Evolution and Cooperation in Noisy Repeated Games

By Drew Fudenberg and Eric Maskin*

The theory of repeated games has been an important tool for understanding how cooperation might arise in a population of self-interested agents. If the prisoner's dilemma, as depicted by the following matrix,

	C	D
C	2,2	−1,3
D	3,−1	0,0

is played only once, the unique equilibrium is for both players to play D, resulting in the inefficient payoffs $(0,0)$. If, instead, the game is repeated infinitely often and the players are not too impatient, there are "cooperative" equilibria in which both players always play C for fear that failure to do so would cause the opponent to "punish" them with D in the future.

Although repeated play permits cooperation as an equilibrium, it does not preclude less cooperative outcomes. Indeed, it is also an equilibrium for both players to use the strategy "always play D." More generally, the Folk Theorems establish that any feasible individually rational payoffs can arise in an equilibrium if the players are sufficiently patient. (See Robert Aumann and Lloyd Shapley, 1976; Ariel Rubinstein, 1979; and our 1986 article.) In our version of the prisoner's dilemma, a payoff vector is individually rational if each player's payoff is positive.

However, not all these equilibria are equally plausible. Specifically, there is a widespread intuition that often the most likely equilibria are those whose payoffs are efficient, and efficiency is typically *assumed* in economic applications of repeated games. In fact, selecting the efficient equilibria is an approach sometimes taken for games in general (see John Harsanyi and Reinhard Selten, 1988). But the intuition for efficiency seems particularly strong for the case of repeated games.

Herein we provide support for this intuition. For a class of repeated games that includes the prisoner's dilemma, we show that efficiency is implied by evolutionary stability if there is a small probability that the players make "mistakes," so that their realized actions are sometimes different from those they intended.

The evolutionary process we have in mind is one in which pairs of strategies from a population are matched at random to play the repeated game, and a strategy has equal chances of being assigned the role of players 1 and 2. If a strategy performs well on average against the population (i.e., its average payoff is comparatively high), the fraction of the population corresponding to that strategy grows; by contrast, the proportion for a low-payoff strategy declines. In biological settings, such a process results from the relative advantage that a successful strategy has in reproducing itself. In economic or sociological applications, evolution may derive from learning: if players try out strategies on an experimental basis and one works well against the population, more and more players are likely to adopt it (as word gets around), whereas a poor strategy will probably be abandoned.

A strategy is evolutionarily stable (ES) if it can ultimately dominate. That is, following John Maynard Smith (1982), it cannot be "invaded" by any mutant strategy. An ES strategy must be a "best response" to itself, because a mutant playing a better response would have a higher average payoff. Moreover, for strategy s to be ES, there cannot be a mutant s' that does better than s does against a population consisting of a high proportion of s's and a small proportion of s''s.

*MIT, Cambridge, MA 02139, and Harvard University, Cambridge, MA 02138, respectively.

Robert Axelrod and William Hamilton (1981) showed that the strategy "always D" is not ES in the repeated prisoner's dilemma with time-average (undiscounted) payoffs. In particular, always D can be invaded by the strategy "tit-for-tat" (that plays C in the first period, and then always plays the same action as its opponent played the previous period). Tit-for-tat can invade because it does as well against always D as always D does against itself (a time-average of 0), and tit-for-tat obtains a payoff of 2 when paired against itself.

Although evolutionary stability rules out always D, the uncooperative outcome can be approximated arbitrarily closely by an ES strategy profile. For example, consider the strategy "Play C in periods $0, n, 2n, \ldots$, and D in all other periods, so long as past play has always conformed to this pattern; if past play has not always conformed, then play D forever afterwards." Call this strategy "mostly D." For large n, mostly D is nearly as uncooperative as always D. Yet it is ES, because if an invader deviates from the pattern, it is punished forever and so obtains a time-average payoff of at most 0. In contrast, mostly D obtains $2/n$ against itself. If the proportion of mutants is sufficiently small, even a mutant that attains an average of 2 when paired against itself cannot overcome this punishment. Thus evolutionary stability by itself has almost no restrictive power in the repeated prisoner's dilemma.

