood cells from transporting much oxygen, greatly reducing fitness.

In a malarial environment, then, the fitness shortfall m_{aa} is positive, reflecting the greatly reduced fitness, but the shortfall m_{Aa} for the mixed genotype is negative, indicating enhanced fitness or "overdominance." Field studies by the World Health Organization produced the estimates $m_{Aa} = -0.6$ and $m_{aa} = 0.8$.

Suppose that the initial prevalence of sickle cell allele is small, say $s_a(0) = 0.001$. Then $\overline{W} \approx W_{AA} = 1$ and $\frac{d\overline{W}}{ds_a} \approx -2m_{Aa} = -2 * (-0.6) = 1.2$. Fisher's equation then says $\Delta s_a = \frac{s_A s_a}{2\overline{W}} \frac{d\overline{W}}{ds_a} \approx \frac{1*s_a}{2*1} * 1.2 = 0.6 s_a$, i.e., at first, the sickle cell gene's share grows exponentially, by almost 60% per period.

To find steady states, we solve $0 = \Delta s_a$. Of course, as usual, we have the steady states $s_a = 0$ and $s_A = 0$. More interestingly, we have an interior steady state if $\frac{d\overline{W}}{ds_a}$ has a root between 0 and 1. Using equation (1.27), the condition is $0 = \frac{d\overline{W}}{ds_a} = -2m_{Aa} + [4m_{Aa} - 2m_{aa}]s_a$, or $s_a^* = m_{Aa}/[2m_{Aa} - m_{aa}] = -0.6/[2 * (-0.6) - 0.8] = 0.3$. Thus, in malarial environments, we get an interior steady state due to overdominance, with about a 30% share for the sickle cell gene. Simulations suggest, and analytical methods introduced in Chapter 3 show, that the steady state is globally stable, as shown in Figure 1.8.

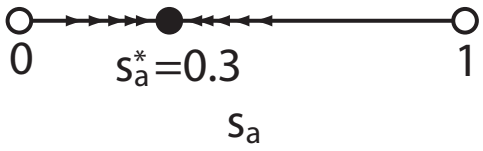

Figure 1.8: Phase portrait for the sickle cell allele, a. The stable interior steady state at $s_a^* = 0.3$ implies genotype shares approximately $s_{aa}^* = 0.3^2 = 0.09 = 9\%, s_{AA}^* = 0.7^2 = 49\%$ and $s_{aA}^* = 2(0.7)(0.3) = 42\%$ in a malarial environment.

1.8 Appendix

Detailed derivation of the Fisher equation. The Fisher equation has intrinsic importance, and its derivation illustrates mathematical tricks (attributed to the evolutionary biologist, J. B. S. Haldane) that we will use over and over in later chapters.

We begin the complete solution of Δs_a with equation (1.23)

$$\overline{W} = 1 - 2s_A s_a m_{Aa} - s_a^2 m_{aa} \tag{1.29}$$

equation (1.24), which uses the fact that $s_a = 1 - s_A$

$$\overline{W} = 1 - 2m_{Aa} s_a + 2m_{Aa} s_a^2 - m_{aa} s_a^2, \tag{1.30}$$

and a slight re-arrangement of equation (1.25) :

$$\Delta s_a = s_a' - s_a = \frac{s_a^2 - m_{aa} s_a^2 + \frac{1}{2}(2s_A s_a - m_{Aa} 2s_A s_a)}{\overline{W}} - s_a \tag{1.31}$$

We need to bring the final s_a over the common denominator $\overline{W}$, but the trick is to use Equation 1.29, not the formula for $\overline{W}$ expressed simply as s_a, because we will ultimately need to pull out a term for s_A and s_a (which together comprise the binomial variance) [N.B., the real trick here is we know where we are going, you can also get there by stumbling through the proof, which we have done in the past, but the quick proof requires forethought]:

$$\Delta s_a = s_a' - s_a = \frac{s_a^2 - m_{aa} s_a^2 + (s_A s_a - m_{Aa} s_A s_a)}{\overline{W}} - \frac{s_a(1 - 2s_A s_a m_{Aa} - s_a^2 m_{aa})}{\overline{W}} \tag{1.32}$$

immediately realize s_a is in every term so you can pull it out front:

$$\Delta s_a = s_a \frac{s_a - m_{aa}s_a + s_A - m_{Aa}s_A - 1 + 2s_A s_a m_{Aa} + s_a^2 m_{aa})}{\overline{W}} \tag{1.33}$$

then realize $s_a + s_A = 1$ which cancels with -1:

$$\Delta s_a = s_a \frac{(-m_{aa}s_a - m_{Aa}s_A + 2s_A s_a m_{Aa} + s_a^2 m_{aa})}{\overline{W}} \tag{1.34}$$

collect the terms in m_{aa}:

$$\Delta s_a = s_a \frac{(-m_{Aa}s_A + 2s_A s_a m_{Aa} + s_a^2 m_{aa} - m_{aa}s_a)}{\overline{W}} \tag{1.35}$$

