Chapter 10: Evolutionary Games

We have so far studied games with many different features — simultaneous and sequential moves, zero and non-zero sum payoffs, strategic moves to manipulate rules of games to come, one-shot and repeated play, and so on. However, one ground rule has remained unchanged in all of our discussions, namely that all the players in all these games are rational — each player has an internally consistent value systems, can calculate the consequences of his strategic choices, and makes the choice that best favors his interests.

In this we merely follow the route taken by most of game theory, which was developed mainly by economists. Economics was founded on the dual assumptions of rational behavior and equilibrium. Indeed, these assumptions have proved useful in game theory. We have obtained quite a good understanding of games in which the players participate sufficiently regularly to have learnt the best choices by experience. The assumptions ensure that a player does not attribute any false naivete to his rivals and thus does not get exploited by these rivals. The theory also gives some prescriptive guidance to players as to how they should play.

However, other social scientists are much more skeptical of the rationality assumption, and therefore of a theory built upon such a foundation. Economists, too, should not take rationality for granted as we pointed out in Chapter 7. The trouble is finding a viable alternative. While we may not wish to impose conscious and perfectly calculating rationality, we do not want to abandon the idea that some strategies are better than others. We do want good strategies to be rewarded with higher payoffs; we do want players to observe or imitate success, and to experiment with new strategies; we do want good strategies to be used more often, and bad strategies less often, as players gain experience playing the game. We find one possible alternative in the biological theory of evolution and evolutionary dynamics and devote this chapter to studying its lessons.

1. The Framework

The process of evolution in biology offers a particularly attractive parallel to the theory of games used by social scientists. This theory rests on three fundamentals: heterogeneity, fitness, and selection. The starting point is that a significant part of animal behavior is genetically determined; a complex of one or more genes (genotype) governs a particular pattern of behavior, called a behavioral phenotype. Natural diversity of the gene pool ensures a heterogeneity of phenotypes in the population. Some behaviors are better suited than others to the prevailing conditions and the success of a phenotype is given a quantitative measure called its fitness. People are used to thinking in terms of the common but misleading phrase “survival of the fittest,” but the ultimate test of biological fitness is not mere survival, it is reproductive success. That is what enables an animal to pass on its genes to the next generation, and perpetuate its phenotype. The fitter phenotypes will then become relatively more numerous in the next generation than the less fit phenotypes. This process of selection is the dynamic that changes the mix of genotypes and phenotypes, and perhaps leads to an eventual stable state.
From time to time, chance produces new genetic mutations. Many of these mutations produce behaviors (that is, phenotypes) that are ill-suited to the environment, and they die out. But occasionally a mutation leads to a new phenotype that is fitter. Then such a mutant gene can successfully invade a population, that is, spread to become a significant proportion of the population.

Biologists call a particular phenotype evolutionary stable if its population cannot be invaded successfully by any mutant. This is a static test, but often a more dynamic criterion is applied: a phenotype is evolutionary stable if, starting from any mixture of phenotypes, the population evolves into a state where this type predominates.¹

The fitness of a phenotype depends on the relationship of the individual animal to its environment; for example, the fitness of a particular bird depends on the aerodynamic characteristics of its wings. It also depends on the whole complex of the proportions of different phenotypes that exist in the environment, how aerodynamic its wings are relative to the rest of its species. Thus the fitness of a particular animal, with its behavioral traits like aggression and sociability, depends on whether others in its species are predominantly aggressive or passive, crowded or dispersed, etc. For our purpose this interaction among phenotypes within a species is the most interesting aspect of the story. Of course, sometimes an individual interacts with members of another species; then the fitness of a particular type of sheep may depend on the traits that prevail in the local population of wolves. We consider this type of interaction as well, but only after we have covered the within-species case.

The basic idea of evolutionary games in biology is that the strategy of an animal in such an interaction - for example to fight or to retreat - is not calculated, but predetermined by its phenotype. The population being a mix of phenotypes, different pairs selected from it will bring to their interaction different combinations of such strategies. Those animals whose strategies are better suited for these interactions, on average against all the phenotypes they might encounter, will have greater evolutionary success. The eventual outcome of the population dynamics, or a strategy that cannot be upset by any successful invasion of another, will be called an “evolutionary stable strategy” (ESS).

Biologists have used this approach very successfully. Combinations of aggressive and cooperative behavior, locations of nesting sites, and many more phenomena that elude more conventional explanations can be understood as the stable outcomes of an evolutionary process of selection of fitter strategies. Interestingly, biologists developed the idea of evolutionary games using the preexisting body of game theory, drawing from its language but modifying its assumption of

¹ Underlying these changes in the proportions of the phenotypes are changes in genotypes, but this branch of biology usually focuses its analysis at the phenotype level and does not explicitly consider the genetic aspects of evolution.
conscious maximizing to suit their needs. Now game theorists are in turn using insights from the research on biological evolutionary games to enrich their own subject.²

Indeed, the theory of evolutionary games seems a ready-made framework for a new approach to game theory, relaxing the assumption of rational behavior.³ According to this view of games, individual players have no freedom to choose their strategy at all. Some are “born” to play one strategy; others, another. The idea is interpreted more broadly than it would be in biology. In human interactions, a strategy may be embedded in a player’s mind for a variety of reasons — not merely genetics but also (and probably more importantly) socialization, cultural background, education, or a rule of thumb based on past experience. The population can consist of a mixture of different individuals with different backgrounds or experiences that embed different strategies into them. Thus, some politicians may be motivated to adhere to certain moral or ethical codes even at the cost of electoral success, while others are mainly concerned with their own re-election; similarly, some firms may pursue profit alone, while others are motivated by social or ecological objectives.

From a population with its heterogeneity of embedded strategies, pairs of players are selected to interact (play the game) repeatedly, with others of the same or different “species”. In each interaction the payoff of each player depends on the strategies of both; this is governed by the usual “rules of the game” and reflected in the game table or tree. The fitness of a particular strategy is defined as its aggregate or average payoff in its matchings with all the strategies in the population. Some strategies will have a higher level of fitness than others and in the next generation, that is, the next round of play, these will be used by more players and will proliferate; strategies with lower fitness will be used by fewer players and will decay or die out. The central question is whether this process of selective proliferation or decay of certain strategies in the population will have an evolutionary stable outcome, and if so, what it will be. In terms of our examples just above, will


³ Indeed, applications of the evolutionary perspective need not stop there. The following joke offers an “evolutionary theory of gravitation” as an alternative to Newton’s or Einstein’s physical theories:

Question: Why does an apple fall from the tree to the earth?
Answer: Originally, apples that came loose from trees went in all directions. But only those that were genetically predisposed to fall to the earth could reproduce.
society end up with a situation in which all politicians are concerned with re-election, and all finns with profit? In this chapter we develop the framework and methods for answering such questions.

Although we use the biological analogy, the reason why the fitter strategies proliferate and the less fit ones die out in socio-economic games differs from the strict genetic mechanism of biology: players who fared well in the last round will transmit the information to their friends and colleagues playing the next round, those who fared poorly in the last round will observe which strategies succeeded better and try to imitate them, and some purposer thinking and revision of previous rules of thumb will occur between successive rounds. Such “social” and “educational” mechanisms of transmission are far more important in most strategic games than any biological genetics and, indeed, this is how the reelection-orientation of legislators or the profit-maximization motive of finns is reinforced. Finally, conscious experimentation with new strategies substitutes for the accidental mutation in biological games.

Outcomes of biological games include an interesting new possibility. A single phenotype need not prevail completely in the eventual outcome of evolutionary dynamics. Two or more phenotypes may be equally fit, and can coexist in certain proportions. Then the population is said to exhibit polymorphism, that is, a multiplicity (poly) of forms (morph). Such a state will be stable if no new phenotype or feasible mutant can achieve a higher fitness against such a population than the fitness of the types that are already present in the polymorphic population.

Polymorphism comes close to the game-theoretic notion of a mixed strategy. However, there is an important difference. To get polymorphism, no individual player need follow a mixed strategy. Each can follow a pure strategy, but the population exhibits a mixture because different individuals pursue different pure strategies.

In this chapter we develop some of these ideas, as usual through a series of illustrative examples. We begin with symmetric games, where the two players are on similar footing, for example two members of the same species competing with each other for food or mates; in a social science interpretation, they could be two elected officials competing for the right to continue in public office. In terms of the payoff table for the game, each could be designated as either the row player or the column player with no difference.

