10

Evolutionarily Stable Strategies

There is but a step between the sublime and the ridiculous.

Leo Tolstoy

In 1973 the biologist John Maynard Smith and the mathematician G. R. Price wrote an article in *Nature* showing how game theory applies to the behavior of animals (Maynard Smith and Price 1973). Maynard Smith went on to write a book on the subject (Maynard Smith 1982), which has become a classic. The idea of applying game theory to animals, and not just the higher primates, but fish, dung beetles, fireflies, and pond scum as well, seemed strange at the time, because game theory had always been the preserve of hyperrationality. Animals hardly fit the bill. Maynard Smith made three critical shifts from traditional game theory. The first is in the concept of a strategy, the second in the concept of equilibrium, and a third in the nature of agent interactions.

Strategy. In classical game theory, *players* have strategy sets from which they choose particular strategies. In biology, *species* have strategy sets (genotypic variants), of which *individuals* inherit one or another variant, perhaps mutated, that they then play in their strategic interactions. This extends nicely to the treatment of culture in human society, in which case we say that *society* has the strategy set (the set of alternative cultural forms) and *individuals* inherit or choose among them.

Equilibrium. In place of the Nash equilibrium, Maynard Smith and Price used the *evolutionarily stable strategy* (ESS) concept. A strategy is an ESS if a whole population using that strategy cannot be invaded by a small group with a mutant genotype. Similarly, a cultural form is an ESS if, upon being adopted by all members of a society (firm, family, etc.), no small group of individuals using an alternative cultural form can invade. We thus move from explaining the actions of individuals to modeling the diffusion of forms of behavior ("strategies") in society.

Player interactions. In place of the one-shot and repeated games of classical game theory, Maynard Smith introduced the notion of the *repeated,* *random pairing of agents* who play strategies based on their genome but not on the previous history of play.

The ESS concept is particularly useful because it says something about the dynamic properties of a system without being committed to any particular dynamic model. As we shall see, however, an evolutionary system with a symmetrical two-player stage game can be dynamically stable without being an ESS (§12.9).

10.1 Evolutionarily Stable Strategies: Definition

Consider a two-player normal form game in which both players have the set $S = \{s_1, \dots, s_n\}$ of pure strategies, and the payoffs to an agent playing $s_i \in S$ and another agent playing $s_j \in S$ are π^1_{ij} for the first and $\pi^2_{ij} = \pi^1_{ji}$ for the second. We call such a game *symmetric in payoffs*. In addition, we assume the agents cannot condition their play on whether they are "player 1" or "player 2." We call such a game *symmetric in strategies*. If a game is symmetric in both payoffs and strategies, we simply call the game *symmetric*. We call $A = (\pi^1_{ij})$ the *matrix* of the symmetric game. Note that A represents only the payoffs for the row player, because the payoffs to the column player are just the transpose of A.

Let $\mathcal{G}$ be a symmetric game with matrix A (we'll call it the *stage game*) and large population of agents. In each period $t = 1, 2, \dots$, agents are randomly paired and they play the stage game $\mathcal{G}$ once. Each agent is of type i for some $s_i \in S$, meaning that the agent uses strategy s_i in the stage game. If the proportion of agents of type j is p_j at a particular time, we say the *state* of the population is $\sigma = p_1 s_1 + \ldots + p_n s_n$. Note that we must have $p_1, \dots, p_n \geq 0$ and $\sum_i p_i = 1$. The payoff at that time to a player of type i when the state of the population is σ is defined by

$$\pi_{i\sigma} = \sum_{j=1}^{n} \pi_{ij} p_j, \tag{10.1}$$

which is the player's expected payoff before being assigned a particular partner. These conditions define a new game, called the *evolutionary game* corresponding to the stage game $\mathcal{G}$.

Suppose the state of the population is σ, and some small subpopulation plays a "mutant" strategy $\tau = q_1 s_1 + \ldots + q_n s_n$, in the sense that q_i is the frequency of pure strategy s_i in this subpopulation. We say the mutant is of "type τ," and has payoff

$$\pi_{\tau\sigma} = \sum_{i,j=1}^{n} q_i \pi_{ij} p_j,$$

when a random member of its population meets a random member of the population σ.

Suppose the state of the population is $\sigma = p_1 s_1 + \ldots + p_n s_n$. The expected payoff to a randomly chosen member of the population is thus just $\pi_{\sigma\sigma}$. If we replace a fraction $\epsilon > 0$ of the population with a "mutant" of type τ, the new state of the population is

$$\mu = (1-\epsilon)\sigma + \epsilon\tau,$$

so the payoff to a randomly chosen nonmutant is

$$\pi_{\sigma\mu} = (1-\epsilon)\pi_{\sigma\sigma} + \epsilon\pi_{\sigma\tau},$$

and the expected payoff to a mutant is

$$\pi_{\tau\mu} = (1-\epsilon)\pi_{\tau\sigma} + \epsilon\pi_{\tau\tau}.$$

We say the mutant type can *invade* the population if $\sigma \neq \mu$ and for all sufficiently small $\epsilon > 0$,

$$\pi_{\tau\mu} \geq \pi_{\sigma\mu},$$

which says that, on average, a mutant does at least as well against the new population as does a nonmutant. We say σ is an *evolutionarily stable strategy* (ESS) if it cannot be invaded by any mutant type, in a sense defined precisely below.

