GAMES OF STRATEGY

THIRD EDITION

Avinash Dixit
Princeton University

Susan Skeath
Wellesley College

David Reiley
University of Arizona and Yahoo! Research

W. W. Norton & Company
New York • London

13

Evolutionary Games

WE HAVE SO FAR STUDIED GAMES with many different features—simultaneous and sequential moves, zero-sum and non-zero-sum payoffs, strategic moves to manipulate rules of games to come, one-shot and repeated play, and even games of collective action in which a large number of people play simultaneously. However, one ground rule has remained unchanged in all of the discussions—namely, that all the players in all these games are rational: each player has an internally consistent value system, can calculate the consequences of her strategic choices, and makes the choice that best favors her interests.

In applying this rule, 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 learned what their best choices are by experience. The assumptions ensure that a player does not attribute any false naiveté to her 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 on such a foundation. Economists, too, should not take rationality for granted, as pointed out in Chapter 5. The trouble is finding a feasible alternative. Although we may not wish to impose

conscious and perfectly calculating rationality on players, we do not want to abandon the idea that some strategies are better than others. We want good strategies to be rewarded with higher payoffs; we want players to observe or imitate success and to experiment with new strategies; we 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 to rationality in the biological theory of evolution and evolutionary dynamics and will study its lessons in this chapter.

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 of this success as meaning the common but misleading phrase "survival of the fittest"; however, the ultimate test of biological fitness is not mere survival, but reproductive success. That is what enables an animal to pass on its genes to the next generation and perpetuate its phenotype. The fitter phenotypes 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 eventually to a 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.

At any time, a population may contain some or all of its biologically conceivable phenotypes. Those that are fitter than others will increase in proportion, some unfit phenotypes may die out, and other phenotypes not currently in the population may try to invade it. Biologists call a configuration of a population and its current phenotypes **evolutionary stable** if the population cannot be invaded successfully by any mutant. This is a static test, but often a more dynamic criterion is applied: a configuration is evolutionary stable if it is

the limiting outcome of the dynamics of selection, starting from any arbitrary mixture of phenotypes in the population.[1]

The fitness of a phenotype depends on the relationship of the individual organism 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 those of the rest of its species. Thus the fitness of a particular animal—with its behavioral traits, such as aggression and sociability—depends on whether other members of its species are predominantly aggressive or passive, crowded or dispersed, and so on. For our purpose, this **interaction** between phenotypes within a species is the most interesting aspect of the story. Sometimes an individual member of a species interacts with members of another species; then the fitness of a particular type of sheep, for example, 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 biological process of evolution finds a ready parallel in game theory. The behavior of a phenotype can be thought of as a *strategy* of the animal in its interaction with others—for example, whether to fight or to retreat. The difference is that the choice of strategy is not a purposive calculation as it would be in standard game theory; rather, it is a genetically predetermined fixture of the phenotype. The interactions lead to *payoffs* to the phenotypes. In biology, the payoffs measure the evolutionary or reproductive fitness; when we apply these ideas outside of biology, they can have other connotations of success in the social, political, or economic games in question.

The payoffs or fitness numbers can be shown in a payoff table just like that for a standard game, with all conceivable phenotypes of one animal arrayed along the rows of the matrix and those of the other along the columns of the matrix. If more animals interact simultaneously—which is called **playing the field** in biology—the payoffs can be shown by functions like those for collective-action games in Chapter 12. We will consider pair-by-pair matches for most of this chapter and will look at the other case briefly in Section 9.

Because the population is a mix of phenotypes, different pairs selected from it will bring to their interactions different combinations of strategies. The actual quantitative measure of the fitness of a phenotype is the average payoff that it gets in all its interactions with others in the population. Those animals with higher fitness will have greater evolutionary success. The eventual outcome of

[1]The dynamics of phenotypes is driven by an underlying dynamics of genotypes but, at least at the elementary level, evolutionary biology focuses its analysis at the phenotype level and conceals the genetic aspects of evolution. We will do likewise in our exposition of evolutionary games. Some theories at the genotypes level can be found in the materials cited in footnote 2.

the population dynamics will be an evolutionary stable configuration of the population.

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 by using the preexisting body of game theory, drawing from its language but modifying the assumption of 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 of inheritance of strategies can be interpreted more broadly in applications of the theory other than in biology. In human interactions, a strategy may be embedded in a player's mind for a variety of reasons—not only genetics but also (and probably more important) socialization, cultural background, education, or a rule of thumb based on past experience. The population can consist of a mixture of different people 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, whereas others are mainly concerned with their own reelection; similarly, some firms may pursue profit alone, whereas others are motivated by social or ecological objectives. We can call each logically conceivable strategy that can be embedded in this way a phenotype for the population of players in the context being studied.

[2]Robert Pool, "Putting Game Theory to the Test," *Science,* vol. 267 (March 17, 1995), pp. 1591–1593, is a good article for general readers and has many examples from biology. John Maynard Smith deals with such games in biology in his *Evolutionary Genetics* (Oxford: Oxford University Press, 1989), chap. 7, and *Evolution and the Theory of Games* (Cambridge: Cambridge University Press, 1982); the former also gives much background on evolution. Recommended for advanced readers are Peter Hammerstein and Reinhard Selten, "Game Theory and Evolutionary Biology," in *Handbook of Game Theory,* vol. 2, ed. R. J. Aumann and S. Hart (Amsterdam: North Holland, 1994), pp. 929–993; and Jorgen Weibull, *Evolutionary Game Theory* (Cambridge: MIT Press, 1995).

[3]Indeed, applications of the evolutionary perspective need not stop with game theory. 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 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.

From a population with its heterogeneity of embedded strategies, pairs of phenotypes are repeatedly randomly selected to interact (play the game) with others of the same or different "species." In each interaction, the payoff of each player depends on the strategies of both; this dependence is governed by the usual "rules of the game" and illustrated in the game table or tree. The *fitness* of a particular strategy is defined as its aggregate or average payoff in its pairings with all the strategies in the population. Some strategies have a higher level of fitness than others; in the next generation—that is, the next round of play—these higher-fitness strategies 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. Occasionally, someone may experiment or adopt a previously unused strategy from the collection of those that are logically conceivable. This corresponds to the emergence of a mutant. If the new strategy is fitter than the ones currently being used, it will start to be used by larger proportions of the population. The central question is whether this process of selective proliferation, decay, and mutation of certain strategies in the population will have an evolutionary stable outcome and, if so, what it will be. In regard to the examples just cited, will society end up with a situation in which all politicians are concerned with reelection and all firms with profit? In this chapter, we develop the framework and methods for answering such questions.

Although we use the biological analogy, the reason that the fitter strategies proliferate and the less fit ones die out in socioeconomic 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 will try to imitate them, and some purposive thinking and revision of previous rules of thumb will take place between successive rounds. Such "social" and "educational" mechanisms of transmission are far more important in most strategic games than any biological genetics; indeed, this is how the reelection orientation of legislators and the profit-maximization motive of firms are reinforced. Finally, conscious experimentation with new strategies substitutes for the accidental mutation in biological games.

Evolutionary stable configurations of biological games can be of two kinds. First, a single phenotype may prove fitter than any others, and the population may come to consist of it alone. Such an evolutionary stable outcome is called **monomorphism**—that is, a single (mono) form (morph). In that case, the unique prevailing strategy is called an **evolutionary stable strategy (ESS).** The other possibility is that two or more phenotypes may be equally fit (and fitter than some others not played), so they may be able to 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 individual players pursue different pure strategies.

The whole setup—the population, its conceivable collection of phenotypes, the payoff matrix in the interactions of the phenotypes, and the rule for the evolution of population proportions of the phenotypes in relation to their fitness—constitutes an evolutionary game. An evolutionary stable configuration of the population can be called an *equilibrium* of the evolutionary game.

In this chapter, we develop some of these ideas, as usual through a series of illustrative examples. We begin with symmetric games, in which 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 the payoff table for the game, each can be designated as the row player or the column player with no difference in outcome.

2 PRISONERS' DILEMMA

Suppose a population is made up of two phenotypes. One type consists of players who are natural-born cooperators; they always work toward the outcome that is jointly best for all players. The other type consists of the defectors; they work only for themselves. As an example, we use the restaurant pricing game described in Chapter 5 and presented in a simplified version in Chapter 11. Here, we use the simpler version in which only two pricing choices are available, the jointly best price of $26 or the Nash equilibrium price of $20. A cooperator restaurateur would always choose $26, whereas a defector would always choose $20. The payoffs (profits) of each type in a single play of this discrete dilemma are shown in Figure 13.1, reproduced from Figure 11.2. Here we call the players simply Row and Column because each can be any individual

		COLUMN	
		20 (Defect)	26 (Cooperate)
ROW	20 (Defect)	288, 288	360, 216
	26 (Cooperate)	216, 360	324, 324

FIGURE 13.1 Prisoners' Dilemma of Pricing ($100s per Month)

restaurateur in the population who is chosen at random to compete against another random rival.

Remember that under the evolutionary scenario, no one has the choice between defecting and cooperating; each is "born" with one trait or the other predetermined. Which is the more successful (fitter) trait in the population?

A defecting-type restaurateur gets a payoff of 288 ($28,800 a month) if matched against another defecting type and a payoff of 360 ($36,000 a month) if matched against a cooperating type. A cooperating type gets 216 ($21,600 a month) if matched against a defecting type and 324 ($32,400 a month) if matched against another cooperating type. No matter what the type of the matched rival, the defecting type does better than the cooperating type.[4] Therefore the defecting type has a better expected payoff (and is thus fitter) than does the cooperating 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 that she will meet another cooperator (and get 324) is x and that she will meet a defector (and get 216) is $(1 - x)$. Therefore a typical cooperator's expected payoff is $324x + 216(1 - x)$. For a defector, the probability of meeting a cooperator (and getting 360) is x and that of meeting another defector (and getting 288) is $(1 - x)$. Therefore a typical defector's expected profit is $360x + 288(1 - x)$. Now it is immediately apparent that

$$360x + 288(1 - x) > 324x + 216(1 - x) \text{ for all } x \text{ between 0 and 1.}$$

Therefore a defector has a higher expected payoff and is fitter than a cooperator. 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.

What if the population initially consists of all defectors? Then in this case no mutant (experimental) cooperator will survive and multiply to take over the population; in other words, the defector population cannot be invaded successfully by mutant cooperators. Even for a very small value of x—that is, when the proportion of cooperators in the population is very small—the cooperators remain less fit than the prevailing defectors, and their population proportion will not increase but will be driven to zero; the mutant strain will die out.

Our analysis shows both that defectors have higher fitness than cooperators and that an all-defector population cannot be invaded by mutant cooperators. Thus the evolutionary stable configuration of the population is monomorphic, consisting of the single strategy or phenotype Defect. We therefore call Defect the evolutionary stable strategy for this population engaged in this dilemma game. Note that Defect is a strictly dominant strategy in the rational behavior analysis of this same game. This result is very general: if a game has a strictly dominant strategy, that strategy will also be the ESS.

[4] In the rational behavior context of the preceding chapters, we would say that Defect is the strictly dominant strategy.

