ERIC RASMUSEN

FOLK THEOREMS FOR THE OBSERVABLE IMPLICATIONS OF REPEATED GAMES*

ABSTRACT. The fact that infinitely repeated games have many different equilibrium outcomes is known as the Folk Theorem. Previous versions of the Folk Theorem have characterized only the payoffs of the game. This paper shows that over a finite portion of an infinitely repeated game, the concept of perfect equilibrium imposes virtually no restrictions on observable behavior. The Prisoner's Dilemma is presented as an example and discussed in detail.

Keywords: Folk Theorem, prisoner's dilemma, repeated games, supergames, tit-for-tat.

1. INTRODUCTION

The Folk Theorem says that an infinitely repeated game has many different equilibrium outcomes. In formalizing this idea, the most important tasks are to define 'many', 'outcome', and 'equilibrium', and to specify how the payoffs of infinite games are calculated. Different formalizations have resulted in a variety of statements of the Folk Theorem, many of which are surveyed by Fudenberg and Maskin (1986).[1] Fudenberg and Maskin pay special attention to the importance of discounting, and the difference between the Nash and subgame perfect equilibrium concepts.[2]

Running throughout Fudenberg and Maskin and the rest of the literature is the definition of 'outcome' as a set of payoffs averaged over the course of the game. The usual versions of the Folk Theorem therefore say that in an infinitely repeated game a large number of equilibrium average payoff combinations exist.[3]

This paper will use a different definition of outcome: that an outcome is a time path of action combinations. Such a definition is useful for two reasons. First, it directly applies to behavior, which is usually what interests the applied modeller. Second, it calls attention to the fact that some equilibria include actions which generate current-period payoffs below the minimax levels. This fact is obscured in

Theory and Decision **32**: 147–164, 1992.

existing Folk Theorems by the definition of 'many', which restricts equilibrium outcomes to average payoffs greater than the minimax levels.

The new theorems in this article are not implied by existing theorems, but neither they nor their proofs are deep, given existing results. I have concentrated on trying to provide applied modellers with some useful theorems and an unusually lengthy discussion of the assumptions. Section 2 lays out the model and the notation. Section 3 states four theorems: a conventional Folk Theorem from Fudenberg and Maskin (1986), and three new theorems, all stated in terms of observed outcomes, but differing in their applicability. Section 4 discusses the theorems and their assumptions, and Section 5 applies them to the repeated Prisoner's Dilemma.

2. THE MODEL

Consider an n-person game with payoffs $\pi: M_1 \times \cdots \times M_n \rightarrow R^n$, where M_i is a finite set whose elements are the moves available to player i, and π is the vector of payoff functions for the n players.[4] This is the *one-shot game*, in distinction to the larger game consisting of repetitions of the one-shot game.

In the mixed extension of the game, player i also has the option of randomizing over M_i, and we will denote a set of moves or lotteries of moves from M_i by a_i. In the one-shot game, a_i would be player i's strategy, but in the repeated game, it is just one period's component of his strategy. Game theory lacks a conventional term for the set of moves plus lotteries of moves; let us call a_i an *action* in this article; an action can therefore be either pure or mixed. An *action combination* is a vector a of one action for each player, $\{a_i\}$, $(i = 1, \ldots, n)$. Given the action combination a, a_{-j} denotes the set of $(n-1)$ actions that excludes a_j.

We will assume that the actions are chosen simultaneously, that they are observed by all players, and that the players observe which mixed actions are being played (rather than having to infer them from the realizations of randomized decisions). There exists some set of actions a_{-i} that minimizes player i's payoff given that he is allowed to choose his best response to a_{-i}. We say that the other players are *minimaxing* player i in this situation. Since the payoffs are in units of non-transferable utility, they can be rescaled. Let us normalize to zero the one-shot payoff for player i when he is being minimaxed, so

$$0 \equiv \min_{a_{-i}} \max_{a_i}$$

Let W denote the convex hull of the set of all possible one-shot payoff combinations and let W^* denote the set of strictly positive n-vectors lying in W (i.e. vectors that exceed the minimax levels in every component). In the mixed extension of the game, all payoff combinations in W are feasible.