We assume that mutants can employ mixed strategies in applying the ESS criterion, because as we shall see later (§12.7), with this assumption evolutionarily stable strategies have powerful dynamic properties. A Nash equilibrium in an evolutionary game can consist of a *monomorphic* population of agents, each playing the same mixed strategy, or a *polymorphic* population, a fraction of the population playing each of the underlying pure strategies in proportion to its contribution to the mixed Nash strategy. The two interpretations are interchangeable under many conditions, and we shall not commit ourselves exclusively to either interpretation. Because the stage game is a one-shot, it is rarely plausible to hold that an individual will play a strictly mixed strategy. Thus, in general, the heterogeneous population interpretation is superior. The heterogeneous mutant τ must then possess some internal mechanism for maintaining the constant frequency distribution $q_1, \ldots, q_n$ from period to period. We relax this assumption when we treat evolutionary games as dynamical systems in chapter 12.

10.2 Properties of Evolutionarily Stable Strategies

Prove the following properties of evolutionarily stable strategies:

a. Strategy $\sigma \in \Delta S$ is an ESS if, for every mutant type $\tau \in \Delta S$, there is an $\epsilon_\tau > 0$ such that for all $\epsilon \in (0, \epsilon_\tau)$ and defining $\mu = (1-\epsilon)\sigma + \epsilon\tau$, we have

$$\pi_{\sigma\mu} > \pi_{\tau\mu}. \tag{10.2}$$

b. We say that $\sigma \in \Delta S$ has a *uniform invasion barrier* if there is some $\epsilon_o \in (0,1)$ such that (10.2) holds for all $\tau \neq \sigma$ and all $\epsilon \in (0, \epsilon_o)$. Strategy σ is an ESS if and only if it has a uniform invasion barrier.

c. Strategy $\sigma \in \Delta S$ is an ESS if and only if, for any mutant type $\tau \in \Delta S$, we have

$$\pi_{\sigma\sigma} \geq \pi_{\tau\sigma},$$

and if $\pi_{\sigma\sigma} = \pi_{\tau\sigma}$, then

$$\pi_{\sigma\tau} > \pi_{\tau\tau}.$$

This says that $\sigma \in \Delta S$ is an ESS if and only if a mutant cannot do better against an incumbent than an incumbent can do against another incumbent, and if a mutant does as well as an incumbent against another incumbent, then an incumbent must do better against a mutant than a mutant does against another mutant. Note here that we are assuming mutants can use mixed strategies.

d. An evolutionarily stable strategy is a Nash equilibrium that is isolated in the set of symmetric Nash equilibria (that is, it is a strictly positive distance from any other symmetric Nash equilibrium).

e. Every strict Nash equilibrium in an evolutionary game is an ESS.

10.3 Characterizing Evolutionarily Stable Strategies

THEOREM 10.1 *Suppose symmetric two-player game* $\mathcal{G}$ *has two pure strategies. Then, if* $\pi_{11} \neq \pi_{21}$ *and* $\pi_{12} \neq \pi_{22}$*,* $\mathcal{G}$ *has an evolutionarily stable strategy.*

PROOF: Suppose $\pi_{11} > \pi_{21}$. Then, pure strategy 1 is a strict Nash equilibrium, so it is an evolutionarily stable strategy. The same is true if $\pi_{22} > \pi_{12}$. So suppose $\pi_{11} < \pi_{21}$ and $\pi_{22} < \pi_{12}$. Then, we can show that the game has a unique completely mixed symmetric equilibrium p, where each player uses strategy 1 with probability $\alpha_p \in (0,1)$. The payoff to strategy 1 against the mixed strategy $(\alpha_p, 1-\alpha_p)$ is then $\alpha_p\pi_{11} + (1-\alpha_p)\pi_{12}$, and the payoff to strategy 2 against this mixed strategy is $\alpha_p\pi_{21} + (1-\alpha_p)\pi_{22}$. Because these must be equal, we find that $\alpha_p = (\pi_{22} - \pi_{12})/\Delta$, where $\Delta = \pi_{11} - \pi_{21} + \pi_{22} - \pi_{12} < 0$. Note that under our assumptions, $0 < \alpha_p < 1$, so there is a unique completely mixed Nash equilibrium $(\alpha_p, 1-\alpha_p)$.

Now let α_q be the probability a mutant player uses pure strategy 1. Because each pure strategy is a best response to α_p, α_q must also be a best response to α_p, so clearly, $\pi_{qp} = \pi_{pp}$. To show that p is an ESS, we must show that $\pi_{pq} > \pi_{qq}$. We have

$$\pi_{pq} = \alpha_p[\pi_{11}\alpha_q + \pi_{12}(1-\alpha_q)] + (1-\alpha_p)[\pi_{21}\alpha_q + \pi_{22}(1-\alpha_q)]$$

and

$$\pi_{qq} = \alpha_q[\pi_{11}\alpha_q + \pi_{12}(1-\alpha_q)] + (1-\alpha_q)[\pi_{21}\alpha_q + \pi_{22}(1-\alpha_q)].$$