The first step in our argument is the observation that mostly D performs quite badly (payoff 0) against itself if one of the players deviates from the prescribed pattern by "mistake." This suggests that, if mistakes are possible, evolution may tend to weed out strategies that impose drastic penalties for deviations. We then exploit the idea that, if punishments are not too drastic, it is relatively easy for inefficient strategies to be invaded by efficient ones, since the penalty for trying to initiate efficient play against an inefficient opponent is not large.

To formalize the notion of a small probability of mistake, we suppose that each player's expected payoff conditional on there never being a mistake is lexicographically more important than his expected payoff conditional on one mistake, which in turn is more important than his payoff conditional on two mistakes, and so on. In this paper, we consider only the class of symmetric two-player games, which includes the prisoner's dilemma and also the well-known "battle of the sexes." We also assume that players use strategies of only finite complexity: those that can be implemented by a finite computer with a finite memory. We show that if, as in the prisoner's dilemma, there is a unique payoff pair that maximizes the sum of players' payoffs, then any ES (pure) strategy must be efficient. If there are multiple pairs that maximize the sum of the payoffs, then ES does not imply efficiency but still imposes restrictions on the set of equilibrium payoffs.

I. Symmetric Repeated Games with Noise

We consider a symmetric two-player game. Hence, both players choose actions from a common (finite) set A, and we can express a player's payoff as $u(a, b)$, where a is his own action and b that of his opponent. For future reference let $\underline{v} = \min_b \max_a u(a, b)$ be a player's *minmax* payoff and let $V = $ Convex Hull$\{(v_1, v_2)|$there exists (a, b) with $u(a, b) = v_1$ and $u(b, a) = v_2\}$ be the set of feasible payoffs.

We are interested in the infinitely repeated game where the above one-shot game is played in each period $t = 0, 1, \ldots$. We suppose that a player's realized action $\hat{a}$ is sometimes different from the action a he intended (so that the realized action is a noisy signal of the one intended), and that payoffs in each period depend only on realized actions. (Alternatively, we could assume that intended and realized actions never differ, but that a player's action might be misperceived by his opponent. In this case, payoffs would depend on both actions and the noise creating the misperception.) Furthermore, a player observes just his opponent's realized action, not the intended one. Thus a player's public history at time t, denoted $h(t)$, is the sequence $[(\hat{a}(0), \hat{b}(0), \ldots, (\hat{a}(t-1), \hat{b}(t-1))]$, where the $\hat{a}$'s and $\hat{b}$'s are his own and his opponent's realized actions, respectively. (Set the history $h(0)$ at

date 0 equal to the null set.) History $h(t)$ has length, denoted by $l(h(t))$, equal to t. Let H be the space of all histories. Note that each sequence of play generates two histories, one for each player. For each $h \in H$, let $\pi(h)$ be the corresponding history for the other player, that is, $\pi(h)$ is obtained by permuting $\hat{a}(\tau)$ and $\hat{b}(\tau)$ for each τ. We use this formulation so that, when players use symmetric strategies, we can denote them both by the same map s: $H \to A$. For example, for either player the strategy tit-for-tat in the repeated prisoner's dilemma can be expressed as follows: for all $h \in H$, $s(h) = C$ if $l(h) = 0$; and $s(h) = \hat{b}(l(h) - 1)$ if $l(h) > 0$. We will restrict attention to pure strategies that depend only on the public history.

We shall assume that players use strategies of only finite complexity. A formal definition of this concept is introduced by Ehud Kalai and William Stanford (1988). Informally, as we indicated in the introduction, a strategy of finite complexity is a finite program that can be run on a finite computer with bounded memory. Most familiar strategies are of this form, including all the repeated prisoner's dilemma strategies mentioned above.

A pair of finitely complex strategies $\delta = (s, s')$ gives rise to a sequence of play $[(\hat{a}(0), \hat{b}(0)), (\hat{a}(1), \hat{b}(1)), \ldots]$ (the $\hat{a}$'s and $\hat{b}$'s correspond to the players using s and s', respectively) that eventually repeats itself (i.e., forms a repetitive cycle). We suppose that players do not discount, that is, they maximize their time-average payoffs. A player's time-average payoff from δ (contingent on there being no mistake) is

$$v(\delta) = \lim_{T \to \infty} (1/T) \sum_{t=0}^{T} u(\hat{a}(t), \hat{b}(t)).$$