A. The Repeated Prisoners' Dilemma

We saw in Chapter 11 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 three times in succession. The overall payoff to a player from such an interaction is the sum of what she gets in the three rounds.

Each individual player is still programmed to play just one strategy, but that strategy has to be a complete plan of action. In a game with three moves, a strategy can stipulate an action in the second or third play that depends on what happened in the first or second 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 as long as you cooperated on the preceding play; and I will defect in all later plays if you defect in an early play" is also a valid strategy; in fact, this last strategy is just tit-for-tat (TFT).

To keep the initial analysis simple, we suppose in this section that there are just two types of strategies that can possibly exist in the population: always defect (A) and tit-for-tat (T). Pairs are randomly selected from the population, and each selected pair plays the game a specified number of times. The fitness of each player is simply the sum of her payoffs from all the repetitions played against her specific opponent. We examine what happens with two, three, and more generally n such repetitions in each pair.

1. TWICE-REPEATED PLAY Figure 13.2 shows the payoff table for the game in which two members of the restaurateur population meet and play against one another exactly twice. If both players are A types, both defect both times, and Figure 13.1 shows that then each gets 288 each time, for a total of 576. If both are T types, defection never starts, and each gets 324 each time, for a total of 648. If one is an A type and the other a T type, then on the first play the A type defects and the T type cooperates, so the former gets 360 and the latter 216. On the second play both defect and get 288. So the A type's total payoff is 360 + 288 = 648, and the T type's total is 216 + 288 = 504.

In the twice-repeated dilemma, we see that A is only weakly dominant. It is easy to see that if the population is all A, then T-type mutants cannot invade, and A is an ESS. But if the population is all T, then A-type mutants cannot do any better than the T types. Does this mean that T must be another ESS, just as it would be a Nash equilibrium in the rational-game-theoretic analysis of this game? The answer is no. If the population is initially all T and a few A mutants enter, then the mutants would meet the predominant T types most of the time and would do as well as T does against another T. But occasionally an A mutant would meet

		COLUMN	
		A	T
ROW	A	576, 576	648, 504
	T	504, 648	648, 648

FIGURE 13.2 Outcomes in the Twice-Repeated Prisoners' Dilemma ($100s)

another A mutant, and in this match she does better than would a T against an A. Thus the mutants have just *slightly* higher fitness than that of a member of the predominant phenotype. This advantage leads to an increase, albeit a slow one, in the proportion of mutants in the population. Therefore an all-T population *can* be invaded successfully by A mutants; T is not an ESS.

Our reasoning relies on two tests for an ESS. First we see if the mutant does better or worse than the predominant phenotype when each is matched against the predominant type. If this primary criterion gives a clear answer, that settles the matter. But if the primary criterion gives a tie, then we use a tie-breaking, or secondary, criterion: does the mutant fare better or worse than a predominant phenotype when each is matched against a mutant? Ties are exceptional and most of the time we do not need the secondary criterion, but it is there in reserve for situations such as the one illustrated in Figure 13.2.[5]

II. THREEFOLD REPETITION Now suppose each matched pair from the (A,T) population plays the game three times. Figure 13.3 shows the fitness outcomes, summed over the three meetings, for each type of player when matched against rivals of each type.

To see how these fitness numbers arise, consider a couple of examples. When two T players meet each other, both cooperate the first time, and therefore both cooperate the second time and the third time as well; both get 324 each time, for a total of 972 each over 3 months. When a T player meets an A player, the latter does well the first time (360 for the A type versus 216 for the T player), but then the T player also defects the second and third times, and each gets 288 in both of those plays (for totals of 936 for A and 792 for T).

The relative fitnesses of the two types depend on the composition of the population. If the population is almost wholly A type, then A is fitter than T (because

[5]This game is just one example of a twice-repeated dilemma. With other payoffs in the basic game, twofold repetition may not have ties. That is so in the husband–wife jail story of Chapter 4. If both the primary and secondary criteria yield ties, neither phenotype satisfies our definition of ESS, and we need to broaden our understanding of what constitutes an equilibrium in the evolutionary game. We consider such a possibility in Section 7 and provide the general theory for dealing with such an outcome in Section 8.

		COLUMN	
		A	T
ROW	A	864, 864	936, 792
	T	792, 936	972, 972

FIGURE 13.3 Outcomes in the Thrice-Repeated Prisoners' Dilemma ($100s)

A types meeting mostly other A types earn 864 most of the time, but T types most often get 792). But if the population is almost wholly T type, then T is fitter than A (because T types earn 972 when they meet mostly other Ts, but A types earn 936 in such a situation). Each type is fitter when it already predominates in the population. Therefore T cannot invade successfully when the population is all A, and vice versa. Now there are two possible evolutionary stable configurations of the population; in one configuration, A is the ESS and, in the other, T is the ESS.

Next consider the evolutionary dynamics when the initial population is made up of a mixture of the two types. How will the composition of the population evolve over time? Suppose a fraction x of the population is T type and the rest, $(1 - x)$, is A type.[6] An individual A player, pitted against various opponents chosen from such a population, gets 936 when confronting a T player, which happens a fraction x of the times, and 864 against another A player, which happens a fraction $(1 - x)$ of the times. This gives an average expected payoff of

$$936x + 864(1 - x) = 864 + 72x$$

for each A player. Similarly, an individual T player gets an average expected payoff of

$$972x + 792(1 - x) = 792 + 180x.$$

Then a T player is fitter than an A player if the former earns more on average; that is, if

$$\begin{aligned} 792 + 180x &> 864 + 72x \\ 108x &> 72 \\ x &> 2/3. \end{aligned}$$

[6]Literally, the fraction of any particular type in the population is finite and can only take on values such as 1/1,000,000, 2/1,000,000, and so on. But, if the population is sufficiently large and we show all such values as points on a straight line, as in Figure 13.4, 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. By the same reasoning, if one individual member goes to jail and is removed from the population, her removal does not change the population's proportions of the various phenotypes.

In other words, if more than two-thirds (67%) of the population is already T type, then T players are fitter and their proportion will grow until it reaches 100%. If the population starts with less than 67% 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 the A players. The evolutionary dynamics move the population toward one of the two extremes, each of which is a possible ESS. The dynamics leads to the same conclusion as the static test of mutants' invasion. This is a common, although not universal, feature of evolutionary games.

Thus we have identified two evolutionary stable configurations of the population. In each one the population is all of one type (monomorphic). For example, if the population is initially 100% T, then even after a small number of mutant A types arise, the population mix will still be more than 66.66 . . . % 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 number of T-type mutants will leave the population mix with less than 66.66 . . . % T, so the A types will be fitter and the mutant T strain will die out. And as we saw earlier, experimenting mutants of type N can never succeed in a population of A and T types that is either largely T or largely A.

What if the initial population has exactly 66.66 . . . % T players (and 33.33 . . . % A players)? Then the two types are equally fit. We could call this *poly*morphism. But it is not really a suitable candidate for an evolutionary stable configuration. 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 accumulate until the ESS with 100% of that type is reached. This is just an application of the secondary criterion for evolutionary stability. We will sometimes loosely speak of such a configuration as an unstable equilibrium, so as to maintain the parallel with ordinary game theory where mutations are not a consideration and a delicately balanced equilibrium can persist. But in the strict logic of the biological process, it is not an equilibrium at all.

This reasoning can be shown in a simple graph that closely resembles the graphs that 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, as shown in Figure 13.4. Along the horizontal axis, we measure the proportion x of T players in the population from 0 to 1. We measure fitness along the vertical axis. Each line shows the fitness of one type. The line for the T type starts lower (at 792 compared with 864 for the A-type line) and ends higher (972 against 936). The two lines cross when $x = 0.66\ldots$. To the right of this point, the T type is fitter, so its population proportion

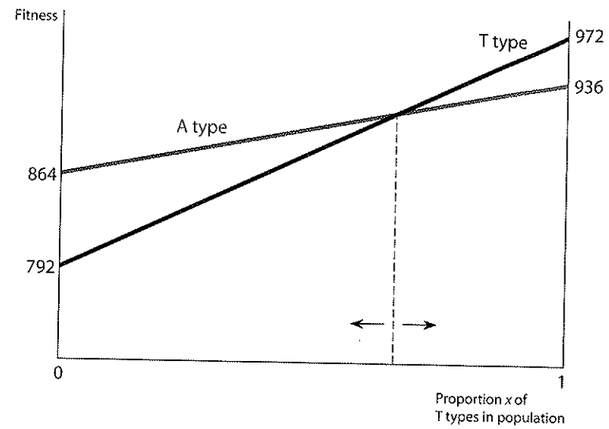

FIGURE 13.4 Fitness Graphs and Equilibria for the Thrice-Repeated Prisoners' Dilemma

increases over time and x *increases* toward 1. Similarly, to the left of this point, the A type is fitter, so its population proportion increases over time and x *decreases* toward 0. Such diagrams often prove useful as visual aids, and we will use them extensively.[7]

B. Multiple Repetitions

What if each pair plays some unspecified number of repetitions of the game? Let us focus on a population consisting of only A and T types in which interactions between random pairs occur n times (where $n > 2$). The table of the total outcomes from playing n repetitions is shown in Figure 13.5. When two A types meet, they always defect and earn 288 every time, so each gets $288n$ in n plays. When two T types meet, they begin by cooperating, and no one is the first to defect, so they earn 324 every time, for a total of $324n$. When an A type meets a T type, on the first play the T type cooperates and the A type defects, and so the A type gets 360 and the T type gets 216; thereafter the T type retaliates against the preceding defection of the A type for all remaining plays, and each gets 288 in all of the remaining $(n - 1)$ plays. Thus the A type earns a total of $360 + 288(n - 1) = 288n + 72$ in n plays against a T type, whereas the T type gets $216 + 288(n - 1) = 288n - 72$ in n plays against an A type.

[7]You should now draw a similar graph for the twice-repeated case. You will see that the A line is above the T line for all values of x less than 1, but the two meet on the right-hand edge of the figure where $x = 1$.

		COLUMN	
		A	T
ROW	A	$288n, 288n$	$288n + 72, 288n - 72$
	T	$288n - 72, 288n + 72$	$324n, 324n$

FIGURE 13.5 Outcomes in the *n*-fold-Repeated Dilemma

If the proportion of T types in the population is x, then a typical A type gets $x(288n + 72) + (1 - x)288n$ on average, and a typical T type gets $x(324n) + (1 - x)(288n - 72)$ on average. Therefore the T type is fitter if

$$x(324n) + (1 - x)(288n - 72) > x(288n + 72) + (1 - x)288n$$
$$36xn > 72$$
$$x > \frac{72}{36n} = \frac{2}{n}.$$

Once again we have two monomorphic ESSs, one all T (or $x = 1$, to which the process converges starting from any $x > 2/n$) and the other all A (or $x = 0$, to which the process converges starting from any $x < 2/n$). As in Figure 13.4, there is also an unstable polymorphic equilibrium at the balancing point $x = 2/n$.