2. Prisoners’ Dilemma

Suppose a population is made up of two phenotypes. One consists of those who are natural born cooperators; they do not confess under questioning. The other type consists of the defectors; they confess readily. The payoffs of each type in a single play of the dilemma are given by the familiar table of the husband-wife game, Figure 8.1, which we reproduce here as Figure 10.1 with the now familiar warning that high numbers in this game are bad. Now we call the players simply Row and Column, because they can be any persons in the population who are randomly arrested and paired against another random rival.
Remember that under the evolutionary scenario, no individual has the choice between confessing and not confessing; each is “born” with one trait or the other predetermined. Which trait is the more successful, or fitter, in the population?

A confessing type gets 10 years if matched against another confessing type, and 1 when matched against a non-confessing type. A non-confessing type gets 25 years if matched against a confessing type, and 3 if matched against another non-confessing type. No matter what the type of the matched rival, the confessing type does better than the non-confessing type. Therefore the confessing type has a higher expected payoff (is fitter) than the non-confessing type, irrespective of the proportions of the two types in the population.

A little more formally, let $x$ be the proportion of cooperators in the population. Consider any one particular cooperator. In a random draw, the probability of his meeting another cooperator (and getting 3 years) is $x$, and that of meeting a defector (and getting 25 years) is $(1-x)$. Therefore a typical cooperator’s expected sentence is $3x + 25(1-x)$. For a defector, the probability of meeting a cooperator (and getting 1 year) is $x$, and that of meeting another defector (and getting 10 years) is $(1-x)$. Therefore a typical defector’s expected sentence is $x + 10(1-x)$. Now it is immediate that

$$x + 10(1-x) < 3x + 25(1-x) \text{ for all } x \text{ between } 0 \text{ and } 1.$$

Therefore a defector is fitter than a cooperator. (Remember that years in jail are bad, so lower numbers represent better outcomes in this game.) This will lead to an increase in the proportion of defectors (a decrease in $x$) from one “generation” of players to the next, until the whole population consists of defectors. Again, we repeat that the dynamics in a strategic game are more likely to be the result of observation, education, or thinking (social influences) than of genetics (biological determinism).

If the population initially consists of defectors, it cannot be invaded successfully by a mutant (experimental) cooperator - even for very small $x$, the cooperators remain less fit than the prevailing defectors, and the mutant strain will die out.

Thus Confession must be the evolutionary stable strategy for this population engaged in this dilemma game. This is a general proposition: if a game has a dominant strategy, that strategy will also be the ESS.
The Role of Repetition

We saw in Chapter 8 how a repetition of the prisoners' dilemma permitted consciously rational players to sustain cooperation for their mutual benefit. Let us see if a similar possibility exists in the evolutionary story. Suppose each chosen pair of players plays the dilemma twice in succession. The overall payoff of a player from such an interaction is the sum of what she gets in the two games.

Each individual is still programmed to play just one strategy, but in a game with two moves, a strategy has to be a complete plan of action, so a strategy can stipulate an action in the second play that depends on what happened in the first play. For example, "I will always cooperate no matter what" and "I will always defect no matter what" are valid strategies. But, "I will begin by cooperating and continue to cooperate on the second play if you cooperated on the first but defect otherwise" is also a valid strategy; it is just tit-for-tat (TFT).

Suppose these are the three types of strategies that can possibly exist in the population. Then we have three types of individuals, those who always confess, those who never confess, and those who play tit-for-tat. Figure 10.2 shows the outcomes (jail years) summed over the two meetings, of each type when matched against rivals of each type. For brevity we have labeled the strategy of always confessing as A, that of never confessing as N, and tit-for-tat as T.4

Figure 10.2 - Outcomes in the twice-repeated dilemma

<table>
<thead>
<tr>
<th></th>
<th>Column</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A</td>
</tr>
<tr>
<td>A</td>
<td>20, 20</td>
</tr>
<tr>
<td>T</td>
<td>35, 11</td>
</tr>
<tr>
<td>N</td>
<td>50, 2</td>
</tr>
</tbody>
</table>

To see how these numbers arise, consider a couple of examples. When two tit-for-tat (T) players meet each other, neither confesses the first time, and therefore neither confesses the second time; both get 3 each time for a total of 6 each. When a T player meets one who always confesses (A), the latter does well the first time (1 year as against 25 for the tit-for-tat player), but then the T player also confesses the second time, and both get 10 (for totals of 11 for A and 35 for T).

What do these numbers tell us about the fitness of the different strategies? The first thing to note is that the naively cooperative strategy of never confessing (N) is not going to do well. It does

---

4 When each pair plays twice, the maximum possible sentence is 50 years! The story makes more sense if we now think of the sentences as months rather than years.
exactly as well as T when the rival is either N or T. But it does worse than T against an A; it lets itself get fooled twice (shame on it). Therefore in any mixed population, T will be fitter than N, and will eventually dominate N. If the population initially consists of only N and A, a mutant T can invade successfully and then eventually dominate N. If the population initially happens to consist entirely of N, then a mutant A does better and can invade successfully. From either point of view, N cannot be an evolutionary stable strategy (ESS).

So let us confine our attention to the other two strategies, A and T. Their relative fitness depends on the composition of the population. If the population is almost wholly A, then A is fitter than T (because A-types meeting mostly other A-types get 20 years of jail most of the time while T-types most often get 35 years). But if the population is almost wholly T, then T is fitter than A (because T-types get only 6 years when they meet mostly other Ts, but A-types get 11 years in such a situation). Each type is fitter when it is already predominant in the population. Therefore T cannot invade successfully when the population is all A, and vice versa. Each of the two strategies, A and T, is an ESS.

Now consider the case in which the population is made up of a mixture of the two types. What will the evolutionary dynamics in the population entail? That is, how will the composition of the population evolve over time? Suppose a fraction x of the population plays T and the rest, (1-x), play A. An individual A-player, pitted against various opponents chosen from such a population, gets 11 when confronting a T-player which happens a fraction x of the times, and 20 against another A-player, which happens a fraction (1-x) of the times. This gives an average of

\[ 11x + 20(1-x) = 20 - 9x \]

Similarly, an individual T-player gets

\[ 6x + 35(1-x) = 35 - 29x \]

Then a T-player is fitter than an A-player if the former serves less in jail time on average, that is, if

\[ 35 - 29x < 20 - 9x, \quad \text{or} \quad 20x > 15, \quad \text{or} \quad x > 0.75 \]

In other words, if more than 75% of the population is already T, then T-players are fitter and their proportions will grow until we reach the ESS with 100% T. If the population starts with less than

---

3Literally, the fraction of any particular type in the population is finite and can only take on values like 1/1,000,000, 2/1,000,000, and so on. But if the population is sufficiently large, and we show all such values by points on a straight line as in the figure below, then these points are very tightly packed together, and we can regard them as forming a continuous line. This amounts to letting the fractions take on any real value between 0 and 1. We can then talk of the population proportion of a certain behavioral type.
75% T, then A-players will be fitter, and the proportion of T-players will go on declining until there are 0% of them, or 100% of A-players.

Thus we have identified two equilibria. In each equilibrium the population is all of one type, therefore each is said to be monomorphic. Both of these equilibria are evolutionary stable. For example, if the population is initially 100% T, then a small invasion of experimenting mutants of A types will leave the population mix with more than 75% T; T will remain the fitter type and the mutant A strain will die out. Similarly, if the population is initially 100% A, then a small invasion of experimenting mutants of T types will leave the population mix with less than 75% T, so the A types will be fitter and the mutant T strain will die out. And, as we saw above, experimenting mutants of type N can never succeed in a population of A and T-types that is either largely T or largely A.

Just in between, if the initial population has exactly 75% T-players (and 25% A-players), then the two types are equally fit. This population proportion is a polymorphic equilibrium, but an unstable one; the population can sustain this delicately balanced outcome only until a mutant of either type surfaces. By chance, such a mutant must arise sooner or later. The mutant's arrival will tip the fitness calculation in favor of the mutant type, and the advantage will cumulate until the ESS with 100% of that type is reached.

This reasoning can be shown in a simple picture that closely resembles the pictures we drew when calculating the equilibrium proportions in a mixed strategy equilibrium with consciously rational players. The only difference is that in the evolutionary context the proportion in which the separate strategies are played is not a matter of choice by any individual player, but a property of the whole population; the diagram for this case is shown in Figure 10.3.

**Figure 10.3 - Fitness graphs and equilibria for twice-repeated dilemma**

![Fitness Graph](image)

Figure 10.3 shows the proportion x of T-players measured from 0 to 1 from left to right in the horizontal direction. The vertical direction measures unfitness (these are jail times, and therefore higher numbers are bad). Each of the two lines show the fitness of one type. The line for the T-type
starts higher (35 as against 20 for the line of the A-type) and ends lower (6 against 11). The two lines cross when x = 0.75. To the right of this point, T is fitter, so x increases toward 1; to the left of this point, A is fitter, so x decreases toward 0. Such diagrams often prove useful as visual aids, and we will use them extensively.