Discounting has two important sources: the rate of time preference, ρ, and the probability that the game is not repeated at the end of the period, denoted ω. Assume both ρ and ω to be constant. We will represent the discount factor – the value of one dollar received after one period – by δ, which equals one if there is no discounting. Because of time preference, a player is indifferent between $1/(1+\rho)$ now and 1 guaranteed to be paid a period later. With probability $(1-\omega)$ the later payment actually takes place, so the player is indifferent between $(1-\omega)/(1+\rho)$ now and the promise of 1 to be paid a period from now conditional upon the game still continuing. The discount factor is therefore

$$\delta = \frac{(1-\omega)}{(1+\rho)} .$$

If there is discounting, the discounted sum of a player's one-shot payoffs is his natural objective function, but if the discount rate is zero, the sum of the payoffs over an infinite number of periods is generally infinite and we must find another maximand. There are two solutions to this problem. The first solution, used by Fudenberg and Maskin, is for players to maximize their average payoffs if there is no discounting and the discounted value otherwise. The second solution is for players to use an overtaking criterion, according to which a player prefers payoff stream X to Y if there exists an integer T such that X's sum of discounted payoffs up to period t exceeds Y's sum for $t > T$. The overtaking criterion applies whether there is discounting or not.

The game to be analyzed is the infinite-repetition of the one-shot game just described, where the overall payoff is defined by the one-shot payoffs, the discount rate, and the overtaking criterion. The task is to characterize the set of subgame-perfect equilibria: Nash strategy combinations such that in each of the infinite number of subgames the remaining components of the strategies remain Nash.

The theorems in the next section will require one of two additional assumptions: the Dimensionality Condition or the Two-Equilibrium Condition.[5]

DIMENSIONALITY CONDITION. *The space W^* of feasible payoff combinations that strictly Pareto-dominate the minimax payoff combinations is nonempty and n-dimensional*:

$$\exists a: \pi(a) > 0\,, \quad \textit{and}$$

$$\forall a: \pi(a) > 0, \forall i, \exists \tilde{a}^i, \varepsilon > 0: \pi(\tilde{a}^i) = (\pi_{-i}(a), \pi_i(a) - \varepsilon) > 0\,.$$

As an example of the Dimensionality Condition failing, consider the n-person coordination game in which player i has the payoff $\pi_i = a$ if every player picks the same action $a \in [0, 1]$, and $\pi_i = 0$ otherwise. If the players coordinate, their common payoff varies depending on the coordinated action, but the Dimensionality Condition fails, because the payoffs vary one-dimensionally and there are $n > 1$ players. Figure 1 contrasts the coordination game with the Prisoner's Dilemma, which does satisfy the Dimensionality Condition.

An alternative to the Dimensionality Condition is the Two-Equilibrium Condition, which says that the repeated game has some stationary perfect equilibrium preferred by each player to some other stationary perfect equilibrium (where the inferior equilibrium can be the same for each player, or different).[6]

TWO-EQUILIBRIUM CONDITION. *Given the discount factor δ, the infinitely repeated game has at least two equilibria: a 'desirable' subgame-perfect equilibrium in which action combination $\bar{a}$ is played each period; and for each player i a 'punishment' subgame-perfect*

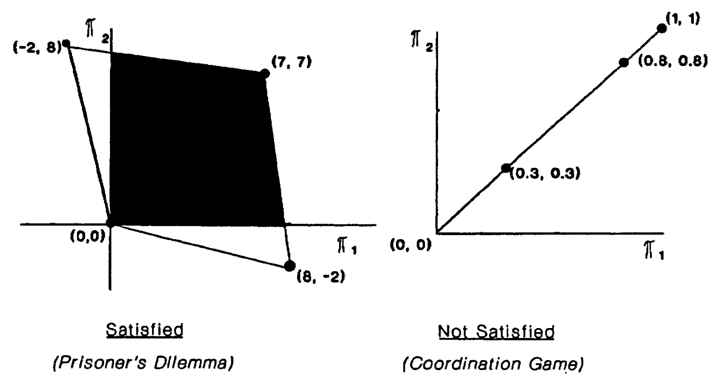

Fig. 1. The Dimensionality Condition.

equilibrium with a lower payoff for him, in which action combination $\underline{a}^i$ *is played each period, where possibly* $\underline{a}^i = \underline{a}^j$ *for players i and j*:

$$\exists \bar{a} : \forall i, \exists \underline{a}^i : \pi_i(\underline{a}^i) < \pi_i(\bar{a}) .$$

The Two-Equilibrium Condition is not a primitive assumption. It requires the game to have *equilibria* with certain properties, rather than requiring the action spaces or payoff functions to have certain properties. Good modelling practice generally calls for making only primitive assumptions, but the Dimensionality Condition might be false or hard to verify for some games for which the Two-Equilibrium Condition clearly is true.