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 2/10, or 0.2. So if the population initially is 20% T players, in a situation where each pair plays 10 repetitions, the proportion of T types will grow until they reach 100%. Recall that when pairs played three repetitions ($n = 3$), the T players needed an initial strength of 67% or more to achieve this outcome, and only two repetitions meant that T types needed to be 100% of the population to survive. (We see the reason for this outcome in our expression for the critical value for x, which shows that when $n = 2$, x must be above 1 before the T types are fitter.) Remember, too, that a population consisting of all T players achieves cooperation. Thus cooperation emerges from a larger range of the initial conditions when the game is repeated more times. 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.

C. Comparing the Evolutionary and Rational-Player Models

Finally, let us return to the thrice-repeated game illustrated in Figure 13.3 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 in which both play A and the other in which both play T. There is also an

equilibrium in mixed strategies, in which T is played 67% of the time and A 33% of the time. The first two are just the monomorphic ESSs that we found, and the third is the unstable polymorphic evolutionary equilibrium. In other words, there is a close relation between evolutionary and consciously rational perspectives on games.

That is not a coincidence. An ESS must be a Nash equilibrium of the game played by consciously rational players with the same payoff structure. To see this, suppose the contrary for the moment. If all players using some strategy—call it S—is not a Nash equilibrium, then some other strategy—call it R—must yield a higher payoff for one player when played against S. A mutant playing R will achieve greater fitness in a population playing S and so will invade successfully. Thus S cannot be an ESS. In other words, if all players using 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 for all players to use S.

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 eventually to a stable strategy, then the outcome must be the same as that resulting from consciously rational play.

Although an ESS must be a Nash equilibrium of the corresponding rational-play game, the converse is not true. We have seen two examples of this. In the twice-repeated dilemma game of Figure 13.2 played rationally, T would be a Nash equilibrium in the weak sense that if both players choose T, neither has any positive gain from switching to A. But in the evolutionary approach A can arise as a mutation and can successfully invade the T population. And in the thrice-repeated dilemma game of Figures 13.3 and 13.4, rational play would produce a mixed-strategy equilibrium. But the biological counterpart to this mixed-strategy equilibrium, the polymorphic state, can be successfully invaded by mutants and is therefore not a true evolutionary stable equilibrium. Thus the biological concept of stability can help us select from a multiplicity of Nash equilibria of a rationally played game.

There is one limitation of our analysis of the repeated game. At the outset, we allowed just two strategies: A and T. 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 the 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 that we place on the set of strategies that can possibly exist in a particular game have important implications for which of these strategies (if any) can be evolutionary stable. In the thrice-repeated prisoners' dilemma example, if we had allowed for a strategy S that cooperated on the first play and defected on the

second and third, then S-type mutants could have successfully invaded an all-T population, so T would not have been an ESS. We develop this possibility further in the exercises at the end of this chapter.

3 CHICKEN

Remember our 1950s youths racing their cars toward one another and seeing who will be the first to swerve to avoid a collision? Now we suppose the players have no choice in the matter: 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 13.6 shows the payoff table for any two such players—say, A and B. (The numbers replicate those we used before in Figure 4.14.)

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, whereas all the Wimps meeting their own types will 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 appropriately 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 who are healthy.

In other words, each type is fitter when it is relatively rare in the population. Therefore each can successfully invade 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

		B	
		Wimp	Macho
A	Wimp	0, 0	−1, 1
	Macho	1, −1	−2, −2

FIGURE 13.6 Payoff Table for Chicken

-1 for a fraction x of the time. Therefore the fitness of a Wimp is $0 \times (1 - x) - 1 \times x = -x$. Similarly, the fitness of a Macho is $1 \times (1 - x) - 2x = 1 - 3x$. The Macho type is fitter if

$$\begin{aligned} 1 - 3x &> -x \\ 2x &< 1 \\ x &< 1/2. \end{aligned}$$

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 1/2, and this 50–50 mix will be the stable polymorphic ESS.

Figure 13.7 shows this outcome graphically. Each straight line shows the fitness (the 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 $-x$, as we saw two paragraphs ago. This is the gently falling line that starts at the height 0 when $x = 0$ and goes to -1 when $x = 1$. The corresponding function for the Macho type is $1 - 3x$. This is the rapidly falling line that starts at height 1 when $x = 0$ and falls to -2 when $x = 1$. The Macho line lies above the Wimp line for $x < 1/2$ and below it for $x > 1/2$, showing that the Macho types are fitter when the value of x is small and the Wimps are fitter when x is large.

Now we can compare and contrast the evolutionary theory of this game with our earlier theory of Chapters 4 and 7, 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

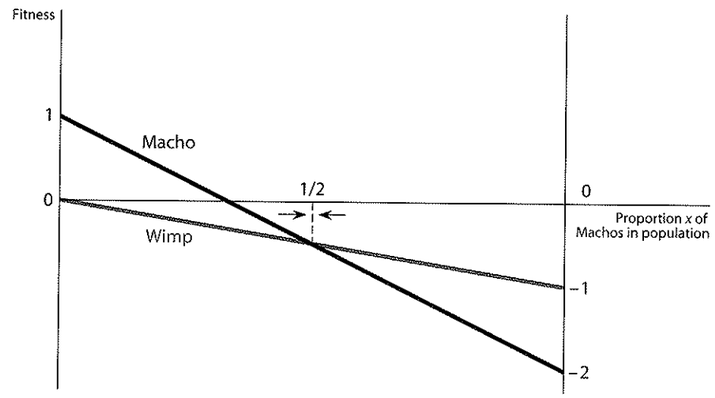

FIGURE 13.7 Fitness Graphs and Polymorphic Equilibrium for Chicken

the other swerves, and one in mixed strategies, where each player goes straight with a probability of 1/2 and swerves with a probability of 1/2.

If the population is truly 100% Macho, then all players are equally fit (or equally unfit). Similarly, in a population of nothing but Wimps, all are equally fit. But these monomorphic configurations are unstable. In an all-Macho population, a Wimp mutant will outscore them and invade successfully.[8] Once some Wimps get established, no matter how few, our analysis shows that their proportion will rise inexorably toward 1/2. Similarly, an all-Wimp population is vulnerable to a successful invasion of mutant Machos, and the process again goes to the same polymorphism. Thus the polymorphic configuration is the only true evolutionary stable outcome.

Most interesting is the connection between the mixed-strategy equilibrium of the rationally played game and the polymorphic equilibrium 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 member of the population uses a pure strategy, but different members use different strategies, and so we see a mixture in the population.[9]

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 see it in its generality later, in Section 6. Indeed, evolutionary stability provides an additional rationale for choosing one of the many Nash equilibria in such rationally played games.

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 emerges as the stable outcome of a natural dynamic process in which each player tries to improve his payoff against the population that he confronts.

4 THE ASSURANCE GAME

Among the important classes of strategic games introduced in Chapter 4, we have studied prisoners' dilemma and chicken from the evolutionary perspective.

[8] *The Invasion of the Mutant Wimps* could be an interesting science-fiction comedy movie.

[9] There can also be evolutionary stable mixed strategies in which each member of the population adopts a mixed strategy. We investigate this idea further in Section 6.E.

That leaves the assurance game. We illustrated this type of game in Chapter 4 with the story of two undergraduates, Harry and Sally, deciding where to meet for coffee. In the evolutionary context, each player is born liking either Starbucks or Local Latte and the population includes some of each type. Here we assume that pairs of the two types, which we classify generically as men and women, are chosen at random each day to play the game. We denote the strategies now by S (for Starbucks) and L (for Local Latte). Figure 13.8 shows the payoff table for a random pairing in this game; the payoffs are the same as those illustrated earlier in Figure 4.12.

If this were a game played by rational strategy-choosing players, there would be two equilibria in pure strategies: (S, S) and (L, L). 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 to coordinate through a convergence of expectations—that is, by finding a focal point.

The rationally played game has a third equilibrium, in mixed strategies, that we found in Chapter 7. In that equilibrium, each player chooses Starbucks with a probability of 2/3 and Local Latte with a probability of 1/3; the expected payoff for each player is 2/3. As we showed in Chapter 7, this payoff is worse than the one associated with the less attractive of the two pure-strategy equilibria, (S, S), because independent mixing leads the players to make clashing or bad choices quite a lot of the time. Here, the bad outcome (a payoff of 0) has a probability of 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 member is hardwired, either to choose S or to choose L. Randomly chosen pairs of such people are assigned to attempt a meeting. Suppose x is the proportion of S types in the population and $(1 - x)$ is that of L types. Then the fitness of a particular S type—her expected payoff in a random encounter of this kind—is $x \times 1 + (1 - x) \times 0 = x$. Similarly, the fitness of each L type is $x \times 0 + (1 - x) \times 2 = 2(1 - x)$. Therefore the S type is fitter when $x > 2(1 - x)$, or for $x > 2/3$. The L type is fitter when $x < 2/3$. At the balancing point $x = 2/3$, the two types are equally fit.

		WOMEN	
		S	L
MEN	S	1, 1	0, 0
	L	0, 0	2, 2

FIGURE 13.8 Payoff Matrix for the Assurance Game

As in chicken, once again the probabilities associated with the mixed-strategy equilibrium that would obtain under rational choice seem to reappear under evolutionary rules as the population proportions in a polymorphic equilibrium. But now this mixed equilibrium is not stable. The slightest chance departure of the proportion x from the balancing point 2/3 will set in motion a cumulative process that takes the population mix farther away from the balancing point. If x increases from 2/3, the S type becomes fitter and propagates faster, increasing x even more. If x falls from 2/3, the L type becomes fitter and propagates faster, lowering x even more. Eventually x will either rise all the way to 1 or fall all the way to 0, depending on which disturbance occurs. 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 midrange 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 is the same type as you—so population proportions tend to move toward the extremes.

Figure 13.9 illustrates the fitness graphs and equilibria for the assurance game; this diagram is very similar to Figure 13.7. 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 13.7 it was the less numerous type.

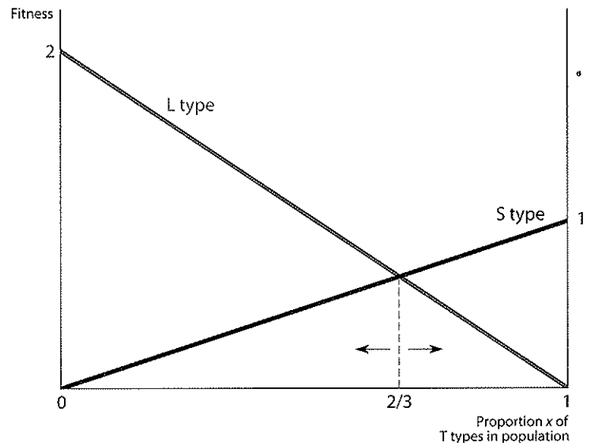

FIGURE 13.9 Fitness Graphs and Equilibria for the Assurance Game

Because each type is less fit when it is rare, only the two extreme monomorphic configurations of the population are possible evolutionary stable states. It is easy to check that both outcomes are ESS according to the static test: 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 there is a positive probability that the players choose clashing strategies. However, the dynamics do not guarantee convergence to the better of the two equilibria when starting from an arbitrary initial mixture of phenotypes—where the population ends up depends on where it starts.