What if each pair plays more than two repetitions of the game? Let us just focus on a population comprised only of A and T-types in which interactions between random pairs occur n (n > 2) times. The table of the total outcomes from playing n repetitions is shown in Figure 10.4. When two A-types meet, they always cheat and get 10 years every time, so each gets 10n over n plays. When two T-types meet, they begin by cooperating, and no one is the first to defect, so they get 3 every time for a total of 3n. When an A-type meets a T-type, on the first play the T-type cooperates and the A-type defects, so the A-type gets 1 and the other gets 25; thereafter the T-type retaliates against the previous defection of the A-type for all remaining plays and each gets 10 in all of the remaining (n-1) plays. Thus, the A-type gets a total of 1+10(n-1) = 10n-9 in n plays against a T-type, while the T-type gets 25+10(n-1) = 10n+15 in n plays against an A. (Please accept these as a metaphor, even though for large n the sentences run longer than lifetimes.)

Figure 10.4 - Outcomes in the n-fold repeated dilemma

<table>
<thead>
<tr>
<th></th>
<th>Column</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A</td>
</tr>
<tr>
<td>Row</td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>10n, 10n</td>
</tr>
<tr>
<td>T</td>
<td>10n+15, 10n-9</td>
</tr>
</tbody>
</table>

If the proportion of T-types in the population is x, then a typical A-type gets x (10n-9) + (1-x) 10n on average, and a typical T-type gets x 3n + (1-x) (10n+15) on average. Therefore the T-type is fitter if

\[ x (10n-9) + (1-x) 10n > x 3n + (1-x) (10n+15) \]

or

\[ x(7n+6) > 15, \text{ or } x > 15/(7n+6). \]

Once again we have two stable monomorphic ESS, one all-T (or x = 1, to which the process converges starting from any x > 15/(7n+6)) and the other all-A (or x = 0, to which the process converges starting from any x < 15/(7n+6)), and an unstable polymorphic equilibrium at the balancing point x = 15/(7n+6).

Notice that the proportion of T at the balancing point depends on n; it is smaller when n is larger. When n = 10, it is 15/76 or approximately 0.20. So, if the population initially has 20% of T-players, in a situation where each pair plays 10 repetitions, their numbers will grow until they reach 100%. When pairs played only two repetitions (n = 2), the T-players needed an initial strength of
75% or more to achieve this outcome. Remember too that a population of all T-players achieves cooperation. Thus cooperation emerges from a larger range of the initial conditions when the game is repeated longer. In this sense, with more repetition, cooperation becomes more likely. What we are seeing is the result of the fact that the value of establishing cooperation increases as the length of the interaction increases.

Finally, let us return to the two-fold repeated game, and instead of using the evolutionary model, consider it played by two consciously rational players. What are the Nash equilibria? There are two in pure strategies, one where both play A, and the other where both play T. There is also one in mixed strategies, where T is played 75% of the time and A 25% of the time. The first two are just the monomorphic ESS we found, and the third is the unstable polymorphic evolutionary equilibrium. In other words, there is a close relationship between evolutionary and consciously rational perspectives on games.

That is not a coincidence. In fact an ESS must be a Nash equilibrium of the game with the same payoff structure played by consciously rational players. To see this, suppose the contrary for the moment. If some strategy, call it S, is not a Nash equilibrium, some other strategy, call it T, must yield a higher payoff when played against S. Then a mutant playing T will achieve greater fitness in a population playing S, and so will invade successfully. Thus S cannot be an ESS. In other words, if S is not a Nash equilibrium, then S cannot be an ESS. This is the same as saying that if S is an ESS, it must be a Nash equilibrium.

Thus the evolutionary approach provides a backdoor justification for the rational approach. Even when players are not consciously maximizing, if the more successful strategies get played more often and the less successful ones die out, and if the process converges to a stable eventual strategy, it must be as if resulting from consciously rational play.

If a game has several Nash equilibria under rational play, the evolutionary approach sometimes offers a supplementary reasoning that helps us select one, or at least narrow down the ambiguity somewhat. The two-fold repeated prisoners' dilemma above had three Nash equilibria, two in pure strategies and one in mixed. The evolutionary dynamics could converge to one of the pure ones depending on the initial proportions in the population, but the mixed equilibrium was unstable. Therefore we could legitimately disregard it and focus on the pure ones.

In concluding this section, we should point out one limitation of our analysis of the repeated game. At the outset we allowed just three strategies, A, T, and N. Nothing else was supposed to exist, or arise as a mutation. In biology, the kinds of mutations that arise are determined by genetic considerations. In social or economic or political games, the genesis of new strategies is presumably governed by history, culture, and experience of the players: the ability of people to assimilate and process information and to experiment with different strategies must also play a role. However, the restrictions we place on the set of strategies that can possibly exist in a particular game has important implications for which of these strategies (if any) can be evolutionary stable. In the twice-repeated Prisoners' Dilemma example above, if we had allowed a strategy S which cooperates on the first play and defects on the second, then S-type mutants could have successfully invaded an all-T population,
and then T would not have been an ESS. We develop this possibility further in an exercise at the end of this chapter.

3. Chicken

Remember our 1950s youth racing their cars toward each other and seeing who will be the first to swerve to avoid a collision. Now we suppose that the players have no choice in the matter, but each is genetically hardwired to be either a wimp (always swerve) or a macho (always go straight). The population consists of a mixture of the two types. Pairs are picked at random every week to play the game. Figure 10.5 shows the payoff table for any two such players, say A and B. The numbers just replicate those we used before (Figures 4.10 and 5.7).

Figure 10.5 - Payoff table for Chicken

<table>
<thead>
<tr>
<th></th>
<th>B</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wimp</td>
<td>0, 0</td>
</tr>
<tr>
<td>Macho</td>
<td>-1, 1</td>
</tr>
</tbody>
</table>

How will the two types fare? The answer depends on the initial population proportions. If the population is almost all wimps, then a macho mutant will win and score 1 lots of times while all the wimps meeting their own types get mostly zeroes. But if the population is mostly macho, then a wimp mutant scores -1, which may look bad but is better than the -2 that all the machos get. You can think of this very fittingly in terms of the biological context and the sexism of the 1950s: in a population of wimps, a macho newcomer will show all the rest to be chickens and so will impress all the girls. But if the population consists mostly of machos, they will be in the hospital most of the time and the girls will have to go for the few wimps that are healthy.

In other words, each type is fitter when it is relatively rare in the population. Therefore each can invade successfully a population consisting of the other type. We should expect to see both types in the population in equilibrium; that is, we should expect an ESS with a mixture, or polymorphism.

To find the proportions of wimps and machos in such an ESS, let us calculate the fitness of each type in a general mixed population. Write x for the fraction of machos and (1-x) for the proportion of wimps. A wimp meets another wimp and gets 0 for a fraction (1-x) of the time, and meets a macho and gets -1 for a fraction x of the time. Therefore the fitness of a wimp is 0 (1-x) - 1 x = - x. Similarly, the fitness of a macho is 1 (1-x) - 2 x = 1- 3 x. The macho type is fitter if

\[ 1 - 3x > -x, \quad \text{or} \quad 2x < 1, \quad \text{or} \quad x < \frac{1}{2}. \]
If the population is less than half macho, then the machos will be fitter and their proportion will increase. On the other hand, if the population is more than half macho, then the wimps will be fitter and the macho proportion will fall. Either way, the population proportion of machos will tend toward $\frac{1}{2}$, and this 50:50 mix will be the stable polymorphic ESS.

Figure 10.6 shows this outcome graphically. Each straight line shows the fitness (expected payoff in a match against a random member of the population) for one type, in relation to the proportion $x$ of machos. For the wimp type, this functional relation showing their fitness as a function of the proportion of the machos is $W(x) = -x$, as we saw above. This is the gently falling line, which starts at the height 0 when $x = 0$, and goes to -1 when $x = 1$. The corresponding function for the macho type is $M(x) = 1 - 3x$. This is the rapidly falling line, which starts at height 1 when $x = 0$, and falls to -2 when $x = 1$. The line $M(x)$ lies above the line $W(x)$ for $x < \frac{1}{2}$ and below it for $x > \frac{1}{2}$, showing that the macho types are fitter when $x$ is small and the wimps are fitter when $x$ is large.

Figure 10.6 - Fitness graphs and polymorphic equilibrium for Chicken

Now we can compare and contrast the evolutionary theory of this game with our earlier theory of Chapters 4 and 5, which was based on the assumption that the players were conscious rational calculators of strategies. There we found three Nash equilibria, two in pure strategies where one player goes straight and the other swerves, and one in mixed strategies, where each player goes straight with probability $\frac{1}{2}$ and swerves with probability $\frac{1}{2}$.

