Econometrica, Vol. 56, No. 2 (March, 1988), 383–396

ON THE THEORY OF INFINITELY REPEATED GAMES WITH DISCOUNTING

BY DILIP ABREU[1]

This paper presents a systematic framework for studying infinitely repeated games *with discounting*, focussing on pure strategy (subgame) perfect equilibria. It introduces a number of concepts which organize the theory in a natural way. These include the idea of an *optimal penal code*, and the related notions of *simple penal codes* and *simple strategy profiles*.

I view a strategy profile as a rule specifying an initial path (i.e., an infinite stream of one-period action profiles), and *punishments* (also paths, and hence infinite streams) for any deviations from the initial path, or from a previously prescribed punishment. An arbitrary strategy profile may involve an infinity of punishments and complex history-dependent prescriptions.

The main result of this paper is that much of this potential strategic complexity is redundant: every perfect equilibrium path is the outcome of some perfect *simple strategy profile*. A simple strategy profile is *independent of history* in the following strong sense: it specifies the *same* player specific punishment after *any* deviation by a particular player. Thus simple strategy profiles have a parsimonious description in terms of $(n+1)$ paths where n is the number of players. Unlike the undiscounted case there is no need to "make the punishment fit the crime." In particular, a player who has a "myopic" incentive to deviate from his own punishment may be deterred from doing so simply by restarting the punishment already in effect.

The key to the above result is that, with discounting, *worst* perfect equilibria *exist* for each player. These define an *optimal penal code*. The notion of a *simple penal code* yields an elementary proof of the existence of an optimal penal code and leads directly to the theorem on the "sufficiency" of simple strategy profiles.

KEYWORDS: Repeated games, discounting, subgame perfect equilibrium, simple strategy profiles, optimal punishments, simple penal codes.

1. INTRODUCTION

THIS PAPER PROVIDES an analytical framework for infinitely repeated games with discounting, introducing a number of concepts which organize the theory in a natural way. I consider, in particular, the set of pure strategy subgame perfect equilibria (see Selten (1965, 1975)) of such games. Progress in understanding these equilibria has been impeded by the fact that, in principle, they may be extremely complex. My results provide a major dimensional simplification: they show that every pure strategy (subgame) perfect equilibrium path is the outcome of some perfect *simple strategy profile*. Such strategy profiles have a very

[1] This paper is based on Section 2 of Abreu (1982), which was also incorporated into my doctoral dissertation at Princeton University. I wish to thank my supervisors, Hugo Sonnenschein and Bobby Willig, for much help and advice. Assistance and comments from Bob Anderson, Jacques Crémer, Andreu Mas-Colell, and Ennio Stacchetti proved very valuable. I enjoyed discussing this research with, and benefited greatly from the suggestions of, among others, Ed Green, Christopher Harris, Vijay Krishna, Eric Maskin, Roger Myerson, and Ariel Rubinstein. I am also grateful to David Kreps and two anonymous referees for detailed and useful comments. Financial support from Princeton University, the Sloan Foundation, and the IMA, University of Minnesota is gratefully acknowledged. Finally I wish to record a special debt to my colleague David Pearce for his substantial help and encouragement at all stages of this research. Of course, errors remain my own.

elementary structure. Moreover, they are extremely tractable; it may be easily checked whether or not they are perfect.

The theory developed here builds on the seminal work of Aumann and Shapley (1976) and Rubinstein (1979) on infinitely repeated games *without discounting*. In particular, it borrows from their work the fundamental idea of punishing a player for not participating in the punishment of another player. In other respects, however, their techniques do not extend; the latter depend in an essential way upon the possibility of trading future losses one-for-one against present gains. With the introduction of discounting, intertemporal trade-offs are more subtle, and detailed shapes of punishments become important. While discounted repeated games are delicate in this respect, realism demands that they be investigated. Indeed, in most economic applications, the assumption of a zero interest rate is inappropriate; we are typically concerned with situations in which the future is less important than the present.

Much of the analysis of this paper is couched in terms of outcome *paths* or *punishments*. A path (or punishment) is an infinite stream of one-period action *profiles*. I view a strategy profile as a rule specifying (or prescribing) an initial path and punishments for any *deviation* from the initial path, or from a previously prescribed punishment. *The remarks to follow are to be understood in terms of this definition*. It is equivalent to the standard one (see Section 4) but the appropriate formulation for the perspective taken here.

The incentive constraints implicit in a *perfect* equilibrium may be quite subtle. For instance a player who is being punished for a deviation may have "myopic" incentives to cheat (i.e., the action stipulated by the equilibrium may not be a single-period best response, in one or more stages of the punishment (path)). If so, perfection requires that he be deterred from cheating by the threat of a *further* punishment. This problem reappears at the next level and one is led into a *hierarchy* of punishments. Reflecting this complexity, an arbitrary strategy profile may be contingent on history in essential and elaborate ways. It may involve an infinity of punishments and a complicated rule specifying which punishment is imposed for any particular deviation from the initial path or from an ongoing punishment.