5 INTERACTIONS ACROSS SPECIES

A final class of strategic games to consider is that of the battle-of-the-sexes game. In Chapter 4 (Figure 4.13), we saw that the battle of the sexes game looks similar to the assurance game in some respects. We differentiate between the two by assuming here that "men" and "women" are still interested in meeting at either Starbucks or Local Latte—no meeting yields each a payoff of 0—but now each type prefers a different café. Thus a premium remains on taking mutually consistent actions, just as in the assurance game. 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 (L, L) to (S, S). The players in the battle game differ in theirs: Local Latte gives a payoff of 2 to women and 1 to men, and Starbucks 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.[10] Effectively, they belong to different species (as indeed men and women often believe of each other).

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

Suppose there is a large population of men and a large population of women. One of each "species" is picked, and the two are asked to attempt a meeting. All men agree among themselves about the valuation (payoffs) of Starbucks, Local Latte, and no meeting. Likewise, all women agree among themselves. But within each population, some members are hard-liners and others are compromisers.

[10]In evolutionary biology, games of this type are labeled "asymmetric" games. Symmetric games are those in which a player cannot distinguish the type of another player simply from observing that player's outward characteristics; in asymmetric games, players can tell each other apart.

A hard-liner will always go to his or her species' preferred café. A compromiser recognizes that the other species wants the opposite and goes to that location, to get along.

If the random draws happen to have picked a hard-liner of one species and a compromiser of the other, the outcome is that preferred by the hard-liner's species. We get no meeting if two hard-liners are paired and, strangely, also if two compromisers are chosen, because they go to each other's preferred café. (Remember, they have to choose independently and cannot negotiate. Perhaps even if they did get together in advance, they would reach an impasse of "No, I insist on giving way to your preference.")

We alter the payoff table in Figure 4.13 as shown in Figure 13.10; what were choices are now interpreted as actions predetermined by type (hard-liner or compromiser).

In comparison with all the evolutionary games studied so far, the new feature here is that the row player and the column player come from different species. Although each species is a heterogeneous mixture of hard-liners and compromisers, there is no reason why the proportions of the 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.

We let x be the proportion of hard-liners among the men and y that among the women. Consider a particular hard-liner man. He meets a hard-liner woman a proportion y of the time and gets a 0, and he meets a compromising woman the rest of the time and gets a 2. Therefore his expected payoff (fitness) is $y \times 0 + (1 - y) \times 2 = 2(1 - y)$. Similarly, a compromising man's fitness is $y \times 1 + (1 - y) \times 0 = y$. Among men, therefore, the hard-liner type is fitter when $2(1 - y) > y$, or $y < 2/3$. The hard-liner men 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. This is not surprising; remember that the games that each species plays are now all against the members of the other species.[11]

		WOMEN	
		Hard-liner	Compromiser
MEN	Hard-liner	0, 0	2, 1
	Compromiser	1, 2	0, 0

FIGURE 13.10 Payoffs in the Battle-of-the-Sexes Game

[11]And this finding supports and casts a different light on the property of mixed-strategy equilibria, that each player's mixture keeps 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, the proportion of each species' type keeps all the surviving types of the others species equally fit.

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

Figure 13.11 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 nor decreases. If the current population proportions lie to the left of this line (that is, $x < 2/3$), y is increasing (moving the population proportion of hard-liner women in the vertically upward direction). If the current proportions lie to the right of AB ($x > 2/3$), then y is decreasing (motion vertically downward). Similarly, the horizontal line CD shows all points where $y = 2/3$, which is the balancing point for x. When the population proportion of hard-liner women is below this line (that is, when $y < 2/3$), the proportion of hard-liner men, x, increases (motion horizontal and rightward) and decreases for population proportions above it, when $y > 2/3$ (motion horizontal and leftward).

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 13.11, for example, the dynamics

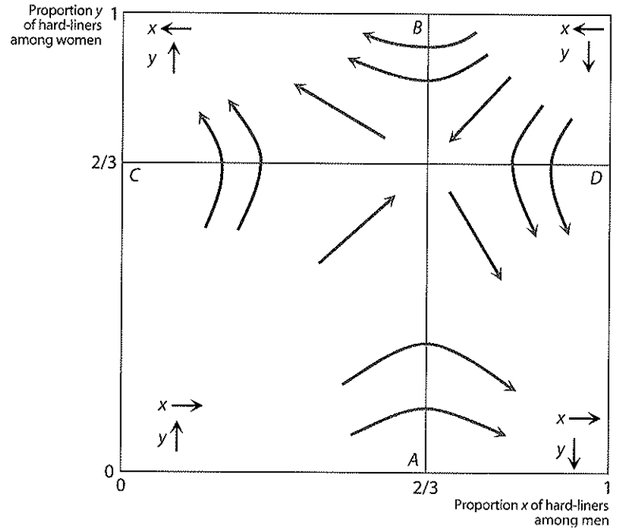

FIGURE 13.11 Population Dynamics in the Battle of the Sexes

entail both y and x 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 paths lead to either the southeast or northwest corners of the diagram; that is, they converge either to (1, 0) or (0, 1). Thus in most cases evolutionary dynamics will lead to a configuration in which one species is entirely hard-line 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 value of x and a larger value of y are more likely to cross the *CD* line first and head for (0, 1)—all hard-line women, $y = 1$—than to hit the *AB* line first and head for (1, 0); similar results follow for a starting position with a small y but a larger x. The species that starts out with more hard-liners will have the advantage of ending up all hard-line 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 outcome in chicken, the polymorphism in the battle of the sexes is unstable. Most chance departures will set in motion a cumulative process that leads to one of the two extreme equilibria; those are the two ESSs for this game. This is a general property—such multispecies games can have only ESSs that are monomorphic for each species.

6 THE HAWK-DOVE GAME

The **hawk-dove game** was the first example biologists studied in their development of the theory of evolutionary games. It has instructive parallels with our analyses so far of the prisoners' dilemma and chicken, so we describe it here 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 avoid a fight. When two Hawk types meet each other, they fight. Each animal is equally likely (probability 1/2) to win and get V or to lose, be injured, and get $-C$. Thus the expected payoff for each is $(V - C)/2$. 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 a 0, whereas the former gets V. Figure 13.12 shows the payoff table.

		B	
		Hawk	Dove
A	Hawk	$(V-C)/2, (V-C)/2$	$V, 0$
	Dove	$0, V$	$V/2, V/2$

FIGURE 13.12 Payoff Table for the Hawk-Dove Game

The analysis of the game is similar to that for the prisoners' dilemma and chicken games, except that the numerical payoffs have been replaced by algebraic symbols. We will compare the equilibria of this game when the players rationally choose to play Hawk or Dove and then compare the outcomes when players are acting mechanically and success is being rewarded with faster reproduction.

A. Rational Strategic Choice and Equilibrium

1. If $V > C$, then the game is a prisoners' dilemma in which the Hawk strategy corresponds to "defect" and Dove corresponds to "cooperate." Hawk is the dominant strategy for each, but (Dove, Dove) is the jointly better outcome.

2. If $V < C$, then it's a game of chicken. Now $(V - C)/2 < 0$ and so Hawk is no longer a dominant strategy. Rather, 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:

$$p(V - C)/2 + (1 - p)V = p \times 0 + (1 - p)V/2$$
$$p = V/C.$$

B. Evolutionary Stability for $V > C$

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 could write the population proportion of the mutant phenotype as m, for mutant, but for clarity in our case we will use 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 will 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 $[d(V/2) + (1 - d) \times 0]$. Because $V > C$, it follows that $(V - C)/2 > 0$. Also, $V > 0$ implies that $V > V/2$. Then, for any value of d between 0 and 1, we have

$$dV + (1 - d)(V - C)/2 > d(V/2) + (1 - d) \times 0,$$

and so the Hawk type is fitter. The Dove mutants cannot successfully invade. The Hawk strategy is evolutionary stable, and the population is monomorphic (all Hawk).

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 is initially all Doves, mutant Hawks can invade and take over. Thus the dynamics confirm that the Hawk strategy is the only ESS. This algebraic analysis affirms and generalizes our earlier finding for the numerical example of the prisoners' dilemma of restaurant pricing (Figure 13.1).

C. Evolutionary Stability for $V < C$

If the initial population is again predominantly Hawks, with a small proportion d of Dove mutants, then each has the same fitness function derived in Section 6.B. When $V < C$, however, $(V - C)/2 < 0$. We still have $V > 0$, and so $V > V/2$. But because d is very small, the comparison of the terms with $(1 - d)$ is much more important than that of the terms with d, so

$$d(V/2) + (1 - d) \times 0 > dV + (1 - d)(V - C)/2.$$

Thus the Dove mutants are fitter than the predominant Hawks and can invade successfully.

But if the initial population is 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 used h for the proportion of the mutant invaders.) The Hawk mutants have a fitness of $[h(V - C)/2 + (1 - h)V]$ compared with $[h \times 0 + (1 - h)(V/2)]$ for the Doves. Again $V < C$ implies that $(V - C)/2 < 0$, and $V > 0$ implies that $V > V/2$. But, when h is small, we get

$$h(V - C)/2 + (1 - h)V > h \times 0 + (1 - h)(V/2).$$

This inequality shows that Hawks are fitter and will successfully invade a Dove population. Thus mutants of each type can invade populations of the other type. The population cannot be monomorphic, and neither pure phenotype can be an ESS. The algebra again confirms our earlier finding for the numerical example of chicken (Figures 13.6 and 13.7).

What happens in the population then when $V < C$? There are two possibilities. In one, every player follows a pure strategy, but the population has a stable mix of players following different strategies. This is the polymorphic equilibrium developed for chicken in Section 13.3. The other possibility is that every player uses a mixed strategy. We begin with the polymorphic case.

D. $V < C$: Stable Polymorphic Population

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

$$h(V - C)/2 + (1 - h)V > (1 - h)(V/2),$$

which simplifies to:

$$\begin{aligned} h(V - C)/2 + (1 - h)(V/2) &> 0 \\ V - hC &> 0 \\ h &< V/C. \end{aligned}$$

The Dove type is then fitter when $h > V/C$, or when $(1 - h) < 1 - V/C = (C - V)/C$. 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 $h = V/C$. This is exactly the probability with which each individual member plays the Hawk strategy in the mixed-strategy Nash equilibrium of the game under the assumption of rational behavior, as calculated in Section 6.A. Again, we have an evolutionary "justification" for the mixed-strategy outcome in chicken.

We leave it to you to draw a graph similar to that in Figure 13.7 for this case. Doing so will require you to determine the dynamics by which the population proportions of each type converge to the stable equilibrium mix.

E. $V < C$: Each Player Mixes Strategies

Recall the equilibrium mixed strategy of the rational-play game calculated earlier in Section 6.A in which $p = V/C$ was the probability of choosing to be a Hawk, while $(1 - p)$ was the probability of choosing to be a Dove. Is there a parallel in the evolutionary version, with a phenotype playing a mixed strategy? Let us examine this possibility. We still have H types who play the pure Hawk strategy and D types who play the pure Dove strategy. But now a third phenotype called M can exist; such a type plays a mixed strategy in which it is a Hawk with probability $p = V/C$ and a Dove with probability $1 - p = 1 - V/C = (C - V)/C$.