Subtracting and simplifying, we get

$$\pi_{pq} - \pi_{qq} = -(\alpha_p - \alpha_q)^2\Delta > 0,$$

which proves we have an ESS. ■

THEOREM 10.2 *Using the same notation, the stage game has a strictly mixed Nash equilibrium if and only if* $\pi_{11} > \pi_{21}$ *and* $\pi_{22} > \pi_{12}$*, or* $\pi_{11} < \pi_{21}$ *and* $\pi_{22} < \pi_{12}$*. The equilibrium is an ESS only if the second set of inequalities holds.*

PROOF: It is easy to check that if there is a mixed strategy equilibrium, the frequency α of pure strategy 1 must satisfy

$$\alpha = \frac{\pi_{22} - \pi_{12}}{\Delta}, \qquad \text{where } \Delta = \pi_{11} - \pi_{21} + \pi_{22} - \pi_{12}.$$

Suppose $\Delta > 0$. Then $0 < \alpha < 1$ if and only if $0 < \pi_{22} - \pi_{12} < \pi_{11} - \pi_{21} + \pi_{22} - \pi_{12}$, which is true if and only if $\pi_{11} > \pi_{21}$ and $\pi_{22} > \pi_{12}$. If $\Delta < 0$, a similar argument shows that $0 < \alpha < 1$ if and only if the other pair of inequalities holds.

Suppose there is a "mutant" that uses pure strategy 1 with probability β. Thus, in general,

$$\begin{aligned}\pi_{\gamma\delta} &= \gamma\delta\pi_{11} + \gamma(1-\delta)\pi_{12} + (1-\gamma)\delta\pi_{21} + (1-\gamma)(1-\delta)\pi_{22} \\ &= \gamma\delta\Delta + \delta(\pi_{21} - \pi_{22}) + \gamma(\pi_{12} - \pi_{22}) + \pi_{22}.\end{aligned}$$

It follows that

$$\pi_{\alpha\alpha} - \pi_{\beta\alpha} = (\alpha - \beta)[\alpha\Delta - (\pi_{22} - a_{12})] = 0,$$

so the equilibrium is an ESS if and only if $\pi_{\alpha\beta} > \pi_{\beta\beta}$. But

$$\begin{aligned}\pi_{\alpha\beta} - \pi_{\beta\beta} &= \alpha\beta\Delta + \beta(a_{21} - a_{22}) + \alpha(a_{12} - a_{22}) + a_{22} \\ &\quad - \beta^2\Delta - \beta(a_{21} - a_{22}) - \beta(a_{12} - a_{22}) - a_{22} \\ &= \beta(\alpha - \beta)\Delta + (\alpha - \beta)(a_{12} - a_{22}) \\ &= (\alpha - \beta)(\beta\Delta + a_{12} - a_{22}) \\ &= (\alpha - \beta)(\beta\Delta - \alpha\Delta) \\ &= -(\alpha - \beta)^2\Delta.\end{aligned}$$

Thus, the equilibrium is an ESS if and only if $\Delta < 0$, which is equivalent to $a_{11} < a_{21}$ and $a_{22} < a_{12}$. This proves the assertion. ■

THEOREM 10.3 *Suppose* $\sigma = \alpha_1 s_1 + \ldots + \alpha_n s_n \in \Delta S$ *is an ESS, where* s_i *is a pure strategy and* $\alpha_i > 0$ *for* $i = 1, \ldots, n$. *Suppose* $\tau = \beta_1 s_1 + \ldots + \beta_n s_n \in \Delta S$ *is also an ESS. Then,* $\beta_i = \alpha_i$ *for* $i = 1, \ldots n$. *In other words, the support of an ESS cannot strictly contain the support of another ESS.*

THEOREM 10.4 *If* $\sigma \in \Delta S$ *is weakly dominated, then* σ *is not an ESS.*

THEOREM 10.5 *An evolutionary game whose stage game has a finite number of pure strategies can have only a finite number of evolutionarily stable strategies.*

PROOF: Suppose there are an infinite number of distinct evolutionarily stable strategies. Then there must be two, say σ and τ, that use exactly the same pure strategies. Now τ is a best response to σ, so σ must do better against τ than τ does against itself. But σ does equally well against τ as τ does against τ. Thus, σ is not an ESS and similarly for τ. ■

By the *distance* between two strategies $\sigma = \sum_i p_i s_i$ and $\tau = \sum_i q_i s_i$, we mean the distance in $\mathbf{R}^n$ between the points $(p_1, \ldots, p_n)$ and $(q_1, \ldots, q_n)$. The following is proved in Hofbauer and Sigmund (1998). Note that the theorem implies that an evolutionarily stable strategy σ is an *isolated Nash equilibrium*, in the sense that there is an $\epsilon > 0$ such that no strategy $\tau \neq \sigma$ within distance ϵ of σ is a Nash equilibrium.