Although the Two-Equilibrium Condition is an alternative to the Dimensionality Condition, they are not equivalent. The Dimensionality Condition does imply the Two-Equilibrium Condition (by virtue of the Folk Theorem itself), but the converse is not true. The coordination game mentioned earlier did not satisfy the Dimensionality Condition, but it does satisfy the Two-Equilibrium Condition, because even as a one-shot game, it has two Nash equilibria (in fact, it has a continuum). To satisfy the Two-Equilibrium Condition, let $\underline{a}^i$ be 'every player chooses 0.3' for every i, and let $\bar{a}$ be 'every player chooses 0.8'.

The Dimensionality and Two-Equilibrium Conditions are useful for constructing strategies that induce the players to punish anyone who deviates from equilibrium. In the prisoner's dilemma with the payoffs of Figure 2 below, the threat can be to play actions yielding $(-2, 8)$ or $(8, -2)$ if player 1 or 2 deviates. In the coordination game, the threat can be to play actions yielding $(0, 0)$ instead of $(1, 1)$ if either player deviates. The coordination game is a harder one in which to maintain punishments, because the act of punishing hurts the punisher.

3. FOUR VERSIONS OF THE FOLK THEOREM

The most important of the several versions of the Folk Theorem in Fudenberg and Maskin (1986) is restated below in the notation of this paper.[7]

1. THE FUDENBERG–MASKIN THEOREM (Fudenberg and Maskin, 1986). *For any* $(w_1^*, \ldots, w_n^*)$ *in* W^*, *there exists a subgame-perfect equilibrium of the infinitely repeated game in which player i's average payoff is* w_i^* *if*

(1) *The Dimensionality Condition holds, and*
(2) *The discount factor is sufficiently close to one.*

The various versions of the folk theorem are generally proved by construction. The equilibrium strategy combination to support a desired outcome has two parts: an equilibrium path, and a punishment to be imposed on a player who deviates. For it to be a perfect equilibrium, the punishment must be credible, which is not immediate if carrying out the punishment imposes costs on the punishing players. The punishment part of the strategy must therefore also specify a punishment for players who deviate by not punishing when they are supposed to.

Fudenberg and Maskin's proof of their theorem does not use the obvious approach to constructing an equilibrium, which is to build an ever-increasing series of punishments for deviation, deviation from punishing, deviation from punishing non-punishing, etc., as Rubinstein

(1979) and Benoit and Krishna (1985) do for undiscounted games. That approach runs into difficulty when payoffs are discounted, because the severity of the punishments cannot increase fast enough as the punishment phase becomes longer. The punishment must be slightly more complicated: the length of the punishment increases if a player deviates, but also whichever player last deviated bears the heaviest punishment. A player who deviates by not punishing immediately starts earning a lower per-period payoff, besides lengthening the number of punishment periods. The Dimensionality Condition is needed to ensure that the deviator can be thus punished more severely than the other players.

The first of the three new versions of the Folk Theorem is the Stationary Actions Theorem, which appears very similar to the Fudenberg–Maskin Theorem. The difference is that the Fudenberg–Maskin Theorem says that any of a range of average payoffs can be sustained in equilibrium; this theorem says that any of a range of stationary action outcomes can be sustained in equilibrium.

2. THE STATIONARY ACTIONS THEOREM. *In the infinitely repeated game, repetition of any action combination* a^* *that yields per-period payoffs in* W^* *is the equilibrium path of some subgame-perfect equilibrium if*

(1) *The Dimensionality Condition holds, and*
(2) *The discount factor is sufficiently close to one.*

The proof of the Stationary Actions Theorem is very close to that of the Fudenberg–Maskin Theorem. Fudenberg and Maskin construct an equilibrium-path strategy that yields a particular average payoff w^*. That equilibrium-path strategy can be the stationary action combination a^*.

The Stationary Actions Theorem only applies to stationary strategies yielding payoffs strictly greater than the minimax payoffs. The range of observable equilibrium outcomes is much greater than this, as the next two theorems show. Both of these next theorems have the same conclusions, but they differ in their assumptions.