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

$$\begin{aligned} p\,\frac{V - C}{2} + (1 - p)V &= \frac{V}{C}\,\frac{V - C}{2} - \frac{C - V}{C}\,V \\ &= -\frac{1}{2}\,\frac{V}{C}(C - V) + \frac{V}{C}(C - V) \\ &= V\,\frac{(C - V)}{2C}. \end{aligned}$$

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

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

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, we call this common payoff K, where $K = V(C - V)/2C$.

But these equalities create a problem when we test M for evolutionary stability. Suppose the population consists entirely of M types and that a few mutants of the H type, constituting a very small proportion h of the total population, invade. Then the typical mutant gets the expected payoff $h(V - C)/2 + (1 - h)K$. 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 instance. She then faces an H type for a fraction h of the interactions; in these interactions she plays H a fraction p of the time and gets $(V - C)/2$, and she plays D a fraction $(1 - p)$ of the time and gets 0. Thus the M type's total expected payoff (fitness) is

$$hp(V - C)/2 + (1 - h)K.$$

Because 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.

Evolutionary stability hinges on whether the original population M type is fitter than the mutant H when each is matched against one of the few mutants. Algebraically, M is fitter than H against other mutant H types when $pV(C - V)/2C = pK > (V - C)/2$. In our example here, this condition holds because $V < C$ (so $(V - C)$ is negative) and because K is positive. Intuitively, this condition tells us that 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. Overall, the M type does better when matched against the mutants.

Similarly, the success of a Dove invasion against the M population depends on the comparison between a mutant Dove's fitness and the fitness of an M type. As before, 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 type 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, thereby gaining pV, and plays D a fraction $(1 - p)$ of these times, thereby gaining $(1 - p)V/2$. The Dove's fitness is then $[dV/2 + (1 - d)K]$, while the fitness of the M type is $d \times [pV + (1 - p)V/2] + (1 - d)K$. The final term in each fitness expression is the same, so a Dove invasion is successful only if $V/2$ is greater than $pV + (1 - p)V/2$. This condition does not hold; the

latter expression includes a weighted average of V and $V/2$ that must exceed $V/2$ whenever $V > 0$. Thus the Dove invasion cannot succeed either.

This analysis tells us that M is an ESS. Thus if $V < C$, the population can exhibit either of two evolutionary stable outcomes. 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 THREE PHENOTYPES IN THE POPULATION

If there are only two possible phenotypes (strategies), we can carry out static checks for ESS by comparing the type being considered with just one type of mutant. We can show the dynamics of the population in an evolutionary game with graphs similar to those in Figures 13.4, 13.7, and 13.9. Now we illustrate how the ideas and methods can be used if there are three (or more) possible phenotypes and what new considerations arise.

A. Testing for ESS

Let us reexamine the thrice-repeated prisoners' dilemma of Section 13.2.A.II and Figure 13.3 by introducing a third possible phenotype. This strategy, labeled N, never defects. Figure 13.13 shows the fitness table with the three strategies—A, T, and N.

To test whether any of these strategies is an ESS, we consider whether a population of all one type can be invaded by mutants of one of the other types. An all-A population, for example, cannot be invaded by mutant N or T types; so A is an ESS. An all-N population can be invaded by type-A mutants, however; N lets itself get fooled thrice (shame on it). So N cannot be an ESS.

What about T? An all-T population cannot be invaded by A. But when faced with type-N mutants, the T types find themselves equally matched; notice that

		COLUMN		
		A	T	N
ROW	A	864, 864	936, 792	1080, 648
	T	792, 936	972, 972	972, 972
	N	648, 1080	972, 972	972, 972

FIGURE 13.13 Thrice-Repeated Prisoners' Dilemma with Three Types ($100s)

the four cells showing T and N competing only with each other show identical payoffs for both phenotypes. In this situation the mutant N types would not proliferate, but they would not die out either. A small proportion of mutants could coexist with the (almost) all-T population. Thus T does not satisfy either of the criteria for being an ESS, but it does exhibit some resistance to invasion.

We recognize the resilience shown by the T type in our example by introducing the concept of a **neutral ESS.**[12] In contrast to the standard ESS, in which a member of the main population needs to be strictly fitter than a mutant in a population with a small proportion of mutants, neutral stability requires only that a member of the main population have at least as high a fitness as does a mutant. Then the mutant proportion does not increase but can stay at an initially small level. This is the case when our all-T population is invaded by a small number of mutant N types. In the game illustrated in Figure 13.13, then, we have one standard ESS, strategy A, and one neutral ESS, strategy T.

Let us consider further the situation when an all-T population is invaded by type-N mutants. If the proportion of mutants is sufficiently small, the two types can coexist happily. But if the mutant population is too large a proportion of the full population, then type-A mutants can invade; A types do well against N but poorly against T. To be specific, consider a population with proportions x of N and $(1 - x)$ of T. The fitness of each of these types is 972. The fitness of a type-A mutant in this population is $936(1 - x) + 1{,}080x = 144x + 936$. This exceeds 972 if $144x > 972 - 936 = 36$, or $x > 1/4$. Thus we can have T as a neutral ESS coexisting with some small proportion of N-type mutants, but only so long as the proportion of Ns is less than 25%.

B. Dynamics

To motivate our discussion of dynamics in games with three possible phenotypes, we turn to another well-known game, rock–paper–scissors (RPS). In rational game-theoretic play of this game, each player simultaneously chooses one of the three available actions, either rock (make a fist), paper (lay your hand flat), or scissors (make a scissorlike motion with two fingers). The rules of the game state that rock beats ("breaks") scissors, scissors beat ("cut") paper, and paper beats ("covers") rock; identical actions tie. If players choose different actions, the winner gets a payoff of 1 and the loser gets a payoff of −1; ties yield both players 0.

For an evolutionary example, we turn to the situation faced by the side-blotched lizards living along the California coast. That species supports three types of male mating behavior, each type associated with a particular throat color. Males with blue throats guard a small number of female mates and fend

[12] Weibull describes neutral stability as a weakening of the standard evolutionary stability criteria in his *Evolutionary Game Theory* (p. 46).

		COLUMN			
		Yellow-throated sneaker	Blue-throated guarder	Orange-throated aggressor	q-mix
ROW	Yellow-throated sneaker	0	−1	1	$-q_2 + (1 - q_1 - q_2)$
	Blue-throated guarder	1	0	−1	$q_1 - (1 - q_1 - q_2)$
	Orange-throated aggressor	−1	1	0	$-q_1 + q_2$

FIGURE 13.14 Payoffs in the Three-Type Evolutionary Game

off advances made by yellow-throated males who attempt to sneak in and mate with unguarded females. The yellow-throated sneaking strategy works well against males with orange throats, who maintain large harems and are often out aggressively pursuing additional mates; those mates tend to belong to the blue-throated males, which can be overpowered by the orange-throat's aggression.[13] Their interactions can be modeled by using the payoff structure of the RPS game shown in Figure 13.14. We include a column for a q-mix to allow us to consider the evolutionary equivalent of the game's mixed-strategy equilibrium, a mixture of types in the population.[14]

Suppose q_1 is the proportion of lizards in the population that are yellow throated, q_2 the proportion of blue throats, and the rest, $(1 - q_1 - q_2)$, the proportion of orange throats. The right-hand column of the table shows each Row player's payoffs when meeting this mixture of phenotypes; that is, just Row's fitness. Suppose, as has been shown to be true in the side-splotched lizard population, that the proportion of each type in the population grows when its fitness is positive and declines when it is negative.[15] Then

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

The proportion of yellow-throated types in the population increases when q_2, the proportion of blue-throated types, is small or when $(1 - q_1 - q_2)$, the

[13] For more information about the side-blotched lizards, see Kelly Zamudio and Barry Sinervo, "Polygyny, Mate-Guarding, and Posthumous Fertilizations As Alternative Mating Strategies," *Proceedings of the National Academy of Sciences,* vol. 97, no. 26 (December 19, 2000), pp. 14427–14432.

[14] One exercise in Chapter 7 considers the rational game-theoretic equilibrium of a version of the RPS game. You should be able to verify relatively easily that the game has no equilibrium in pure strategies.

[15] 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. In the exercises, we develop the dynamics more rigorously for readers with sufficient mathematical training.

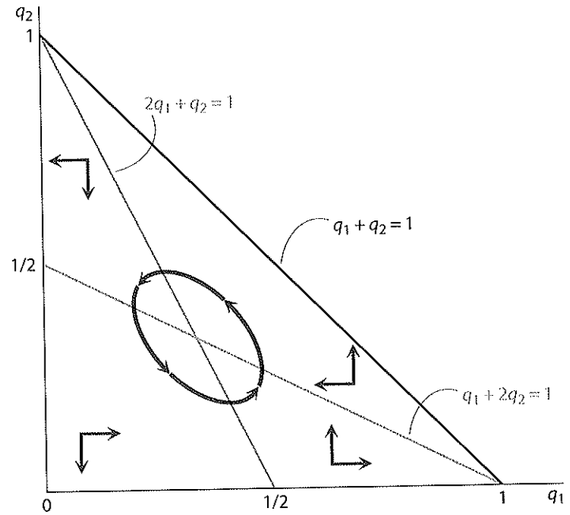

FIGURE 13.15 Population Dynamics in the Evolutionary RPS Game

proportion of orange-throated types, is large. This makes sense; yellow throats do poorly against blue throats but well against orange throats. Similarly, we see that

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

Blue-throated males do better when the proportion of yellow-throated competitors is large or the proportion of orange-throated types is small.

Figure 13.15 shows graphically the population dynamics and resulting equilibria for this game. 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, q_1 (the proportion of yellow-throated players) increases; for combinations above this line, q_1 decreases. The second, steeper line is $2q_1 + q_2 = 1$, which is the balancing line for q_2. To the right of this line (when $2q_1 + q_2 > 1$), q_2 increases; to the left of the line (when $2q_1 + q_2 < 1$), q_2 decreases. Arrows on the diagram show directions of motion of these population proportions; red curves show typical dynamic paths. The general idea is the same as that of Figure 13.13.

On each of the two gray lines, one of q_1 and q_2 neither increases nor decreases. Therefore the intersection of the two lines represents the point

where q_1, q_2, and therefore also $(1 - q_1 - q_2)$, are all constant; this point thus corresponds to a polymorphic equilibrium. It is easy to check that here $q_1 = q_2 = 1 - q_1 - q_2 = 1/3$. These proportions are the same as the probabilities in the rational mixed-strategy equilibrium of the RPS game.

Is this polymorphic outcome stable? In general, we cannot say. The dynamics indicate paths (shown in Figure 13.15 as a single ellipse) that wind around it. Whether these paths wind in a decreasing spiral toward the intersection (in which case we have stability) or in an expanding spiral away from the intersection (indicating instability) depends on the precise response of the population proportions to the fitnesses. It is even possible that the paths circle as drawn, neither approaching nor departing from the equilibrium.