Much of this potential complexity is in fact redundant. As indicated earlier we need only consider *simple strategy profiles*. These are simple in that they are history-independent in the following strong sense: they specify the *same* punishment Q^i for *any* deviation, after any (previous) history, by player i. They are thus defined by the initial path, n player-specific punishments, where n is the number of players, and the simple rule described above.

A central concept of this paper is the notion of an *optimal penal code*. An optimal penal code is an n-vector of perfect strategy profiles, the ith strategy profile of which yields the ith player at least as low a payoff as does any other perfect equilibrium. Optimal penal codes provide a simple criterion to determine whether a path is the outcome of a perfect equilibrium; their existence underlies the results obtained here. A *simple penal code* is an n-vector of strategy profiles defined by an n-vector of punishments $(Q^1, \ldots, Q^n)$. The initial path of the kth strategy profile is Q^k, and Q^i is imposed if player i deviates (singly) from any ongoing punishment Q^j. The notion of a simple penal code permits an elementary proof of the existence of an optimal penal code. The key result is that there exists a simple penal code which is an optimal penal code. The sufficiency of simple strategy profiles follows as an easy consequence.

The assumption that payoffs are discounted is an essential element of the proof that an optimal penal code exists. Without discounting, existence is not assured: a player may be minmaxed (i.e., forced down to his one-period individually rational payoff) for T periods for any finite T, but possibly not forever. This nonexistence is related to the fact that the Aumann-Shapley-Rubinstein strategies involve punishments which are history-dependent and whose severity depends on the pattern of previous deviations.

My results make it unnecessary to contemplate complex hierarchies of punishments; in no sense is there any need to "make the punishment fit the crime." Optimality might demand that a deviant "cooperate" in his own punishment (see Abreu (1986)). The theorem on simple penal codes implies that he may be persuaded to do so simply by threatening him with the punishment already in effect. This appears paradoxical. How can a player be deterred from cheating when he is already being punished as severely as possible? The appropriate resolution is that in such situations, the early stages of an optimal punishment must be more unpleasant than the remainder. The results do not imply the absence of subtle interactions. Rather, they clarify that all relevant complexity resides entirely in the structure of the n punishments $(Q^1, \ldots, Q^n)$.

I present my results for *pure* strategy equilibria of the repeated game. All *mixed* strategy equilibria could be obtained within the framework developed here if players in the stage game have access to randomizing devices the outcomes of which become common knowledge concurrently with the players' actions. In this case a player's strategy set in the stage game could simply be taken to be his set of mixed strategies. If randomizing devices are observed with a lag, or not at all, my results do not extend, as they depend on deviations being detected with certainty before the next period's play. *Unobservable* mixed strategies lead to a significant change in the nature of the analysis, and this presents a very attractive area for further research. It should be understood in what follows that the equilibria to which we refer are pure strategy equilibria, though the theory also encompasses the "observable" mixed-strategy case.[2]

Earlier work on repeated games with discounting (see, for instance, the classic paper by Friedman (1971)) has concentrated on paths supportable by Cournot-Nash punishments, i.e., players revert to single-period Cournot-Nash equilibrium behavior forever if a deviation occurs. Cournot-Nash reversion, while subgame perfect, is not in general optimal, and will therefore only suffice to support a limited range of perfect equilibrium behavior. The theory developed here is

[2] I would like to thank Bob Anderson for first bringing up the issue of mixed strategies and pointing out that these might lead to more severe punishments. A referee stressed the relevance of "observable" mixed strategies.

relevant only when one is attempting to design perfect penal codes more severe than Cournot-Nash reversion. Such an exercise is critical for any study of *extremal* equilibria. Examples of the latter include *maximal* collusion in a repeated oligopolistic game or *efficient* contracts in repeated principal-agent problems.

The next section provides a simple example which illustrates some of the main ideas. The reader might find it useful to return to it after reading Sections 3–5, which contain the formal analysis. Section 6 concludes.

2. AN EXAMPLE

Consider the two-person simultaneous game below:

			Player 2		
			L	M	H
G_1	Player 1	L	10, 10	3, 15	0, 7
		M	15, 3	7, 7	-4, 5
		H	7, 0	5, -4	-15, -15

Roughly speaking, G_1 may be thought of as a symmetric, discrete, quantity-setting duopoly game in which each firm may choose a "low," "medium," or "high" output level. It has a unique Nash equilibrium (M, M).

Denote by $G_1^\infty(\delta)$ the infinitely repeated game associated with the stage game G_1, where δ is the discount factor used to evaluate payoffs. Suppose one wished to support the "collusive" outcome (L, L) in a perfect equilibrium of the repeated game. The standard way to attempt to do so is to revert to the one-shot Nash equilibrium forever if a deviation occurs. More generally, let σ^i be a perfect equilibrium of $G_1^\infty(\delta)$, which we would like to think of as a "punishment" for player i. Now, one could support (L, L) if the gain from cheating were outweighed by the loss associated with reverting to σ^i rather than continuing with (L, L), i.e., if

$$5 \leqslant \frac{\delta}{1-\delta}10 - \tilde{v}_i(\sigma^i) \qquad (i = 1, 2)$$

where $\tilde{v}_i(\sigma^i)$ is the repeated game payoff to player i from the strategy profile σ^i. (Notice that first period payoffs are discounted.) As may be checked, for $\delta < 5/8$, Cournot-Nash reversion is not severe enough to support (L, L). Fix $\delta = 4/7$. The lower are $\tilde{v}_1(\sigma^1)$ and $\tilde{v}_2(\sigma^2)$, the easier it is to support "collusion," and one might hope that the *optimal penal code* $(\underline{\sigma}^1, \underline{\sigma}^2)$ will be severe enough to do so.