and pull out $s_a m_{aa}$ in the last term:

$$\Delta s_a = s_a \frac{(-m_{Aa}s_A + 2s_A s_a m_{Aa} + s_a m_{aa}(-1 + s_a))}{\overline{W}} \tag{1.36}$$

realize $-s_A = -1 + s_a$:

$$\Delta s_a = s_a \frac{(-m_{Aa}s_A + 2s_A s_a m_{Aa} - s_a s_A m_{aa})}{\overline{W}} \tag{1.37}$$

pull out s_A to the front:

$$\Delta s_a = s_a s_A \frac{(-m_{Aa} + 2s_a m_{Aa} - s_a m_{aa})}{\overline{W}} \tag{1.38}$$

substitute in $\frac{1}{2}\frac{d\overline{W}}{ds_a}$:

$$\Delta s_a = \frac{s_a s_A}{2\overline{W}} \frac{d\overline{W}}{ds_a} \tag{1.39}$$

which is the classic solution of Δs_a. In the problem set you can practice these algebraic skills on the simplex by solving for Δs_A.

1.9 Exercises

1. Insert spreadsheet exercise here for simple replicator equation.

2. Insert a problem recycled from an old exam here.

3. Show that the continuous replicator equation (12) is unchanged when you replace the **w** matrix by another matrix **v** whose first column is shifted by any constant c_1, i.e., $v_{11} = w_{11} + c_1$ and $v_{21} = w_{21} + c_1$. Then show that it is still unchanged by shifting the second row by any constant c_2. *Extra credit*: show that for the continuous replicator equation for an n dimensional state, each column vector can be shifted by a different constant without affecting replicator dynamics, so it is without loss of generality to assume that **w** has zero entries along the main diagonal, i.e., that fitness is normalized relative to the common type.

4. The last problem suggests the conjecture that discrete replicator dynamics (1.6) are unchanged when any row is multiplied by a constant; this would justify assuming that **W** had 1s down the main diagonal. Show by example that this is not so.

5. (a) Section 1.7.3 showed that there is a unique interior steady state in the overdominance case ($m_{Aa} < 0 < m_{aa}$) of the Hardy-Weinberg single locus model. Show that the same is true in the underdominance case ($m_{Aa} > 0 > m_{aa}$).
   (b) Adapt the spreadsheets in exercise (1) above to confirm that the interior steady state is stable in the overdominance case but that it is unstable (and the corner steady states are stable) in the underdominance case.

6. Try solving Equation (**??**). This will allow you to practice a variety of substitutions of shares on a simplex.

1.10 Notes and Readings

Chapter endnotes provide additional details, tangents and guides to the literature.

Section 1. For more on Russia's transition in the 1990s, see (Friedman, 2008, Ch 5).

Section 1.2.2. For more discussion of this point, see any personal financial advisor or a textbook such as xxx

Section 1.3. Mention the meme controversy; many authors prefer to speak of social or cultural selection or evolution...cites in old reading lists.

[also mention jumping genes as an exception to the general rule that genetic transmission is vertical.]

Section 1.6. Here is the formal definition of stability. A steady state s^* is (Lyapunov) *stable* if $\forall \epsilon > 0$ $\exists \delta > 0$ s.t. $|s(0) - s^*| < \delta \Rightarrow |s(t) - s^*| < \epsilon$ $\forall t > 0$. There are bizarre examples of states that are asymptotically stable but not stable. You are very unlikely to encounter them in applications.

Section 1.7. In many cases, select components of the total genome of even diploid organisms exhibit haploid inheritance. For example, the Y-chromosome of males behaves as a haploid genetic element in the population of males. Likewise, mitochrondria, which are passed down matrilineally from mother to progeny behaves as a haploid in the population of females. There are unfortunately no cases in which the diploid component of the genome behaves in ways that even approximate haploids. This is because the genomes of diploids undergo segregation and recombination and this greatly complicates the algebra of games for diploids. However, in the case in which males and females share the same game, we can follow the fate of the additive effect of alleles from one generation to the next in a simple case, outlined below to illustrate the kind of algebraic complexity entailed in diploid games. We illustrate the simplification afforded by focusing on the gene frequencies and their additive effects on phenotype in the next three sections. The structure of all population genetic models necessarily requires diploid genetics for diploid organisms like animals and plants. Classic evolutionary game theory that originates with the publication Maynard Smith (1982) typically ignores this detail. Maynard Smith (1982) was however, very clear that he only considered the broad class of phenotypic models, which ignores this diploid structure to genes.

The theoretical development of the field of population genetics (Provine 1971) includes an incredibly rich set of tools for analyzing the dynamics and stability of frequency dependent and general games of selection.

Readings.

collect and insert biblio.

G.H. Hardy and A? Weinberg..

Field studies by the World Health Organization

Comments on virus study, etc etc.