3. THE DIMENSIONALITY THEOREM. *In the infinitely repeated game, any action combinations observed over any finite number of repetitions T can be generated by a subgame-perfect equilibrium if*

(1) *The Dimensionality Condition holds, and*
(2) *The discount factor is sufficiently close to one.*

4. THE TWO-EQUILIBRIUM THEOREM. *In the infinitely repeated game, any action combinations observed over any finite number of repetitions T can be generated by a subgame-perfect equilibrium if*

(1) *The Two-Equilibrium Condition holds, and*
(2) *The discount factor is sufficiently close to one.*

The proof of the Dimensionality Theorem is based on the Fudenberg–Maskin Theorem, which says that there exist equilibria that yield per-period payoff vectors π^1 and π^2 with $\pi^1 > \pi^2$, equilibria we shall call E^1 and E^2. Let the observed sequence of actions up to period T be denoted by a^T. An equilibrium that generates a^T is for player i to (a) play his components of a^T up to period T, unless someone deviates, (b) play according to E^1 after T if no deviations have yet occurred, and (c) play according to E^2 immediately and forever if any player deviates. Player i has no incentive to deviate from (a) because whatever T-period gain might be received is outweighed, if $\delta = 1$, by an infinite number of periods in which π_i^2 is earned instead of π_i^2 – and by continuity, the gain is outweighed for some $\delta < 1$ as well. Player i has no incentive to deviate in (b) and (c) because E^1 and E^2 were chosen to be subgame perfect equilibria.

Proof of the Two-Equilibrium Theorem is very similar, although without the Dimensionality Condition one cannot invoke the Fudenberg–Maskin Theorem. But the Two-Equilibrium Condition does specify that there exist equilibria E^1 in which $\bar{a}$ is played and E^{2i} in which $\underline{a}^i$ is played, with $\pi_i^1 > \pi_i^{2i}$. The argument then proceeds as in the previous paragraph, but with the E^2 in phase (c) replaced by the E^{2i} appropriate to player i.

4. COMPARING THE THEOREMS

The Folk Theorem tells us that any pattern of behavior observed over a finite number of periods of a supergame can be supported by some equilibrium. Nothing can be excluded. What use is such a negative result? One use is to warn that models with an infinite number of periods can be misleading. It is not enough simply to find the perfect equilibrium of such a game; the equilibrium must be supported by further arguments, since the equilibrium is far from unique. The theorem's value is to provoke close scrutiny of infinite horizon models so that a modeller cannot escape showing not only that the equilibrium he favors exists, but why it is better than the multitude of other equilibria. He must go beyond satisfaction of the standard technical criteria.

The Fudenberg–Maskin Theorem characterizes the entire game, but the characterization says nothing about any finite part of the game. A given average payoff can be generated by any of a wide array of actions, because the action in any one period has an imperceptible effect on the average payoff. The Dimensionality and Two-Equilibrium Theorems, on the other hand, although they characterize any finite part of the game, say little about the payoffs. Since only finite amounts of time are observable, the modeller may care only about a subset of the periods of play, even if the player has a low enough discount rate that such a subset is unimportant to him. One of the most important lessons of the Folk Theorem is that some models are empty because they have no observable implications. Since average payoffs of an infinite game are never observed, previous versions of the Folk Theorem have not really made this point.

It is also important to realize that the first T periods of an equilibrium might look nothing like the rest of it. A casual reading of the Fudenberg–Maskin Theorem and the Stationary Actions Theorem might conclude that although any payoff in W^* is possible, the players will at least choose actions that guarantee them their minimax payoffs in every period. One might also wonder why any but the Stationary Actions Theorem is needed, since it characterizes both the entire game and every part of the game. But it considers only stationary strategies, just as the Fudenberg–Maskin Theorem only considers average payoffs. While it justifies a wide class of equilibria – stationary equilibria with payoffs exceeding the minimax level – there exist other equilibria which are nonstationary. In particular, the Dimensionality and Two-Equilibrium Theorems justify equilibria with payoffs below the minimax level in some periods. These equilibria are nonstationary, and the low payoffs are only transitory, so they are excluded from consideration by the first two theorems.