My results imply that in looking for optimal penal codes it suffices to restrict attention to *simple penal codes*, as described in the introduction (see also Definitions 1 and 3 of Sections 4 and 5 respectively). For a two player game, such penal codes are defined by a pair of outcome paths (Q^1, Q^2), and this restriction may be exploited together with the particular structure of the one-shot game to determine the optimal punishment paths explicitly.

In our example it turns out that

$$\underline{Q}^1 = \{(M, H), (L, M), (L, M), \ldots\},$$
$$\underline{Q}^2 = \{(H, M), (M, L), (M, L), \ldots\},$$

define an optimal simple penal code $(\underline{\sigma}^1, \underline{\sigma}^2)$. In words, the play described by $\underline{Q}^1$ is: play (M, H) for one period (i.e., player 1 plays M and player 2 plays H) followed by (L, M) thereafter. Similarly for $\underline{Q}^2$: play (H, M) for one period followed by (M, L) thereafter. The strategy profile $\underline{\sigma}^1$ specifies that play proceeds according to $\underline{Q}^1$ until some player i deviates singly from this arrangement. If player 1 deviates, his "punishment path" $\underline{Q}^1$ is restarted: (M, H) is stipulated in the period after the deviation, followed by (L, M) thereafter. If player 2 deviates, $\underline{Q}^2$ is imposed following the deviation. In general the response to any deviation (by player i alone) from whatever path is in force is to impose $\underline{Q}^i$ starting the period after the deviation.[3] The description of $\underline{\sigma}^2$ is identical except that we now start with $\underline{Q}^2$.

Note that when player 1 is being punished, if he doesn't play M in the first round, player 2 will play H following the deviation and will continue to do so until player 1 plays M. Only by "taking his medicine" and playing M can player 1 move play to the more attractive part of the path $\underline{Q}^1$, namely, rounds $2, 3, \ldots$. Why does player 2 play H when he would rather play M against M by player 1? Because if he does not do so, the regime will change to $\underline{Q}^2$ and player 2 will himself be punished. And so on. These remarks are meant to convey the flavor of the rather subtle incentive structure of these equilibria. To pin things down exactly, Proposition 1 of Section 4 is very helpful—perfection may be verified by checking "one-shot" deviations alone (i.e., deviations followed by conformity with the strategy profile in question). It is useful now to note that $\tilde{v}_i(\underline{\sigma}^i) = 0$, i.e., the path $\underline{Q}^i$ yields player i a payoff of zero (when $\delta = 4/7$ which was assumed earlier). Thus a one-shot deviation is followed by a post-deviation payoff of zero. This turns out to be a sufficient deterrent. For instance, suppose $\underline{Q}^1$ is in force and the players are in the first period of the path. If player 2 deviates (optimally) he gets a payoff of $7\delta = 4$ (I discount first period payoffs) today and a future payoff of zero; if he conforms he gets 5 today and 15 thereafter, i.e., a present discounted value of $(100/7) > 4$. If, on the other hand, player 1 deviates optimally, he gets zero today and a zero payoff in the future which, as noted earlier, is exactly his payoff from conforming. The other cases may be checked directly. Thus $\underline{\sigma}^i$, $i = 1, 2$ is a perfect equilibrium. Since player i can guarantee himself a payoff of zero by playing L in all contingencies, he cannot receive a lower payoff in any perfect equilibrium. Hence $(\underline{\sigma}^1, \underline{\sigma}^2)$ is an optimal penal code. Finally, note that $(\underline{\sigma}^1, \underline{\sigma}^2)$ indeed supports (L, L).

[3] Simultaneous deviations do not affect future play, that is, they go unpunished. See Section 4 for a discussion of why this is appropriate.

I have developed a discrete example with numbers that "work" for expositional reasons. In a continuous setting one could construct optimal simple penal codes using outcome paths with the same structure as $\underline{Q}^1, \underline{Q}^2$ above, for a neighborhood of values of δ. Now M and H would be defined as functions of δ. For an extensive application of this paper in conventional continuous-action oligopoly models, see Abreu (1986).

3. NOTATION AND DEFINITIONS

The notation and definitions presented below are adapted from Rubinstein (1979).