Evidence suggests that the side-splotched lizard population is cycling around the evenly split polymorphic equilibrium point, with one type being slightly more common for a period of a few years but then being overtaken by its stronger competitor. Whether the cycle is approaching the stable equilibrium remains a topic for future study. At least one other example of an RPS-type interaction in an evolutionary game entails three strains of food-poisoning-related *E. coli* bacteria. Each strain displaces one of the others but is displaced by the third, as in the three-type game described earlier. Scientists studying the competition among the three strains have shown that a polymorphic equilibrium can persist if interactions between pairs stay localized, with small clumps of each strain shifting position continuously.[16]

8 SOME GENERAL THEORY

We now generalize the ideas illustrated in Section 6 to get a theoretical framework and set of tools that can then be applied further. This generalization unavoidably requires some slightly abstract notation and a bit of algebra. Therefore we cover only monomorphic equilibria in a single species. Readers who are adept at this level of mathematics can readily develop the polymorphism cases with two species by analogy. Readers who are not prepared for this material or interested in it can omit this section without loss of continuity.[17]

[16]The research on *E. coli* is reported in Martin Nowak and Karl Sigmund, "Biodiversity: Bacterial Game Dynamics," *Nature*, vol. 418 (July 11, 2002), p. 138. If the three strains were forcibly dispersed on a regular basis, a single strain could take over in a matter of days; the "winning" strain out-multiplied a second strain, which could quickly kill off the third.

[17]Conversely, readers who want more details can find them in Maynard Smith, *Evolution and the Theory of Games*, especially pp. 14–15. John Maynard Smith is a pioneer in the theory of evolutionary games.

We consider random matchings from a single species whose population has available strategies I, J, K. . . . Some of them may be pure strategies; some may be mixed. Each individual member is hardwired to play just one of these strategies. We let $E(\mathrm{I}, \mathrm{J})$ denote the payoff to an I player in a single encounter with a J player. The payoff of an I player meeting another of her own type is $E(\mathrm{I}, \mathrm{I})$ in the same notation. We write $W(\mathrm{I})$ 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 type. 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

$$W(\mathrm{I}) = mE(\mathrm{I}, \mathrm{J}) + (1 - m)E(\mathrm{I}, \mathrm{I}),$$

and the fitness of a mutant is

$$W(\mathrm{J}) = mE(\mathrm{J}, \mathrm{J}) + (1 - m)E(\mathrm{J}, \mathrm{I}).$$

Therefore the difference in fitness between the population's main type and its mutant type is

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

Because m is very small, the main type's fitness will be higher than the mutant's if the second half of the preceding expression is positive; that is,

$$W(\mathrm{I}) > W(\mathrm{J}) \quad \text{if } E(\mathrm{I}, \mathrm{I}) > E(\mathrm{J}, \mathrm{I}).$$

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 $W(\mathrm{I}) < W(\mathrm{J})$, owing to $E(\mathrm{I}, \mathrm{I}) < E(\mathrm{J}, \mathrm{I})$, the J-type mutants will invade successfully, and an all-I population cannot be evolutionary stable.

However, it is possible that $E(\mathrm{I}, \mathrm{I}) = E(\mathrm{J}, \mathrm{I})$, as indeed happens if the population initially consists of a single phenotype that plays a strategy of mixing between the pure strategies I and J (a monomorphic equilibrium with a mixed strategy), as was the case in our final variant of the Hawk-Dove game (Section 6.E). Then the difference between $W(\mathrm{I})$ and $W(\mathrm{J})$ is governed by how each type fares against the mutants.[18] When $E(\mathrm{I}, \mathrm{I}) = E(\mathrm{J}, \mathrm{I})$, we get $W(\mathrm{I}) > W(\mathrm{J})$ if $E(\mathrm{I}, \mathrm{J}) > E(\mathrm{J}, \mathrm{J})$. This is the **secondary criterion** for the evolutionary stability of I, to be invoked only if the primary one is inconclusive—that is, only if $E(\mathrm{I}, \mathrm{I}) = E(\mathrm{J}, \mathrm{I})$.

If the secondary condition is invoked—because $E(\mathrm{I}, \mathrm{I}) = E(\mathrm{J}, \mathrm{I})$—there is the additional possibility that it may also be inconclusive. That is, it may also be the case that $E(\mathrm{I}, \mathrm{J}) = E(\mathrm{J}, \mathrm{J})$. This is the case of *neutral stability* introduced in Section 7. If both the primary and secondary conditions for the evolutionary stability of I are inconclusive, then I is considered a neutral ESS.

Note that the primary criterion carries a punch. It says that if the strategy I is evolutionary stable, then for all other strategies J that a mutant might try, $E(\mathrm{I}, \mathrm{I}) \geq E(\mathrm{J}, \mathrm{I})$. 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, everyone playing I would be a Nash equilibrium. *Evolutionary stability thus implies Nash equilibrium of the corresponding rationally played game!*[19]

This is a remarkable result. If you were dissatisfied with the rational behavior assumption underlying the theory of Nash equilibria given in earlier chapters and you came to the theory of evolutionary games looking for a better explanation, you would find that it yields the same results. The very appealing biological description—fixed nonmaximizing 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 stability is static rather than dynamic. It only checks whether the configuration of the population (monomorphic, or polymorphic in just the right proportions) that 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 classification correctly and some type of mutant that she overlooked could actually arise, that mutant might invade successfully and destroy the supposed equilibrium. Our remark at the end of the twice-played prisoners' dilemma in Section 2.A warned of this possibility, and you will see in the exercises how it can arise. Finally, in Section 7 we saw how evolutionary dynamics can fail to converge at all.

[18] If the initial population is polymorphic and m is the proportion of J types, then m may not be "very small" any more. The size of m is no longer crucial, however, because the second term in $W(\mathrm{I}) - W(\mathrm{J})$ is now assumed to be zero.

[19] In fact, the primary criterion is slightly stricter than the standard definition of Nash equilibrium, which conforms more closely to that of neutral stability.

9 PLAYING THE FIELD

We have thus far looked at situations where each game is played between just two players who are randomly chosen from the population. There are other situations, however, when the whole population plays at once. In biology, a whole flock of animals with a mixture of genetically determined behaviors may compete for some resource or territory. In economics or business, many firms in an industry, each following the strategy dictated by its corporate culture, may compete all with all.

Such evolutionary games stand in the same relation to the rationally played collective-action games of Chapter 12 as do the pair-by-pair played evolutionary games of the preceding sections to the rationally played two-person games of Chapters 4 through 8. Just as we converted the expected payoff graphs of those chapters into the fitness diagrams in Figures 13.4, 13.7, and 13.9, we can convert the graphs for collective-action games (Figures 12.6 through 12.8) into fitness graphs for evolutionary games. For example, consider an animal species all of whose members come to a common feeding ground. There are two phenotypes: one fights for food aggressively, and the other hangs around and sneaks what it can. If the proportion of aggressive ones is small, they will do better; but, if there are too many of them, the sneakers will do better by ignoring the ongoing fights. This will be a collective chicken game whose fitness diagram will be exactly like Figure 12.7. Because no new principles or techniques are required, we leave it to you to pursue this idea further.

10 EVOLUTION OF COOPERATION AND ALTRUISM

Evolutionary game theory rests on two fundamental ideas: first, that individual organisms 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 (in which individual animals behave in a way that yields greater benefit to everyone in a group) and even altruism (in which individual animals 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?

Biologists use a fourfold classification of the ways in which cooperation and altruism can emerge among selfish animals (or phenotypes or genes). Lee Dugatkin names the four categories (1) family dynamics, (2) reciprocal transactions, (3) selfish teamwork, and (4) group altruism.[20]

The behavior of ants and bees is probably the easiest to understand as an example of family dynamics. All the individual members of an ant colony or a beehive are closely related and have genes in common to a substantial extent. All worker ants in a colony are full sisters and therefore have half their genes in common; 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 have a quarter of their genes in common. 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 the underlying genes of those groups whose members exhibit such behavior (phenotype) will proliferate. The idea that evolution ultimately operates at the level of the gene has had enormous implications for biology, although it has been misapplied by many people, just as Darwin's original idea of natural selection was misapplied.[21] 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.

Reciprocal altruism can arise among unrelated individual members of the same or different species. This behavior is essentially an example of the resolution of prisoners' dilemmas through repetition in which the players use strategies that are remarkably like tit-for-tat. For example, 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." A more fascinating, although gruesome, example is that of vampire bats, who share blood with those who have been unsuccessful in their own hunting. In an experiment in which bats from different sites were brought together and selectively starved, "only bats that were on the verge of starving (i.e., would die within twenty-four hours without a meal) were given blood by any other bat in the experiment. But, more to the point, individuals were given a blood meal only from bats they already knew from their site. . . . Furthermore, vampires were much more likely to regurgitate blood to the specific individual(s)

[20]See his excellent exposition in *Cheating Monkeys and Citizen Bees: The Nature of Cooperation in Animals and Humans* (Cambridge: Harvard University Press, 1999).

[21]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 cited in this section, is Matt Ridley, *The Origins of Virtue* (New York: Penguin, 1996). We should also point out that we do not examine the connection between genotypes and phenotypes in any detail or the role of sex in evolution. Another book by Ridley, *The Red Queen* (New York: Penguin, 1995), gives a fascinating treatment of this subject.

from their site that had come to their aid when they needed a bit of blood."[22] Once again, it is not to be supposed that each animal consciously calculates whether it is in its individual interest to continue the cooperation or to defect. Instead, the behavior is instinctive.

Selfish teamwork arises when it is in the interests of each individual organism to choose cooperation when all others are doing so. In other words, this type of cooperative behavior applies to the selection of the good outcome in assurance games. Dugatkin argues that populations are more likely to engage in selfish teamwork in harsh environments than in mild ones. When conditions are bad, the shirking of any one animal in a group could bring disaster to the whole group, including the shirker. Then in such conditions each animal is crucial for survival, and none shirk so long as others are also pulling their weight. In milder environments, each may hope to become a free rider on the others' effort without thereby threatening the survival of the whole group, including itself.

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 brings us to the idea of group altruism, which suggests that we should see some cooperation even among individual members of a group who are not close relatives. We do indeed find instances of it. Groups of predators such as wolves are a case in point, and groups of apes often behave like extended families. Even among species of prey, cooperation arises when individual fishes in a school take turns looking out for predators. And cooperation can also extend across species.

The general idea is that a group whose members behave cooperatively is more likely to succeed in its interactions with other groups than one whose members seek benefit of free-riding within the group. If, in a particular context of evolutionary dynamics, between-group selection is a stronger force than within-group selection, then we will see group altruism.[23]

An instinct is hardwired into an individual organism's brain by genetics, but reciprocity and cooperation can arise from more purposive thinking or experimentation within the group and can spread by socialization—through explicit instruction or observation of the behavior of elders—instead of genetics. The relative importance of the two channels—nature and nurture—will differ from one species to another and from one situation to another. One would expect socialization to be relatively more important among humans, but there are instances of its role among other animals. We cite a remarkable one. The expedition that Robert F. Scott led to the South Pole in 1911–1912 used teams of Siberian dogs. This group of dogs, brought together and trained for this specific purpose, developed

[22]Dugatkin, *Cheating Monkeys*, p. 99.