THEOREM 10.6 *Strategy $\sigma \in \Delta S$ is an ESS if and only if there is some $\epsilon > 0$ such that $\pi_{\sigma\tau} > \pi_{\tau\tau}$ for all $\tau \in \Delta S$ within distance ϵ of σ.*

PROOF: Suppose σ is an ESS, so for any $\tau \neq \sigma$, there is an $\tilde{\epsilon}(\tau)$ such that

$$\pi_{\tau,(1-\epsilon)\sigma+\epsilon\tau} < \pi_{\sigma,(1-\epsilon)\sigma+\epsilon\tau} \quad \text{for all } \epsilon \in (0, \tilde{\epsilon}(\tau)). \tag{10.3}$$

In fact, we can choose $\tilde{\epsilon}(\tau)$ as follows. If (10.3) holds for all $\epsilon \in (0,1)$, then let $\tilde{\epsilon}(\tau) = 1$. Otherwise, let $\tilde{\epsilon}$ be the smallest $\epsilon > 0$ such that (10.3) is violated and define

$$\tilde{\epsilon}(\tau) = \frac{\pi_{\sigma\sigma} - \pi_{\tau\sigma}}{\pi_{\tau\tau} - \pi_{\tau\sigma} - \pi_{\sigma\tau} + \pi_{\sigma\sigma}}.$$

It is easy to check that $\tilde{\epsilon}(\tau) \in (0,1]$ and (10.3) are satisfied. Let $T \subset S$ be the set of strategies such that if $\tau \in T$, then there is at least one pure strategy used in σ that is not used in τ. Clearly, T is closed and bounded, $\sigma \notin T$, $\tilde{\epsilon}(\tau)$ is continuous, and $\tilde{\epsilon}(\tau) > 0$ for all $\tau \in T$. Hence, $\tilde{\epsilon}(\tau)$ has a strictly positive minimum ϵ^* such that (10.3) holds for all $\tau \in T$ and all $\epsilon \in (0, \epsilon^*)$.

If τ is a mixed strategy and s is a pure strategy, we define $s(\tau)$ to be the weight of s in τ (that is, the probability that s will be played using τ). Now consider the neighborhood of s consisting of all strategies τ such that $|1-s(\tau)| < \epsilon^*$ for all pure strategies s. If $\tau \neq s$, then $\epsilon^* > 1-s(\tau) = \epsilon > 0$

for some pure strategy s. Then $\tau = (1-\epsilon)s + \epsilon r$, where $r \in T$. But then (10.3) gives $\pi_{r\tau} < \pi_{s\tau}$. If we multiply both sides of this inequality by ϵ and add $(1-\epsilon)\pi_{s\tau}$ to both sides, we get $\pi_{\tau\tau} < \pi_{s\tau}$, as required. The other direction is similar, which proves the assertion. ■

THEOREM 10.7 *If $\sigma \in \Delta S$ is a completely mixed evolutionarily stable strategy (that is, it uses all pure strategies with positive probability), then it is the unique Nash equilibrium of the game and $\pi_{\sigma\tau} > \pi_{\tau\tau}$ for all $\tau \in \Delta S$, $\tau \neq \sigma$.*

PROOF: If σ is completely mixed, then for any $tau \in S$, $\pi_{\sigma\sigma} = \pi_{\tau\sigma}$, because any pure strategy has the same payoff against σ as σ does against σ. Therefore, any mixed strategy has the same payoff against σ as σ has against σ. For similar reasons, $\pi_{\sigma\tau} = \pi_{\sigma\sigma}$. Thus, σ is an ESS and if τ is any other strategy, we must have $\pi_{\sigma\tau} > \pi_{\tau\tau}$. ■

10.4 A Symmetric Coordination Game

Consider a two-player pure coordination game in which both players win $a > 0$ if they both choose Up, and they win $b > 0$ if they both choose Down, but they get nothing otherwise. Show that this game has a mixed-strategy equilibrium with a lower payoff than either of the pure-strategy equilibria. Show that this game is symmetric, and the mixed-strategy equilibrium is not an ESS. Show that there are, however, two ESSs. This example shows that sometimes adding the ESS requirement eliminates implausible and inefficient equilibria.

10.5 A Dynamic Battle of the Sexes

The battle of the sexes (§3.9) is not symmetric, and hence the concept of an evolutionarily stable strategy does not apply. However, there is an obvious way to recast battle of the sexes so that it becomes symmetric. Suppose when two players meet, one is randomly assigned to be player 1, and the other player 2. A pure strategy for a player can be written as "xy," which means "if I am Alfredo, I play x, and if I am Violetta, I play y." Here x stands for Opera and y stands for Gambling. There are thus four pure strategies, OO, OG, GO, GG. This game is symmetric, and the normal form matrix (only the payoff to player 1 is shown) is

	OO	OG	GO	GG
OO	3/2,3/2	1,1/2	1/2,1	0,0
OG	1/2,1	0,0	1,2	1/2,1/2
GO	1,1/2	2,1	0,0	1,1/2
GG	0,0	1,1/2	1/2,1	3/2,3/2

Let $\alpha \geq 0$, $\beta \geq 0$, $\gamma \geq 0$ and $\delta = 1 - \alpha - \beta - \gamma \geq 0$ be the fraction of players who use strategy OO, OG, GO, and GG, respectively (or, equivalently, let $(\alpha, \beta, \gamma, \delta)$ be the mixed strategy of each player). Show that there are two pure-strategy Nash equilibria, OO and GG, and for each $\alpha \in [0, 1]$, there is a mixed-strategy Nash equilibrium $\alpha\text{OO} + (1/3 - \alpha)\text{OG} + (2/3 - \alpha)\text{GO} + \alpha\text{GG}$. Show that the payoffs to these equilibria are 3/2 for the pure-strategy equilibria and 2/3 for each of the mixed-strategy equilibria. It is easy to show that the first two equilibria are ESSs, and the others are not–they can be invaded by either OO or GG.