Let $v(\delta|h)$ be a player's expected payoff from δ conditional on history h occurring, where $l(h) = \tau$. That is, if $[(\hat{a}(\tau), \hat{b}(\tau)), (\hat{a}(\tau+1), \hat{b}(\tau+1)), \ldots]$ is the sequence of action pairs induced by δ after history h,

$$v(\delta|h) = \lim_{T \to \infty} (1/T) \sum_{\tau}^{T+\tau} u(\hat{a}(t), \hat{b}(t)).$$

To compute the expected payoff conditional on one mistake requires a probability distribution on when mistakes occur. For our results, we can accommodate any distribution with the properties that, conditional on each mistake, each player has a positive probability of making it and each period has a positive probability of being the one in which it occurs. For $m = 1, 2, \ldots$, let $v^m(\delta)$ be a player's expected payoff conditional on there having been m mistakes. By convention, $v^0(\delta) = v(\delta)$.

We call $v^m(\delta)$ a player's "mth-order" payoff from the strategy pair δ. Our lexicographic assumption implies that, for any two pairs δ and $\hat{\delta}$, if $m < m'$ any difference between $v^m(\delta)$ and $v^m(\hat{\delta})$ outweighs a difference between $v^{m'}(\delta)$ and $v^{m'}(\hat{\delta})$ in a player's performance. As we mentioned, this formalizes the idea that mistakes are improbable.

The strategies constructed in the papers cited above on the Folk Theorem extend to our "noisy" repeated game, so that for any payoff vector $(v, v) \in V$ with $v \geq \underline{v}$ there is a subgame-perfect equilibrium whose payoffs contingent on no mistakes are (v, v). We will see that considering ES strategies considerably restricts the set of equilibrium payoffs.

II. Evolutionary Stability

As described informally in the introduction, a strategy s is ES if no other strategy s' can invade a population consisting of a large fraction of s's and a small proportion of s''s. We shall suppose that the proportion of s''s, although small, is big relative to the probability of a mistake. Hence, given the lexicographic preferences we have assumed, s' can invade even if it performs worse to the mth-order against s than s does itself, provided that s' fares better to a lower order against itself than s does against s'.

Formally, s is *evolutionary stable* provided that, for all finitely complex strategies s', if there exists m^* such that $v^m(s', s') = v^m(s, s')$ for all $m < m^*$, then there exists $m^{**} \leq m^*$ such that

$$(*) \qquad v^{m^{**}}(s, s) \geq v^{m^{**}}(s', s)$$

and $v^m(s, s) \geq v^m(s', s)$ for all $m < m^{**}$.

Furthermore, if $v^{m*}(s', s') > v^{m*}(s, s')$ then (*) is strict.

This definition reduces to the standard one (see Maynard Smith) if the probability of mistakes is literally zero.[1] In that case, only preferences to the 0th order matter, and so s' can invade only if either it performs strictly better than s does against s, or it performs equally well but fares strictly better than s against s'.

We shall say that payoff vector $(v, v') \in V$ is *efficient* if it maximizes the sum of players' payoffs on the set V (i.e., $(v, v') \in \arg\max_{(u, u')}\{u + u' | (u, u') \in V\}$). A strategy s is efficient if $(v(s, s), v(s, s))$ is efficient. Note that the sum of payoffs is relevant because, as we discussed, a strategy has an equal chance of being assigned either player's role.

THEOREM 1: *Let* $\underline{u} = \min\{u | \text{there exists } u'$ *such that* (u, u') *is efficient*$\}$. *If a finitely complex strategy s is ES, then $v(s, s) \geq \underline{u}$. Thus if there is a unique efficient pair (u^*, u^*), $v(s, s) = u^*$.*

PROOF:

Note first that, for any history h, $(v(s, s|h), v(s, s|\pi(h))) \in V$. Because s is finitely complex, there is a history h^* that minimizes $v(s, s|h)$. We claim that there exists v' such that $(v(s, s|h^*), v')$ is efficient. If not, then in particular $(v(s, s|h^*), v(s,s|\pi(h^*)))$ is inefficient. Choose a sequence of action pairs $\{(a^*(t), b^*(t))\}_{t=0}^{\infty}$ whose time-average payoffs are efficient. Fix actions $a \neq s(h)$ and $b \neq s(\pi(h)),$ $(s(\pi(h)), a))$ (where $(\pi(h), (s(\pi(h)), a))$ is the history $\pi(h)$ followed by the action pair $\pi(a, s(\pi(h))) = (s(\pi(h)), a)$. Let s' be a strategy that coincides with s expected after histories h^* and $\pi(h^*)$. After h^*, define s' so