In one sense we still have all three of those equilibria. If the population is truly 100% macho, then they are all equally fit (or equally unfit) and the situation can persist. Similarly a population of nothing but wimps can continue. So we still have two equilibria of the pure type. But they are unstable. In an all-macho population, a wimp mutant will outscore them and invade successfully. Once some wimps get established, no matter how few, the above logic shows that their proportion will rise inexorably toward half. Similarly, an all-wimp population is vulnerable to a successful invasion of mutant machos, and the process again goes to the polymorphic ESS.

---

6 The Invasion of the Mutant Wimps could be an interesting science fiction movie.
Most interesting is the connection between the mixed-strategy equilibrium of the rationally played game and the polymorphic ESS of the evolutionary game. The mixture proportions in the equilibrium strategy of the former are exactly the same as the population proportions in the latter: a 50:50 mixture of wimp and macho. But the interpretations differ: in the rational framework, each player mixes his own strategies; in the evolutionary framework, every individual uses a pure strategy, but different individuals use different strategies so we see a mixture in the population.7

This correspondence between Nash equilibria of a rationally-played game and stable outcomes of a game with the same payoff structure when played according to the evolutionary rules is a very general proposition, and we will see it in its generality in Section 6 below.

When we looked at Chicken from the rational perspective, the mixed strategy equilibrium seemed puzzling. It left open the possibility of costly mistakes. Each player went straight one time in two, so one time in four they collided. The pure strategy equilibria avoided the collisions. At that time this may have led you to think that there was something undesirable about the mixed strategy equilibrium, and you may have wondered why we were spending time on it. Now you see the reason. The seemingly strange equilibrium in fact emerges as the stable outcome of a natural dynamic process where each player tries to improve his payoff against the population he confronts.

4. The Assurance Game

Of the triad of important classes of strategic games we introduced in Chapter 4, we have studied Prisoners’ Dilemma and Chicken from the evolutionary perspective. That leaves the Assurance game. We illustrated this type of game in Chapter 4 using the story of two superpowers deciding whether to build more nuclear arms. In the evolutionary context it makes no sense to have pairs chosen from a large population of superpowers playing at arm races time and again, so we use a different example.

Suppose that pairs of players are drawn from the U.S. population. The two members of a pair, whose identities are unknown to each other, are supposed to meet somewhere in New York City, but the exact location is not specified for them. There are two types in the population: theater-fans and movie-buffs. When the theater-fans think of New York City, they think of nothing but the Theater District and Times Square, more specifically and that is where they will go. The movie-buffs, remembering a triad of famous film rendezvous (or attempted rendezvous) – in *An Affair to Remember*, *Sleepless in Seattle* and, of course, *King Kong* -- will go to the most obvious meeting place (to them) at the top of the Empire State Building.

If the two people in a particular pair go to different places and fail to meet, each of them will get a payoff of 0. If both are theater-fans and go to Times Square, each will get 1, while if both are

---

7 There can also be evolutionary stable mixed strategies, where each player mixes and the population shows a mixture corresponding to the actual realizations from all the players’ mixtures. We leave this for more advanced treatments.
movie-buffs and go to the top of Empire State Building, each will get 2. The payoffs are higher when both go to the top of the Empire State Building because it is a smaller location and meeting there is easier, not to mention safer. Figure 10.7 shows the payoff table for a random pairing in this game. As before, we call the two players A and B. For the two types and their strategies, we let T denote the theater type and its Times Square Strategy, and M denote the movie type and its Empire State Building strategy.

Figure 10.7 - Payoff matrix for the meeting game

<table>
<thead>
<tr>
<th></th>
<th>B</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>T</td>
</tr>
<tr>
<td>A T</td>
<td>1, 1</td>
</tr>
<tr>
<td>A M</td>
<td>0, 0</td>
</tr>
</tbody>
</table>

If this were a game played by rational strategy-choosing players, there would be two equilibria in pure strategies: (T, T) and (M, M). The latter is better for both players. If they communicate and coordinate explicitly, they can settle on it quite easily. But if they are making the choices independently, they need tacit coordination through a convergence of expectations – that is, a focal point.

The game has a third equilibrium, in mixed strategies, although we did not consider this equilibrium in our analysis of the arms race in Chapter 4. The equilibrium mixtures are easy to find, though, using the method of keeping the opponent indifferent. If A chooses T with probability p and M with probability (1-p), B gets the expected payoff p from choosing T and 2(1-p) from choosing M. Her indifference requires p = 2 (1-p), or p = 2/3. The expected payoff for each player is then 1(2/3) + 0(1/3) = 2/3. This is worse than the payoff associated with the less attractive of the two pure strategy equilibrium, (T, T). The reason is similar to what we found for the mixed strategy equilibrium of Chicken in Chapter 5. When the two players mix their strategies independently, they make the clashing or bad choices quite a lot of the time. Here the bad outcome (payoff 0) occurs when A chooses T and B chooses M or the other way round; the probability of this is 2 (2/3) (1/3) = 4/9 – the two players go to different meeting places almost half the time.

What happens when this is an evolutionary game? In the large population, each individual is “hard-wired,” some to choose T and the others to choose M. Randomly chosen pairs of such people are assigned to go to New York City and to attempt a meeting. Suppose x is the proportion of the T-types, and (1-x) that of the M-types, in the population. Then the fitness of a particular T-type – her expected payoff in a random encounter of this kind – is x 1 + (1-x) 0 = x. Similarly, the fitness of each M-type is x 0 + (1-x) 2 = 2 (1-x). Therefore the T-type is fitter when x > 2 (1-x), or for x > 2/3. The M-type is fitter when x < 2/3. At the balancing point x = 2/3, the two are equally fit.
As in Chicken, once again, the probabilities associated with the mixed strategy equilibrium that would obtain under rational choice seem to reappear as the population proportions in a polymorphic equilibrium under evolutionary rules. But now this mixed equilibrium is not stable. The slightest chance departure of the proportion \( x \) from the balancing point \( \frac{2}{3} \) will set in motion a cumulative process that takes the population mix farther away from the balancing point. If \( x \) increases from \( \frac{2}{3} \), the T-type becomes fitter and propagates faster, increasing \( x \) even more. If \( x \) falls from \( \frac{2}{3} \), the M-type becomes fitter and propagates faster, lowering \( x \) even more. Depending on which disturbance occurs, eventually \( x \) will either rise all the way to 1, or fall all the way to 0. The difference is that in Chicken each type was fitter when it was rarer, so the population proportions tended to move away from the extremes and toward a mid-range balancing point. In contrast, in the Assurance game each type is fitter when it is more numerous; the risk of failing to meet falls when more of the rest of the population are the same type as you – so population proportions tend to move toward the extremes.

Figure 10.8 illustrates the fitness graphs and equilibria for the meeting version of the assurance game; this diagram is very similar to Figure 10.6. The two lines show the fitness of the two types in relation to the population proportion. The intersection of the lines gives the balancing point. The only difference is that away from the balancing point, the more numerous type is the fitter, whereas in Figure 10.6 it was the less numerous type.

Each type being less fit when it is rare, only the two extreme monomorphic equilibria survive. It is easy to check that both are evolutionary stable: an invasion by a small mutant population of the other type will die out because the mutants, being rare, will be less fit.

Thus in assurance or coordination games, unlike in Chicken, the evolutionary process does not preserve the bad equilibrium where the players choose clashing strategies. But the process cannot guarantee that it will lead to the better of the two extreme pure strategy equilibria. That depends on the initial mix in the population.
5. Interaction Across Species

The Battle of the Two Cultures game of Chapter 4 (Figure 4.12) looks similar to the assurance game in some respects. In our example, a university faculty was voting on the use of some vacant space. The humanities faculty prefer to have a theater, and the science faculty prefer a lab. If the two agree on one use, that is what is done. If each insists on its preferred use, the result will be an impasse, nothing will be done, and that will be the worst outcome for both. Thus there is a premium on taking mutually consistent actions, just as in the assurance example. But the consequences of the two possible mutually consistent actions differ. The types in the assurance game do not differ in their preferences; both prefer (M,M) to (T,T). The faculty in the Battle game differ in theirs: the theater gives a payoff of 2 to the humanists and 1 to the scientists, and the lab the other way around. These preferences distinguish the two types. In the language of biology, they can no longer be considered random draws from a homogeneous population of animals. Effectively, they belong to different species (as indeed humanities and science faculties often believe of each other).

To study such games from an evolutionary perspective, we must extend our methodology to the case where the matches are between randomly drawn members of different species or populations. We develop the Battle of the Two Cultures example to illustrate how this is done.