All four theorems specify that the game be infinitely repeated, not finitely repeated many times. For some games, similar results can be obtained for finite repetitions. This is not true of all games: Moreaux (1985) has shown that if a one-shot game has just one Nash equilibrium, then those actions form the unique perfect equilibrium outcome of the finitely repeated game. But Benoit and Krishna (1985) show that when there is no discounting and the one-shot game has multiple equilibria, the repeated game has many equilibrium outcomes. The reasoning is similar to the Two-Equilibrium Theorem above: when a game has two equilibria, other equilibria can be constructed using the differential between the two fundamental equilibria as a punishment for deviation.

4.1. The Assumptions

Having discussed the theorems, it will now be easier to explain why the various assumptions are important. Allowing discounting is important to show that there is no discontinuity at the discount factor of one. If a particular undiscounted game has a large number of equilibrium outcomes, the theorems say that any outcome can survive adding a little discounting to the model. True, the equilibria are discontinuous at some discount factor less than one (as will be seen in the Prisoner's Dilemma in Section 5). But this is acceptable because a repeated game with heavy discounting should have the same properties as the one-shot game, and some discount factor must divide heavy from non-heavy. A little discounting, however, seems to be unimportant, in strong contrast with making the number of repetitions infinite instead of merely large. The unimportance of a little discounting also contrasts with discounting's importance in bargaining games, which can have a continuum of perfect equilibria with zero discounting, but a unique equilibrium with a small amount (Rubinstein, 1982).

A constant probability of the game ending at each repetition has the same effect as time preference because it makes present payments preferable to future payments. A difference in interpretation is that the ending probability ensures that the game ends in finite time with probability one, or, phrased less dramatically, that the expected number of repetitions is finite.[8] The game behaves like a discounted infinite game because the expected number of future repetitions never diminishes, no matter how many have already occurred. The game still has no 'last period'. Imposing a last period, no matter how far beyond the expected number of repetitions, would still radically change the game.[9] Game G_1 that lasts 8 periods is very different from Game G_2 that lasts an *expected* 8 periods because of an ending probability of about 0.17 per period; after 7 periods, G_1 is a 1-period game, while G_2, if it has not yet ended, is still expected to last 8 more periods.

All of the theorems could be adapted to apply even if the rate of time preference and the ending probability varied over time, so long as both remained sufficiently small. If they do vary, simply take ρ and ω to be the maximum values of the discount rate and the ending probability. The proofs hold *a fortiori*: if the maximum values are small enough, any outcome is possible; otherwise, further computation is necessary to see whether the appropriate average future rates are small enough. The point is especially important with respect to the ending probability: if the ending probability approaches one over time (as we might expect of a mortality rate), the Folk Theorem fails to apply, even if the probability never reaches one.

In some games, the question arises of whether correlated strategies matter to the Folk Theorem. A player employing a correlated strategy makes his action contingent on a signal which some or all players can observe (Aumann, 1974). Such signals matter because they may allow players to coordinate more effectively in minimaxing a deviator. This affects the Fudenberg–Maskin and Stationary Actions Theorems because the feasible payoff set, W^*, depends on the minimax strategies. Disallowing correlated strategies would also restrict the applicability of the Dimensionality and Two-Equilibrium Theorems, but in a different way. Those theorems employ pure strategies for the first T periods, but the use of correlated strategies can increase the severity of the out-of-equilibrium punishments, so it affects the number of periods, T, and the discount factor, δ, for which a particular pattern of actions can be supported in equilibrium. All the theorems remain valid if correlated strategies are disallowed, but their range of applicability shrinks.

5. AN EXAMPLE: THE PRISONER'S DILEMMA

The Prisoner's Dilemma illustrates many of these points. Let us use the payoffs shown in Figure 2. The minimax payoff is 0, which each player can guarantee himself by playing *Fink*, and the maximum possible payoff is 8. The game also allows worse-than-minimax payoffs: a player who chooses *Cooperate* when the other player chooses *Fink* receives -2.