The Stage Game

The stage game is denoted $G = (\{S_i\}_{i=1}^n; \{\pi_i\}_{i=1}^n)$ where $N = \{1, \dots, n\}$ is the set of players, S_i is a *pure* strategy set for player i, and π_i: $S_1 \times S_2 \times \cdots \times S_n \to R$ is player i's payoff function. Assume that S_i contains at least two elements. Elements of S_i are denoted q_i and are referred to as *actions*. Set $S \equiv S_1 \times S_2 \times \cdots \times S_n$, $q \equiv (q_1, q_2, \dots, q_n)$, and $\pi \equiv (\pi_1, \pi_2, \dots, \pi_n)$. Note that whenever player-subscripted symbols are used, the corresponding unsubscripted symbol refers to a Cartesian product or a vector, depending on context. I assume that G is *simultaneous* in order to abstract from problems of perfection in the stage game.

The Repeated Game

Let $G^\infty(\delta)$ denote the supergame with discounting obtained by repeating G infinitely often, and evaluating payoffs in terms of the discount factor $\delta \in (0,1)$. A *pure* strategy for player i is denoted σ_i. Each σ_i is a sequence of functions $\sigma_i(1), \sigma_i(2), \dots, \sigma_i(t), \dots,$ one for each period t. The function for period t determines player i's action at t as a function of the actions of *all* players in all previous periods. Formally, $\sigma_i(1) \in S_i$ and for $t = 2, 3, \dots, \sigma_i(t)$: $S^{t-1} \to S_i$. Player i's strategy set is denoted Σ_i, and $\Sigma \equiv \Sigma_1 \times \Sigma_2 \times \cdots \times \Sigma_n$ denotes the set of strategy profiles.

A *stream* of action *profiles* $\{q(t)\}_{t=1}^\infty$ is referred to as a *path* or *punishment* and is denoted by Q. Let $\Omega \equiv S^\infty$ denote the set of paths. Any strategy *profile* $\sigma \in \Sigma$ generates a path denoted $Q(\sigma) = \{q(\sigma)(t)\}_{t=1}^\infty$, and defined inductively as follows: $q(\sigma)(1) = \sigma(1)$, and $q(\sigma)(t) = \sigma(t)(q(\sigma)(1), \dots, q(\sigma)(t-1))$. Let v_i: $\Omega \to R$ define the ith player's payoff from a path $Q \in \Omega$: $v_i(Q) = \sum_{t=1}^\infty \delta^t \pi_i(q(t))$, and let $\tilde{v}_i$: $\Sigma \to R$ be the ith player's payoff function: $\tilde{v}_i(\sigma) = v_i(Q(\sigma))$. Note that I am discounting to the beginning of period 1, and that period t payoffs are received at the end of period t. I assume that all players have the same discount factor; this assumption is made for notational convenience and plays no role in the analysis.

Let $H=(q(1),\ldots,q(t))\in S^t$ denote an arbitrary t-period history, and $\sigma|_H\in\Sigma$ the strategy profile induced by σ on the subgame following H. For definitions of Nash equilibrium and (subgame) perfect equilibrium, see Rubinstein (1979).

I assume *throughout* that the set of payoffs of the stage game is bounded:

ASSUMPTION 1: $\{\pi(s)|s\in S\}$ *is bounded.*

4. SIMPLE STRATEGY PROFILES

This section analyzes a particularly simple class of strategy profiles, *simple strategy profiles*. Before giving a definition, I discuss how strategy profiles may be described in terms of paths and *deviations* from previously specified paths. In general such a formulation might be unnecessarily cumbersome; for simple strategy profiles, however, it is the natural way to proceed.

Implicit in any strategy profile is a notion of initial and subsequent deviations from behavior specified by σ. In the absence of deviations, conformity with σ results in the infinite sequence of action profiles given by the path $Q(\sigma)$. Any particular deviation (different players might deviate, at distinct times and to varying extents) from $Q(\sigma)$ leads into a particular subgame; σ induces a strategy profile and, therefore, a path on this subgame. A subsequent deviation from this secondary path is possible; this leads to another subgame, another induced strategy profile and another path which becomes the new standard with respect to which conformity with σ is determined; and so on, for higher order deviations.

Thus a strategy profile may be viewed as a rule specifying: (i) an initial path $Q(\sigma)\in\Omega$; (ii) paths $Q''\in\Omega$ after every particular deviation from an ongoing path $Q'\in\Omega$, depending on all that came before that ongoing path.

A noninitial path may be thought of as a "punishment." It is imposed for deviating from an ongoing path. The latter may be the initial path or itself a punishment.

Observe that Q'' may depend not only on Q' and the *particular* deviation from it, but more generally on the entire history of previously prescribed paths or punishments and deviations from them. Indeed, the punishments which the Aumann-Shapley-Rubinstein strategies employ are history dependent in a non-trivial way. Punishments are tailored to "fit the crime," and the argument that such strategies are perfect depends critically on this construction.

A simple strategy profile on the other hand, is completely defined by an $(n+1)$-vector of paths $(Q^0,Q^1,\ldots,Q^n)$ and a very simple rule. The initial path is Q^0, and each Q^i, $i\in N$, is a player-specific punishment. Any deviation by player i *alone* from any ongoing prescribed path (the initial path or one of the n punishments) is responded to by imposing Q^i. Simultaneous deviations are ignored (i.e., go unpunished). Since the equilibrium notions employed in this paper are noncooperative, it is sufficient to restrict attention to the deterrence of uncoordinated (among players) deviations: punishments designed for groups of simultaneously defecting players are irrelevant.