Suppose there is a large population of scientists and a large population of humanists. One of each “species” is picked, and the two are asked to cast secret ballots for the lab or the theater. The votes are compared, and the comparison determines the outcome – lab, theater or impasse – according to the rule stated above.

All scientists are agreed among themselves about the valuation (payoffs) of the lab, the theater and the impasse. Likewise, all humanists are agreed among themselves. But within each population, some individuals are hardliners and others are compromisers. A hardliner will always vote for her species’ preferred use. A compromiser recognizes that the other species wants the opposite, and votes for that so as to get along.

If the random draws happen to have picked a hardliner of one species and a compromiser of the other, the outcome is that preferred by the hardliner’s species. We have an impasse if two hardliners meet, and strangely, also if two compromisers meet, because they vote for each other’s preferred use. (Remember they have to vote in secret, and cannot negotiate. Perhaps even if they did meet, they would reach an impasse of “No, I insist on giving way to your preference.”)

We reproduce the payoff table of Figure 4.12 as Figure 10.9; of course, what were choices of votes are now interpreted as actions predetermined by your type (hardliner or compromiser).

In comparison with all the evolutionary games we have studied so far, the new feature here is that the row player and the column player come from different species. While each species is a heterogeneous mixture of hardliners and compromisers, there is no reason why the proportions of the mixture of types should be the same in both species. Therefore we must introduce two variables to represent the two mixtures and study the dynamics of both.
So let $x$ be the proportion of hardliners among the scientists, and $y$ that among the humanists. Consider a particular hardliner scientist. She meets a hardliner humanist a proportion $y$ of the time and gets 0, and meets a compromising humanist for the rest of the time and gets 2. Therefore her expected payoff (fitness) is $y \cdot 0 + (1-y) \cdot 2 = 2(1-y)$. Similarly, a compromising scientist’s fitness is $y \cdot 1 + (1-y) \cdot 0 = y$. Among scientists, therefore, the hardliner type is fitter when $2(1-y) > y$, or $y < 2/3$. The hardliner scientists will reproduce faster when they are fitter; that is, $x$ increases when $y < 2/3$.

Note the new, and at first sight surprising, feature of the outcome: the fitness of each type within a given species depends on the proportion of types found in other species. Of course there is nothing surprising here—after all, now the games each species plays are all against the members of the other species.  

Similarly, considering the other species, we have the result that the hardliner humanists are fitter, so $y$ increases, when $x < 2/3$. To understand the result intuitively, note that it says that the hardliners of each species do better when the other species does not have too many hardliners of its own, because then they meet compromisers of the other species quite frequently.

Figure 10.10 shows the dynamics of the configurations of the two species. Each of $x$ and $y$ can range from 0 to 1, so we have a graph with a unit square and $x$ and $y$ on their usual axes. Within that, the vertical line AB shows all points where $x = 2/3$, the balancing point at which $y$ neither increases or decreases. Then if the current population proportions lie to the left of this line so $x < 2/3$, $y$ increases (moving the population proportion of hardliner humanists in the vertically upward direction) while $y$ decreases (motion vertically downward) when the current proportions lie to the right of AB and $x > 2/3$. Similarly, the horizontal line CD shows all points where $y = 2/3$, which is the balancing point for $x$. The proportion of hardliner scientists, $x$, increases for population proportions of hardliner humanists below this line, for $y < 2/3$, (motion horizontal and rightward) and decreases for population proportions above it, when $y > 2/3$ (motion horizontal and leftward).

---

8 And this finding supports and casts a different light on the property of mixed strategy equilibria, that each player’s mixture kept the other player indifferent among her pure strategies. Now we can think of it as saying that in a polymorphic evolutionary equilibrium of a two-species game, each species’ type proportions keep all the surviving types of the other species equally fit.
When we combine the motions of $x$ and $y$, we can follow their dynamic paths to determine the location of the population equilibrium. From a starting point in the bottom left quadrant of Figure 10.10, for example, the dynamics entail $y$ and $x$ both increasing. This joint movement (to the northeast) continues until either $x = 2/3$ and $y$ begins to decrease (motion now to the southeast) or $y = 2/3$ and $x$ begins to decrease (motion now to the northwest). Similar processes in each quadrant yield the curved dynamic paths shown in the diagram. The vast majority of these lead to either the northwest or southeast corners of the diagram; that is, they converge either to $(1, 0)$ or $(0, 1)$. Thus, evolutionary dynamics will lead in most cases to a configuration where one species is entirely hardline and the other is entirely compromising. Which species will be which type depends on the initial conditions. Note that the population dynamics starting from a situation with a small $x$ and a larger $y$ are more likely to cross the CD line first and head for $(0, 1)$ – all hardline humanists, $y = 1$ – than to hit the AB line first and head for $(0, 1)$; similar results follow for a starting position with a small $y$ but a larger $x$. The species that starts out with more hardliners will have the advantage of ending up all-hardline and getting the payoff of 2.

If the initial proportions are balanced just right, the dynamics may lead to the polymorphic point $(2/3, 2/3)$. But unlike the polymorphic ESS in Chicken, the polymorphic equilibrium in the Battle of the Two Cultures is unstable. Any chance departure will set in motion a cumulative process that leads to one of the two extreme equilibria; those are the two ESS for this game.

We invite you to draw your own conclusions about scientists and humanists, or indeed about the original “battle of the sexes” game we mentioned in Chapter 4, which has the same payoff structure.
6. The Hawk-Dove Game

The Hawk-Dove game was the first example studied by biologists in their development of the theory of evolutionary games. It has instructive parallels with our analyses above of the Prisoners’ Dilemma and Chicken, so we describe it to reinforce and improve your understanding of the concepts.

The game is played, not by birds of these two species, but by two animals of the same species, and Hawk and Dove are merely the names for their strategies. The context is competition for a resource. The Hawk strategy is aggressive, and fights to try to get the whole resource of value $V$. The Dove strategy is to offer to share, but to shirk from a fight. When two Hawk-types meet each other, they fight. Such fights always end in a “draw,” so the animals eventually share the resource anyway. But in the process they incur a total cost of fighting, $C$. Thus each gets payoff $(V-C)/2$. When two Dove-types meet, they share without a fight, so each gets $V/2$. When a Hawk-type meets a Dove-type, the latter retreats and gets 0, while the former gets $V$. Figure 10.11 shows the payoff table.

**Figure 10.11 - Payoff table for the Hawk-Dove game**

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Hawk</td>
<td>Dove</td>
</tr>
<tr>
<td>Hawk</td>
<td>$(V-C)/2, (V-C)/2$</td>
<td>$V, 0$</td>
</tr>
<tr>
<td>Dove</td>
<td>$0, V$</td>
<td>$V/2, V/2$</td>
</tr>
</tbody>
</table>

The analysis of the game is similar to that for the Prisoners’ Dilemma and Chicken games above, with the numerical payoffs replaced by algebraic symbols. We will compare the equilibria of this game when the players rationally choose to play Hawk or Dove, and the outcomes when players are acting mechanically and success is being rewarded by faster reproduction.

**Rational strategic choice and equilibrium**

[1] If $V > C$, this is a Prisoners’ Dilemma. Hawk is the dominant strategy for each, but (Dove, Dove) is the jointly better outcome.

[2] If $V < C$, this is a game of Chicken. There are two pure strategy Nash equilibria, (Hawk, Dove) and (Dove, Hawk). There is also a mixed-strategy equilibrium, where B’s probability $p$ of choosing Hawk is such as to keep A indifferent:

$$p \frac{(V-C)}{2} + (1-p) V = p 0 + (1-p) V/2 \quad \text{or} \quad p = V/C.$$
We start with an initial population predominantly of Hawks, and test whether it can be invaded by mutant Doves. Following the convention used in analyzing such games, we write the population proportion of the mutant phenotype as \( m \), for mutant, or in our case as \( d \) for mutant Dove; the population proportion of Hawks is then \( 1-d \). Then in a match against a randomly drawn opponent, a Hawk will meet a Dove a proportion \( d \) of the time and get \( V \) on each of those occasions, and meet another Hawk a proportion \( 1-d \) of the time and get \( (V-C)/2 \) on each of those occasions. Therefore the fitness of a Hawk is \( dV + (1-d)(V-C)/2 \). Similarly, the fitness of one of the mutant doves is \( dV/2 + (1-d)0 \). With \( V > C \), it follows that \( (V-C)/2 > 0 \). Also, \( V > 0 \) implies \( V > V/2 \). Then for any \( d \) between 0 and 1, we have

\[
dV + (1-d)(V-C)/2 > dV/2 + (1-d)0,
\]

so the Hawk type is fitter. The Dove mutants cannot invade, and the Hawk strategy is evolutionary stable.