10.6 Symmetrical Throwing Fingers

Similarly, although throwing fingers (§3.8) is not a symmetric game, and hence the concept of an evolutionarily stable strategy does not apply, there is an obvious way to recast throwing fingers so that it becomes symmetric. Suppose when two players meet, one is randomly assigned to be player 1, and the other player 2. A pure strategy for a player can be written as "xy," which means "if I am player 1, I show x fingers, and if I am player 2, I show y fingers." There are thus four pure strategies, 11, 12, 21, and 22. Show that this game is symmetric, and derive the normal form matrix (only the payoff to player 1 is shown)

	11	12	21	22
11	0,0	−1,1	1,−1	0,0
12	1,−1	0,0	0,0	−1,1
21	−1,1	0,0	0,0	1,−1
22	0,0	1,−1	−1,1	0,0

Let $\alpha \geq 0$, $\beta \geq 0$, $\gamma \geq 0$ and $\delta = 1 - \alpha - \beta - \gamma \geq 0$ be the fraction of players who use strategy 11, 12, 21, and 22, respectively (or, equivalently, let $(\alpha, \beta, \gamma, \delta)$ be the mixed strategy of each player). Show that a Nash equilibrium is characterized by $\alpha = 1/2-\gamma$, $\beta = \gamma$ (which implies $\delta = \alpha$). It is easy to show that any such Nash equilibrium can be invaded by any distinct strategy $(\alpha', \beta', \gamma', \delta')$ with $\alpha' = 1/2 - \gamma'$, $\beta' = \gamma'$, so there is no evolutionarily stable strategy for throwing fingers.

10.7 Hawks, Doves, and Bourgeois

THEOREM 10.8 *The mixed-strategy equilibrium in the hawk-dove game (§3.10) is an ESS.*

PROOF: The payoff to H is $\alpha(v - w)/2 + (1 - \alpha)v = v - \alpha(v + w)/2$, and the payoff to D is $(1 - \alpha)(v/2) = v/2 - \alpha(v/2)$. These are equated when $\alpha = v/w$, which is < 1 if $w > v$. To show that this mixed-strategy equilibrium is an ESS, note that $\pi_{11} = (v-w)/2$, $\pi_{21} = 0$, $\pi_{22} = v/2$, and $\pi_{12} = v$. Thus $\pi_{11} = (v - w)/2 < 0 = \pi_{21}$ and $\pi_{22} = v/2 < v = \pi_{12}$, so the equilibrium is an ESS. ■

Note that in the hawk-dove-bourgeois game (§6.41), the bourgeois strategy is a strict Nash equilibrium, and hence is an ESS.

10.8 Trust in Networks II

We now show that the completely mixed Nash equilibrium found in trust in networks (§6.23) is not an ESS and can be invaded by trusters. In case you think this means this equilibrium is dynamically unstable, think again! See Trust in Networks III (§12.10).

For specificity, we take $p = 0.8$. You can check that the equilibrium has inspect share $\alpha^* \approx 0.71$ trust share $\beta^* \approx 0.19$, and defect share $\gamma^* \approx 0.10$. The payoff to the equilibrium strategy s is $\pi_{ss} \approx 0.57$. The payoff to trust against the equilibrium strategy is of course $\pi_{ts} = \pi_{ss} \approx 0.57$, but the payoff to trust against itself is $\pi_{tt} = 1$, so trust can invade.

10.9 Cooperative Fishing

In a certain fishing village, two fisherman gain from having the nets put out in the evening. However, the fishermen benefit equally whether or not they share the costs of putting out the nets. Suppose the expected catch is ν, the cost of putting out the nets to each is c_1 if each fisherman does it alone, and the cost to each is $c_2 < c_1$ if they do it together. We assume $\nu/2 > c_1$, so it is worthwhile for a fisherman to put out the nets even if he has to do it alone. But because $c_2 < c_1$, he prefers help. On the other hand, by free-riding on the first fisherman's effort (that is, by not helping), the other fisherman gets $\nu/2$ anyway.

	Put Out	Free Ride
Put Out	$\frac{\nu}{2} - c_2, \frac{\nu}{2} - c_2$	$\frac{\nu}{2} - c_1, \frac{\nu}{2}$
Free Ride	$\frac{\nu}{2}, \frac{\nu}{2} - c_1$	0,0

Figure 10.1. Cooperative fishing

Figure 10.1 shows the normal form game, where each fisherman has the available strategies put out (P) and free ride (F)? We can show there is a unique mixed-strategy equilibrium and that this strategy is an ESS. It is easy to see there are no pure-strategy symmetric equilibria, because $\nu/2 > c_1$. There are two pure-strategy asymmetric equilibria, FP and PF. Consider a mixed-strategy equilibrium where a fraction α of the population plays P. The payoff to P is then

$$\alpha\left(\frac{\nu}{2} - c_2\right) + (1-\alpha)\left(\frac{\nu}{2} - c_1\right) = \frac{\nu}{2} - [\alpha c_2 + (1-\alpha)c_1].$$

The payoff to F is simply $\alpha\nu/2$. Equating the two payoffs, we get

$$\alpha = \frac{\frac{\nu}{2} - c_1}{\frac{\nu}{2} + c_2 - c_1}.$$

Note that we have $0 < \alpha < 1$, so this is a strictly mixed Nash equilibrium. Is this mixed strategy, which we will call M, an evolutionarily stable strategy? We have $\pi_{11} = v/2 - c_2$, $\pi_{21} = v/2$, $\pi_{22} = 0$, and $\pi_{12} = v/2 - c_1$. Thus $\pi_{11} = v/2 - c_2 < v/2 = \pi_{21}$ and $\pi_{22} = 0 < v/2 - c_1 = \pi_{12}$, so the equilibrium is an ESS.