The Prisoner's Dilemma is a good example of the difference between a finite and an infinite number of repetitions. No matter how large the number of repetitions, if it is finite the outcome (*Fink*, *Fink*) is the unique perfect equilibrium outcome (and, in fact, the unique Nash outcome). The argument for this, older than Selten (1978) but known as the Chainstore Paradox after his article, is as follows. In the last period, *Fink* is the dominant strategy, since that subgame is identical to the one-shot game. In the two-period subgame starting in the next-to-last period, it is common knowledge that both players will fink in the last period, so cooperating has no reputational or other advantages, and both players fink. In the subgame starting yet one period earlier, both players know that present cooperation will not

		Column	
		Cooperate	Fink
	Cooperate	7, 7	−2, 8
Row			
	Fink	8, −2	0, 0

Payoffs to: (*Row*, *Column*).

Fig. 2. The Prisoner's Dilemma.

induce future cooperation, so they fink then also, and the argument can be carried back to the first period.

The argument for why (*Fink*, *Fink*) is the unique Nash outcome (not just the unique perfect outcome) is different. Subgames are unimportant to whether a strategy combination is a Nash equilibrium, but we can rule out successive classes of strategies from being Nash. No strategy in the class which calls for *Cooperate* in the last period can be Nash, because the same strategy except with *Fink* in the last period would dominate it. But if both players have strategies calling for finking in the last period, then no strategy which does not call for finking in the next-to-last period is Nash, because the same strategy except for *Fink* in the next-to-last period would dominate it. The argument can be carried back to the first period, ruling out any class of strategies that does not call for finking everywhere along the equilibrium path.[10]

The finitely repeated Prisoner's Dilemma thus has a simple set of perfect equilibria. The theorem of Benoit and Krishna (1985) mentioned earlier fails to apply, because the one-shot Prisoner's Dilemma has a unique Nash equilibrium; the theorem of Moreaux (1985) therefore does apply, which is another way to know that (*Fink*, *Fink*) is the unique perfect equilibrium of the finitely repeated game.

The infinitely repeated Prisoner's Dilemma has a large set of equilibria. To apply the four theorems presented for infinite games, we must see which conditions are satisfied. The Dimensionality Condition is satisfied because convexifying the four payoff combinations (to allow for mixed strategies) generates a two-dimensional space of which W^* in Figure 1 is a subset. The Two-Equilibrium Condition is satisfied because Pareto-ranked equilibria can be easily constructed for the repeated game. One equilibrium is repetition of (*Fink*, *Fink*), since any Nash equilibrium of a one-shot game is a perfect equilibrium of the infinitely repeated game. A second equilibrium strategy is (*Cooperate until someone finks*; *then always fink*), which is known as the grim strategy. The grim strategy supports an equilibrium with cooperation as the outcome, so its payoff combination of $(7,7)$ Pareto-dominates the payoff combination of $(0,0)$ from *Always Fink*. Hence, the Two-Equilibrium Condition is satisfied.

The grim strategy does not support an equilibrium unless the discount factor is close enough to one. We can calculate the critical value of the discount rate for the parameters of this game. The equilibrium payoff is the current payoff of 7 plus the value of the rest of the game, which is $7/r$. If Row deviates by finking, he receives a current payoff of 8, but the value of the rest of the game falls to $0/r$. The critical value of the discount rate at which finking becomes profitable is found by solving the equation $7+7/r=8+0/r$, which yields $r=7$ and $\delta=0.125$. With the payoffs of Figure 2, finking is not tempting at moderate discount rates.[11]

Since both the Dimensionality Condition and the Two-Equilibrium Condition are satisfied for discount factors between 0.125 and 1, all four Folk Theorems apply to the Prisoner's Dilemma. The Fudenberg–Maskin Theorem says that any average payoff in the darkly shaded area in Figure 1 (excluding the axes, but including the outer boundary) can be achieved in a perfect equilibrium. As explained earlier, this says nothing about the particular actions, because a given average payoff can be achieved via many different combinations of *Fink* and *Cooperate*.

The Stationary Actions Theorem does describe possible equilibrium actions over the entire game and any T-period subset. It implies, for example, that there exists a perfect equilibrium in which both players always choose *Cooperate* and gain the payoff of 7. It is subject to the same limitations as the Fudenberg–Maskin Theorem, however; it only applies to strategies that generate payoffs lying in the darkly shaded area of Figure 1. Moreover, it only describes stationary strategies.