[1]Actually it reduces to something slightly weaker than the standard definition. According to the latter, s is ES provided that, for all s', $v(s, s) \geq v(s', s)$ and, if this holds with equality, then $v(s, s') > v(s', s')$. When mistakes have zero probability, we (like Axelrod and Hamilton) deem s to be ES even when the last inequality is weak. However, given the noise in our model, this discrepancy may well be an artifact of our no-discounting assumption.

that it plays a for two periods. If the opponent plays b in the second period after h^* (i.e., in period $l(h^*)+1$), s' plays the sequence $\{a^*(t)\}$ beginning in the third period after h^* (i.e., in period $l(h^*)+2$, it plays $a^*(0)$, in period $l(h^*)+3$ it plays $a^*(1)$, etc.). If the opponent fails to play b in period $l(h^*)+1$, then s' coincides with s beginning in period $l(h^*)+2$. After $\pi(h^*)$, define s' so that it plays $s(\pi(h^*))$ the first period (i.e., in period $l(\pi(h^*))$). If the opponent plays a in that first period s' plays b in the next period (period $l(\pi(h^*))+1$), and subsequently plays the sequence $\{b^*(t)\}$ (i.e., it plays $b^*(0)$ in period $l(\pi(h^*))+2$, $b^*(1)$ in period $l(\pi(h^*))$ $+3$, etc.). If the opponent fails to play a in period $l(\pi(h^*))$, s' coincides with s thereafter.

We will show that s' can invade. We first claim that

$$(1) \qquad v(s', s|h^*) = v(s, s|h^*).$$

By our choice of h^* and because s' reverts to s after h^* if its opponent does not play according to s', $v(s', s|h^*) \geq v(s, s|h^*)$. Now if this last inequality is strict, let s'' be the strategy that coincides with s' after h^* and otherwise coincides with s. Then $v(s'', s|h^*)$ $> v(s, s|h^*)$, and so if m^* is the number of "mistakes" that must occur if (s, s) gives rise to h^*, $v^{m*}(s'', s) > v^{m*}(s, s)$. But by definition of s'', $v^{m}(s'', s) = v^{m}(s, s)$ and $v^{m}(s'', s'') = v^{m}(s, s'')$ for all $m < m^*$. Thus, s'' can invade, contradicting the evolutionary stability of s. We conclude that (1) must hold.

Because s' plays like s after $\pi(h^*)$ unless its opponent plays according to s',

$$(2) \quad v(s', s|\pi(h^*)) = v(s, s|\pi(h^*))$$

and $\qquad v(s, s|h^*) = v(s, s'|h^*).$

Now, by choice of s', $v(s', s'|h^*) +$ $v(s', s'|\pi(h^*))$ is efficient. Hence, from (1) and because there exists no v' such that $(v(s, s|h^*), v')$ is efficient, $v(s', s'|h^*) +$ $v(s', s'|\pi(h^*)) > v(s', s|h^*) + v(s, s'|\pi(h^*))$. But from (1) and (2) this last inequality is

equivalent to

$$(3) \quad v(s',s'|h^*) + v(s',s'|\pi(h^*))$$

$$> v(s,s'|h^*) + v(s,s'|\pi(h^*)).$$

Now (1), (2), and (3) imply that

$$(4) \quad v^{m^*}(s',s') > v^{m^*}(s,s')$$

and $\quad v^{m^*}(s',s) = v^{m^*}(s,s).$

Our choice of s' implies, moreover, that

$$(5) \quad v^m(s',s') = v^m(s,s') = v^m(s',s)$$

$$= v^m(s,s) \text{ for } m < m^*.$$

But from (4) and (5) we infer that s' can invade, contradicting the evolutionary stability of s.

We conclude that there must exist v' such that $(v(s,s|h^*), v')$ is efficient after all. The theorem follows from the definition of h^*.