10.10 Evolutionarily Stable Strategies Are Not Unbeatable

It is easy to show that x is an ESS in the game shown in the diagram. We shall see later that it is also asymptotically stable in the replicator dynamic (§12.9). Nevertheless, it is not an *unbeatable strategy*, in the sense of always having the highest payoff when invaded by multiple mutants.

	x	y	z
x	1,1	1,1	0,0
y	1,1	0,0	1,0
z	0,0	0,1	0,0

a. Show that x is an ESS.
b. Show that if a fraction ϵ_y of y-players and a fraction $\epsilon_z > \epsilon_y$ of z-players simultaneously invade, then y has a higher payoff than x.
c. Is the average payoff to the invaders higher than the payoff to x?

To complicate the picture, some game theorists have *defined* the ESS concept as "unbeatability" in the preceding sense. In a famous article, Boyd and Lorberbaum (1987) showed that "no pure strategy is an ESS in the repeated prisoner's dilemma game," and Farrell and Ware (1989) extended this by showing that no mixed strategy using a finite number of pure strategies is an ESS. Finally, Lorberbaum (1994) extended this to all nondeterministic strategies, and Bendor and Swistak (1995) showed that, for a sufficiently low discount rate, no pure strategy is an ESS in any nontrivial repeated game. In all cases, however, the ESS criterion is interpreted as "unbeatability" in the preceding sense. But unbeatability is not a very important concept, because it has no interesting dynamic properties. Be sure you understand how invasion by a pair of pure mutant strategies is not the same as being invaded by a single mixed strategy, and also be able to explain the intuition behind the preceding example.

10.11 A Nash Equilibrium That Is Not an EES

	1	2
1	1,1	1,1
2	1,1	2,2

Suppose agents consume each other's products but not their own. An agent can produce one or two units per period at cost 1, and then he meets another consumer. They can agree to exchange either one or two units. The utility of consumption is 2 per unit consumed. The first strategy is thus "exchange equal for equal, but at most one unit of the good," and strategy two is "exchange equal for equal, but at most two units of the good." The payoff matrix is shown in the diagram. Show that one of the Nash equilibria consists of evolutionarily stable strategies, but the other does not. What does this say about the ESS criterion and the elimination of weakly dominated strategies?

10.12 Rock, Paper, and Scissors Has No ESS

A Nash equilibrium that is not an ESS may nevertheless be quite important. Consider, for instance, rock, paper, and scissors (§6.25). Show that the unique, completely mixed Nash equilibrium to this game is not an ESS. We will see later (§12.14) that under a replicator dynamic, rock, paper, and scissors traces out closed orbits around the equilibrium (1/3, 1/3, 1/3), as suggested in the example of the lizard *Uta stansburiana* (§6.25).

10.13 Invasion of the Pure-Strategy Mutants

	1	2	3
1	1,1	1,1	1,1
2	1,1	0,0	a,a
3	1,1	a,a	0,0

It is possible that a Nash equilibrium be immune to invasion by any *pure* strategy mutant but not by an appropriate *mixed*-strategy mutant. This is the case with respect to the game in the diagram if one assumes $a > 2$. Here the first strategy is an ESS if only pure-strategy mutants are allowed, but not if mixed strategy mutants are allowed. Show that a mixed strategy using pure strategies 2 and 3 each with probability 1/2 can invade a Nash equilibrium consisting of strategy 1 alone.

Is there an evolutionarily stable strategy using pure strategies 2 and 3? Because mutants are normally considered to be rare, it is often plausible to restrict consideration to single mutant types or to mixed strategies that include only pure strategies used in the Nash equilibrium, plus at most one mutant.

10.14 Multiple Evolutionarily Stable Strategies

Using the matrix in the diagram, show that there are two evolutionarily stable strategies, one using the first two rows and columns, and the second using the second and third strategies. Show that there is also a completely mixed Nash equilibrium that is stable against invasion by pure strategies but is not an ESS. Prove the latter either by finding a mixed strategy that does invade. Hint: Try one of the evolutionarily stable strategies or use a previously proved theorem. If you want to cheat, look up a paper by Haigh (1975), which develops a simple algorithm for determining whether a Nash equilibrium is an ESS.

	1	2	3
1	5,5	7,8	2,1
2	8,7	6,6	5,8
3	1,2	8,5	4,4

10.15 Evolutionarily Stable Strategies in Finite Populations

Consider a population in which agents are randomly paired in each period and each pair plays a 2 × 2 game. Let $r_{\mu\nu}$ be the payoff to playing μ when your partner plays ν. Let $r(\mu)$ and $r(\nu)$ be the expected payoffs to an μ-type and a ν-type agent, respectively.