The same holds true for any population proportion of Doves, for all values of \( d \). Therefore, from any initial mix, the proportion of Hawks will grow and they will predominate. In addition, if the population initially has all Doves, mutant Hawks can invade and take over. Thus the Hawk strategy is the only ESS. This confirms and extends our finding for the Prisoners’ Dilemma above.

**Evolutionary stability: Case \( V < C \)**

If the initial population is again predominantly Hawks, with a small proportion \( d \) of Dove mutants, each has the same fitness function derived above. When \( V < C \), however, \( (V-C)/2 < 0 \). We still have \( V > 0 \), so \( V > V/2 \). But since \( d \) is very small, the comparison of the terms with \( 1-d \) is much more important than that of the terms involving \( d \), so

\[
dV/2 + (1-d)0 > dV + (1-d)(V-C)/2.
\]

Thus the Dove mutants have higher fitness than the predominant Hawks. They can invade successfully.

But if the initial population has almost all Doves, then we must consider whether a small proportion \( h \) of Hawk mutants can invade; note that because the mutant is now a Hawk, we have followed convention and used \( h \) (instead of \( m \)) for the proportion of the mutant (Hawk) type. The Hawk mutants have fitness \( h(V-C)/2 + (1-h)V \) compared to \( h0 + (1-h)V/2 \) for the Doves. Again \( V < C \) implies \( (V-C)/2 < 0 \), and \( V > 0 \) implies \( V > V/2 \); but when \( h \) is small, we get

\[
h(V-C)/2 + (1-h)V > h0 + (1-h)V/2.
\]

Thus the Hawks have greater fitness and successfully invade a Dove population. Thus, mutants of each type can invade populations of the other type; there is no pure ESS. This also confirms our finding above for Chicken.

What happens in the population then when \( V < C \)? There are two possibilities. In one, every individual follows a pure strategy, but the population has a stable mix of players following different strategies. This is the polymorphic equilibrium we developed for Chicken in Section 10.3. The other possibility is that every individual uses a mixed strategy. We begin with the polymorphic case.
Ch. 10: Evolutionary Games

$V < C$: Stable polymorphic population

When the population proportion of Hawks is $h$, the fitness of a Hawk is $h \left( V-C \right) / 2 + (1-h) V$, and the fitness of a Dove is $0 + (1-h) V/2$. The Hawk type is fitter if:

$$\frac{h (V-C)}{2} + (1-h) V > \frac{(1-h) V}{2},$$

which simplifies to:

$$\frac{h (V-C)}{2} + (1-h) V > 0, \text{ or } V - h C > 0, \text{ or } h < \frac{V}{C}. \tag{9.1}$$

The Dove type is then fitter when $h > V/C$, or when $(1-h) < 1 - \frac{V}{C} = \frac{(C-V)}{C}$. Thus each type is fitter when it is rarer. Therefore we have a stable polymorphic equilibrium at the balancing point, where the proportion of Hawks in the population is $h = \frac{V}{C}$. As we calculated just above, this is exactly the probability with which each individual plays the Hawk strategy in the mixed strategy Nash equilibrium of the game under the assumption of rational behavior. Again we have an evolutionary “justification” for the mixed strategy outcome in Chicken.

We leave it to you as an exercise to draw a graph similar to Figure 10.6 for this case. Doing so will require that you determine the dynamics by which the population proportions of each type converge to the stable equilibrium mix.

$V < C$: Each individual mixes strategies

Recall the equilibrium mixed strategy of the rational play game calculated above in which $p = \frac{V}{C}$ was the probability of choosing to be a Hawk while $(1-p)$ was the probability of choosing to be a Dove. Let us examine whether this strategy, call it $M$, is an ESS. Now there are three types to consider: $H$-types who play the pure Hawk strategy, $D$-types who play the pure Dove strategy, and $M$-types who play the mixture in the proportions $p = \frac{V}{C}$ and $(1-p) = 1 - \frac{V}{C}$ or $(C-V)/C$.

When an $H$ or a $D$ meets an $M$, their expected payoffs depend on $p$, the probability that $M$ is playing $H$, and $(1-p)$, the probability that $M$ is playing $D$. Then they each get $p$ times their payoff against an $H$ plus $(1-p)$ times their payoff against a $D$. So, when an $H$-type meets an $M$-type, she gets the expected payoff

$$p \left( \frac{V-C}{2} \right) + (1-p) \frac{V}{2} = \left( \frac{V-C}{2} \right) + \frac{(C-V)}{C} \frac{V}{2} = V \left( \frac{C-V}{2C} \right) = \frac{V}{2C}.$$

And when a $D$-type meets an $M$-type, she gets

$$p 0 + (1-p) \frac{V}{2} = \left( \frac{C-V}{2} \right) \frac{V}{2} = \frac{V}{2C}.$$

The two fitnesses are equal. This should not be a surprise; the proportions of the mixed strategy are determined to achieve exactly this equality. Then an $M$-type meeting another $M$-type also gets the same expected payoff. For brevity of future reference, call this common payoff $K$, $K = V \left( \frac{C-V}{2C} \right)$.

But these equalities create a problem when we test $M$ for evolutionary stability. Suppose the population consists of all $M$-types, and a few mutants of the $H$-type, comprising a very small proportion $h$ of the total population, invade. Then the typical mutant gets the expected payoff
To calculate the expected payoff of an M-type, note that she faces another M-type a fraction \((1-h)\) of the time and gets \(K\) in each of these. She then faces an H-type for a fraction \(h\) of the interactions; in these she plays \(H\) a fraction \(p\) of the time and gets \((V-C)/2\), and plays \(D\) a fraction \((1-p)\) of the time and gets 0. Thus her total expected payoff (fitness) is
\[
h \frac{p(V-C)}{2} + (1-h) K.
\]

Since \(h\) is very small, the fitnesses of the M-types and the mutant H-types are almost equal. The point is that when there are very few mutants, both the H-type and the M-type meet only M-types most of the time, and in this interaction the two have equal fitness as we just saw.

Therefore evolutionary stability hinges on the distinction of whether the original population M-type is fitter than the mutant H when each is matched against one of the few mutants, that is, whether \(p \frac{V(C-V)}{2C} = p K > (V-C)/2\). This is true in this case, because \(V < C\), so \((V-C)\) is negative, while \(K\) is positive. In words, an H-type mutant will always do badly against another H-type mutant because of the high cost of fighting, but the M-type fights only part of the time and therefore suffers this cost only a fraction \(p\) of the time.

Similarly, the success of a Dove invasion against the M population depends on the comparison between the mutant Dove’s fitness: \([d \frac{V}{2} + (1-d) K]\) and the fitness of an M-type: \(d \left[pV+(1-p)V/2\right] + h K\). As above, the mutant faces another D a fraction \(d\) of the time and faces an M a fraction \((1-d)\) of the time. An M-type also faces another M a fraction \((1-d)\) of the time; but a fraction \(d\) of the time, the M faces a D and plays H a fraction \(p\) of these times (gaining \(pV\)) and plays D a fraction \((1-p)\) of these times (gaining \((1-p)V/2\)). Then a Dove invasion is only successful if \(V/2\) is greater than \(pV+(1-p)V/2\). This condition does not hold, since the latter expression, being a weighted average of \(V\) and \(V/2\) must exceed the former whenever \(V > 0\). Thus the Dove invasion cannot succeed either.

This analysis tells us that M is an ESS. The case of \(V < C\) can have two equilibria. One entails a mixture of types (a stable polymorphism) and the other entails a single type that mixes its strategies in the same proportions that define the polymorphism.

7. Some General Theory

We now generalize the ideas illustrated above, and get a theoretical framework and set of tools that can then be applied further. This unavoidably requires some slightly abstract notation and a bit of algebra. Therefore we cover only the case of monomorphic equilibria in a single species. Readers who are adept at this level of mathematics can readily develop the cases of polymorphism.
and two species by analogy. Readers who are not prepared for or interested in this material can omit this section without loss of continuity.  

We consider random matchings from a single species whose population has available strategies I, J, ... Some of these may be pure strategies; some may be mixed. Each individual is hardwired to play just one of these strategies. We denote by $E(I,J)$ the payoff to an I-player in single encounter with a J-player. The payoff of an I-player meeting another of her own type is $E(I,I)$ in the same notation. We write $W(I)$ for the fitness of an I-player. This is just her expected payoff in encounters with randomly picked opponents, when the probability of meeting a type is just the proportion of this type in the population.