[23]Group altruism used to be thought impossible according to the strict theory of evolution that emphasizes selection at the level of the gene, but the concept is being revived in more sophisticated formulations. See Dugatkin, *Cheating Monkeys*, pp. 141–145 for a fuller discussion.

within a few months a remarkable system of cooperation and sustained it by 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."[24]

This is an encouraging account of how cooperative behavior can be compatible with evolutionary game theory and one that suggests that dilemmas of selfish actions can be overcome. Indeed, scientists investigating altruistic behavior have recently reported experimental support for the existence of such *altruistic punishment,* or *strong reciprocity* (as distinguished from reciprocal altruism), in humans. Their evidence suggests that people are willing to punish those who don't pull their own weight in a group setting, even when it is costly to do so and when there is no expectation of future gain. This tendency toward strong reciprocity may even help to explain the rise of human civilization if groups with this trait were better able to survive the traumas of war and other catastrophic events.[25] Despite these findings, strong reciprocity may not be widespread in the animal world. "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."[26] In other words, precisely the conditions that our theoretical analysis in Section 2.D of Chapter 11 identified as being necessary for a successful resolution of the repeated prisoners' dilemma are seen to be relevant in the context of evolutionary games.

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, phenotypes with higher *fitness* survive repeated *interactions* with others to reproduce and to increase their representation in the population. A population containing one or more phenotypes in certain proportions is called *evolutionary stable* if it cannot be *invaded* successfully by other, *mutant* phenotypes or if it is the limiting outcome of the dynamics of proliferation of fitter phenotypes. If one phenotype maintains its dominance in the population

[24]Apsley Cherry-Garrard, *The Worst Journey in the World* (London: Constable, 1922; reprinted New York: Carroll and Graf, 1989), pp. 485–486.

[25]For the evidence on altruistic punishment, see Ernst Fehr and Simon Gachter, "Altruistic Punishment in Humans," *Nature,* vol. 415 (January 10, 2002), pp. 137–140.

[26]Ridley, *Origins of Virtue,* p. 83.

when faced with an invading mutant type, that phenotype is said to be an *evolutionary stable strategy,* and the population consisting of it alone is said to exhibit *monomorphism.* If two or more phenotypes coexist in an evolutionary stable population, it is said to exhibit *polymorphism.*

When the theory of evolutionary games is applied more generally to nonbiological games, the strategies followed by individual players 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.

Evolutionary games may have payoff structures similar to those analyzed in Chapters 4 and 7, 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, and so there is a polymorphic equilibrium; in the assurance game, types are less fit when rare, and so the polymorphic configuration is unstable and the equilibria are at the extremes. When play is between two different types of members of each of two different species, a more complex but similarly structured analysis is used to determine equilibria.

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 performed 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

evolutionary stable (495)
evolutionary stable strategy (ESS) (498)
fitness (495)
genotype (495)
hawk-dove game (516)
interaction (496)
invasion by a mutant (495)
monomorphism (498)
mutation (495)
neutral ESS (522)
phenotype (495)
playing the field (496)
polymorphism (498)
primary criterion (526)
secondary criterion (526)
selection (495)

SOLVED EXERCISES

S1. Two travelers buy identical handcrafted souvenirs and pack them in their respective suitcases for their return flight. Unfortunately, the airline manages to lose both suitcases. Because the airline doesn't know the value of the lost souvenirs, it asks each traveler to report independently a value. The airline agrees to pay each traveler an amount equal to the minimum of the two reports. If one report is higher than the other, the airline takes a penalty of $20 away from the traveler with the higher report and gives $20 to the traveler with the lower report. If the reports are equal to one another, there is no reward or penalty. Neither traveler remembers exactly how much the souvenir cost, so that value is irrelevant; each traveler simply reports the value that her type determines she should report.

There are two types of travelers. The High type always reports $100, and the Low type always reports $50. Let h represent the proportion of High types in the population.

(a) Draw the payoff table for the game played between two travelers selected at random from the population.
(b) Graph the fitness of the High type, with h on the horizontal axis. On the same figure, graph the fitness of the Low type.
(c) Describe all of the equilibria of this game. For each equilibrium, state whether it is monomorphic or polymorphic and whether it is stable.

S2. In Section 7.A, we considered testing for ESSs (evolutionary stable strategies) in the thrice-repeated restaurant-pricing prisoners' dilemma.
(a) Explain completely (using Figure 13.13) why an all-type-A population cannot be invaded by either N- or T-type mutants.
(b) Explain why an all-N-type population can be invaded by type A mutants, and to what extent it can be invaded by type T mutants. Relate this explanation to the concept of neutral stability in the chapter.
(c) Finally, explain why an all-T-type population cannot be invaded by type A mutants but can be invaded by mutants that are type N.

S3. Consider a population in which there are two phenotypes: natural-born cooperators (who do not confess under questioning) and natural-born defectors (who confess readily). If two members of this population are drawn at random, their payoffs in a single play are the same as those of the husband-wife prisoners' dilemma game of Chapter 4, reproduced below. In repeated interactions there are two strategies available in the population, as there were in the restaurant-dilemma game of Section 13.2. The two strategies are A (always confess) and T (play tit-for-tat, starting with not confessing).

		COLUMN	
		Confess	Not
ROW	Confess	10 yr, 10 yr	1 yr, 25 yr
	Not	25 yr, 1 yr	3 yr, 3 yr

(a) Suppose that a pair of players plays this dilemma twice in succession. Draw the payoff table for the twice-repeated dilemma.
(b) Find all of the ESSs in this game.
(c) Now add a third possible strategy, N, which never confesses. Draw the payoff table for the twice-repeated dilemma with three possible strategies and find all of the ESSs of this new version of the game.

S4. In the assurance (meeting-place) game in this chapter, the payoffs were meant to describe the value of something material that the players gained in the various outcomes; they could be prizes given for a successful meeting, for example. Then other individual persons 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 we can make a more biological interpretation. Suppose the column players are always female and the row players always male. When two of these players meet successfully, they pair off, and their children are of the same type as the parents. Therefore the types would proliferate or die off as a result of successful or unsuccessful meetings. The formal mathematics of this new version of the game makes it a "two-species game" (although the biology of it does not). Thus, the proportion of S-type females in the population—call this proportion x—need not equal the proportion of S-type males—call this proportion y.
(a) Examine the dynamics of x and y by using methods similar to those used in the chapter for the battle-of-the-sexes game.
(b) Find the stable outcome or outcomes of this dynamic process.

S5. Recall from Exercise S1 the travelers reporting the value of their lost souvenirs. Assume that a third traveler phenotype exists in the population. The third traveler type is a mixer; she plays a mixed strategy, sometimes reporting a value of \$100 and sometimes reporting a value of \$50.
(a) Use your knowledge of mixed strategies in rationally played games to posit a reasonable mixture for the mixer phenotype to use in this game.
(b) Draw the three-by-three payoff table for this game when the mixer type uses the mixed strategy that you found in part (a).
(c) Determine whether the mixer phenotype is an ESS of this game. (Hint: Test whether a mixer population can be invaded successfully by either the High type or the Low type.)

S6. Consider a simplified model in which everyone gets electricity either from solar power or from fossil fuels, which are both in relatively inelastic supply. (In the case of solar power, think of the required equipment as being in inelastic supply.) The upfront costs of using solar energy are high, so when the price of fossil fuels is low (that is, when few people are using fossil fuels and there is a high demand for solar equipment), the cost of solar can be prohibitive. On the other hand, when many individuals are using fossil fuels, the demand for them (and thus the price) is high, whereas the demand (and thus the price) for solar energy is relatively lower. Assume the payoff table for the two types of energy consumers to be as follows:

		COLUMN	
		Solar	Fossil fuels
ROW	Solar	2, 2	3, 4
	Fossil fuels	4, 3	2, 2

(a) Describe all possible ESS of this game in terms of s, the proportion of solar users, and explain why each is either stable or unstable.
(b) Suppose there are important economies of scale in producing solar equipment, such that the cost savings increase the payoffs in the (solar, solar) cell of the table to (y, y) where $y > 2$. How large would y need to be for the polymorphic equilibrium to have $s = 0.75$?

S7. There are two types of racers—tortoises and hares—who race against one another in randomly drawn pairs. In this world, hares beat tortoises every time without fail. If two hares race they tie, and they are completely exhausted by the race. When two tortoises race they also tie, but they enjoy a pleasant conversation along the way. The payoff table is as follows (where $c > 0$):

		COLUMN	
		Tortoise	Hare
ROW	Tortoise	c, c	$-1, 1$
	Hare	$1, -1$	0, 0

(a) Assume that the proportion of tortoises in the population, t, is 0.5. For what values of c will tortoises have greater fitness than hares?
(b) For what values of c will tortoises be fitter than hares if $t = 0.1$?
(c) If $c = 1$, will a single hare successfully invade a population of pure tortoises? Explain why or why not.
(d) In terms of t, how large must c be for tortoises to have greater fitness than hares?

(e) In terms of c, what is the level of t in a polymorphic equilibrium? For what values of c will such an equilibrium exist? Explain.

S8. Consider a population with two types, X and Y, with a payoff table as follows:

		COLUMN	
		X	Y
ROW	X	2, 2	5, 3
	Y	3, 5	1, 1

(a) Find the fitness for X as a function of x, the proportion of X in the population, and the fitness for Y as a function of x.

Assume that the population dynamics from generation to generation conform to the following model:

$$x_{t+1} = x_t \times F_{Xt} / [x_t \times F_{Xt} + (1 - x_t) \times F_{Yt}],$$

where x_t is the proportion of X in the population in period t, x_{t+1} is the proportion of X in the population in period $t+1$, F_{Xt} is the fitness of X in period t, and F_{Yt} is the fitness of Y in period t.

(b) Assume that x_0, the proportion of X in the population in period 0, is 0.2. What are F_{X0} and F_{Y0}?

(c) Find x_1, using x_0, F_{X0}, F_{Y0}, and the model given above.

(d) What are F_{X1} and F_{Y1}?

(e) Find x_2 (rounded to five decimal places).

(f) What are F_{X2} and F_{Y2} (rounded to five decimal places)?

S9. Consider an evolutionary game between green types and purple types with a payoff table as follows:

		COLUMN	
		Green	Purple
ROW	Green	a, a	4, 3
	Purple	3, 4	2, 2

Let g be the proportion of greens in the population.

(a) In terms of g, what is the fitness of the purple type?

(b) In terms of g and a, what is the fitness of the green type?

(c) Graph the fitness of the purple types against the fraction g of green types in the population. On the same diagram, show three lines for the fitness of the green types when $a = 2$, 3, and 4. What can you conclude from this graph about the range of values of a that guarantees a stable polymorphic equilibrium?

(d) Assume that a is in the range found in part (c). In terms of a, what is the proportion of greens, g, in the stable polymorphic equilibrium?

S10. Prove the following statement: "If a strategy is strictly 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 of all types."