Suppose there are n agents, m of which play the "mutant" strategy μ, the rest playing the "incumbent" strategy ν. Then we have

$$\begin{aligned} r(\mu) &= \left(1 - \frac{m-1}{n}\right) r_{\mu\nu} + \frac{m-1}{n} r_{\mu\mu} \\ r(\nu) &= \left(1 - \frac{m}{n}\right) r_{\nu\nu} + \frac{m}{n} r_{\nu\mu}. \end{aligned}$$

It follows that

$$\begin{aligned} r(\nu) - r(\mu) = &\left(1 - \frac{m}{n}\right)(r_{\nu\nu} - r_{\mu\nu}) \\ &+ \frac{m}{n}(r_{\nu\mu} - r_{\mu\mu}) + \frac{1}{n}(r_{\mu\mu} - r_{\mu\nu}). \end{aligned} \tag{10.4}$$

We say a strategy ν is *noninvadable* if there is an $\epsilon > 0$ such that for all feasible mutants $\mu \neq \nu$ and all positive m such that $m/n < \epsilon$, $r(\nu) > r(\mu)$. When this condition fails, we say ν is *invadable*. We say a strategy ν is *Nash* if ν is a best reply to itself or, equivalently, if there is a Nash equilibrium in which only ν is played. We say a strategy ν is *evolutionarily stable* if there is a population size n such that ν is noninvadable for all populations of size n or greater.

While it is obviously possible for a Nash strategy to be invadable, it is also possible for a non-Nash strategy to be noninvadable, even by a Nash strategy. To see this, let $r_{\mu\nu} = 0$, $r_{\nu\mu} = n$, $r_{\mu\mu} = n + 1$, and $r_{\nu\nu} = -1$. Then ν is not Nash, because $r_{\mu\nu} > r_{\nu\nu}$, μ is Nash because $r_{\mu\mu} > r_{\mu\nu}$ and $r_{\mu\mu} > r_{\nu\mu}$. But, $r(\nu) - r(\mu) = 1/n > 0$ for any m.

THEOREM 10.9 *Strategy ν is evolutionarily stable if and only if ν is a Nash strategy, and for any μ that is a best reply to ν, ν is a better reply to μ than μ is to itself, or if μ is as good a reply to itself as ν is to μ, then μ is a better reply to itself than μ is to ν.*

PROOF: Suppose ν is evolutionarily stable but is not Nash. Then there is some μ such that $r_{\nu\nu} < r_{\mu\nu}$. Let $m = 1$. Then for sufficiently large n we have $r(\nu) < r(\mu)$ in

$$\begin{aligned} r(\nu) - r(\mu) = {} & \left(1 - \frac{m}{n}\right)(r_{\nu\nu} - r_{\mu\nu}) \\ & + \frac{m}{n}(r_{\nu\mu} - r_{\mu\mu}) + \frac{1}{n}(r_{\mu\mu} - r_{\mu\nu}). \qquad (10.5) \end{aligned}$$

Hence, ν must be Nash. Now suppose ν is evolutionarily stable and $r_{\nu\nu} = r_{\mu\nu}$ but $r_{\nu\mu} < r_{\mu\mu}$. Equation (10.5) becomes

$$r(\nu) - r(\mu) = \frac{1}{n}\left\{m[r_{\nu\mu} - r_{\mu\mu}] + [r_{\mu\mu} - r_{\mu\nu}]\right\}.$$

Given $\epsilon > 0$, choose $\overline{m}$ so that the term in brackets is negative, and then choose n so that $\overline{m}/n < \epsilon$. Then $r(\nu) < r(\mu)$ for all positive $m \le \overline{m}$, which is a contradiction. So suppose in addition to $r_{\nu\nu} = r_{\mu\nu}$ and $r_{\nu\mu} = r_{\mu\mu}$, we have $r_{\mu\mu} < r_{\mu\nu}$. Then clearly $r(\nu) - r(\mu) = [r_{\mu\mu} - r_{\mu\nu}]/n < 0$, again a contradiction. This proves that the stated conditions are necessary. We can reverse the argument to prove the conditions are sufficient as well. ■

The forgoing conditions are not those of Maynard Smith, which state that ν is evolutionarily stable if and only if ν is Nash and for any μ that is a best reply to n, ν is a better reply to μ than μ is to itself or, equivalently, for any mutant μ, either $r_{\nu\nu} > r_{\mu\nu}$, or $r_{\nu\nu} = r_{\mu\nu}$ and $r_{\nu\mu} > r_{\mu\mu}$. However, we can derive Maynard Smith's conditions by letting $m, n \to \infty$ in (10.4) in such a manner that $m/n = \epsilon$, but the limit argument cannot be used to conclude that $r(\nu) > r(\mu)$ in the "large finite" case.

To see this, note that in the limit we have

$$r(\nu) - r(\mu) = (1-\epsilon)[r_{\nu\nu} - r_{\mu\nu}] + \epsilon[r_{\nu\mu} - r_{\mu\mu}].$$

The conclusion follows immediately from this equation. The limit argument cannot be used to conclude that $r(\nu) > r(\mu)$ in the "large finite" case if $r_{\nu\nu} = r_{\mu\nu}$ and $r_{\nu\mu} = r_{\mu\mu}$.

Let $p = m/n, a = r_{\nu\nu} - r_{\mu\nu}, b = r_{\nu\mu} - r_{\mu\mu} - r_{\nu\nu} + r_{\mu\nu}$, and $c = r_{\mu\mu} - r_{\mu\nu}$. Then (10.4) becomes

$$r(\nu) - r(\mu) = \frac{1}{n}(na + mb + c). \tag{10.6}$$

Suppose ν can be invaded by mutant strategy μ, and the system follows any dynamic in which a strategy with a higher payoff increases in frequency. Then, if $n(a+b) > c$, μ will invade until μ is the largest integer less than $-(na+c)/b$. Otherwise μ will invade until ν is extinct.