The Dimensionality and Two-Equilibrium Theorems characterize a particular T-period portion of the game, over which the outcome can be any pattern of *Fink* and *Cooperate*. Outside of those T periods, the outcome takes whatever form necessary to support the desired pattern. If, for example, the discount rate is zero, the theorems say that the pattern (*Fink*, *Fink*; *Cooperate*, *Fink*; *Fink*, *Cooperate*; *Fink*, *Cooperate*; *Cooperate*, *Fink*) is an equilibrium outcome for the first five periods – despite the fact that (*Fink*, *Cooperate*) yields a worse-than-minimax payoff to player 2. One equilibrium which generates this outcome is for the players to adopt the pattern's actions as their strategies for the first five periods, to cooperate in the succeeding periods, and to always fink if either player deviates. This equilibrium also shows that many equilibria can map to a single average payoff. The limiting average payoff is $(7, 7)$, which is identical to that for the grim strategy equilibrium. Under the grim strategy, both players always cooperate, but the difference in the first five periods is irrelevant to the limit of the average payoff as the number of periods becomes infinite.

None of these versions of the Folk Theorem say that any strategy can be part of an equilibrium, only that a wide range of actions and payoffs can be. The strategy Tit-for-Tat, for example, calls for a player to start by playing *Cooperate* and then in period t to imitate the other player's action of period $t-1$. The equilibrium outcome is (*Cooperate*, *Cooperate*) in each period, which is indeed supported by a perfect equilibrium, but Tit-for-Tat itself is not perfect for the prisoner's dilemma of Figure 2. Consider the out-of-equilibrium subgame in which player Row has finked and Column has cooperated in period $t-1$. Tit-for-Tat calls for alternating (*Cooperate*, *Fink*) with (*Fink*, *Cooperate*), a pattern which generates low payoffs for both players. But player Column would gain by deviating from Tit-for-Tat, ignoring Row's *Fink* so that the actions would return to (*Cooperate*, *Cooperate*) in every period, which has higher payoffs. Tit-for-Tat is therefore not perfect, even though it would support (*Cooperate*, *Cooperate*) if it were.[12]

Although the outcome (*Fink*, *Fink*) for every repetition can be generated by a perfect equilibrium, this is also unimplied by any of the theorems. The Fudenberg–Maskin Theorem does not characterize actions, and even if it did, neither it nor the Stationary Actions Theorem apply to equilibria with average payoffs of zero, which is outside of W^*. The Dimensionality and Two-Equilibrium Theorems are not so limited, but they only say that (*Fink*, *Fink*) could be observed for T periods. Thus, the outcome of infinitely repeated (*Fink*, *Fink*) must be justified by construction.

6. SUMMARY

During a finite portion of an infinitely repeated n-player game, a great variety of action patterns can be observed in perfect equilibrium. This is true even if payoffs are discounted and the game has some probability of ending after every period. In the new theorems presented here, one of two technical conditions must be satisfied: the modeller must either discover at least two perfect equilibria by construction, or show that the payoff space has the dimensionality of the number of players. This done, the theorems tell him that the equilibrium can be characterized by a large number of stationary action combinations or an even larger number of action combinations for a limited number of periods.

NOTES

* I would like to thank an anonymous referee, Sushil Bikhchandani, David Hirshleifer, David Levine, Thomas Voss, and participants in the UCLA Game Theory Seminar for helpful comments.

[1] Aumann (1981) is an earlier survey, which contains the Aumann–Shapley (1977) proof of a perfect equilibrium Folk Theorem. More recent papers include Abreu (1988), Benoit and Krishna (1985, 1987), Cave (1987), Friedman (1984), and Guth, Leininger, and Stephan (1988). Chapters 4 and 5 of Rasmusen (1989) discusses applications.

[2] Another category of results is the extension of the Folk Theorem to more complicated kinds of repeated games such as finitely repeated games of incomplete information (Fudenberg and Maskin, 1986) and games with a state variable which changes from repetition to repetition (Lockwood, 1988).

[3] One version not stated in terms of payoffs is in the unpublished Rubinstein (1977), which applies to stationary strategies without discounting.

[4] The finiteness of M_i avoids the technical problems of working with the mixed extensions of games with continuous action sets. Finiteness also helps ensure existence of equilibrium in the one-shot game.

[5] These are not necessary conditions. Fudenberg and Maskin note that in two-player games, for example, neither the Dimensionality nor the Two-Dimension condition need be satisfied for their theorem's conclusion to be valid.

[6] Note that the Two-Equilibrium Conditions also implies that the difference between the two payoffs increases as the discount rate declines, since the per-period payoffs are constant. Hence, as the discount rate declines, the punishment increases.