DEFINITION 1: Let $Q^i \in \Omega$, $i = 0, 1, \ldots, n$. The *simple strategy profile* $\sigma(Q^0, Q^1, \ldots, Q^n)$ specifies: (i) play Q^0 until some player deviates singly from Q^0; (ii) for any $j \in N$, play Q^j if the jth player deviates singly from Q^i, $i = 0, 1, \ldots, n$, where Q^i is an ongoing previously specified path; continue with Q^i if no deviations occur or if two or more players deviate simultaneously.

Note for later use that $\tilde{v}(\sigma(Q^0, Q^1, \ldots, Q^n)) = v(Q^0)$.

The notion of a *one-shot deviation* appears frequently below. A one-shot deviation from $\sigma \in \Sigma$ (by some player j), involves an initial deviation from $Q(\sigma)$ (by player j), followed by conformity with σ thereafter.

Given our interest in perfect equilibria, the "simplicity" of a strategy profile must be judged in terms of how easy it is to verify whether it is perfect. Perfection is in general a difficult criterion to apply because (i) we need to check that the strategy profiles induced by *all* histories are Nash equilibria; and (ii) it is a nontrivial exercise to determine whether any given strategy profile is a Nash equilibrium. In particular, for any player, both a single deviation from the initial path, and all conceivable patterns of successive deviations from subsequently prescribed paths, must not yield a higher payoff.

In terms of the perfection criterion, simple strategy profiles are very simple indeed. We have to consider only $(n+1)$ induced strategy profiles and need only check one-shot deviations from each of these. This is the content of Proposition 1. It asserts that the simple strategy profile $\sigma(Q^0, Q^1, \ldots, Q^n)$ is perfect if and only if no one-shot deviation by any player $j \in N$ from Q^i, $i = 0, 1, \ldots, n$, yields him a higher payoff, given that he and all other players will conform with Q^j after the deviation.

That one-shot deviations suffice follows from an important principle of dynamic programming. This is the criterion of "unimprovability." (See the seminal work of Howard (1960) and, for an account of more recent elaborations, Whittle (1983).) For arbitrary σ this criterion is still cumbersome to apply, as it requires that one-shot deviations from paths induced after *all* histories need to be checked. Simple strategy profiles are simple precisely because there are at most $(n+1)$ distinct paths to consider.

Why is Proposition 1 true? The argument is straightforward as the following verbal discussion suggests. Consider $\sigma(Q^0, Q^1, \ldots, Q^n)$ and suppose that no one-shot deviation by any player from any of the $(n+1)$ paths yields him a higher payoff. The way simple strategy profiles are constructed, any deviation by player $j \in N$ results in Q^j being imposed. Suppose that Q^j has just been imposed. Since Q^j is restarted after any deviation by j and since one-shot deviations do not yield a higher payoff, no finite sequence of deviations by j from Q^j will. However, given discounting and a uniform upper bound on payoffs (Assumption 1), if an infinite sequence of deviations is profitable, then a large enough finite sequence will be as well. Thus, under our hypothesis, player j cannot do better than conform with Q^j once it is imposed. But now it is clear that player j will conform with Q^i, $i = 0, 1, \ldots, n$, since a deviation from Q^i

results in Q^j being imposed, from which, we have just argued, no further deviation is profitable.

The following notation is convenient:

$$\alpha_j(q_j^*, q_{-j}) \equiv \pi_j(q_j^*, q_{-j}) - \pi_j(q) \quad \text{where } q \in S \text{, and}$$

$$q_{-j} \equiv (q_1, \ldots, q_{j-1}, q_{j+1}, \ldots, q_n); \quad \text{and}$$

$$v_j(Q; t+1) \equiv \sum_{s=1}^{\infty} \delta^s \pi_j(q(t+s)).$$

That is, $\alpha_j(q_j^*, q_{-j})$ is the change in player j's stage-game payoff when he plays q_j^* instead of q_j and the other players play according to q, and $v_j(Q; t+1)$ is the present discounted value of player j's payoffs in periods $t+1$ to ∞ along the path Q. The subscript "$-j$" is used consistently to denote a profile with a missing jth element. Paths and the action profiles of which they are comprised are associated in a natural way: $Q^i \equiv \{q^i(t)\}_{t=1}^{\infty}$, $\underline{Q}^i \equiv \{\underline{q}^i(t)\}_{t=1}^{\infty}$, etc. Proposition 1 can now be stated.

PROPOSITION 1: *Under Assumption* 1, *the simple strategy profile* $\sigma(Q^0, Q^1, \ldots, Q^n)$ *is a perfect equilibrium if and only if*

(1) $$\alpha_j\left(q_j^*, q_{-j}^i(t)\right) \leqslant v_j(Q^i; t+1) - v_j(Q^j)$$

for all $q_j^* \in S_j/q_j^i(t)$, $j \in N$, $i = 0, 1, \ldots, n$, *and* $t = 1, 2, \ldots$.