In the case of partial invasion in the preceding example, we say μ is *quasi-evolutionarily stable*. Note that μ is quasi-evolutionarily stable with respect to ν for very large n if and only if μ and ν are part of a completely mixed Nash equilibrium (assuming there are no other feasible pure strategies).

10.16 Evolutionarily Stable Strategies in Asymmetric Games

Many situations can be modeled as evolutionary games, except for the fact that the two players are not interchangeable. For instance, in one-card two-round poker with bluffing (§6.21), the player going first has a different set of strategies from the player going second. Yet, despite the lack of symmetry, we simulated the game quite nicely as an evolutionary game. Analogous situations include interactions between predator and prey, boss and worker, male and female, incumbent and intruder, among a host of others.

This is not simply a technicality; it makes no sense to say that a "mutant meets its own type" when the game is asymmetric, so the ESS criterion has no meaning. The obvious way around this problem is to define a homogeneous set of players who in each period are paired randomly, one of the pair being randomly assigned to be player 1, and the other to be player 2. We may call this the "symmetric version" of the asymmetric evolutionary game. However, *an evolutionarily stable strategy in the symmetric version*

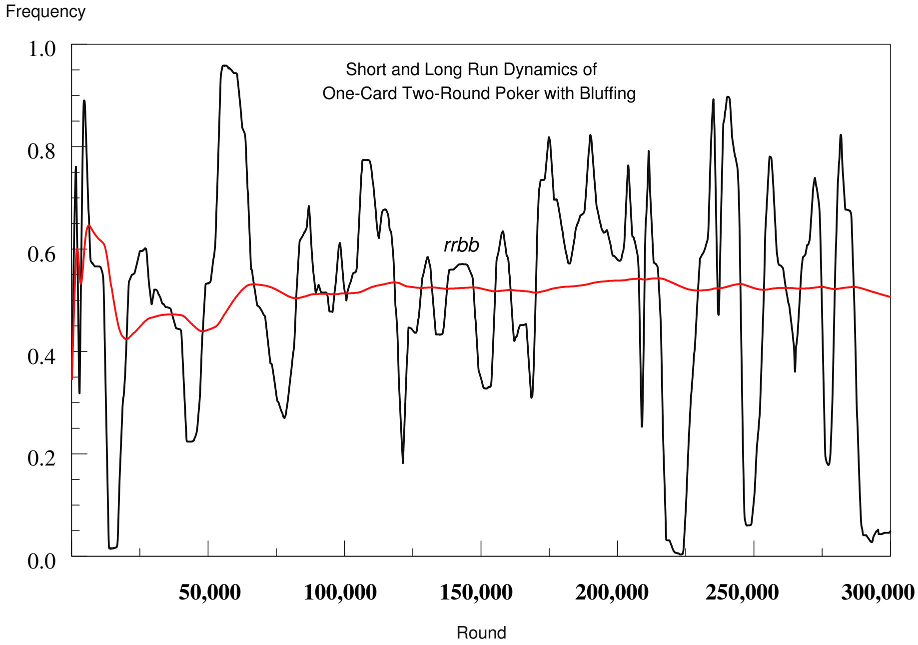

Figure 10.2. An agent-based Model of one-card two-round poker with bluffing

of an asymmetric evolutionary game must be a strict Nash equilibrium; that is, each type in the asymmetric game must use exactly *one* pure strategy (Selten 1980). To see this, suppose there is a Nash equilibrium $u = (\sigma_1, t_2)$ of the symmetric version, where a player uses strictly mixed strategy σ_1 when assigned to be player 1 and uses t_2 (pure or mixed) when assigned to be player 2. Consider a mutant that uses $v = (s_1, t_2)$, where s_1 is a pure strategy that appears with positive weight in σ_1. Then v does as well against u as u does against itself, and v does as well against v as u does against v. All this is true because the payoff to s_1 against t_2 in the asymmetric game is equal to the payoff to σ_1 against t_2 by the fundamental theorem (§3.6).

It follows that *mixed-strategy Nash equilibria in asymmetric evolutionary games are never evolutionarily stable in the symmetric version of the game*. As we shall see later, this situation reflects the fact that mixed-strategy Nash equilibria in asymmetric evolutionary games with a replicator dynamic are never asymptotically stable (§12.17). Some game theorists consider this a weakness of evolutionary game theory (Mailath 1998), but in fact it reflects a deep and important regularity of social interaction. In asymmetric evolutionary games, the frequency of different types of behavior goes through periodic cycles through time.

For a dramatic example of this important insight, we return to our model of one-card two-round poker with bluffing (figure 6.6). In this agent-based model, I lowered the mutation rate to 1% and ran the model for 300,000 periods. The results are shown in figure 10.2 for one of the player 1 types, $rrbb$ (bluff all the way). Note that the *average frequency* of each strategy settles down to the theoretically predicted equilibrium value, but the *period-to-period frequencies* fluctuate wildly in the medium run. Strategy $rrbb$, which has the equilibrium frequency of about 50%, sometimes goes for thousands of periods with frequency under 5% or over 90%.