Because $(2,2)$ is the unique efficient payoff pair in the prisoner's dilemma, Theorem 1 implies that any finitely complex ES strategy must give rise to the cooperative outcome in the repeated game. The qualification "finitely complex" is, however, important. For example, consider the following infinitely complex strategy s^0: "alternate between D and C; if either player deviates from this pattern, switch to one in which C is played only every third period; if either player deviates from *this* pattern, switch to one in which C is played only every fourth period, etc." Strategy s^0 is ES because, regardless of the history, any attempt to deviate from (s^0, s^0) is always punished by having one's payoff reduced by a positive amount. (For a finitely complex strategy, by contrast there must always be a history after which further deviations no longer reduce one's payoff.) But (s^0, s^0) results in the inefficient payoffs $(1,1)$. Indeed, evolutionary stability does not restrict the Folk Theorem payoffs at all if infinitely complex strategies are possible.

Even for finitely complex strategies, Theorem 1 does not necessarily imply efficiency,

TABLE 1—MODIFIED BATTLE OF THE SEXES

	a	b	c	d
a	0,0	4,1	0,0	0,0
b	1,4	0,0	0,0	0,0
c	0,0	0,0	0,0	0,0
d	0,0	0,0	0,0	2,2

as we noted earlier. Consider the modified version of the battle of the sexes shown in Table 1. Theorem 1 rules out only payoffs less than 1 in the repeated game (which, however, is still considerably more restrictive than the Folk Theorem, which permits payoffs as low as zero). Indeed, as the following converse of Theorem 1 establishes, any feasible payoff greater than 1 can arise from ES strategies.

THEOREM 2: *Let $\underline{u}$ be defined as in Theorem 1. If $(v,v) \in V$ and $v > max\{\underline{v}, \underline{u}\}$, then there exists a finitely complex ES strategy s such that $v(s,s) = v$, provided there exists a finitely complex strategy with these payoffs.*

Rather than prove this theorem in full generality, we shall simply exhibit ES strategies for the prisoner's dilemma and modified battle of the sexes. In the former, consider the following modification of tit-for-tat: "play C the first period and thereafter play C in a given period if and only if either both players played C or both played D the previous period." Observe that continuation payoffs are always efficient: a mistake triggers a single period of "punishment" and then a return to cooperation. Note that tit-for-tat itself is not ES, because a mistake will trigger the (inefficient) infinite cycle (C, D) (D, C), (C, D) and so on. In the modified battle of the sexes, consider the strategy "Play d the first period and subsequently play d as long as in every past period either both players played d or neither did. If a single player deviates from d, then henceforth that player plays b and his opponent plays a." Even though this strategy is not efficient, it is ES, since any attempt to promote greater efficiency will be punished forever; the punishments are stable because the strategies at that point form an efficient equilibrium all of whose continuation equilibria are efficient.

III. Extensions

In this paper we have assumed no discounting, infinitesimal noise, pure strategies, symmetry across players, and equilibrium configurations where only a single strategy is played. The intuitive reasons behind our results, however, seem quite general, and so in a forthcoming paper we expect to relax all these assumptions. Our work builds on an extensive literature, but, for lack of space, we must also postpone discussion of previous work to the sequel.

REFERENCES

Aumann, Robert and Shapley, Lloyd, "Long-Term Competition: A Game-Theoretical Analysis," mimeo., 1976.

Axelrod, Robert and Hamilton, William, "The Evolution of Cooperation," *Science*, November 1981, *211*, 1390–96.

Fudenberg, Drew and Maskin, Eric, "The Folk Theorem in Repeated Games with Discounting or with Incomplete Information," *Econometrica*, May 1986, *54*, 533–56.

Harsanyi, John and Selten, Reinhard, *A General Theory of Equilibrium Selection in Games*, Cambridge: MIT Press, 1988.

Kalai, Ehud and Stanford, William, "Finite Rationality and Interpersonal Complexity in Repeated Games," *Econometrica*, March, 1988, *56*, 391–410.

Maynard Smith, John, *Evolution and the Theory of Games*, Cambridge: Cambridge University Press, 1982.

Rubinstein, Ariel, "Equilibrium in Supergames with the Overtaking Criterion," *Journal of Economic Theory*, January, 1979, *21*, 1–9.