PROOF: Let σ denote $\sigma(Q^0, Q^1, \ldots, Q^n)$. From the point of view of player j, σ_{-j} defines a stationary discounted Markov decision problem with states $q_{-j}^k(t)$, $k = 0, 1, \ldots, n$, $t = 1, 2, \ldots$. Transition probabilities are given by:

$$\text{prob}\left(q_{-j}^k(t+1) | q_{-j}^k(t), q_j\right) = 1 \quad \text{if} \quad q_j = q_j^k(t)$$

and

$$\text{prob}\left(q_{-j}^j(1) | q_{-j}^k(t), q_j\right) = 1 \quad \text{if} \quad q_j \neq q_j^k(t).$$

Since σ_j specifies the action $q_j^k(t)$ in state $q_{-j}^k(t)$, the inequalities (1) assert that σ_j is unimprovable. Hence (see Theorem 2.1, Chapter 24, Whittle (1983), or Proposition 7, Chapter 6, Bertsekas (1976)) σ_j is optimal, i.e., $\sigma_j|_H$ is a best response to $\sigma_{-j}|_H$ for all histories $H \in S^{t'}$, $t' = 0, 1, 2, \ldots$. This establishes sufficiency. Necessity follows directly from the fact that all states are reachable after some history. *Q.E.D.*

5. OPTIMAL PENAL CODES

In this Section I define an *optimal penal code*, and establish my main result: There exists a *simple penal code* which is an *optimal penal code*. Thus optimal penal codes are shown to exist and are also characterized. An important implication of this theorem is that a path Q^0 is the outcome of a perfect equilibrium if and only if it is the outcome of some perfect simple strategy profile. Hence the perfect equilibrium paths of $G^{\infty}(\delta)$ may be completely analyzed in terms of simple profiles.

Let Σ^p denote the set of perfect equilibrium strategy profiles of $G^{\infty}(\delta)$. The associated sets of perfect equilibrium paths and payoffs are denoted $\Omega^p \equiv \{Q(\sigma) | \sigma \in \Sigma^p\}$ and $V = \{v(Q) | Q \in \Omega^p\}$, respectively.

ASSUMPTION 2: *Σ^p is nonempty.*

A simple condition which implies the above is that the one-shot game G has a pure strategy Nash equilibrium $q^e \in S$. Denote by σ^e the strategy profile according to which player i plays q_i^e after all histories. It may be easily checked that σ^e is a perfect equilibrium.[4]

DEFINITION 2: An *optimal penal code* is an n-vector of strategy profiles $(\underline{\sigma}^1, \ldots, \underline{\sigma}^n)$ such that for all i,

$$\underline{\sigma}^i \in \Sigma^p \quad \text{and} \quad \tilde{v}_i(\underline{\sigma}^i) = \min\{\tilde{v}_i(\sigma) | \sigma \in \Sigma^p\}.$$

Note that $\underline{\sigma}^i$ will in general be different from $\underline{\sigma}^j$, $j \neq i$; the "worst" equilibrium from one player's point of view need not be "worst" from another's.

Let $\underline{v}_i \equiv \inf\{\tilde{v}_i(\sigma) | \sigma \in \Sigma^p\}$. Of course, if an optimal penal code exists, $\underline{v}_i = \tilde{v}^i(\underline{\sigma}^i)$.

Optimal penal codes are important because they lead to a simple characterization of the set of perfect equilibrium paths; see Proposition 4 below.

I now define a simple penal code.

DEFINITION 3: Let $\sigma^i(Q^1, \ldots, Q^n) \equiv \sigma(Q^i, Q^1, \ldots, Q^n)$. The *simple penal code* *SPC* $(Q^1, \ldots, Q^n)$ is the n-vector of strategy profiles $(\sigma^1(Q^1, \ldots, Q^n), \ldots, \sigma^n(Q^1, \ldots, Q^n))$.

Notice that a simple penal code is defined by an n-vector of outcome paths (as opposed to $(n+1)$ for a simple strategy profile) and that the components of the n-vector of simple profiles which define a simple penal code differ only in the initial path they prescribe.

Simple penal codes, like simple strategy profiles, are appropriately named. In particular, by Proposition 1, every strategy profile of *SPC* $(Q^1, \ldots, Q^n)$ is perfect if and only if no one-shot deviation by any player j from Q^i, $i \in N$, yields him a higher payoff given that he and all other players will conform with Q^j after the deviation.

DEFINITION 4: *SPC* $(Q^1, \ldots, Q^n)$ is an *optimal simple penal code* if it is an optimal penal code.

[4] Weaker sufficient conditions which exploit the intertemporal structure of the problem could also be formulated. This issue is, however, best addressed in the context of the particular features of the one-shot game G in question, and is somewhat peripheral to the main objectives of the present paper.

This is offered to avoid legitimate confusion; an optimal simple penal code is optimal among the class of all penal codes, simple or otherwise.

Lemma 1 leads to the main theorem.