UNSOLVED EXERCISES

U1. Consider a survival game in which a large population of animals meet and either fight over or share a food source. There are two phenotypes in the population: one always fights, and the other always shares. For the purposes of this question, assume that no other mutant types can arise in the population. Suppose that the value of the food source is 200 calories and that caloric intake determines each player's reproductive fitness. If two sharing types meet one another, they each get half the food, but if a sharer meets a fighter, the sharer concedes immediately, and the fighter gets all the food.

(a) Suppose that the cost of a fight is 50 calories (for each fighter) and that when two fighters meet, each is equally likely to win the fight and the food or to lose and get no food. Draw the payoff table for the game played between two random players from this population. Find all of the ESSs in the population. What type of game is being played in this case?

(b) Now suppose that the cost of a fight is 150 calories for each fighter. Draw the new payoff table and find all of the ESSs for the population in this case. What type of game is being played here?

(c) Using the notation of the Hawk-Dove game of Section 13.6, indicate the values of V and C in parts (a) and (b) and confirm that your answers to those parts match the analysis presented in the chapter.

U2. Suppose that a single play of a prisoners' dilemma has the following payoffs:

		PLAYER 2	
		Cooperate	Defect
PLAYER 1	Cooperate	3, 3	1, 4
	Defect	4, 1	2, 2

In a large population in which each member's behavior is genetically determined, each player will be either a defector (that is, always defects in any play of a prisoners' dilemma game) or a tit-for-tat player. (In multiple rounds of a prisoners' dilemma, she cooperates on the first play, and on any subsequent play she does whatever her opponent did on the preceding play.) Pairs of randomly chosen players from this population will play "sets" of n single plays of this dilemma (where $n \geq 2$). The payoff to each player in one whole set (of n plays) is the sum of her payoffs in the n plays.

Let the population proportion of defectors be p and the proportion of tit-for-tat players be $(1 - p)$. Each member of 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.

(a) Show in a two-by-two table the payoffs to a player of each type when, in one set of plays, each player meets an opponent of each of the two types.
(b) Find the fitness (average payoff in 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 in this exercise? (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 more repetition (larger values of n) facilitate the evolution of cooperation?

U3. Suppose that in the twice-repeated prisoners' dilemma of Exercise S3, a fourth possible type (type S) also can exist in the population. This type does not confess on the first play and confesses on the second play of each episode of two successive plays against the same opponent.

(a) Draw the four-by-four fitness table for the game.
(b) Can the newly conceived type S be an ESS of this game?
(c) In the three-types game of Exercise S3, A and T were both ESS, but T was only neutrally stable because a small proportion of N mutants could coexist. Show that in the four-types game here, T cannot be ESS.

U4. Following the pattern of Exercise S4, analyze an evolutionary version of the tennis point game (Figure 4.15). Regard servers and receivers as separate species, and construct a figure like Figure 13.11. What can you say about the ESS and its dynamics?

U5. Recall from Exercise U1 the population of animals fighting over a food source worth 200 calories. Assume that, as in part (b) of that exercise, the cost of a fight is 150 calories per fighter. Assume also that a third phenotype exists in the population. That phenotype is a mixer; it plays a mixed strategy, sometimes fighting and sometimes sharing.
(a) Use your knowledge of mixed strategies in rationally played games to posit a reasonable mixture for the mixer phenotype to use in this game.
(b) Draw the three-by-three payoff table for this game when the mixer type uses the mixed strategy that you found in part (a).
(c) Determine whether the mixer phenotype is an ESS of this game. (Hint: Test whether a mixer population can be invaded successfully by either the fighting type or the sharing type.)

U6. Consider an evolutionary version of the game between Baker and Cutler, from Exercise U1 of Chapter 11. This time Baker and Cutler are not two individuals but two separate species. Each time a Baker meets a Cutler, they play the following game. The Baker chooses the total prize to be either \$10 or \$100. The Cutler chooses how to divide the prize chosen by the Baker: the Cutler can choose either a 50:50 split or a 90:10 split in the Cutler's own favor. The Cutler moves first, and the Baker moves second.

There are two types of Cutlers in the population: type F chooses a fair (50:50) split, whereas type G chooses a greedy (90:10) split. There are also two types of Bakers: type S simply chooses the large prize (\$100) no matter what the Cutler has done, whereas type T chooses the large prize (\$100) if the Cutler chooses a 50:50 split, but the small prize (\$10) if the Cutler chooses a 90:10 split.

Let f be the proportion of type F in the Cutler population, so that $(1 - f)$ represents the proportion of type G. Let s be the proportion of type S in the Baker population, so that $(1 - s)$ represents the proportion of type T.
(a) Find the fitness of the Cutler types F and G in terms of s.
(b) Find the fitness of the Baker types S and T in terms of f.

(c) For what value of s are types F and G equally fit?
(d) For what value of f are types S and T equally fit?
(e) Use the answers above to sketch a graph displaying the population dynamics. Assign f as the horizontal axis and s as the vertical axis.
(f) Describe all of the equilibria of this evolutionary game, and indicate which ones are stable.

U7. Recall Exercise S7. Hares, it turns out, are very impolite winners. Whenever hares race tortoises they mercilessly mock their slow-footed (and easily defeated) rivals. The poor tortoises leave the race not only in defeat, but with their tender feelings crushed by the oblivious hares. The payoff table is thus:

		COLUMN	
		Tortoise	Hare
ROW	Tortoise	c, c	$-2, 1$
	Hare	$1, -2$	$0, 0$

(a) For what values of c are tortoises fitter than hares if t, the proportion of tortoises in the population, is 0.5? How does this compare with the answer in Exercise S7, part (a)?
(b) For what values of c are tortoises fitter than hares if $t = 0.1$? How does this compare with the answer in Exercise S7, part (b)?
(c) If $c = 1$, will a single hare successfully invade a population of pure tortoises? Explain why or why not.
(d) In terms of t, how large must c be for tortoises to be fitter than hares?
(e) In terms of c, what is the level of t in a polymorphic equilibrium? For what values of c will such an equilibrium exist? Explain.
(f) Will the polymorphic equilibria found to exist in part (e) be stable? Why or why not?

U8. **(Use of spreadsheet software recommended)** This problem explores more thoroughly the generation-by-generation population dynamics seen in Exercise S8. Since the math can quickly become very complicated and tedious, it is much easier to do this analysis with the aid of a spreadsheet.

Again, consider a population with two types, X and Y, with a payoff table as follows:

		COLUMN	
		X	Y
ROW	X	2, 2	5, 3
	Y	3, 5	1, 1

Recall that the population dynamics from generation to generation are given by:

$$x_{t+1} = x_t \times F_{Xt} / [x_t \times F_{Xt} + (1 - x_t) \times F_{Yt}],$$

where x_t is the proportion of X in the population in period t, x_{t+1} is the proportion of X in the population in period $t+1$, F_{Xt} is the fitness of X in period t, and F_{Yt} is the fitness of Y in period t.

Use a spreadsheet to extend these calculations to many generations. [Hint: Assign three horizontally adjacent cells to hold the values of x_t, F_{Xt}, and F_{Yt}, and have each successive row represent a different period ($t = 0, 1, 2, 3, \ldots$). Use spreadsheet formulas to relate F_{Xt} and F_{Yt} to x_t and x_{t+1} to x_t, F_{Xt}, and F_{Yt} according to the population model given above.]
(a) If there are initially equal proportions of X and Y in the population in period 1 (that is, if $x_0 = 0.5$), what is the proportion of X in the next generation, x_1? What are F_{X1} and F_{Y1}?
(b) Use a spreadsheet to extend these calculations to the next generation, and the next, and so on. To four decimal places, what is the value of x_{20}? What are F_{X20} and F_{Y20}?
(c) What is x^*, the equilibrium level of x? How many generations does it take for the population to be within 1% of x^*?
(d) Answer the questions in part (b), but with a starting value of $x_0 = 0.1$.
(e) Repeat part (b), but with $x_0 = 1$.
(f) Repeat part (b), but with $x_0 = 0.99$.
(g) Are monomorphic equilibria possible in this model? If so, are they stable? Explain.

U9. Consider an evolutionary game between green types and purple types, with a payoff table as follows:

		COLUMN	
		Green	Purple
ROW	Green	a, a	b, c
	Purple	c, b	d, d

In terms of the parameters a, b, c, and d, find the conditions that will guarantee a stable polymorphic equilibrium.

U10. **(Optional, for mathematically trained students)** In the three-type evolutionary game of Section 7.B and Figure 13.14, let $q_3 = 1 - q_1 - q_2$ denote the proportion of the orange-throated aggressor types. Then the dynamics of the population proportions of each type of lizard can be stated as

$$q_1 \text{ increases if and only if } -q_2 + q_3 > 0$$

and

$$q_2 \text{ increases if and only if } q_1 - q_3 > 0.$$

We did not state this explicitly in the chapter, but a similar rule for q_3 is

$$q_3 \text{ increases if and only if } -q_1 + q_2 > 0.$$

(a) Consider the dynamics more explicitly. Let the speed of change in a variable x in time t be denoted by the derivative dx/dt. Then suppose

$$dq_1/dt = -q_2 + q_3,\ dq_2/dt = q_1 - q_3, \text{ and } dq_3/dt = -q_1 + q_2.$$

Verify that these derivatives conform to the preceding statements regarding the population dynamics.

(b) Define $X = (q_1)^2 + (q_2)^2 + (q_3)^2$. Using the chain rule of differentiation, show that $dX/dt = 0$, that is, show that X remains constant over time.

(c) From the definitions of the entities, we know that $q_1 + q_2 + q_3 = 1$. Combining this fact with the result from part (b), show that over time, in three-dimensional space, the point (q_1, q_2, q_3) moves along a circle.

(d) What does the answer to part (c) indicate regarding the stability of the evolutionary dynamics in the colored-throated lizard population?

14

Mechanism Design

JAMES MIRRLEES WON THE NOBEL PRIZE in Economics in 1996 for his pioneering work on optimal nonlinear income taxation and related policy issues. Many noneconomists, and some economists too, found his work difficult to understand. But *The Economist* magazine gave a brilliant characterization of the broad importance and relevance of the work. It said that Mirrlees showed us "how to deal with someone who knows more than you do."[1]

We have already seen some of the ways in which such asymmetric information affects the analysis of games, in Chapter 9. But the underlying problem for Mirrlees differed slightly from the situations we considered earlier. In his work, one player (the government) needed to devise a set of rules so that the other players' (the taxpayers') incentives were aligned with the first player's goals. Models with this general framework, in which a less-informed player works to create motives for the more-informed player to take actions beneficial to the less informed, now abound and are relevant to a wide range of social and economic interactions. Generally, the less-informed player is called the *principal* while the more-informed is called the *agent;* hence these models are termed *principal-agent* models. And the process that the principal uses to devise the correct set of incentives for the agent is known as **mechanism design.**

In Mirrlees's model, the government seeks a balance between efficiency and equity. It wants the more productive members of society to contribute effort to increase total output; it can then redistribute the proceeds to benefit the poorer

[1] "Economics Focus: Secrets and the Prize," *The Economist,* October 12, 1996.