[7] This is Theorem 2 in Fudenberg and Maskin (1986, p. 544).

[8] It does not matter if lump sum payments are made to players whenever the game ends. They have no control over the ending time, so such payments are irrelevant to their behavior unless there are income effects.

[9] The similarity of infinite models to stationary possibly-ending models is also well-known in the context of speculative bubbles (Blanchard, 1979).

[10] Although *Always fink* is a Nash strategy for the finitely repeated game, it is not necessarily a dominant strategy, except in the one-shot game. A dominant strategy is a strictly best response to any other strategy, but *Always Fink* might not (depending on the discount factor) be the best response to various non-Nash strategies such as (*Cooperate until the other player finks, then fink for the rest of the game*).

[11] The critical discount rate does not depend on Row's one-shot payoff when he cooperates and Column finks, a value of −2 in Figure 2. That number is irrelevant to Row's decision as to whether or not to deviate from the cooperative equilibrium, because it cannot be reached if Row alone deviates.

[12] Tit-for-tat is not perfect for the Prisoner's Dilemma with the parameters of Figure 2, but for very special discount rates it can be perfect in Prisoner's Dilemmas with other parameters.

REFERENCES

Abreu, Dilip: 1988, 'On the theory of infinitely repeated games with discounting', *Econometrica*, **56**, 383–396.

Aumann, Robert: 1974, 'Subjectivity and correlation in randomized strategies', *Journal of Mathematical Economics*, **1**, 67–96.

Aumann, Robert: 1981, 'Survey of repeated games', in *Essays in Game Theory and Mathematical Economics in Honor of Oscar Morgenstern*, Robert Aumann (ed.), Mannheim: Bibliographisches Institut.

Aumann, Robert and Shapley, Lloyd: 1977, 'Long-term competition – a game theoretic analysis', mimeo, February.

Benoit, Jean-Pierre and Krishna, Vijay: 1985, 'Finitely repeated games', *Econometrica*, **53**, 905–22.

Benoit, Jean-Pierre and Krishna, Vijay: 1987, 'Nash equilibria of finitely repeated games', *International Journal of Game Theory*, **16**, 197–204.

Blanchard, Olivier: 1979, 'Speculative bubbles, crashes, and rational expectations', *Economics Letters*, **3**, 387–389.

Cave, Jonathan: 1987, 'Equilibrium and perfection in discounted supergames', *International Journal of Game Theory*, **16**, 15–41.

Friedman, James: 1984, 'Trigger strategy equilibria in finite horizon supergames', mimeo, University of North Carolina.

Fudenberg, Drew and Maskin, Eric: 1986, 'The folk theorem in repeated games with discounting and with incomplete information', *Econometrica*, **54**, 533–54.

Guth, Werner, Leininger, Wolfgang, and Stephan, Gunter: 1988, 'On supergames and folk theorems: A conceptual discussion', Game Theory Working Paper No. 19, University of Bielefeld, October.

Kreps, David, Milgrom, Paul, Roberts, John, and Wilson, Robert: 1982, 'Rational cooperation in the finitely repeated prisoner's dilemma', *Journal of Economic Theory*, 1988, **27**, 245–52.

Lockwood, Ben: 1988, 'The folk theorem in dynamic games with and without discounting', Birkbeck College mimeo, May.

Moreaux, Michel: 1985, 'Perfect Nash equilibria in finite repeated game and uniqueness of Nash equilibrium in the constituent game', *Economics Letters*, **17**, 317–20.

Rasmusen, Eric: 1989, *Games and Information*, Oxford: Basil Blackwell.

Rubinstein, Ariel: 1977, 'Equilibrium in supergames', Hebrew University CRMEGT Research Memorandum No. 25, May.

Rubinstein, Ariel: 1979, 'Equilibrium in supergames with the overtaking criterion', *Journal of Economic Theory*, **21**, 1–9.

Rubinstein, Ariel: 1982, 'Perfect equilibrium in a bargaining model', *Econometrica*, **50**, 97–110.
Selten, Reinhard: 1978, 'The chain store paradox', *Theory and Decision*, **9**, 127–59.

John E. Anderson Graduate School of Management
University of California at Los Angeles
405 Hilgard Avenue
Los Angeles, CA 90024, U.S.A.