LEMMA 1: *If Q is a perfect equilibrium path, then*

$$\alpha_j\big(q_j^*, q_{-j}(t)\big) \leqslant v_j(Q; t+1) - \underline{v}_j \tag{2}$$

for all $q_j^* \in S_j$, $j \in N$, $t = 1, 2, \ldots$.

PROOF: Consider $\sigma \in \Sigma^p$ such that $Q(\sigma) = Q$. Let H^* denote the history $(q(1), \ldots, q(t-1), (q_j^*, q_{-j}(t)))$. Let σ_j^* denote the strategy $\sigma_j^*(s) = \sigma_j(s)$, $s \neq t$, $\sigma_j^*(t)(H) = q_j^*$, for all $H \in S^{t-1}$. Since σ is a Nash equilibrium,

$$\begin{aligned}&\delta^{-t}\big[\tilde{v}_j(\sigma_j^*, \sigma_{-j}) - \tilde{v}_j(\sigma)\big]\\ &\qquad = \alpha_j\big(q_j^*, q_{-j}(t)\big) - \big[v_j(Q; t+1) - \tilde{v}_j(\sigma|_{H^*})\big] \leqslant 0.\end{aligned}$$

Finally, since $\sigma \in \Sigma^p$, $\sigma|_{H^*} \in \Sigma^p$, and $\tilde{v}_j(\sigma|_{H^*}) \geqslant \underline{v}_j$. *Q.E.D.*

Proposition 2 establishes the existence of an optimal penal code. A comment of Jacques Crémer (1983) greatly simplified my original proof, and recently Harris (1984) has, independently of Crémer, provided a very similar argument. All these proofs exploit in an essential way the idea of a simple penal code and the one-shot deviation criterion of Proposition 1. The following standard assumptions, which imply Assumption 1, are used in the proof of Proposition 2.

ASSUMPTION 3: *S is a compact topological space.*

ASSUMPTION 4: *π: $S \to R^n$ is continuous.*

PROPOSITION 2: *Under Assumptions* 2 *to* 4, *an optimal simple penal code exists.*

PROOF: By Assumption 2, $V \neq \varnothing$ and $\underline{v}_i$ is well defined, $i \in N$. Consider $\{Q^{i\eta}\}_{\eta=1}^{\infty}$, such that $Q^{i\eta} \in \Omega^p$ and $\lim_{\eta \to \infty} v_i(Q^{i\eta}) = \underline{v}_i$. Endow $\Omega = S^\infty$ with the product topology. By (A3) and (A4), v: $\Omega \to R^n$ is continuous and by Tychonoff's theorem, Ω is compact. Assume w.l.o.g. that $\{Q^{i\eta}\}$ is a convergent sequence and let $\underline{Q}^i = \lim Q^{i\eta}$. Since v_i is continuous, $v_i(\underline{Q}^i) = \underline{v}_i$. To complete the proof I show that $\sigma^k(\underline{Q}^1, \ldots, \underline{Q}^n)$ is a perfect equilibrium, for all $k \in N$.

Suppose not. Then by Proposition 1,

$$\alpha_j\big(q_j^*, \underline{q}_{-j}^i(t)\big) > v_j\big(\underline{Q}^i; t+1\big) - \underline{v}_j, \quad \text{for some } i, j, t, q_j^* \neq \underline{q}_j^i(t).$$

Since $Q^{i\eta} \to \underline{Q}^{i}$, by continuity for η large enough,

$$\alpha_j\left(q_j^*, q_{-j}^{i\eta}(t)\right) > v_j\left(Q^{i\eta}; t+1\right) - \underline{v}_j.$$

Hence, by Lemma 1, $Q^{i\eta} \notin \Omega^p$, a contradiction. *Q.E.D.*

Proposition 3 expresses the main idea of the paper. It indicates how an optimal simple penal code may be directly constructed from an arbitrary optimal penal code. To do so one simply extracts the *equilibrium* path of each strategy profile of an optimal penal code; the *n*-vector of paths so obtained defines an optimal simple penal code.

PROPOSITION 3: *Let* $(\underline{\sigma}^1, \ldots, \underline{\sigma}^n)$ *be an optimal penal code and* $\underline{Q}^i \equiv Q(\underline{\sigma}^i)$, $i \in N$. *Then* $(\underline{Q}^1, \ldots, \underline{Q}^n)$ *defines an optimal simple penal code.*

PROOF: Lemma 1 applied to $\underline{Q}^i$ yields the inequalities (2), all $i \in N$. Since $\underline{v}_j = v_j(\underline{Q}^j)$, Proposition 1 now implies that $\sigma^i(\underline{Q}^1, \ldots, \underline{Q}^n)$ is a perfect equilibrium, for all $i \in N$. Since $\tilde{v}_i(\sigma^i) = \underline{v}_i$, the proof is complete. *Q.E.D.*

The centrality of optimal penal codes is attested to by the next result which indicates how an optimal penal code may be used to characterize the set of perfect equilibrium paths. It provides a "converse" to Lemma 1 under the additional assumption that an optimal penal code exists. Note that the hypotheses of Proposition 2 are sufficient, but not necessary, for existence.