Suppose the population is all-I. We consider whether this can be an evolutionary stable configuration. To do so, we imagine that the population is invaded by a few J-type mutants, so the proportion of mutants in the population is a very small number, $m$. Now the fitness of an I-type is:

$$W(I) = m E(I,J) + (1-m) E(I,I)$$

and the fitness of a mutant is:

$$W(J) = m E(J,J) + (1-m) E(J,I)$$

Therefore, the difference in the fitness of the population's main type and the mutant type is:

$$W(I) - W(J) = m [E(I,J)-E(J,J)] + (1-m) [E(I,I)-E(J,I)].$$

Since $m$ is very small, we will have

$$W(I) > W(J)$$

if $E(I,I) > E(J,I)$, which guarantees that the second half of the $W(I)-W(J)$ expression is positive; the fact that $m$ is small guarantees that the second term determines the sign of the overall inequality. Then the main type in the population cannot be invaded; it is fitter than the mutant type when each is matched against a member of the main type. This forms the **primary criterion** for evolutionary stability. Conversely, if $W(I) < W(J)$, due to $E(I,I) < E(J,I)$, the type-J mutants will invade successfully and an all-I population cannot be evolutionary stable.

However, it is possible that $E(I,I) = E(J,I)$, as indeed happens if the initial situation is an equilibrium where the population is mixing between I and J (a monomorphic equilibrium with a mixed strategy), as was the case in our final variant of the Hawk-Dove game above. Then the difference between $W(I)$ and $W(J)$ is governed by how each type fares against the mutants. If $E(I,I) = E(J,I)$, we get $W(I) > W(J)$ if $E(I,J) > E(J,J)$. This is the **secondary criterion** for evolutionary stability.

---


10 If the initial population is polymorphic and m the proportion of J-types, then m may not be “very small” any more. The size of m is no longer crucial, however, since the second term in $W(I) - W(J)$ is now assumed to be zero.
evolutionary stability of I, to be invoked only if the primary one is inconclusive – that is, only if $E(I,I) = E(J,I)$.

The primary criterion carries a punch. It says that the strategy I is evolutionary stable if, for all other strategies J that a mutant might try, we have $E(I,I) > E(J,I)$. This means that I is the best response to itself. In other words, if the members of this population suddenly started playing as rational calculators, playing I would be a Nash equilibrium. *Evolutionary stability in the above-defined sense implies Nash!*

This is a remarkable result. If you came to the theory of evolutionary games with some sense of dissatisfaction with the rational behavior assumption that underlies our earlier theory of Nash equilibria, you find it yields the same results. The very appealing biological description – fixed non-maximizing behavior, but selection in response to resulting fitness – does not yield any new outcomes. If anything, it provides a backdoor justification for Nash equilibrium. When a game has several Nash equilibria, the evolutionary dynamics may even provide a good argument for choosing among them.

However, your reinforced confidence in Nash equilibrium should be cautious. Our definition of “evolutionary stable” is static rather than dynamic. It only checks that the configuration of the population (monomorphic, polymorphic in just the right proportions) which we are testing for equilibrium cannot be successfully invaded by a small proportion of mutants. It does not test whether, starting from an arbitrary initial population mix, all the unwanted types will die out and the equilibrium configuration will be reached. And the test is carried out for those particular classes of mutants that are deemed logically possible; if the theorist has not specified this correctly and some type of mutant she had overlooked can actually arise, that mutant might invade successfully and destroy the supposed equilibrium. Our remark at the end of the twice-repeated Prisoners' Dilemma warned you of this possibility, and Exercise 2 below shows how it can arise. Finally, in the next section we show how evolutionary dynamics can fail to converge at all.

### 8. Dynamics with Three Types in the Population

If there are only two possible phenotypes (strategies) I and J, we can show the dynamics of the population in an evolutionary game using figures similar to 10.6. If there are three or more possible phenotypes, matters can get more complicated. As a final illustration that will lead you into more advanced study of evolutionary games, we show the dynamics of the Rock-Paper-Scissors game from an evolutionary viewpoint.

In Chapter 7 we considered mixed strategies in this game from a rational choice perspective. Figure 10.12 reproduces the payoff table. In the evolutionary context, players are not choosing mixtures, but there can be a mixture of types in the population.
Suppose $q_1$ is the proportion of types in the population that are playing Rock, $q_2$ the proportion of Paper-playing types, and the rest, $(1-q_1-q_2)$ play Scissors. The last column of the table shows each Row type's payoffs when meeting this mixture of strategies; that is just her fitness. Suppose the proportion of each type in the population grows when its fitness is positive and declines when it is negative. Then

$$q_1 \text{ increases if, and only if, } -q_2 + (1-q_1-q_2) > 0 \text{ or } q_1 + 2q_2 < 1.$$ 

The proportion of Rock-playing types in the population increases when $q_2$, the proportion of Paper-playing types, is small or when $(1-q_1-q_2)$, the proportion of Scissors-playing types, is large. This makes sense; Rock players do poorly against Paper players but well against Scissors players. Similarly, we see that

$$q_2 \text{ increases if, and only if, } q_1 - (1-q_1-q_2) > 0 \text{ or } 2q_1 + q_2 > 1.$$ 

Paper players do better when the proportion of Rock players is large or the proportion of Scissors players is small.

Figure 10.13 shows the population dynamics and resulting equilibria for this game graphically. The triangular area defined by the axes and the line $q_1 + q_2 = 1$ contains all the possible equilibrium combinations of $q_1$ and $q_2$. There are also two straight lines within this area. The first is $q_1 + 2q_2 = 1$ (the flatter one), which is the balancing line for $q_1$; for combinations of $q_1$ and $q_2$ below this line, our fitness calculations above tell us that $q_1$, the proportion of Rock players, increases while for combinations above this line $q_1$ decreases. The second line is $2q_1 + q_2 = 1$ (the steeper one), which is the balancing line for $q_2$. To the right this line, when $2q_1 + q_2 > 1$, $q_2$ increases and to the left of the line $q_2$ decreases. Arrows on the diagram show directions of motion of these population proportions; curves labeled by arrows show typical dynamic paths. The general idea is the same as that of Figure 10.10.

---

11 A little more care is necessary to ensure that the three proportions sum to 1, but that can be done, and we hide the mathematics so as to convey the ideas in a simple way.
Each of the two straight lines consists of points where one of $q_1$ and $q_2$ neither increases nor decreases. Therefore their intersection represents the point where $q_1$, $q_2$, and therefore also $(1-q_1-q_2)$, are all constant; that is, this intersection corresponds to a polymorphic equilibrium. It is easy to check that at this point $p_1 = p_2 = 1-p_1-p_2 = 1/3$. Thus its proportions are the same as the probabilities in the rational mixed strategy equilibrium we found in Chapter 7.

Is this polymorphic outcome stable? We cannot say. The dynamics shows paths that wind around it. Figure 10.13 illustrates the path winding in an ellipse around the polymorphic equilibrium point. Whether the dynamic paths actually wind in a decreasing spiral toward the intersection (in which case we have stability) or in an expanding spiral away from the intersection (instability) depends on the precise response of the population proportions to the fitnesses. It is even possible that the paths circle around the equilibrium as we have drawn, neither approaching it nor departing from it.

9. The Evolution of Cooperation and Altruism

Evolutionary game theory rests on two fundamental ideas: first, that individuals are engaged in games with others in their own species or with members of other species, and second, that the genotypes that lead to higher-payoff (fitter) strategies proliferate while the rest decline in their proportions of the population. These ideas suggest a vicious struggle for survival like that depicted by some interpreters of Darwin who understood “survival of the fittest” in a literal sense and who conjured up images of a “nature red in tooth and claw.” In fact nature shows many instances of cooperation (where individual animals behave in a way that yields greater benefit to everyone in a group) or even altruism (where individuals incur significant costs in order to benefit others). Beehives and ant colonies are only the most obvious examples. Can such behavior be reconciled with the perspective of evolutionary games?
The behavior of ants and bees is probably the easiest to understand in these terms. All the individuals in an ant colony or a beehive are closely related and therefore share genes to a substantial extent. All worker ants in a colony are full sisters and therefore share half their genes; therefore the survival and proliferation of one ant’s genes is helped just as much by the survival of two of its sisters as by its own survival. All worker bees in a hive are half-sisters and therefore share a quarter of their genes. Of course an individual ant or bee does not make a fine calculation of whether it is worthwhile to risk its own life for the sake of two or four sisters, but those groups whose members exhibit such behavior (phenotype) will see the underlying genes proliferate. The idea that evolution ultimately operates at the level of the gene has had profound implications for biology, although it has been misapplied by many people just as Darwin’s original idea of natural selection was misapplied. The interesting idea is that a “selfish gene” may prosper by behaving unselfishly in a larger organization of genes, such as a cell. Similarly a cell and its genes may prosper by participating cooperatively and accepting their allotted tasks in a body. The next step goes beyond biology and into sociology: a body (and its cells and ultimately its genes) may benefit by behaving cooperatively in a collection of bodies, namely a society.