PROPOSITION 4: *Suppose an optimal penal code exists. Then* $Q^0 \in \Omega^p$ *if and only if*

$$\alpha_j\left(q_j^*, q_{-j}^0(t)\right) \leqslant v_j\left(Q^0; t+1\right) - \underline{v}_j \quad \text{for all } q_j^* \in S_j,\ j \in N,\ t = 1, 2, \ldots.$$

PROOF: Necessity follows from Lemma 1. To prove sufficiency, let $(\underline{Q}^1, \ldots, \underline{Q}^n)$ define an optimal simple penal code. Lemma 1 applied to $\underline{Q}^i$ yields the inequalities (2), for all $i \in N$. Together with the inequalities above, Proposition 1 now implies that $\sigma(Q^0, \underline{Q}^1, \ldots, \underline{Q}^n)$ is a perfect equilibrium. *Q.E.D.*

The two preceding proofs directly imply that simple strategy profiles suffice to obtain all perfect equilibrium paths.

PROPOSITION 5: *Suppose an optimal penal code exists. Then* $Q^0 \in \Omega^p$ *if and only if there exist* $Q^i \in \Omega$, $i \in N$, *such that* $\sigma(Q^0, Q^1, \ldots, Q^n)$ *is a perfect equilibrium.*

6. CONCLUSIONS

This paper provides a general framework for analyzing the set of perfect equilibrium paths of infinitely repeated games with discounting. Essential to this framework is the concept of an optimal penal code, and the two related notions of a simple penal code and simple strategy profile. Conditions are given under which there exists a simple penal code which is an optimal penal code. An implication of this result is that discounted games may be completely analyzed in terms of simple strategy profiles. Simple strategy profiles, like simple penal codes, are "simple" in the relevant sense: they may be easily checked to be perfect. In particular, only one-shot deviations from at most $(n+1)$ outcome paths need be considered.

The theory developed here does not apply to the undiscounted case. Optimal penal codes are in general not defined there, and simple strategy profiles do not suffice to generate all perfect equilibrium outcome paths.

The results of Abreu (1986) for a class of oligopolistic supergames, demonstrate how the general framework presented here can be exploited, and suggest that the optimality and simplicity approach could be usefully applied to other repeated economic models with discounting. In particular, the somewhat unimaginative reliance on Cournot-Nash reversion to support cooperative behavior does not seem tenable on theoretical or pragmatic grounds. These results also illustrate an important general point which was mentioned in earlier sections. Optimality might *require* that a deviant "cooperate" (in a one-period sense) in his own punishment. Whereas in the undiscounted model, the possibility of "cooperative" deviants is a redundant nicety, it is often critical in a world with discounting. This point does not seem to have been appreciated before. It is in a sense counterintuitive, particularly in view of the result on simple penal codes. The puzzle is resolved by recognizing that optimal punishment paths may be highly nonstationary: specifically the early stages of an optimal punishment must yield the player being punished a lower (average) payoff than the subsequent stages. This property has a very sharp expression in the oligopolistic quantity-setting games studied in Abreu (1986).

Analogues to the theorems established here ought to appear in any model with discounting and a "repeated" structure. Finally, the conceptualization of strategy profiles in terms of paths and deviations from prescribed paths should prove useful in other contexts.

Department of Economics, Harvard University, Cambridge, MA 02138, U.S.A.

Manuscript received September, 1984; final revision received December, 1986.

REFERENCES

ABREU, D. (1982): "Repeated Games with Discounting: A General Theory and an Application to Oligopoly," mimeo, Princeton.

——— (1986): "Extremal Equilibria of Oligopolistic Supergames," *Journal of Economic Theory*, 39, 191–225.

AUMANN, R. J., AND L. SHAPLEY (1976): "Long Term Competition—A Game Theoretic Analysis," mimeo.

BERTSEKAS, D. (1976): *Dynamic Programming and Stochastic Control*. New York: Academic Press.

CRÉMER, J. (1983): Personal Communication.

FRIEDMAN, J. W. (1971): "A Non-cooperative Equilibrium for Supergames," *Review of Economic Studies*, 28, 1–12.

HARRIS, C. (1984): "A Note on the Existence of Optimal Simple Penal Codes," mimeo, Oxford.

HOWARD, R. (1960): *Dynamic Programming and Markov Processes*. New York: M.I.T. and John Wiley and Sons.

RUBINSTEIN, A. (1979): "Equilibrium in Supergames with the Overtaking Criterion," *Journal of Economic Theory*, 21, 1–9.

SELTEN, R. (1965): "Spieltheoretische Behandlung eines Oligopolmodells mit Nachfrageträgheit," *Zeitschrift fur die Gesamte Staatswissenshaft*, 121, 301–324 and 667–689.

——— (1975): "Reexamination of the Perfectness Concept for Equilibrium Points in Extensive Games," *International Journal of Game Theory*, 4, 22–55.

WHITTLE, P. (1983): *Optimization Over Time: Dynamic Programming and Stochastic Control*, Vol. II. New York: John Wiley and Sons.