This suggests that we should see cooperation among individuals who are less closely related, and indeed we do find instances of this. Groups of predators like wolves are a case in point; groups of apes often behave like extended families. Even among species of prey, cooperation arises when individuals in a school of fish take turns to look out for predators. Cooperation also extends across species. Some small fish and shrimp thrive on parasites that collect in the mouths and gills of some large fish; the large fish let the small ones swim unharmed through their mouths for this “cleaning service.”

Many of these situations can be seen as resolutions of Prisoners’ Dilemmas through repetition. The mutually beneficial outcome is sustained by strategies that are remarkably like tit-for-tat. It is not to be supposed that each animal consciously performs the calculation of whether it is in its individual interest to continue the cooperation or whether it would do better to defect. Instead, the behavior is instinctive. Reciprocity seems to come automatically to many animals, including humans.

An instinct is hardwired into an individual’s brain by genetics, but reciprocity and cooperation can also arise from more purposive thinking or experimentation within the group, and spread by socialization – by way of explicit instruction or observation of the behavior of elders – instead of genetics. The relative importance of the two channels – nurture and nature – will differ from one species to another and from one situation to another. One would expect socialization to be

---

12 In this very brief account we cannot begin to do justice to all the issues and the debates. An excellent popular account, and the source of many examples we cite below, is Matt Ridley, *The Origins of Virtue*. New York: Penguin, 1996. We should also point out that we have not examined the connection between genotypes and phenotypes in any detail, and in particular have not examined the role of sex in evolution. Another book by Matt Ridley, *The Red Queen*, New York: Penguin Books, 1995 gives a fascinating treatment of this subject.
relatively more important among humans, but there are instances of its role among other animals. We cite a remarkable one. The expedition which Robert F. Scott led to the South Pole in 1911-12 used teams of Siberian dogs. This group of dogs, brought together and trained for this specific purpose, developed within a few months a remarkable system of cooperation and sustained it using punishment schemes. "They combined readily and with immense effect against any companion who did not pull his weight, or against one who pulled too much ... their methods of punishment always being the same and ending, if unchecked, in what they probably called justice, and we called murder."

While this is an encouraging account of how cooperative behavior can be compatible with evolutionary game theory, you should not conclude that all animals have overcome dilemmas of selfish actions in this way. "Compared to nepotism, which accounts for the cooperation of ants and every creature that cares for its young, reciprocity has proved to be scarce. This, presumably, is due to the fact that reciprocity requires not only repetitive interactions, but also the ability to recognize other individuals and keep score." In other words, precisely the conditions that our theoretical analysis of Chapter 8 identified as being necessary for a successful resolution of the Prisoners' Dilemma are seen to be relevant in the context of evolutionary games. In the next chapter we will consider when and how the purposive design of institutions, organizations, and incentive mechanisms can help humans overcome the numerous dilemmas of collective or cooperative action which our societies face.

Summary

The biological theory of evolution parallels the theory of games used by social scientists. Evolutionary games are played by behavioral phenotypes with genetically predetermined, rather than rationally chosen, strategies. In an evolutionary game, fitter phenotypes survive repeated interactions with others to reproduce and to increase their representation in the population. If a phenotype maintains its dominance in the population when faced with an invading mutant type, that phenotype's strategy is called evolutionary stable. If two or more types are equally fit, the population exhibits a polymorphism.

When the theory of evolutionary games is used more generally for non-biological games, the strategies followed by individuals are understood to be standard operating procedures or rules of thumb, instead of being genetically fixed. The process of reproduction stands for more general methods of transmission including socialization, education and imitation; and mutations represent experimentation with new strategies.


14 Ridley, Virtue, op. cit., p. 83.
Evolutionary games may have payoff structures similar to those analyzed in chapter 5, including the Prisoners' Dilemma and Chicken. In each case, the evolutionary stable strategy mirrors either the pure strategy Nash equilibrium of a game with the same structure played by rational players or the proportions of the equilibrium mixture in such a game. In a Prisoners’ Dilemma, “always defect” is evolutionary stable; in Chicken, types are fitter when rare so there is a polymorphic equilibrium; in the Assurance version, types are less fit when rare so the polymorphic outcome is unstable and the equilibria are at the extremes. When play occurs between two different types of members of each of two different species, a more complex but similarly structured analysis is used to determine equilibria.

A game known as the “Hawk-Dove” game is the classic biological example. Analysis of this game parallels that of the Prisoners' Dilemma and Chicken versions of the evolutionary game; evolutionary stable strategies depend on the specifics of the payoff structure. The analysis can also be done when more than two types interact or in very general terms. This theory shows that the requirements for evolutionary stability yield an equilibrium strategy that is equivalent to the Nash equilibrium obtained by rational players.

**Key Terms for Chapter 10**

- evolutionary implies Nash
- evolutionary stability
- evolutionary stable strategy
- fitness
- genotype
- interaction
- invasion of a population by a mutant
- monomorphism
- mutation
- phenotype
- polymorphism
- primary criterion
- secondary criterion
- selection

**Exercises**

1. Prove the following statement: “If a strategy is strongly dominated in the payoff table of a game played by rational players, then in the evolutionary version of the same game it will die out no matter what the initial population mix. If a strategy is weakly dominated, it may coexist with some other types but not in a mixture involving all types.”

2. Suppose that a single play of a Prisoners' Dilemma has the following payoffs:
In a large population of individuals where each individual's behavior is genetically determined, each player will be either a defector (always defects in any play of a Prisoners' Dilemma game) or a tit-for-tat player (in any sequence of Prisoners' Dilemmas, cooperates on the first play, and on any subsequent play does what her opponent did on the previous play). Pairs of randomly chosen players from this population will play "sets" of $n$ (where $n \geq 1$) single plays of this dilemma. The payoff to each player over one whole set ($n$ plays) is the sum of her payoffs over the $n$ plays.

(a) Show in a 2-by-2 table the payoffs to an individual of each type when, over one set of plays, each meets an opponent of each of the two types.

<table>
<thead>
<tr>
<th>Player 1</th>
<th>Cooperate</th>
<th>Defect</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cooperate</td>
<td>3,3</td>
<td>1,4</td>
</tr>
<tr>
<td>Defect</td>
<td>4,1</td>
<td>2,2</td>
</tr>
</tbody>
</table>

Let the population proportion of defectors be $p$ and the proportion of tit-for-tat players be $(1-p)$. Each individual in the population plays sets of dilemmas repeatedly, matched against a new randomly chosen opponent for each new set. A tit-for-tat player always begins each new set by cooperating on its first play.

(b) Find the fitness (average payoff over one set against a randomly chosen opponent) for a defector.

(c) Find the fitness for a tit-for-tat player.

(d) Use the answers to parts (b) and (c) to show that when $p > (n-2)/(n-1)$, the defector type has greater fitness, and that when $p < (n-2)/(n-1)$, the tit-for-tat type has greater fitness.

(e) If evolution leads to a gradual increase in the proportion of the fitter type in the population, what are the possible eventual equilibrium outcomes of this process for the population described above? (That is, what are the possible equilibria and which are evolutionary stable?) Use a diagram with the fitness graphs to illustrate your answer.

(f) In what sense does longer repetition (larger $n$) facilitate the evolution of cooperation?

3. Consider the twice-repeated Prisoners' Dilemma of Figure 10.2. Suppose that as well as the three strategy types A, T, and N present there, a fourth type S can also exist. This type cooperates on the first play and cheats on the second play of each episode of two successive plays against the same opponent.
(a) Show the 4-by-4 fitness table for this game.

(b) If the population initially has some of each of the four types, show that the types N, T, and S will die out in that order.

(c) Show that once a strategy (N, T, or S) has died out, as in part (b), a small proportion of mutants of that type that reappeared in the future could not successfully invade the population at that time.

(d) What is the ESS in this game? Why is it different from that found in the text?

4. In the Assurance (meeting) game in this chapter, the payoffs were meant to describe the value of something material gained by the players in the various outcomes; they could be prizes given upon a successful meeting, for example. Then other individuals in the population might observe the expected payoffs (fitness) of the two types, see which was higher, and gradually imitate the fitter strategy. Thus the proportions of the two types in the population would change. But a more biological interpretation can be made. Suppose the row players are always female, and the column players always male. When two of these players meet successfully, they pair off, and their children are of the same type as the parents. Thus the types would proliferate or die off as a result of meetings. The formal mathematics of this new version of the game makes it a "two-species game" (although the biology of it does not, of course). Thus the proportion of T-type females in the population – call this proportion $x$ – need not equal the proportion of T-type males – call this proportion $y$.

Examine the dynamics of $x$ and $y$ using methods similar to those used in the text for the Battle of the Two Cultures game. Find the stable outcome or outcomes of this dynamic process.